



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ ÇOK AŞAMALI TESTLERDE ÇOK KATEGORİLİ MADDE  
ORANININ YETENEK KESTİRİMİNE ETKİSİ

Hasibe YAŞI SARI

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye ... En İyiyeye ...*



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ ÇOK AŞAMALI TESTLERDE ÇOK KATEGORİLİ MADDE  
ORANININ YETENEK KESTİRİMİNE ETKİSİ

THE EFFECT OF POLYTOMOUS ITEM RATIO ON ABILITY ESTIMATION IN  
MULTISTAGE TESTS

Hasibe YAŞI SARI

Doktora Tezi

Ankara, 2023

## Öz

Çalışmanın amacı, karma testlerin kullanıldığı bireyselleştirilmiş çok aşamalı testlerde (BÇAT) çok kategorili madde oranının farklı koşullarda yetenek kestirimine etkisini incelemektir. Çalışma simülasyona dayalı bir araştırmadır. PISA 2018 uygulamasında okuma becerileri alanında kullanılan iki ve çok kategorili maddelerden kestirilen madde parametreleri kullanılarak, bireylerin yetenek parametreleri ve madde havuzu oluşturulmuştur. BÇAT koşulları; panel desenleri ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3"), test uzunlukları (20, 40 ve 60), yönlendirme yöntemleri (Doğru Sayısı (DS) ve Maksimum Fisher Bilgisi (MFB)) ve çok kategorili madde oranıdır (%10, %30 ve %50). Simülasyon veriler, BÇAT deseni ve analizler, WinGen, Cplex ve R Studio programındaki 'mstR' paketi ile elde edilmiştir. Çalışmada toplam 108 koşul, 100 replikasyon incelenmiştir. Simülasyonlar sonucunda RMSE, ortalama mutlak yanlılık ve korelasyon değerleri hesaplanmıştır. Araştırma sonucunda, testlerdeki çok kategorili madde oranı %10'dan %50'ye çıktığında, ortalama mutlak yanlılık ve RMSE değerlerinin düştüğü, korelasyon değerlerinin ise arttığı görülmektedir. Test uzunluğu arttıkça RMSE ve ortalama mutlak yanlılık değerleri düşerken, korelasyon değerlerinin arttığı görülmektedir. Yönlendirme yöntemleri açısından MFB, DS yönlendirme yöntemine göre daha iyi performans göstermiştir. Genel olarak üç aşamalı panel desenleri iki aşamalı panel desenlerine göre manidar düzeyde iyi sonuçlar vermiştir. 1-2 ve 1-4 panel desenlerinde ise hangi yönlendirme yönteminin kullanıldığına bir önemi bulunmamaktadır.

**Anahtar sözcükler:** madde tepki kuramı, bireye uyarlanmış çok aşamalı test, yönlendirme yöntemleri, karma testler, pisa 2018

### Abstract

The aim of the study is to examine the effect of polytomous item ratio on ability estimation in different conditions in multi-stage tests (MST) using mixed tests. The study is a simulation based research. In the PISA 2018 application, the ability parameters of the individuals and the item pool were created by using the item parameters estimated from the dichotomous and polytomous items obtained in the field of reading skills. MST conditions; panel patterns ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" and "1-3-3"), test lengths (20, 40 and 60), routing methods (Number Correct (NC) and Maximum Fisher Information (MFB)) and polytomous item ratio (10%, 30% and 50%). Simulation data, MST pattern and analysis were obtained with the help of WinGen, Cplex and "mstR" package in R Studio program. A total of 108 conditions and 100 replications were examined in the study. As a result of the simulations; RMSE, mean absolute bias and correlation values were calculated. As a result of the research, it is seen that when the ratio of polytomous items in the tests increases from 10% to 50%, the mean absolute bias and RMSE values decrease, while the correlation values increase. As the test length increases, RMSE and mean absolute bias values decrease, while correlation values increase. In terms of routing methods, MFB performed better than the NC routing method. In general, three-stage panel designs gave significantly better results than two-stage panel designs. In 1-2 and 1-4 panel designs, it does not matter which routing method is used.

**Keywords:** item response theory, computerized adaptive multi-stage test, routing methods, mixed tests, pisa 2018

## Teşekkür

Öncelikle, çok değerli bilgi ve tecrübeleri ile doktora eğitimim boyunca bana rehberlik eden, tez yazma sürecim boyunca anlayışlı ve yardımsever tavırlarıyla beni destekleyen, her anlamda örnek aldığım kıymetli danışmanın Prof. Dr. Hülya KELECİOĞLU'na teşekkür ederim.

Tez yazma ve savunma sürecimde, kıymetli önerilerini ve bilgilerini benimle paylaşan ve bana yol gösteren değerli jüri üyesi hocalarım Prof. Dr. Burcu ATAR'a, Doç. Dr. Kübra ATALAY KABASAKAL'a, Doç. Dr. C. Deha DOĞAN'a ve Doç. Dr. Beyza AKSU DÜNYA'ya teşekkür ederim. Ayrıca doktora eğitimim boyunca babacan tavırlarıyla, hem akademik hem de insani anlamda örnek aldığım kıymetli hocalarım Prof. Dr. Selahattin GELBAL ve Prof. Dr. Nuri DOĞAN başta olmak üzere, emeği geçen tüm hocalarıma teşekkür ederim.

Bu uzun ve meşakkatli yolda başta annem Zehra YAHŞİ olmak üzere, babama ve kardeşlerime tüm destekleri için, bana her zaman güvendikleri ve inandıkları için teşekkür ederim.

Manevi desteklerini esirgemeyen, kâh ağlayıp kâh güldüğümüz, isimlerini sayamadığım tüm arkadaşlarıma teşekkür ederim.

Son ve en önemlisi, tüm süreçte yanımda olan, yaptığım her şeyi onlar için yaptığımı bir kez daha dile getirdiğim, beni çok seven ve hep motive eden canım kızlarım Erva Bilge'ye, Elvin Neva'ya ve destekleri için eşim H.İbrahim SARI'ya teşekkür ederim.

Tezimi başta kızlarım olmak üzere, tüm kız çocuklarına ithaf ediyorum. Gelecek sizlerle güzelleşecek...

**İçindekiler**

Öz.....	i
Abstract.....	ii
Teşekkür.....	iii
Tablolar Dizini.....	v
Şekiller Dizini.....	vii
Simgeler ve Kısaltmalar Dizini.....	viii
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	7
Araştırma Problemi.....	8
Sınırlılıklar.....	9
Tanımlar.....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	10
Madde Tepki Kuramı.....	10
Bireye Uyarlanmış Çok Aşamalı Testler.....	12
İlgili Araştırmalar.....	16
Bölüm 3 Yöntem.....	20
Araştırma Yöntemi.....	20
Araştırmanın Deseni.....	20
Verilerin Üretilmesi.....	23
Veri Toplama Araçları.....	27
Verilerin Analizi.....	31
Bölüm 4 Bulgular ve Yorumlar.....	33
Alt Problemlere İlişkin Bulgular ve Yorumlar.....	33
Genel Bulgular ve Yorumlar.....	57
Bölüm 5 Sonuç, Tartışma ve Öneriler.....	69
Sonuçlar.....	69
Tartışma.....	74

Öneriler .....	76
Kaynaklar .....	79
EK-A: Testteki Çok Kategorili Madde Yüzdesi %10 Olduğunda Kesme Noktalarının Grafikleri.....	lxxxii
EK-B: Testteki Çok Kategorili Madde Yüzdesi %30 Olduğunda Kesme Noktalarının Grafikleri.....	lxxxv
EK-C: Testteki Çok Kategorili Madde Yüzdesi %50 Olduğunda Kesme Noktalarının Grafikleri.....	lxxxviii
EK-Ç: Araştırma Etik Komisyon İzin Muafiyeti Formu .....	xcı
EK-D: Etik Beyanı.....	xcii
EK-E: Doktora Tez Çalışması Orijinallik Raporu .....	xciii
EK-F: Dissertation Originality Report .....	xciv
EK-G: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı .....	xcv
<i>*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir. ....</i>	<i>xcv</i>

### Tablolar Dizini

<b>Tablo 1</b> <i>Araştırma Koşulları</i> .....	22
<b>Tablo 2</b> <i>Madde Parametrelerinin Betimsel İstatistikleri</i> .....	24
<b>Tablo 3</b> <i>400 ve 200 Maddelik Havuzdaki Madde Parametrelerinin Betimsel İstatistikleri</i> .	26
<b>Tablo 4</b> <i>Modüllerdeki Madde Sayıları</i> .....	30
<b>Tablo 5</b> <i>Tüm Koşullardaki Kesme Noktaları</i> .....	30



<b>Tablo 6</b> Çok Kategorili Madde Oranı %10 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları .....	34
<b>Tablo 7</b> Çok Kategorili Madde Oranı %30 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları .....	41
<b>Tablo 8</b> Çok Kategorili Madde Oranı %50 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları .....	50
<b>Table 9</b> Tüm Koşullardaki Ortalama Mutlak Yanlılık Bulguları .....	58
<b>Tablo 10</b> Ortalama Mutlak Yanlılık İçin Faktöryel ANOVA Sonuçları .....	59
<b>Table 11</b> Tüm Koşullardaki Ortalama RMSE Bulguları .....	62
<b>Tablo 12</b> Ortalama RMSE İçin Faktöryel ANOVA Sonuçları .....	63
<b>Table 13</b> Tüm Koşullardaki Korelasyon Bulguları .....	65
<b>Tablo 14</b> Korelasyon İçin Faktöryel ANOVA Sonuçları .....	66

## Şekiller Dizini

<b>Şekil 1</b> <i>Modül, aşama, yol ve panel deseni bileşenleri</i> .....	13
<b>Şekil 2</b> <i>Örnek Bireye Uyarlanmış Çok Aşamalı Test Uygulaması</i> .....	14
<b>Şekil 3</b> <i>Örnek BÇAT Deseni</i> .....	15
<b>Şekil 4</b> <i>Üretilen Yetenek Seviyelerinin Dağılımı</i> .....	23
<b>Şekil 4</b> <i>PISA Verilerine Ait Test Bilgi Fonksiyonları</i> .....	25
<b>Şekil 5</b> <i>Madde Havuzlarına Ait Test Bilgi Fonksiyonları</i> .....	26
<b>Şekil 6</b> <i>BÇAT Simülasyonunda Kullanılan Panel Desenleri</i> .....	28
<b>Şekil 7</b> <i>1-3 Panel Deseni, %10 Çok Kategorili Madde Oranı ve 40 Maddelik Testin Yönlendirme Modülü Bilgisi ve Toplam Doğru Sayıları</i> .....	34
<b>Şekil 8</b> <i>Testteki Çok Kategorili Madde Oranı %10 İken Ortalama Mutlak Yanlılık Değerleri</i> .....	35
<b>Şekil 9</b> <i>Testteki Çok Kategorili Madde Oranı %10 İken RMSE Değerleri</i> .....	37
<b>Şekil 10</b> <i>Testteki Çok Kategorili Madde Oranı %10 İken Korelasyon Değerleri</i> .....	38
<b>Şekil 11</b> <i>Testteki Çok Kategorili Madde Oranı %30 İken Ortalama Mutlak Yanlılık Değerleri</i> .....	42
<b>Şekil 12</b> <i>Testteki Çok Kategorili Madde Oranı %30 İken RMSE Değerleri</i> .....	44
<b>Şekil 13</b> <i>Testteki Çok Kategorili Madde Oranı %30 İken Korelasyon Değerleri</i> .....	46
<b>Şekil 14</b> <i>Testteki Çok Kategorili Madde Oranı %50 İken Ortalama Mutlak Yanlılık Değerleri</i> .....	51
<b>Şekil 15</b> <i>Testteki Çok Kategorili Madde Oranı %50 İken RMSE Değerleri</i> .....	53
<b>Şekil 16</b> <i>Testteki Çok Kategorili Madde Oranı %50 İken Korelasyon Değerleri</i> .....	55
<b>Şekil 17</b> <i>Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Ortalama Mutlak Yanlılık Üzerindeki Etkileşim Etkisi</i> .....	60
<b>Şekil 18</b> <i>Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Ortalama RMSE Üzerindeki Etkileşim Etkisi</i> .....	64
<b>Şekil 19</b> <i>Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Korelasyon Üzerindeki Etkileşim Etkisi</i> .....	68

### **Simgeler ve Kısaltmalar Dizini**

**BBT:** Bireyselleştirilmiş Bilgisayarlı Test

**BÇAT:** Bireye Uyarlanmış Çok Aşamalı Test

**DS:** Doğru Sayısı

**MTK:** Madde Tepki Kuramı

**MFB:** Maksimum Fisher Bilgisi

## Bölüm 1

### Giriş

Bu bölümde problem durumu, araştırmanın amacı ve önemi, araştırma problemi, alt problemler, sayıltılar, sınırlılıklar ve tanımlar yer almaktadır.

#### Problem Durumu

Teknoloji; sanayi, iletişim, eğitim gibi alanlar başta olmak üzere her geçen gün büyüyerek hayatımızla bütünleşmiştir. Teknoloji kullanımının yaygınlaşmasıyla bilgisayarın eğitim ve ölçme değerlendirmedeki yeri de artmıştır. Eğitimde bilgisayarlı ölçme ve değerlendirme uygulamaları 1980'li yıllardan beri söz konusu olmaktadır (Weiss, 1982; Weiss & Kindsburry, 1984). Amerika Birleşik Devletleri başta olmak üzere dünya genelinde bilgisayarlı bireye uyarlanmış ve bireyselleştirilmiş çok aşamalı testlerin kullanımı yaygınlaşmaktadır. 2020 yılında tüm dünyada yaşanan COVID-19 salgını ile öğrencilerin uzaktan eğitim almaları, ölçme ve değerlendirmenin uzaktan yapılması eğitimde ölçme değerlendirmede teknoloji kullanımının önemini bir kez daha ortaya çıkarmıştır.

Bireyselleştirilmiş testler ilk kez 1900'lerin başlarında Binet & Simon (1905) zekâ testinde uygulanmıştır (Weiss, 1982). Binet Zeka Testi, her bir zihinsel yaş seviyesinde 10 madde olacak şekilde toplam 210 maddeden oluşmaktadır. Testte her bir bireye uygulanan maddelerin zorluk seviyeleri test işlemi sırasında belirlenmektedir. Testi alacak bireylerle ilgili önsel bilgiye göre 10 madde seçilerek her birey için farklı bir başlangıç noktası belirlenmektedir. Bireyler maddeleri doğru yanıtlama oranlarına göre bir üst zihinsel seviyeyi ölçen maddelere ya da bir alt zihinsel seviyeyi ölçen maddelere geçmektedir. Bu süreç test uygulayıcısının belirlediği sonlandırma kuralına kadar devam eder. Binet Zeka Testi'nin uygulama aşamaları incelendiğinde bireyden bireye farklılık gösterdiği yani bireye uyarlandığı görülmektedir.

Bireyselleştirilmiş bilgisayarlı test (BBT); önceden kalibre edilmiş bir madde havuzu kullanılarak, bireyin yeteneğine ilişkin önceki tahminlere ya da belirlenen yeteneğe dayalı olarak başlayan, bireyin yanıtladığı madde sonrası kestirilen yeteneğe göre yöneltilecek maddeleri seçmek için bir madde seçim kuralının kullanıldığı ve önceden belirlenmiş bir sonlandırma kuralına göre sonlandırılan bilgisayarda uygulanan testlerdir (Wang, 2017; Weiss, 1982). Reckase (1974) tarafından bir parametrelili lojistik (1 PL) model için önerilen bilgisayar programı BBT'nin MTK'ya dayalı ilk örneklerindedir.

BBT uygulamalarının geleneksel testlere göre pek çok avantajları bulunmaktadır. Eğitimde ölçme değerlendirilmede kullanılan bireyselleştirilmiş bilgisayarlı testlerde (BBT) madde tepki kuramının (MTK) sağladığı en büyük avantaj, bireyin yeteneği ile madde güçlüğü arasında ilişki kurulmasıdır. Geleneksel kâğıt-kalem testi uygulamalarındaki sabit test uzunluğunun aksine BBT uygulamalarında her bir bireye özgü test uzunlukları bulunmaktadır (Weiss, 1983). Her birey için farklı ve geleneksel teste göre daha az madde kullanılır. Ancak BBT uygulamaları test uzunluklarının farklı olması, farklı sorular sorulması ve bireyin sorulara geri dönememesi gibi nedenlerde eleştirilmektedir. Sonuçların hemen açıklanması geri bildirim açısından oldukça önemlidir. Ancak bireye uyarlanmış testler BBT ile sınırlı değildir.

Amerika'daki Eğitimsel Test Hizmetleri (Educational Testing Service, [ETS]) şirketi 1993 yılında, ilk geniş ölçekli BBT uygulamasını Lisansüstü Kayıt Sınavı (Graduate Record Examination, [GRE]) ile gerçekleştirmiştir. 2004 yılında Mali Müşavirler (Certified Public Accountants, CPA) Sınavı (Breithaupt & Hare, 2007), 2011'de ise GRE BÇAT şeklinde uygulanmaya başlamıştır. Massachusetts Yetişkin Yeterlilik Testi (The Massachusetts Adult Proficiency Test), Hukuk Fakültesi Kabul Sınavı (The Law School Admission Council), Yabancı Dil Olarak İngilizce Sınavı (Test of English as a Foreign Language, TOEFL), Ulusal Mimari Kayıt Kurulları Sınavı (National Council of Architectural Registry Boards, NCARB), Ulusal Eğitim İlerleme Değerlendirme (National Assessment of

Educational Progress, NAEP), and ABD Tıbbi Lisans İncelemesi (U.S. Medical Liensure Examination, USMLE) BÇAT şeklinde uygulanan sınavlardan bazılarıdır (Hendrickson, 2007).

Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment – PISA), 2015 yılına kadar sadece kağıt kaleme dayalı geleneksel test olarak uygulanmaktayken, 2015 yılı itibariyle geleneksel testin yanı sıra bilgisayara dayalı test olarak da uygulanmaya başlanmıştır. BÇAT yüksek katılımı, yaklaşık 40 ülkede uygulanan Yetişkin Yetkinliklerinin Uluslararası Değerlendirmesi Programı (Programme for the International Assessment of Adult Competencies-PIAAC) için 2012'den itibaren uygulanmaya başlamıştır (Yamamoto, vd. 2019). Dolayısı ile PISA 2018 katılım oranı yüksek sınavlar arasında, bilgisayar tabanlı platformda gerçekleştirilen ikinci uygulama olmuştur. Güncel olarak 2023'te Amerika'da üniversite eğitimi almak isteyenlerin girdiği, standartlaştırılmış geniş ölçekli bir test olan Uluslararası Eğitim Yetenek Testi (Scholastic Aptitude Test – SAT) BÇAT şeklinde dijital ortamda uygulanmıştır. 2024'te ise Amerika genelinde uygulanan Eğitim Yetenek Testi de BÇAT şeklinde uygulanacaktır. Bireye uyarlanmış çok aşamalı test (BÇAT) aşamalardan oluşmuş, bireye uyarlanabilir testlerdir. BÇAT uygulamalarının; bireye uyarlama, test geliştiricisine test formlarını önceden inceleme ve testi alan bireylere yanıtlarını kontrol etme imkanları sunması popülaritesini artmaktadır.

BÇAT'da uygulama sırasında, sınava giren her bireye rastgele bir panel atanır ve o panelin 1. aşama sorularıyla teste başlanır. Bu aşamadaki modülü tamamladıktan sonra testten puan alınır. Alınan puana göre bireyin yeteneği kestirilir ve bir sonraki aşamada en uygun zorluk seviyesindeki modüle yönlendirilir. Bu süreç bireylerin final yetenekleri kestirilene kadar devam eder. Bireyin her aşamadan bir modül içeren testten geçtiği rota "yol" olarak adlandırılır (Zheng, vd., 2012).

BÇAT ve BBT karşılaştırıldığında, BÇAT; maddelerin uygulamadan önce tasarlanıp teste entegre edilebilmesi ve her aşamada puanlanabilmesi, test geliştiricilerin test yapısının kalitesi ve son testin uygulanması üzerinde daha fazla kontrol sahibi olmalarını sağlar. Böylece modüllerdeki maddeler arasındaki yerel bağımsızlığın ve dolayısıyla testin tek boyutluluğunun daha iyi güvence altına alınması, madde seçimi ve içerik etkileri üzerinde kontrolün artması, içerik ve diğer özelliklerin kontrolü, test güvenliğinin artması, maddeleri incelenme kolaylığı ve daha az veri yönetimi sağlanır (Hendrickson, 2007). Aşamalar paralel paneller arasında yer değiştirebilmektedir. BBT'de madde düzeyinde bireye uyarılma yapılırken, BÇAT'da modül düzeyinde uyarılma söz konusudur. Bu durum BÇAT uygulamalarında bireylerin maddeler arasında dönüş yapabilmelerini ve test hazırlayıcının test içeriğini daha iyi kontrol edebilmesini sağlamaktadır (Sarı, vd., 2016). Geleneksel BBT'ler, test sürecinde bireylere sıralı uygulama için bireysel yeteneklerine göre ayrı ayrı uyarlanabilirken; BÇAT, mini testlerden oluşan madde kümeleri kullanan benzer bir yaklaşımdır (Zenisky, vd., 2010). BÇAT, bu şekilde geleneksel testler (kâğıt-kalem ve bilgisayar temelli) ve BBT'lerin dezavantajlarını ortadan kaldırmaya çalışmaktadır (Hendrickson, 2007). BÇAT da dahil olmak üzere bireyselleştirilmiş testlerde, maddelerin yönlendirilebilmesi için test maddelerinin psikometrik özelliklerine ihtiyaç vardır (Colvin, 2014). BBT her maddeyi anında seçerken, BÇAT'ta uygulamadan önce tüm maddeler ve testler belirlenir. Her iki uygulamada da madde havuzundaki madde sayısının oldukça fazla ve madde güçlüğü ranjının geniş olması gerekmektedir. BÇAT uygulamalarında madde havuzlarındaki madde sayısının, test için gerekli madde sayısının 1.5 katı olması tavsiye edilmektedir (Wang, 2017). Yani madde havuzu her iki uygulamada da oldukça önemlidir (Weiss & Kingsbury, 1984). BBT ve BÇAT uygulamalarındaki ölçme doğruluğu, madde havuzunun kalitesi ile doğrudan ilişkilidir (Han & Guo, 2011). BÇAT'lar geniş yetenek seviyesi aralığında ve zor testlerde geleneksel testlere kıyasla daha verimli ve hassas ölçüm yapılmasını sağlar (Hendrickson, 2007; Lord, 1971). BÇAT'lar yetenek seviyesi düşük

bireyleri cesaretlendirirken, yetenek seviyesi yüksek bireylerin sıkılmasını önler. Böylece bireylerin motivasyonu yükselir ve testlere katılım oranı artar. BÇAT'lar genellikle daha kısa test uzunluğu ve puanların daha kısa sürede raporlanmasını sağlar (Weiss, 1982).

BÇAT diğer testlerden farklı olarak modül, aşama, panel, yol, test birleştirme ve yönlendirme gibi yeni terminolojilerin de ortaya çıkmasını sağlamıştır. Modül, aşama ve panel birbiri içerisinde kümelenmiş bileşenlerdir. Maddeler bir araya gelerek, madde takımı olarak da isimlendirilen modülleri oluşturur. Modüllerdeki maddeler farklı konularda olsalar bile benzer psikometrik özellikler göstermelidir. Modüller ölçülmek istenen özelliğe göre kolay, orta ve zor olacak şekilde maddelerden oluşmaktadır. Bir ya da daha fazla modül bir araya gelerek aşamaları oluşturur. BÇAT uygulamaları, ilk aşama yönlendirme modülü olacak şekilde, en az iki aşamadan oluşur. Birden fazla aşamadan oluşan paralel testlerin oluşturduğu desenlere panel denir. BÇAT uygulamaları, madde havuzu hazırlama, panel yapısını belirleme, test birleştirme, yönlendirme ve puanlama aşamalarından oluşmaktadır. Maddeleri psikometrik özelliklerine ve testte kullanılış amacına göre bir araya getirerek, modül, aşama ve panelleri oluşturma işlemi test birleştirme olarak isimlendirilmektedir.

Modüller arasındaki yolları oluşturmak için puanlama ve yönlendirme yapılmalıdır. BÇAT'ta sonraki modülü seçmek için iki temel yaklaşım vardır. Birincisi, uygun bir optimal seçim kriteri ya da toplam test puanını belirlemektir. İkincisi ise, modüller arasında uygun kesme değerleri belirlemektir (Magis, vd., 2017, s.130). Birinci aşamanın, diğer adıyla yönlendirme modülünün istatistiksel özellikleri tüm testin ölçme hassasiyeti üzerinde büyük bir etkiye sahiptir (Kim & Plake, 1993). Ancak Dallas (2014) BÇAT bileşenlerinden biri olan yönlendirme yöntemlerinin literatürde gereken ilgiyi görmediğine dikkat çekmiştir. Literatürde BÇAT tasarlama yönelik test birleştirme yöntemleri ile ilgili pek çok çalışma ile karşılaşılrken (Doğruöz, 2018; Luo & Kim, 2018; Luecht & Nungester, 1998) yönlendirme yöntemleri araştırmalarda ve uygulamalarda nispeten daha az çalışılmıştır.



Literatürü incelediğimizde BÇAT'da yönlendirme yöntemleri ile ilgili yapılmış çalışmalarda genellikle 'Yaklaşık Maksimum Bilgi' (YMB) (Approximate Maximum Information-AMI) ve 'Tanımlanmış Evren Aralıkları' (TEA) (Defined Population Intervals - DPI) yönlendirme yöntemlerinin (Kim, vd., 2010; Wang, 2017; Zenisky, 2004) kullanıldığı görülmektedir. Madde tepki kuramına dayalı Maksimum Fisher Modül Bilgisi (*maximum Fisher module information* -MFI), Maksimum Olasılık Ağırlıklı Modül Bilgisi (*maximum likelihood weighted module information*-MLWMI), Maksimum Sonsal Ağırlıklı Modül Bilgisi (*maximum posterior weighted module information* -MPWMI), Maksimum Modül Kullback-Leibler Bilgisi (*maximum module Kullback-Leibler information*-MKL), Maksimum Sonsal Modül Kullback-Leibler Bilgisi (*maximum posterior module Kullback-Leibler information*-MKLP) ve rastgele (random) olmak üzere altı yönlendirme yöntemi mevcuttur (Boztunç-Öztürk, 2019; Sarı & Raborn, 2018).

Geniş ölçekli sınavlar başarı sınavları olup, bu sınavlarda çoktan seçmeli maddeler kullanılmaktadır. Bu maddeler 0-1 olacak şekilde iki kategorili puanlanmaktadır. Bu maddelerin yanı sıra açık uçlu maddeler, yapılandırılmış cevaplar gibi çok kategorili puanlanan madde türleri bulunmaktadır. Hem iki hem de çok kategorili puanlanan maddelerden oluşan testler karma format testler olarak adlandırılır. Son yıllardaki PIACC (2011-2017), PISA (2018) gibi geniş ölçekli sınavların bazı bölümlerinde karma format testler kullanılmaktadır (Yamamoto vd., 2019). Karma format testler, hem uluslararası geniş ölçekli testlerde yaygınlaşması hem de test bilgi miktarının iki kategorililere göre daha yüksek olması açısından oldukça önemlidir (Rosa, vd., 2001). Literatürü incelediğimizde iki kategorili ve çok kategorili maddelerden oluşan testlerde farklı yönlendirme yöntemleri kullanılarak (Kim, vd., 2010; Kim, vd., 2013; Wang, 2017; Zenisky, 2004) yapılan çalışmalar mevcuttur. Ancak literatürde karma (mixed format) testler üzerinde daha az sayıda çalışma bulunmaktadır (Kim, vd., 2013). Karma testlerin yanı sıra kullanılan çok kategorili maddelerin oranları da yetenek kestiriminin doğruluğu açısından oldukça önemlidir (Kim & Dodd, 2014). Karma testlerin öğrenci yeteneğinin

hesaplanmasında daha yüksek miktarda bilgi sağlaması, gelecekte bu yapıdaki testlere olan ilgiyi arttıracaktır. Ayrıca otomatik test birleştirme yöntemlerinden biri olan doğrusal programlama yaklaşımına göre, test birleştirme yöntemlerinin farklı koşullarda nasıl sonuç verdiğiine ilişkin karma testlerle yapılan sınırlı sayıda çalışma bulunmaktadır (Park, vd., 2014; Park, 2015). İlgili araştırmalar incelendikten sonra, literatürdeki boşluklardan yola çıkarak, doğrusal programlama test birleştirme yöntemi kullanılarak, karma testlerde çeşitli panel desenlerinde, farklı yönlendirme yöntemleri ve farklı oranlardaki çok kategorili maddeleri içeren bir çalışmanın literatüre katkı sağlayacağı kararına varılmıştır.

### **Araştırmanın Amacı ve Önemi**

Araştırmanın amacı, bireye uyarlanmış çok aşamalı karma testlerdeki çok kategorili madde oranlarının yetenek kestirimine etkisini farklı koşullarda incelemektir.

Literatürde BÇAT'ın farklı koşullar altında (yönlendirme yöntemleri, test uzunlukları vb.) incelendiği pek çok çalışma bulunmaktadır (Doğruöz, 2018; Kim vd., 2010; Luecht & Nungester, 1998; Sarı & Raborn, 2018; Wang, 2017; Zenisky, 2004). Ancak, karma testler için farklı yönlendirme koşulları altındaki çalışmalarının oldukça sınırlı olduğu görülmektedir (Kim vd., 2012; Kim vd., 2013). Bu yönüyle çalışmanın literatüre önemli katkılar sunacağı düşünülmektedir. Araştırma verisi olarak, geniş ölçekli PISA 2018 okuma becerileri alanında gerçek BÇAT uygulamasında kullanılmış iki ve çok kategorili verilerin madde parametrelerine dayalı simülasyon verileri kullanılmıştır. Bu nedenle gelecek yıllarda yapılacak PISA sınavlarına ve diğer uygulamalara katkı sağlaması açısından da önem taşımaktadır. Ulusal literatürde ise karma testler ve BÇAT'ın çalışıldığı bir çalışmaya rastlanmamıştır. Ayrıca ülkemizde araştırma düzeyinde gerçek BBT çalışmaları yapılmıştır (Cikrikci, vd., 2020; Şenel & Kutlu, 2018). Gelecekte BÇAT uygulamalarının da gerçekleştirilebileceği aşikârdır. Araştırma bu bağlamda gelecekte ülkemizde uygulanacak çalışmalara da katkı sağlayacaktır.

## Araştırma Problemi

Araştırma problemi: "Karma testten oluşturulmuş bireye uyarlanmış çok aşamalı test (BÇAT) uygulamalarında çok kategorili madde oranları değiştiğinde (%10, %30 ve %50) elde edilen yetenek kestirimlerinin doğruluğu farklı test uzunluğu (20, 40 ve 60), panel deseni ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl bir değişim göstermektedir?". Bu bağlamda çalışmada 108 (6x3x2x3) farklı koşul incelenmiştir.

**Alt problemler.** Problem cümleleri kapsamında alt problemler verilmiştir.

1. Testteki çok kategorili madde oranı %10 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?

2. Testteki çok kategorili madde oranı %30 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?

3. Testteki çok kategorili madde oranı %50 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?

## Sınırlılıklar

Araştırmada içerik dengeleme ve madde kullanım sıklığı göz ardı edilmiştir.

Araştırmada kullanılan gerçek veriler PISA 2018 okuma becerileri alanı testindeki iki kategorili ve çok kategorili maddelerin parametreleri ile sınırlıdır.

## Tanımlar

**Modül:** Farklı güçlük düzeylerindeki birden fazla maddenin oluşturduğu kümedir.

**Aşama:** Bir ya da daha fazla modülün bir araya gelmesiyle oluşan, testi alan her bireyin tamamlaması gereken işlem sırasıdır.

**Panel:** Birden fazla aşamadan oluşan paralel testlerin oluşturduğu desendir.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde madde tepki kuramı ve bireye uyarlanmış çok aşamalı testler olmak üzere iki farklı konu hakkında bilgi verilmiştir. Ayrıca literatürdeki karma testlerin kullanıldığı bireye uyarlanmış çok aşamalı testlerle ilgili araştırmalara yer verilmiştir.

#### Madde Tepki Kuramı

Klasik test kuramı (KTK) ve madde tepki kuramı (MTK) bireylerin yetenek seviyesi gibi örtük özelliklerinin tahmin edilmesinde kullanılan farklı model yaklaşımlarıdır. KTK, MTK'ya göre daha eski test yaklaşımı olmasına rağmen, ürettiği madde ve kişi parametrelerinin daha kesin sonuçlar vermesi nedeniyle son yıllarda daha yaygın olarak kullanılmaktadır. Bu avantajıyla birlikte, çok fazla sayıda örneklem gerektirmesi ve güçlü varsayımlara sahip olması nedeniyle dezavantajları da bulunmaktadır. MTK bireyin gözlenen performansı ile bu performansın altında yatan gözlenemeyen örtük özellik ya da yetenekler arasındaki ilişkiyi ifade eden bir kuramdır. Örtük özellik kuramı ya da madde tepki kuramı olarak adlandırılır (Hambleton & Swaminathan, 1985). Ölçülen bir özelliğin ya da yeteneğin test maddesine verilen tepkiyle olan ilişkisini tanımlayan ilişki modelidir. MTK farklı yetenek seviyesindeki bireylerin maddeleri doğru cevaplama olasılığını lojistik regresyon eğrisi ile modelleyerek açıklamaya çalışır.

Madde tepki kuramının tek boyutluluk ve yerel bağımsızlık olmak üzere iki temel varsayımı bulunmaktadır. Tek boyutluluk, madde tepkileri arasındaki ortak varyansı yeterli düzeyde açıklayan tek bir gizil özelliğinin bulunduğu varsayılmıştır (Embretson & Reise, 2000). Ölçülen özelliğin tek boyutlu bir yapıya sahip olmasından kasıt, tüm test maddelerinin tek bir özelliği ölçmesidir. Yerel bağımsızlık ise aynı yetenek düzeyinde bir maddeyi doğru cevaplama olasılığının diğer maddeleri cevaplama olasılığından bağımsız olması durumudur (Embretson & Reise, 2000; Hambleton & Swamington, 1985). Bireyin

bir maddeye verdiđi tepkilerin diđer maddelere verdiđi tepkilerden bađımsız olmasıdır. Bir maddeye dođru ya da yanlıř cevap verme olasılıđının bir bařka maddeye verdiđi cevabı etkilememesidir.

MTK'de maddelerin cevap örüntüsüne, testin boyut sayısına ve parametrik olma durumuna göre kullanılan çeřitli modeller vardır. Maddeler dođru-yanlıř, 0-1 gibi iki kategorili olarak puanlandıđında kullanılan MTK modelleri; bir parametrelili lojistik (1 PL, Rasch, 1960), iki parametrelili lojistik (2 PL, Birnbaum, 1968), üç parametrelili lojistik (3 PL, Birnbaum, 1968) ve dört parametrelili lojistik (4 PL) modellerdir. BBT ve BÇAT uygulamalarında genellikle 3PL model tercih edilmektedir (Erdem-Kara, 2019; Green, vd., 1984).

0-1 gibi iki kategoriden daha fazla kategori içeren çok kategorili puanlanan maddeler de bulunmaktadır. Çok kategorili maddelerde sınırları ayıran eřik kategorileri bulunmaktadır. Eřik kategorilerine ait sınır tepki fonksiyonu sayısı, kategori sayısından bir eksiktir. Çok kategorili madde yanıtları iki kategorili hale getirilerek deđerlendirilir. Çok kategorili maddeleri iki kategorili madde řeklinde gösterirken iki modelleme yaklařımından yararlanılır. Bunlar; Rasch modelleri ve Thurstone / Samejima modelleridir (Ostini & Nering, 2006). MTK bađlamında, çok kategorili maddeler için; kısmi puan modeli (KPM), dereceleme ölçeđi (DÖM), nominal tepki (NTM) ve ařamalı tepki modelleri (ATM) geliřtirilmiřtir (De Ayala, 2009). KPM, kısmen dođru yanıtların olduđu bařarı testlerinde madde yanıtlarını tanımlamada kullanılmaktadır (Embretson & Reise, 2000). KPM'de madde ayırt edicilikleri -yani a parametresi- maddeler arasında eřit kabul edilir. Bu nedenle Rasch modelin geniřletilmiř hali olarak da bilinmektedir. Genelleřtirilmiř kısmi puan modeli (GKPM) ise, KPM'de tüm maddelerin ayırt edicilik parametresinin modele dâhil edilmesiyle genelleřtirilmiř halidir (Han & Hambleton, 2014). Ařamalı tepki modeli ise sıralı kategorik yanıtların olduđu durumlarda kullanılması uygun olan iki parametrelili lojistik modelin uzantısıdır (Embretson & Reise, 2000).

MTK'de bir diğ er önemli konu ise test bilgisidir. Test bilgi fonksiyonu, madde bilgisi kullanılarak, bireylerin yetenek düzeylerinin ölçüm hassasiyetlerine karar verilmesinde önemli rol oynar. Belirli bir yetenek düzeyinde bilgi fonksiyonu ne kadar yüksek olursa, madde bireyin o düzeydeki yeteneğini de o kadar hassas şekilde ölçer (Embretson & Reise, 2000). Test bilgi fonksiyonunun tersi ölçmenin standart hatasını (ÖSH) vermektedir. ÖSH ile belirli yetenek düzeylerindeki ölçme hassasiyeti değerlendirilebilmektedir (Yang, 2016). Çok kategorili modellerde, madde karakteristik eğrisi her bir kategori için ayrı ayrı oluşturulur ve kategori sayısının bir eksiği kadar eşik değeri bulunur. Bu değ erleri göstermek amacıyla kategori tepki fonksiyonları kullanılmaktadır.

### **Bireye Uyarlanmış Çok Aşamalı Testler**

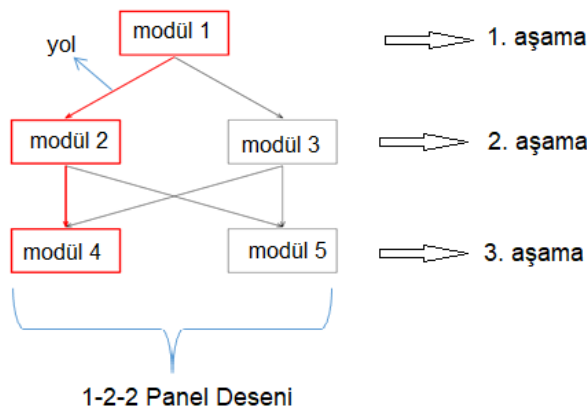
Bireye uyarlanmış çok aşamalı test (BÇAT), kâğıt kalem testlerinin ya da bilgisayar tabanlı testlerin bireye uyarlanabilir olarak uygulandığı ve bir birim olarak puanlandığı testlerdir. Bu testler, madde düzeyinde bireyselleştirilmiş bilgisayarlı testlerin dezavantajlarının birçoğunu ortadan kaldırmakta, yeterlilik ölçeğinde daha verimli ve hassas ölçümün yanı sıra zaman tasarrufu da sağlayarak uyarlanabilir testin tüm avantajlarını ortaya koymaktadır (Hendrickson, 2007).

Alanyazında BÇAT farklı isimler ile adlandırılmıştır. Bu adlandırmalar arasında iki aşamalı test (two-stage testing, Adema, 1990), bilgisayarlı yeterlik testi (computerized mastery testing, Lewis & Sheehan, 1990), çok aşamalı test (multistage testing, Patsula, 1999), yığı nlı çok aşamalı bireyselleştirilmiş test (bundled multistage adaptive test, Luecht, 2003), bilgisayarlaştırılmış çok aşamalı test (computer adaptive multistage testing, Hambleton & Xing, 2006; Patsula, 1999) ve çok aşamalı bireyselleştirilmiş test (multistage adaptive testing) yer almaktadır (Zheng, vd., 2012). Bu araştırmada ise BÇAT, Bireye Uyarlanmış Çok Aşamalı Test (Multistage Adaptive Test) olarak adlandırılmıştır.

BÇAT panel, aşama ve modül olmak üzere birbiri içerisinde kümelenmiş bileşenlerden oluşur. Birleştirilmiş BÇAT'ın temel yapısı "panel" olarak adlandırılır. Bir panel birkaç aşamadan oluşur ve her aşama farklı zorluk seviyelerini temsil eden "modül" olarak adlandırılan birkaç testlet içerir. Modüller madde takımı (testlet) (Wainer & Kiely, 1987) olarak da adlandırılabilir. Bir ya da daha fazla modül bir araya gelerek aşamaları oluşturur. Modül ve yönlendirme kuralları birleşerek panelleri oluştururlar. Bir panelin birkaç farklı yolu olabilir. Modül, aşama, yol ve panel deseni bileşenleri Şekil 1'de gösterilmiştir.

### Şekil 1

*Modül, aşama, yol ve panel deseni bileşenleri*



Şekil 1'de görüldüğü üzere 1-2-2 panel deseninden oluşan bir BÇAT tasarımı bulunmaktadır. Bu tasarım 3 aşama ve 5 modülden oluşmaktadır. Aşamalar arası geçiş rotasını gösteren farklı yollar olabilir. Bu tasarımda olası yollardan bir tanesi kırmızı çizgi ile gösterilmiştir.

Hendrickson (2007)'in bireyselleştirilmiş çok aşamalı test uygulaması ise Şekil 2'de gösterilmiştir.



## Şekil 2

### Örnek Bireye Uyarlanmış Çok Aşamalı Test Uygulaması

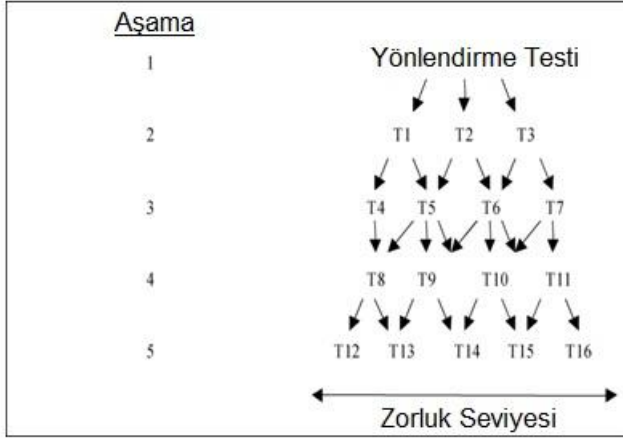
<u>1. Aşama</u>	
Adım 1	Yönlendirme yöntemleri, geniş aralıklı ya da ortalama güçlükte birinci aşama testi
Adım 2	Puan testi
Adım 3	Yetenek kestirimi
Adım 4	Sıradaki testin seçilmesi
<u>Sonraki Aşamalar</u>	İstenen toplam test sayısı elde edilene kadar Adım 1'den 4'e uygun zorluk derecesinde testlerle tekrarlayın
Bitiş	Final yetenek kestirimi

Şekil 2'de görüldüğü üzere 1. aşamada yönlendirme yöntemleri, geniş aralıklı ya da ortalama güçlükte birinci aşama sorularıyla teste başlanır. Daha sonra adım adım testten puan alınır. Alınan puana göre yetenek kestirilir. Bireyin yeteneğine göre sıradaki aşamaya geçilir. Bu süreç bireylerin final yetenekleri kestirilene kadar devam eder. BÇAT geliştirme aşamasında madde havuzu, modül, panel yapılandırılması, aşama, test birleştirme, puanlama ve yönlendirme yöntemi gibi BÇAT bileşenleri ile ilgili temel ölçütler belirlenmelidir. Madde havuzu bireyselleştirilmiş bilgisayarlı testlerde olduğu gibi BÇAT için de oldukça önemlidir. Madde havuzunun niteliği Georgiadou, vd. (2006)'e göre, iki ölçüt dikkate alınarak değerlendirilmektedir. Öncelikle madde havuzu, uygulamanın varsayımlarını yerine getirecek büyüklükte maddeye sahip olmalıdır. İkincisi, madde havuzundaki maddeler yeteneğin her düzeyi için yeterli bilgiyi sağlayacak özelliklere sahip olmalıdır. Güçlük düzeyleri farklı maddelerden oluşan modüller önceden oluşturulduğu için test geliştiriciler uygulamaya başlamadan önce tüm testi görebilir.

Şekil 3 ilk aşamada yönlendirme testi ile başlayan beş aşamalı test desenine sahip, her aşamada üç ile beş madde arasında değişen madde takımlarından (T1-T16) oluşan BÇAT örneği verilmiştir (Hendrickson, 2007).

### Şekil 3

#### Örnek BÇAT Deseni



BÇAT'larda önemli unsurlardan bir tanesi de test birleştirmedir. Manuel olarak modülleri birleştirmek oldukça zor ve zaman alıcıdır (Luecht & Nungester, 1998). Daha kullanışlı ve doğru ölçümler yapan BÇAT çoklu panel desenleri oluşturmak için genellikle otomatik test birleştirme teknikleri kullanılır. Bu teknik bilgisayar algoritmaları kullanılarak gerçekleştirilir. Otomatik test birleştirme yöntemi ile panel yapısı, test uzunluğu, içerik kısıtlamaları ve maruz kalma oranı kontrolleri gibi kısıtlamalarla istenen panel desenleri oluşturulur. En popüler otomatik test birleştirme algoritmalarından ikisi doğrusal programlama ve sezgiseldir (Park, 2015). Doğrusal programlama, bir fonksiyonu optimize eden ve kısıtlamaları en iyi şekilde karşılayan çözüm kümesini, karma tam sayılı doğrusal programlama ile ortaya koyar. Luecht & Nungester (1998), modül ve panellerin en uygun şekilde bir araya getirilmesi için 'yukarıdan-aşağıya' ve 'aşağıdan-yukarıya' test birleştirme yöntemlerini geliştirmişlerdir. Bu yöntemler de otomatik test birleştirme yöntemlerinde olduğu gibi bilgisayar yazılımı aracılığıyla gerçekleştirilir. Test birleştirme yöntemleri açısından alanyazın incelendiğinde 'yukarıdan-aşağıya' ve 'aşağıdan-yukarıya' test

birleştirme yöntemlerinin yaygın kullanıldığı görülmektedir (Doğruöz, 2018; Luo & Kim, 2018; Zheng, vd., 2012).

Yönlendirme metotları, BÇAT uygulanması sırasında yapılan puanlamalar kullanılarak modüller arasında yol (path) oluşturmak amacıyla kullanılır. Doğru sayısı puanlama (Number-correct) veya MTK'ye göre puanlama ile bireyin uygun modüle yönlendirileceği noktada kesme puanına karar verilmektedir. Yaklaşık maksimum bilgi yöntemi ve tanımlanan evren aralıkları yöntemi literatürde sıklıkla kullanılan yönlendirme yöntemleridir (Kim, vd., 2010; Wang, 2017; Zenisky, 2004). Son yıllarda mstR (Magis, vd., 2018) paketinin kullanılmasıyla maksimum Fisher bilgisi (MFI), maksimum olasılık ağırlıklı bilgiler (MLWI), maksimum posterior ağırlıklı bilgiler (MPWI), Kullback – Leibler (KL) ve posterior Kullback – Leibler (KLP) yönlendirme yöntemleri de kullanılmaya başlamıştır (Boztunç-Öztürk, 2019; Sarı & Raborn, 2018).

### **İlgili Araştırmalar**

Bu bölümde, BÇAT uygulamalarının ve BÇAT bağlamında karma testlerin incelendiği çalışmalara yer verilmiştir.

### ***BÇAT Uygulamalarının İncelendiği Araştırmalar***

Patsula (1999) çalışmasında, kâğıt-kalem testi, farklı BBT ve BÇAT desenlerinden elde edilen yetenek kestirimlerinin doğruluklarını karşılaştırmıştır. Çalışmasında sabit toplam test uzunluğu ve maddeye maruz kalma kontrol altına alındığında, aşamaların sayısını (2 ve 3), aşama başına modül sayısını (1–3 ve 1–3–3) ve her modülde (6, 12, 18, 24) ölçme hassasiyetini en üst düzeye çıkarmak için yönlendirme modülü için madde sayısı ile sonraki aşamalardaki madde sayısını araştırmıştır. Çalışma sonuçlarına göre, BBT'ler en doğru yetenek kestirimini üretmiş; aşama sayısı ve modül sayısının artırılması ise BÇAT sonuçlarını BBT sonuçlarına yaklaştırmıştır. BÇAT'larda aşama sayısının ikiden üçe çıkarılması yetenek kestirimindeki hata miktarını azaltmıştır. Modül sayısının üçten beşe çıkarılması ise yetenek kestirimlerinin doğruluğunu artırmıştır.

Wang (2017) çalışmasında BÇAT tasarımını 1-2-3 ve 1-3-3 panel deseni; yaklaşık maksimum bilgi (Approximate Maximum Information, AMI) ve tanımlanmış popülasyon aralıkları (Defined Population Intervals, DPI) yönlendirme yöntemleri; 45 ve 60 test uzunluğu ve ileri ve geri test birleştirme yöntemleri olmak üzere 16 koşulda incelemiştir. Her koşula karşılık gelen BBT sonucuyla karşılaştırmıştır. BÇAT'ın ilgili BBT'ye karşı performansını değerlendirmek için bir simülasyon çalışması yapılmıştır. Sonuçlarda, BÇAT ve BBT arasındaki ölçümler benzerlik göstermiştir. Bu da adil bir karşılaştırma olduğunu göstermiştir.

Doğruöz (2018) çalışmasında, bireyselleştirilmiş çok aşamalı testlerde (BÇAT) farklı koşulları kendi aralarında ve sınıflandırma testi açısından incelemiştir. Çalışmanın yukarıdan-aşağıya test birleştirme yöntemine göre oluşturulan simülasyonundan elde edilen sonuçlara göre örneklemelere uygulanan BÇAT'ın modül uzunluğu arttığında ve panel deseni iki aşamalıdan üç aşamalıya doğru geçtiğinde ortalama hata değerleri düşmüştür. Modül uzunluğu arttığında panel desenlerinin yanlılığı düşmüştür. Örneklem büyüklüğünün artması tüm koşullarda yanlılığı azaltmıştır. Çalışmanın aşağıdan-yukarıya test birleştirme yöntemine göre elde edilen sonuçlarına baktığımızda küçük ve büyük örneklemelerde modül uzunluğu ve panel deseni arttıkça ortalama hata değerinin azaldığı görülmüştür. Her iki örnekte, kısa ve orta uzunluktaki modülde "1-2", "1-2-2" ve "1-2-3" panel desenlerinden oluşan BÇAT'lar; %30, %50 ve %70 geçme oranlarında oldukça iyi düzeyde bireyleri doğru sınıflandırmıştır.

Lim (2019), araştırmasında bireye uyarlanmış çok aşamalı testlerde (BÇAT) yukarıdan aşağıya test birleştirme yöntemi kullanıldığında en iyi sonuç veren deseni bulmayı amaçlamıştır. Bu amaç doğrultusunda BÇAT'ın optimal tasarımını keşfetmek için; BÇAT'ların oluşturulması, desenlerinin sistematik olarak araştırılması ve performanslarının değerlendirilmesi aşamalarının her birinde etkili bir strateji uyguladığı iki çalışma yapmıştır. İlk aşama için yukarıdan aşağıya yöntemini uygulamıştır. İkinci aşama için,

parametrelenmiş deseni sistematik olarak arařtırmanın bir yolunu kullanmıřtır. Üçüncü ařama için, yeni bir analitik deęerlendirme yöntemi önermiřtir. Birinci çalıřmada, BÇAT için yeni bir analitik deęerlendirme yöntemi önermiřtir. Simülasyon sonuçları, yeni analitik yöntemin bir BÇAT'ın Monte Carlo simülasyon yönteminden daha kesin ölçüm özellikleri ürettiğini göstermiřtir. İkinci çalıřmada ise, belirli bir test içerięi ele alındığında, üç verimli stratejiyi uygulayarak optimal ölçüm özelliklerini gösteren bir BÇAT tasarımı bulmak için bir süreç önermiřtir. BÇAT'ın optimum tasarımlarını oluşturmak için daha karmařık ve gerçekçi test kořulları ile genellenebileceęi önerilmiřtir.

### ***BÇAT Baęlamında Karma Testlerin İncelendięi Arařtırmalar***

Kim vd. (2012) çalıřmalarında sınıflandırma kesinlięi ve test güvenlięi özelinde, karma testlerin kullanıldıęı farklı BÇAT panel desenleri ve BBT kořulları incelemiřlerdir. Simülasyon çalıřmalarında BBT baęlamında iki madde maruz kalma kontrolü, üç geęme oranı ve BÇAT baęlamında birinci ařamada üç test bilgi fonksiyonu seviyesi, üç test bilgi fonksiyonu merkezi ve üç geęme oranı (%40, %50 ve %60) olmak üzere toplam 33 kořul bulunmaktadır. Tüm kořullar geęme oranlarına göre test kesinlięi ve madde maruz kalma kontrolü açısından deęerlendirilmiřtir. Sonuç olarak, 1. ařamadaki test bilgi fonksiyonu seviyeleri arttıkça, simülasyonlar daha fazla sayıda doęru sınıflandırma yapmıřtır. Ayrıca tüm BÇAT kořullarında, BBT maksimum bilgi kořullarına göre daha yüksek test güvenlięi elde edilmiřtir.

Park vd. (2014) çalıřmalarında GKPM'ye göre kalibre ettikleri karma format testten oluřan maddelerle, BÇAT uygulamasında yeni bir havuz kullanım yöntemini arařtırmıřlardır. Madde havuzu, 244'ü iki kategorili, 113'ü üç kategorili ve 67'si dört kategorili olmak üzere 424 maddeden oluřan ulusal bir testin parametrelerine göre üretilmiřtir. Birinci BÇAT uygulamasına dayalı olarak doęrusal programlama (LP) modeline göre, üç yeniden test birleřtirme yapılmıřtır. Her bir test birleřtirme için madde deęiřme oranı 0.22, 0.44 ve 0.66'dir. Sınıflandırma testinde geęme oranları %30, %50 ve

%70 dikkate alınmıştır. Sonuçlar, uygulanan BÇAT yeniden birleştirme koşullarının, istenen BÇAT yapısını korurken, madde havuzu kullanım oranlarını arttırdığını göstermiştir. Tüm koşullar sınıflandırma kararının kesinliği açısından eşit derecede iyi performans göstermiştir.

Kim vd. (2013) kısmi kredi modeline dayalı BÇAT uygulamalarında farklı yönlendirme yöntemleri ile panel tasarımlarını sınıflama testi bağlamında karşılaştırmışlardır. Üç yönlendirme yöntemi, dört panel yapısı, iki test uzunluğu ve üç geçiş oranı olacak şekilde 72 koşuldaki oluşan simülasyonlar yapılmıştır. Sonuç olarak farklı yönlendirme yöntemleri ve BÇAT panel tasarımlarının KKM'ye dayalı madde havuzunu kullanarak iyi performans gösterdiği görülmüştür. Daha uzun test uzunluğu, daha yüksek doğruluk sağlamıştır. %50 geçme oranı en düşük doğruluk seviyesi göstermiştir.

Park vd. (2017) simülasyon yapmadan BÇAT performansını tahmin etmek için BÇAT test bilgisi yönteminin kullanılabilirliğini incelemişlerdir. BÇAT test bilgisi yönteminde, bireylerin yetenek düzeyi seviyelerinde yetenek tahminlerinin standart hataları analitik olarak türetilmektedir. Bu yöntemin geçerliğini test etmek amacıyla elde edilen standart hatalar simülasyon sonuçları ile karşılaştırılmıştır. Ölçüm kesinliğini ve sınıflandırma doğruluğunu karşılaştırmak amacıyla iki farklı simülasyon yapılmıştır. Simülasyonlarda kullanılan madde havuzu 1996 Ulusal Eğitimde İlerleme Değerlendirme (National Assessment of Educational Progress, NAEP) sınavının bilim alt grubundaki maddelerden oluşmaktadır. Kullanılan karma testler genelleştirilmiş kısmi kredi modeline göre kalibre edilmiştir. Sonuç olarak, BÇAT test bilgisi yöntemi BÇAT performansını ölçüm kesinliği ve sınıflandırma bağlamında etkin bir şekilde öngörmüştür.

### **Bölüm 3**

#### **Yöntem**

Bu bölüm, araştırma yöntemi, çalışma grubu, veri toplama süreci, veri toplama araçları ve verilerin analizi olmak üzere beş bölümden oluşmaktadır.

#### **Araştırma Yöntemi**

Araştırmanın amacı, bireye uyarlanmış çok aşamalı karma testlerdeki çok kategorili madde oranlarının yetenek kestirimine etkisini farklı koşullarda incelemektir. Araştırmada veriler, PISA 2018 yılı uygulamasına katılan bireylerin okuma becerileri alanındaki doğru yanlış (iki kategorili) ve açık uçlu (çok kategorili) maddelere ait parametrelere dayalı olarak simülasyon yöntemi ile üretilmiştir. Simülasyon ile araştırma bulgularından ve deneyimlerden yararlanarak üretilmiş sistemleri geliştirilecek yeni koşullar belirlenmiştir. Araştırma bu yönüyle simülasyona dayalı deneysel geliştirme araştırmasıdır. Araştırma ve deneysel geliştirme (AR-GE) araştırmaları; temel araştırma, uygulamalı araştırma ve deneysel geliştirme araştırmaları olmak üzere üç temel faaliyetten oluşur. Araştırma veya deneyimden elde edilen mevcut bilgilerden yararlanarak yeni ürünler ve yeni süreçler üretmeye ya da üretilmiş olanları önemli ölçüde geliştirmeye yönelmiş sistemli araştırmalar deneysel geliştirme araştırmalarıdır (Frascati Kılavuzu, 2002: 78). Sosyal bilimlerde elde edilen bilgiyi, sınama ve değerlendirme amaçları için yürütülen demonstrasyon projeleri deneysel geliştirme araştırmaları kapsamında değerlendirilir. Ayrıca literatürde simülasyon çalışmaları deneysel araştırmalar bağlamında ele alınmaktadır (Feinberg & Rubright, 2016; Harwell, vd.,1996).

#### **Araştırmanın Deseni**

Bu kısımda simülasyon çalışmasında değiştirilen (manipüle edilen) ve sabit tutulan (manipüle edilmeyen veya değiştirilmeyen) koşullar açıklanmıştır.

Çalışmadaki sabit koşullar örneklem büyüklüğü (10.000 birey), bireylerin yetenek seviyeleri dağılımı (normal dağılım,  $N(0,1)$ ) ve BÇAT sonunda hesaplanan yetenek seviyeleri için kullanılan yetenek kestirme metodudur (beklenen sonsal dağılım).

Çalışmadaki manipüle edilen koşullar ise BÇAT panel deseni, toplam test uzunluğu, yönlendirme metodu ve çok kategorili maddelerin toplam test uzunluğundaki oranıdır. Mevcut literatür incelendiğinde gerek simülasyon çalışmalarında gerekse gerçek BÇAT uygulamalarında en sık kullanılan koşullar dikkate alınarak belirlenmiştir.

Panel deseninde; Patsula (1999) "1-3" ve "1-3-3", Zenisky (2004) "1-3-3", "1-2-3", "1-3-2" ve "1-2-2", Sarı ve Raborn (2018) "1-3", "1-2-2", "1-2-3" ve "1-3-3" olmak üzere; "1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3" olacak şekilde 6 farklı panel deseni kullanılmıştır. Test uzunluğunu için Wang (2017) 45 ve 60; Sarı & Raborn (2018) 30 ve 60 test uzunluğu kullanmıştır. Mevcut literatürden yola çıkarak test uzunluğu 20, 40 ve 60 olarak kısa, orta ve uzun test uzunluğunu temsil edecek şekilde 3 farklı test uzunluğu belirlenmiştir.

Yönlendirme metodu olarak; Maksimum Fisher Bilgisi (MFB) ve Doğru Sayısı (DS) olmak üzere 2 yönlendirme yöntemi belirlenmiştir. Çok kategorili madde oranının toplam test uzunluğuna oranı ise %10, %30 ve %50 olarak belirlenmiştir. Örneğin çok kategorili madde oranı %10 olduğunda; toplam test uzunluğu 20 ise 2'si çok kategorili maddelerden, 18'i iki kategorilili maddelerden oluşmuştur. Tüm koşullar birbiriyle çaprazlanarak, çalışma 108 (6x3x2x3) farklı koşul altında incelenmiştir. Tüm koşullar için simülasyon 100 replikasyon olacak şekilde tekrarlanmıştır. Araştırmanın manipüle edilen koşulları Tablo 1'de özetlenmiştir.



**Tablo 1**

## Araştırma Koşulları

Koşul	Seviye Sayısı	Alt Seviyeler
		1-2
		1-3
		1-4
Panel Deseni	6	1-2-2
		1-2-3
		1-3-3
		20 madde
Test Uzunluğu	3	40 madde
		60 madde
Yönlendirme Yöntemi	2	Maksimum Fisher Bilgisi
		Doğru sayısı
		%10
Çok Kategorili Madde Yüzdesi	3	%30
		%50
Toplam	6x3x2x3=108	

## Verilerin Üretilmesi

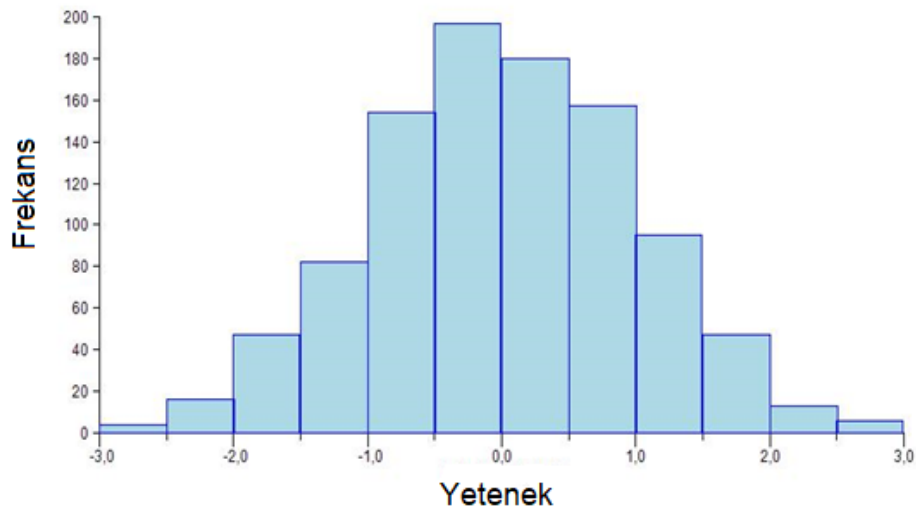
PISA 2018’de okuma becerileri alanında bilgisayar ortamında çok aşamalı test uygulanmıştır. Simülasyonda, bu alandaki iki ve çok kategorili madde parametrelerine dayalı üretilen veriler kullanılmıştır. Verilere Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu’nun (International Association for the Evaluation of Education Achievement, OECD) internet sitesinden erişilmiştir. Bu aşamada WinGen3 (Han, 2007) programı ile 200 çok kategorili ve 400 iki kategorili olacak şekilde madde parametreleri ve 10.000 bireye ait yetenek seviyeleri (teta- $\theta$ ) üretilmiştir. Aşamalar ayrıntılı olarak açıklanmıştır.

### 1.Aşama: Yetenek Seviyelerinin Üretilmesi

İlk aşamada WinGen3 programı ile 10.000 kişilik örnekleme ilişkin yetenek seviyeleri ( $\theta$ ) ortalaması 0, standart sapması 1 olan normal dağılıma göre ( $N(0,1)$ ) üretilmiştir. Yetenek seviyelerinin üretilmesinde 100 iterasyon yapılmıştır. Simülasyonla üretilen bireylerin yetenek düzeylerinde her bir maddeye yanıt verme olasılığı, GKPM’ye göre hesaplanmıştır. 100 iterasyon sonucu üretilen tüm yetenek seviyelerinin histogram grafiği Şekil 4’te verilmiştir.

### Şekil 4

Üretilen Yetenek Seviyelerinin Dağılımı



Şekil 4'te görüldüğü üzere yetenek parametreleri (-3,+3) arasında değişmektedir. Örneklemin geniş bir yetenek seviyesine hitap eden normal dağılım gösterdiği görülmektedir.

## **2.Aşama: Verilerin Elde Edilmesi ve Madde Analizlerinin Yapılması**

Araştırmada PISA 2018 yılı okuma becerileri alanında BÇAT uygulamasında kullanılan madde parametrelerinin özelliklerine dayalı simülasyon çalışması gerçekleştirilmiştir. PISA 2018 Teknik Rapor Final Ek A'ya göre okuma becerileri alanında bilgisayar çok aşamalı testte 72 trend ve 172 yeni olmak üzere 244 madde kullanılmıştır. Bu maddelerden 223'ü iki kategorili ve 21'i çok kategorilidir. Bu bağlamda araştırmada da BÇAT olarak uygulanmış, iki kategorili ve çok kategorili (açık uçlu) maddelerin parametreleri kullanılmıştır. Madde parametrelerinin betimsel istatistikleri ücretsiz bir ara yüz ile çalışan R programı (R Development Core Team, 2018) kullanılarak yapılmıştır. Orijinal maddelerin betimsel istatistikleri Tablo 2'de verilmiştir.

**Tablo 2**

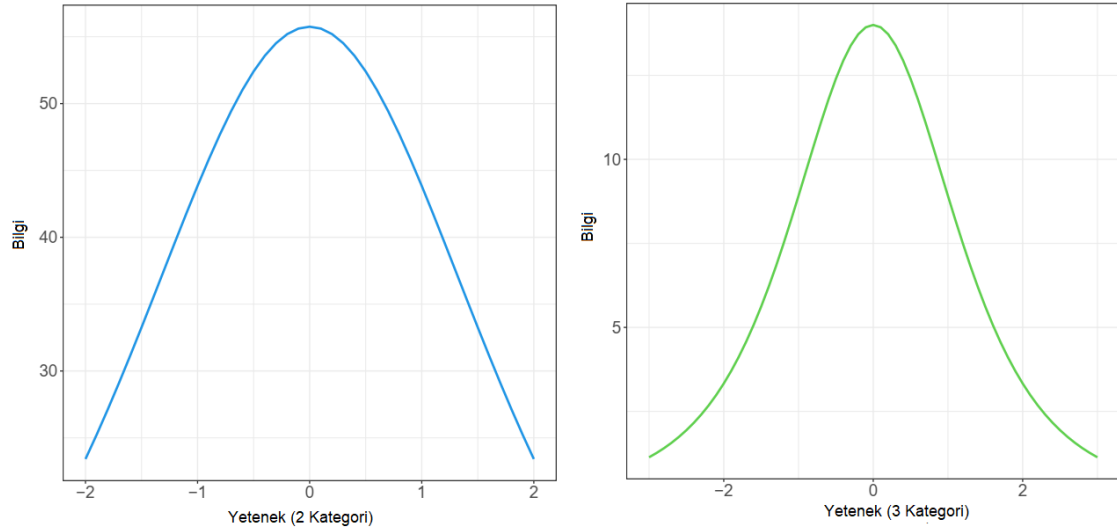
Madde Parametrelerinin Betimsel İstatistikleri

	2 kategori		3 kategori		
	a	b	a	b1	b2
Min.	0.15	-1.91	0.40	-0.40	-2.02
Max.	1.83	2.66	1.25	2.07	2.02
Ortalama	1.01	0.03	0.72	0.65	0.20

PISA 2018 yılı okuma becerileri alanında BÇAT uygulamasında kullanılan 223'ü iki kategorili ve 21'i üç kategorili maddeden oluşan madde havuzlarının test bilgi fonksiyonları Şekil 4'te gösterilmiştir.

## Şekil 4

### PISA Verilerine Ait Test Bilgi Fonksiyonları



### 3.Aşama: Madde Havuzunun Üretilmesi

Bu aşamada kestirilen madde parametreleri kullanılarak madde havuzları üretilmiştir. İki kategorili maddelerin madde parametreleri 2 parametrelili (2 PL) MTK modeli ile, çok kategorili maddeler ise Genelleştirilmiş Kısmi Puan Modeli (GKPM) (Generalized Partial Credit Model - GPCM) ile WinGen3 Versiyon 3.1.15.455 (Han, 2007) programında üretilmiştir. Madde havuzundaki maddelerin parametreleri üretilirken PISA 2018 yılı okuma becerileri alanında kullanılmış madde parametrelerinin dağılımı baz alınmıştır. Madde parametrelerinin betimsel istatistikleri Tablo 2’de verilmiştir.

Literatürde yapılmış BÇAT çalışmalarında madde havuzu büyüklüğü genellikle 200 ile 600 madde arasında değişmektedir (Lim, 2019; Wang, vd., 2016; Xing & Hambleton, 2004; Zheng, vd., 2012). Karma testlerin kullanıldığı BÇAT çalışmalarında ise 424 madde kullanılmıştır (Kim, vd., 2012; Park vd., 2014). Kim, vd. (2012) çalışmalarında 424 maddelik havuzundaki maddelerin 244’ü (% 57.55) iki kategorili, 113’ü (% 26.65) üç kategorili ve 67’si (%15.80) dört kategorili olacak şekildedir.

Bu araştırmada 400’ü iki kategorili ve 200’ü üç kategorili olmak üzere 600 maddelik madde havuzu üretilmiştir. Çok kategorili maddeler uygulanan tüm testin %10, %30 ve

%50'si olacak şekilde panel oluşturma aşamasında modüllere dağıtılmıştır. Üretilen 400 ve 200 maddelik havuzdaki maddelerin betimsel istatistikler Tablo 3'te verilmiştir.

**Tablo 3**

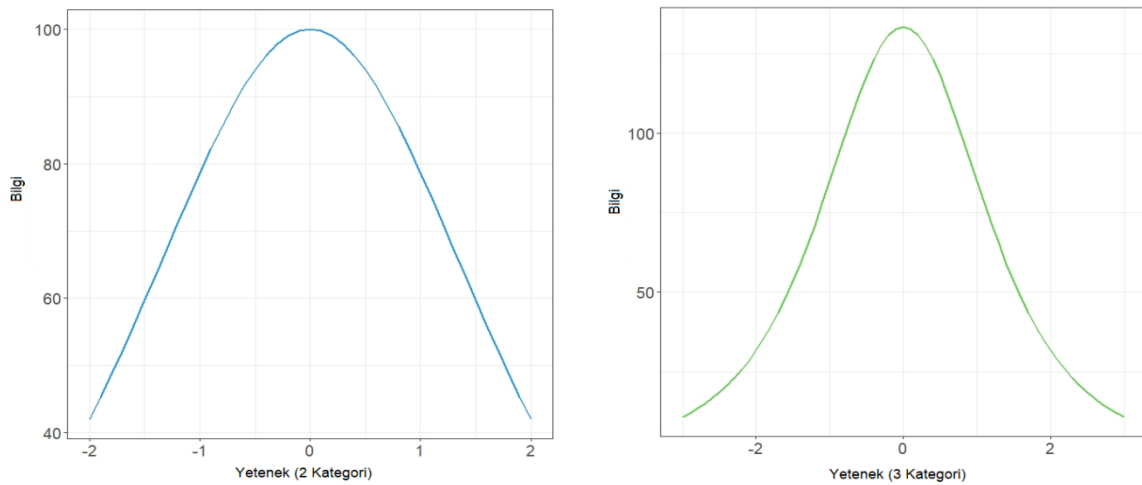
*400 ve 200 Maddelik Havuzdaki Madde Parametrelerinin Betimsel İstatistikleri*

	400 madde		200 madde		
	a	b	a	b1	b2
Min.	0.15	-1.90	0.40	-0.40	-0.34
Max.	1.82	2.65	1.25	1.98	2.07
Ortalama	0.97	0.44	0.85	0.42	1.21

Tablo 2 ve Tablo 3'te gösterilen PISA ve simülasyon verilerine ait betimsel istatistiklerini incelediğimizde, simülasyonla üretilen madde havuzu gerçek veriye göre çok daha fazla olduğu için, istatistiklerde göz ardı edilecek düzeyde farklılıklar olduğu görülmektedir. Üretilen 400 iki kategorili ve 200 üç kategorili maddeden oluşan madde havuzlarının test bilgi fonksiyonları Şekil 5'te verilmiştir.

**Şekil 5**

*Madde Havuzlarına Ait Test Bilgi Fonksiyonları*



Şekil 5'te görüldüğü üzere 400 iki kategorili maddeden oluşan madde havuzu, 200 üç kategorili maddeden oluşan madde havuzundan daha az bilgi vermektedir. Karma testlerin iki kategorili maddelerden oluşan testlere göre daha fazla madde bilgisine sahip olması, testin geçerliğini ve puanların güvenilirliğini arttırmaktadır (Ercikan, vd., 1998). PISA verileri ile simülasyon verilerinin test bilgi fonksiyonlarının dağılımını incelediğimizde, madde sayısı fazla olduğu için simülasyondan elde edilen test bilgi miktarının daha yüksek olduğu görülmektedir.

### **Veri Toplama Araçları**

Veri toplama aracı BÇAT simülasyonu ile oluşturulmuştur. BÇAT simülasyonu modüller ve panellerden oluşmaktadır. 1-2, 1-3, 1-4, 1-2-2, 1-2-3, 1-3-3 olmak üzere toplam altı panel deseni bulunmaktadır. Toplam test uzunluğu 20, 40 ve 60 olmak üzere üç test uzunluğu bulunmaktadır. Modüllerdeki madde sayıları toplam test uzunluğu ve panel desenlerine göre değişkenlik göstermektedir. Tablo 4'te modüllerdeki madde sayıları gösterilmiştir. Modüller ve paneller IBM CPLEX Optimization Studio -20.1.0 versiyon- (ILOG, Inc., 2006) programı ile oluşturulmuştur.

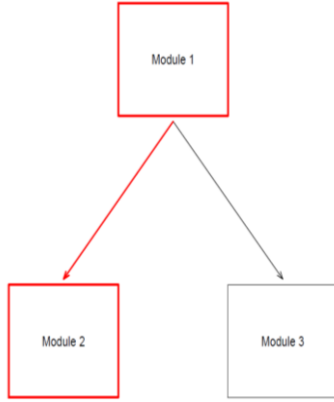
### **BÇAT Simülasyonu**

BÇAT'ta maddelerin farklı desenler için panellere seçilmesi aşamasında yapılan işlemler şu şekildedir. İlk aşamada maddelerin kolay, orta ve zor şeklinde panellere yerleştirilebilmesi için maddelerin bilgi fonksiyonları hesaplanmıştır. Çok kategorili maddeler GKPM, iki kategorili maddeler 2PLM ile hesaplanmıştır.

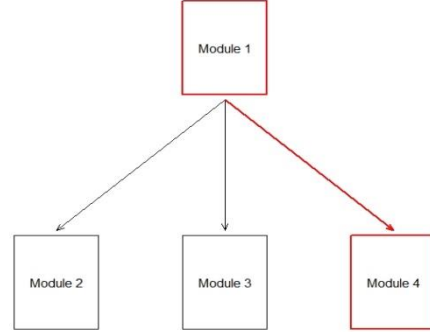
BÇAT simülasyonunda 1-2, 1-3, 1-4, 1-2-2, 1-2-3, 1-3-3 olmak üzere toplam altı adet panel deseni ve belirli yollar bulunmaktadır. Çalışmada kullanılan panel desenleri ve bazı olası yollar Şekil 6'da gösterilmiştir.

**Şekil 6***BÇAT Simülasyonunda Kullanılan Panel Desenleri*

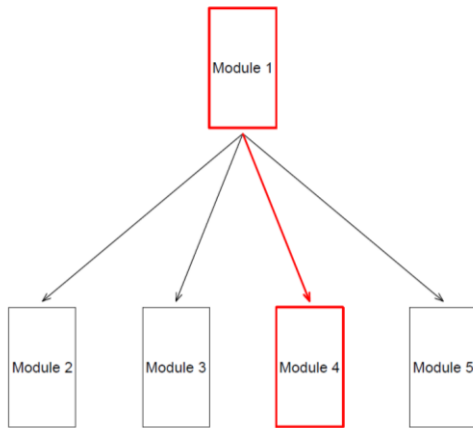
1-2 Panel Deseni



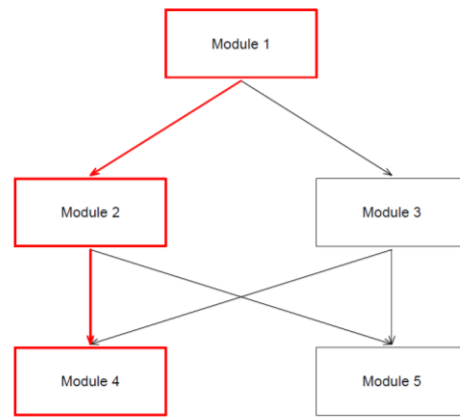
1-3 Panel Deseni



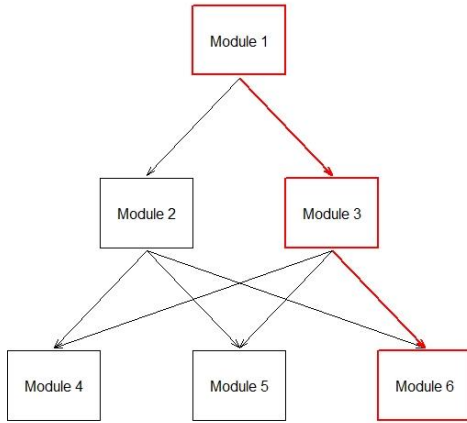
1-4 Panel Deseni



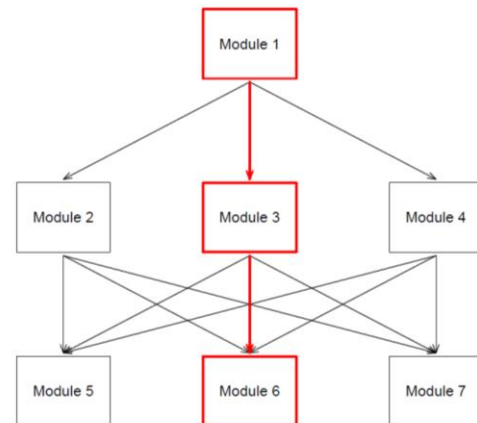
1-2-2 Panel Deseni



1-2-3 Panel Deseni



1-3-3 Panel Deseni



Panellerin oluşturulması ve test derleme işlemi IBM CPLEX optimizasyon programı kullanılarak yapılmıştır. Çalışmada bireylerin panellere ataması rastgele yapılmıştır. Paneller oluşturulurken aşağıdan yukarıya test birleştirme yöntemi kullanılmıştır. Öncelikle modül bilgilerini maksimum yapacak şekilde, CPLEX programında madde bilgi düzeyleri kolay -1, düşük orta -0.3, orta 0, yüksek orta +0.3 ve zor +1 olacak şekilde kodlar yazılmıştır ( $\theta = (-1, -0.3, 0, +0.3, +1)$ ). 1-2, 1-3, 1-2-2, 1-2-3 ve 1-3-3 panel desenleri için;  $\theta=-1$  noktasında yüksek bilgi veren maddeler kolay modüle,  $\theta=0$  noktasında yüksek bilgi veren maddeler orta zorlukta modüle ve  $\theta=1$  noktasında yüksek bilgi veren maddeler zor modüle olacak şekilde belirlenmiştir. 1-4 panel deseni için;  $\theta=-1$  noktasında yüksek bilgi veren maddeler kolay modüle,  $\theta=-0.3$  noktasında yüksek bilgi veren maddeler düşük orta modüle,  $\theta=0$  noktasında yüksek bilgi veren maddeler orta zorlukta modüle,  $\theta=0.3$  noktasında yüksek bilgi veren maddeler yüksek orta zorlukta modüle ve son olarak  $\theta=1$  noktasında yüksek bilgi veren maddeler zor modüle gidecek şekilde belirlenmiştir. Her modülün özelliğine ve sayısına göre en yüksek bilgiyi veren maddeler modüllere atanmıştır. Panellerdeki ve modüllerdeki madde sayıları Tablo 4'te ayrıntılı bir şekilde gösterilmiştir.



**Tablo 4***Modüllerdeki Madde Sayıları*

		Çok kategorili madde oranı									
		%10			%30			%50			
Test											
Uzunluğu		20	40	60	20	40	60	20	40	60	
/ Aşama											
Panel Dizaynı	1-2	1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,
		2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-
			1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15
		1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,
		2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-
			1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15
	1-3	1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,
		2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-
			1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15
		1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,
		2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-
			1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15
1-4	1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,	
	2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-	
		1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15	
	1.aşama	1*/9,	2*/18,	3*/27,	3*/7,	6*/14,	9*/21,	5*/5,	10*/10,	15*/15,	
	2.aşama	1*/9-	2*/18-	3*/27-	3*/7-	6*/14-	9*/21-	5*/5-	10*/10-	15*/15-	
		1*/9	2*/18	3*/27	3*/7	6*/14	9*/21	5*/5	10*/10	15*/15	
1-2-2	1.aşama	0*/6,	0*/12,	0*/20,	2*/4,	4*/8,	6*/14,	4*/2,	6*/6,	10*/10,	

	2.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
	3.aşama	1*/6,	2*/12,	3*/17,	2*/5,	4*/10,	6*/14,	3*/4,	7*/7,	10*/10,
		1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
		1*/6	2*/12	3*/17	2*/5	4*/10	6*/14	3*/4	7*/7	10*/10
		0*/6,	0*/12,	0*/20,	2*/4,	4*/8,	6*/14,	4*/2,	6*/6,	10*/10,
	1.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
1-2-3	2.aşama	1*/6,	2*/12,	3*/17,	2*/5,	4*/10,	6*/14,	3*/4,	7*/7,	10*/10,
	3.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
		1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
		1*/6	2*/12	3*/17	2*/5	4*/10	6*/14	3*/4	7*/7	10*/10
		0*/6,	0*/12,	0*/20,	2*/4,	4*/8,	6*/14,	4*/2,	6*/6,	10*/10,
	1.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
	2.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
1-3-3		1*/6,	2*/12,	3*/17,	2*/5,	4*/10,	6*/14,	3*/4,	7*/7,	10*/10,
		1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
	3.aşama	1*/6-	2*/12-	3*/17-	2*/5-	4*/10-	6*/14-	3*/4-	7*/7-	10*/10-
		1*/6	2*/12	3*/17	2*/5	4*/10	6*/14	3*/4	7*/7	10*/10

\*ile gösterilenler modüllerdeki çok kategorili madde sayısı

Kolay, orta ve zor olarak belirlenen maddelerin seçiminde toplam test uzunluğu ve çok kategorili madde oranı göz önünde bulundurularak maddelerin derlenmesi gerçekleştirilmiştir. Örneğin 1-2 deseni, 60 madde ve %10 çok kategorili madde oranı özel koşulu için,  $\theta=0$  noktasında (orta zorlukta) yüksek bilgi veren %10'u (3 tanesi) çok kategorili olacak şekilde 30 madde seçilip ve ilk aşamadaki yönlendirme modülüne

konulmuştur. Benzer şekilde  $\theta=-1$  noktasında (kolay zorlukta) yüksek bilgi veren %10'u (3 tanesi) çok kategorili olacak şekilde 30 madde seçilip ikinci aşamadaki kolay modüle konulmuştur. Son olarak  $\theta=1$  noktasında (zor düzeyde) yüksek bilgi veren %10'u (3 tanesi) çok kategorili olacak şekilde 30 madde seçilip ikinci aşamadaki zor modüle konulmuştur.

CPLEX programı ile modüller ve paneller oluşturulduktan sonra, BÇAT simülasyonları R programı ile devam etmiştir. BÇAT'a ilişkin simülasyon çalışması 'mstR'(Magis, vd., 2018) R paketi kullanılarak yapılmıştır. Aşamalar sonunda hesaplanan yetenek kestirimleri, toplam madde test bilgisinin hesaplanması ve test derleme işleminde olduğu gibi; bireylerin tüm maddelere verdiği yanıtlar GKPM kullanılarak hesaplanmıştır. Yeteneklerin hesaplanmasında beklenen sonsal dağılım (Expected a Priori-EAP) yöntemi kullanılmıştır. Beklenen sonsal dağılım yönteminde, yeteneklerin seviyesi sonsal dağılım fonksiyonun ortalaması ile elde edilmektedir. Bireyselleştirilmiş bilgisayarlı testlerde, optimal madde havuzu kullanıldığında maksimum olabilirlik ve beklenen sonsal yöntemlerinin benzer düzeyde iyi sonuçlar verdiği görülmektedir. Optimal olmayan test koşullarında ya da madde havuzu sınava girenlerle uygun şekilde eşleşmediğinde, beklenen sonsal dağılım yönteminin kullanılması avantajlı olmaktadır (Gorin, vd., 2005).

BÇAT uygulamalarında çeşitli yönlendirme yöntemleri kullanılmaktadır. Yaygın olarak kullanılan yönlendirme yöntemleri; doğru sayısı (number correct), yetenek düzeyi tahmini ve maksimum Fisher bilgisidir (Weissman, vd., 2007). Bu araştırmada bir aşamadan diğerine geçmek için Maksimum Fisher Bilgisi (MFB) ve Doğru Sayısı (DS) yönlendirme yöntemleri kullanılmıştır. MFB yöntemi en sık kullanılan yönlendirme yöntemlerindedir (Sari & Raborn, 2018). MFB yönlendirme yönteminde bir modülde maksimum miktarda bilgi sağlanarak bir sonraki aşamada uygun modüle yönlendirilir. Birey birinci modüldeki (yönlendirme modülü) maddeleri aldıktan sonra geçici yetenek tahmini yapılır. Daha sonra bu işlem her bir aşamadaki modüller için maksimum bilgi

veren modüller seçilecek şekilde kümülatif olarak devam eder. İki kategorili bir madde için MFB yönlendirme yöntemi aşağıdaki denklem ile gösterilmektedir (Weissman, vd., 2007).

$$FI_j(\theta) = - \sum_{x \in \{0,1\}} f_j(x|\theta) \frac{\partial^2}{\partial \theta^2} \ln f_j(x|\theta), \quad (1)$$

Denklemden;  $f_j(x|\theta)$ , j maddesi için  $\theta$  yetenek seviyesinde x yanıtının olma olasılığı ve  $\hat{\theta}$ , gizil yetenek tahminidir.  $J_b$ , modül b'de bulunan maddelerin dizin kümesini temsil edecek şekilde, yerel bağımsızlık varsayımına göre modül için tahmini FI,  $\hat{\theta}$  tahmininde değerlendirilen maddeler için FI'nın toplamıdır; yani,

$$FI^b = \sum_{j \in J_b} FI_j(\hat{\theta}). \quad (2)$$

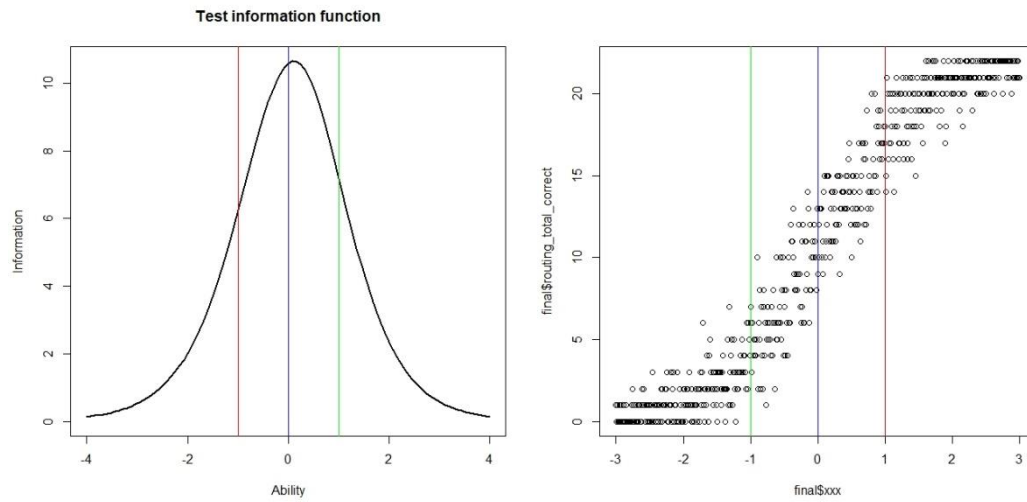
Sıkça kullanılan bir diğer yönlendirme yöntemi ise doğru sayısı (number correct) yönlendirme yöntemidir (Weissman, vd., 2007; Zenisky, vd., 2010). Doğru sayısı, BBT'de MTK yetenek kestiriminde yaygın olarak kullanılan alternatif bir yöntemdir (Armstrong, 2002). Doğru sayısına dayalı yönlendirmede, arka planda bir puan dağılımı elde etmek için MTK modeli kullanılır. Doğru sayısı (DS) yöntemi gerçek doğru sayısı puanlarına göre ya da yetenek seviyelerine göre yapılabilmektedir. Araştırmalar doğru sayısı puanlarına göre uygulanan DS yönlendirme yönteminin etkililiği ile  $\theta$  tahmini için tasarlanmış BÇAT sonuçlarının uyumlu olduğunu göstermektedir (Armstrong, 2002). Doğru sayısı kriteri test uygulayıcıları tarafından BÇAT dizaynı sırasında belirlenir. Doğru sayısı yönteminde her bir modül için kümülatif doğru sayısı puanına (cut score point) bakılarak sınava giren bireyler aşamalar arasında yönlendirilir. Örneğin bu araştırmada 1-3 panel deseni %10 çok kategorili madde olacak şekilde 40 maddelik testte; birinci aşamada (yönlendirme modülü) 20 maddede 5'ten daha az doğru yapanlar 2. aşamada kolay modüle, 5 madde ile

15 madde arasında doğru yapanlar orta modüle, 16 ve daha fazla maddeyi doğru yanıtlayan bireyler ise zor modüle yönlendirilmiştir.

Doğru sayısı yönteminde kesme noktalarındaki doğru sayısını belirlerken, her bir aşamada bireylerin maksimum Fisher bilgisine göre elde ettikleri modül bilgi fonksiyonları ve toplam doğru sayıları bulunmuştur. En yüksek bilgi veren doğru sayıları, modüller arasında kesme noktası olarak belirlenmiştir. Örneğin; 1-3 panel deseni %10 çok kategorili madde olacak şekilde 40 maddelik testte, birinci aşamada bireylerin -1, 0 ve 1 yetenek seviyelerinde modül (test) bilgi fonksiyonları ve bu yetenek düzeyi noktalarında elde ettikleri kümülatif doğru sayıları belirlenmiştir. 2.aşamada kolay, orta ve zor olmak üzere üç modül bulunmaktadır. Grafikler incelenerek kesme puanları 5 ve 16 olarak belirlenmiştir. 2.aşamada 5'ten daha az doğru yapanlar kolay modüle, 5 madde ile 15 madde arasında doğru yapanlar orta modüle, 16 ve daha fazla maddeyi doğru yanıtlayan bireyler ise zor modüle yönlendirilmiştir. 1-3 panel deseni, %10 çok kategorili madde oranı ve 40 maddelik testin yönlendirme modülü bilgisine ve toplam doğru sayılarına ait grafikler Şekil 7'de gösterilmiştir.

## Şekil 7

*1-3 Panel Deseni, %10 Çok Kategorili Madde Oranı ve 40 Maddelik Testin Yönlendirme Modülü Bilgisi ve Toplam Doğru Sayıları*



Testteki çok kategorili madde yüzdesi %10 olduğunda kullanılan kesme noktalarının grafikleri Ek-A'da, %30 olduğunda kullanılan kesme noktalarının grafikleri Ek-B'de ve %50 olduğunda kullanılan kesme noktalarının grafikleri Ek-C'de gösterilmiştir.

mstR'da belirlenmiş kesme puanlarına göre modül seçimi yapılırken, keskomutu, her bir aşamadaki ardışık modül çiftleri arasındaki eşik sayısı kadar satır ve üç sütun içeren bir matris tarafından sağlanmaktadır. Bu matrisin her satırı, önce iki modülü ve ardından kesme puanını (doğru sayısını) temsil eder.  $c(a, b, c)$  şeklinde gösterecek olursak, a ve b ardışık modülleri, c ise doğru sayısını göstermektedir. Yani,  $c(2, 3, 5)$  satırı, 2. ve 3. modüller arasındaki kesme eşiğinin 5 olduğunu gösterir. Örneğin, 1-2 panel deseni %10 çok kategorili madde olacak şekilde 20 maddelik test için kesme puanlarını gösterdiğimizde ( $c(2, 3, 5)$ ), 1.aşama uygulandıktan sonra bir sonraki modülün modül 2 ve 3 arasında seçilmesi gerekmektedir. Kesme puanı yani doğru sayısı 1.modülde 5'ten küçükse modül 2 (kolay), 5'e eşit ve büyükse modül 3 (zor) seçilir.

Aşamalardaki modül sayısı ikiden fazla ise (1-3, 1-4, 1-2-3.. panel deseni gibi) eşik yetenek seviyeleri değerleri belirlenir. Örneğin, 1-3 panel deseni %10 çok kategorili madde olacak şekilde 20 maddelik test için, bu desende toplam dört modül bulunmaktadır. 1.modül 1.aşamada bulunan yönlendirme modülüdür. Diğer üç modül ise 2 (kolay), 3 (orta) ve 4 (zor) numaralı modüllerdir. 2.aşamada 2, 3 ve 4 numaralı modüller arasında kesme puanı belirlenirken, modül 2 ve 3 ile modül 3 ve 4 arasında ayırım yapmak için -1, 0 ve 1 yetenek düzeyi eşik değeri olarak ayarlanmıştır. Bu tasarımda kesme puanları,  $c(2,3,3)$  ve  $c(3,4,9)$  olarak belirlenmiştir. Bu durumda yönlendirme modülünde doğru sayısı 3'ten küçükse modül 2, 9'a eşit ya da büyükse modül 4 ve doğru sayısı 3'ten 8'e kadar ise modül 3 seçilir. Tüm koşullardaki kesme noktaları ise Tablo 5'te verilmiştir.

**Tablo 5***Tüm Koşullardaki Kesme Noktaları*

		Çok kategorili madde yüzdesi								
		%10			%30			%50		
Madde sayısı		20	40	60	20	40	60	20	40	60
Panel deseni	1-2	(2,3,5)	(2,3,11)	(2,3,17)	(2,3,6)	(2,3,13)	(2,3,19)	(2,3,7)	(2,3,14)	(2,3,21)
	1-3	(2,3,3)	(2,3,5)	(2,3,7)	(2,3,3)	(2,3,5)	(2,3,8)	(2,3,3)	(2,3,6)	(2,3,8)
		(3,4,9)	(3,4,16)	(3,4,26)	(3,4,11)	(3,4,20)	(3,4,30)	(3,4,12)	(3,4,24)	(3,4,33)
	1-4	(2,3,2)	(2,3,4)	(2,3,6)	(2,3,3)	(2,3,5)	(2,3,7)	(2,3,3)	(2,3,5)	(2,3,7)
		(3,4,5)	(3,4,12)	(3,4,16)	(3,4,6)	(3,4,13)	(3,4,18)	(3,4,7)	(3,4,13)	(3,4,20)
		(4,5,9)	(4,5,18)	(4,5,25)	(4,5,10)	(4,5,19)	(4,5,30)	(4,5,11)	(4,5,23)	(4,5,32)
	1-2-2	(2,3,4)	(2,3,6)	(2,3,11)	(2,3,4)	(2,3,8)	(2,3,11)	(2,3,5)	(2,3,9)	(2,3,14)
		(4,5,7)	(4,5,15)	(4,5,19)	(4,5,9)	(4,5,16)	(4,5,24)	(4,5,10)	(4,5,19)	(4,5,28)
	1-2-3	(2,3,3)	(2,3,6)	(2,3,10)	(2,3,4)	(2,3,7)	(2,3,12)	(2,3,5)	(2,3,9)	(2,3,13)
		(4,5,3)	(4,5,7)	(4,5,15)	(4,5,4)	(4,5,9)	(4,5,15)	(4,5,5)	(4,5,11)	(4,5,15)
		(5,6,10)	(5,6,20)	(5,6,28)	(5,6,12)	(5,6,23)	(5,6,33)	(5,6,14)	(5,6,26)	(5,6,35)
	1-3-3	(2,3,2)	(2,3,3)	(2,3,5)	(2,3,2)	(2,3,3)	(2,3,4)	(2,3,2)	(2,3,5)	(2,3,5)
		(3,4,5)	(3,4,10)	(3,4,15)	(3,4,6)	(3,4,12)	(3,4,20)	(3,4,7)	(3,4,14)	(3,4,22)
		(5,6,4)	(5,6,10)	(5,6,14)	(5,6,5)	(5,6,10)	(5,6,14)	(5,6,5)	(5,6,10)	(5,6,15)
		(6,7,10)	(6,7,19)	(6,7,27)	(6,7,12)	(6,7,22)	(6,7,33)	(6,7,14)	(6,7,26)	(6,7,38)

## Verilerin Analizi

Çalışmadaki sabit koşullar örneklem büyüklüğü (10.000 birey), bireylerin yetenek seviyeleri dağılımı (normal dağılım,  $N(0,1)$ ), BÇAT sonunda hesaplanan yetenek seviyeleri için kullanılan yetenek kestirme metodudur (beklenen sonsal dağılım, EAP). Çalışmadaki manipüle edilen koşullar ise BÇAT panel dizaynı (1-2,1-3,1-4,1-2-2,1-2-3 ve 1-3-3), toplam test uzunluğu (20, 40 ve 60), yönlendirme metodu (Maksimum Fisher Bilgisi ve doğru sayısı) ve çok kategorili maddelerin toplam test uzunluğundaki oranıdır (%10, %30 ve %50). BÇAT analizleri R programlama dilinde bulunan mstR (Magis, vd., 2018) paketi kullanılarak gerçekleştirilmiştir.

Alt problemlere uygun koşullarda simülasyon verileri elde edildikten sonra elde edilen yetenek kestiriminin kesinliği incelemek amacıyla hata kareleri ortalaması (RMSE), ortalama mutlak yanlılık (mean absolute bias) ve gerçek ve hesaplanan yetenek seviyeleri arasındaki korelasyonlar hesaplanmıştır. Bireylerin yanıtlarını üretmek için kullanılan ve normal dağılımdan üretilen birey yetenek parametreleri, gerçek yetenek parametresi olarak; simülasyon sonunda elde edilen yetenek seviyeleri ise kestirilen yetenek seviyeleri olarak kabul edilmiştir. Yanlılık, gerçek ve kestirilen değer arasındaki farkın ortalamasının kareköküdür. Yanlılık ( $\bar{e}$ ) aşağıdaki gibi formüle edilmektedir:

$$\bar{e} = \sqrt{\frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2}{N}} \quad (3)$$

Ortalama mutlak yanlılık (Mean Absolute Error, MAE) ise, gerçek ve kestirilen değer arasındaki her bir farkın değişmeyen büyüklüğünün mutlak değeridir.

$$MAE = [n^{-1} \sum_{i=1}^n |\hat{\theta}_j - \theta_j|] \quad (4)$$

RMSE, gerçek ve kestirilen değer arasındaki farkın kareleri ortalamasıdır. Hata kareleri ortalaması aşağıdaki gibi formüle edilmektedir.



$$RMSE = \sqrt{\frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2}{N}} \quad (5)$$

Gerçek ( $\theta$ ) ve hesaplanan ( $\hat{\theta}$ ) yetenek seviyeleri arasındaki korelasyon ( $p(\hat{\theta}_j, \theta_j)$ ) ise aşağıdaki gibi formüle edilmektedir.

$$p(\hat{\theta}_j, \theta_j) = \frac{cov(\hat{\theta}_j, \theta_j)}{\sigma_{\hat{\theta}_j} \sigma_{\theta_j}} \quad (6)$$

Yukarıda verilen denklemlerde  $\hat{\theta}_j$  j kişisi için hesaplanan yetenek seviyesini;  $\theta_j$  kişisi için gerçek yetenek seviyesini; N toplam kişi sayısını;  $\sigma_{\hat{\theta}_j}$  ve  $\sigma_{\theta_j}$  ise j kişinin sırasıyla hesaplanan ve gerçek yetenek seviyelerinin standart hata değerlerini temsil etmektedir.

Ortalama mutlak yanlılık ve RMSE değerinin düşük olması, korelasyon değerinin yüksek olması yetenek kestirimlerinin doğruluğunun yüksek olması anlamına gelmektedir. Bulguların değerlendirilmesi için bu değerler tüm iterasyonlar için ayrı ayrı hesaplanmıştır. Her bir koşul için toplam değerler iterasyon sayısına bölünerek ortalama değerler rapor edilmiştir.

Ortalama mutlak yanlılık, RMSE ve korelasyon değerlerini elde ettikten sonra, koşul değişkenleri arasında anlamlı bir fark olup olmadığını belirlemek için her bir koşulun 100 iterasyon sonucu alınarak SPSS'de varyans analizi (ANOVA) yapılmıştır. Faktöryel ANOVA sonucunda elde edilen etki büyüklükleri incelenmiştir. Ortalama mutlak yanlılık, RMSE ve korelasyon değerleri ayrı ayrı bağımlı değişken, diğer tüm koşullar (çok kategorili madde oranı, test uzunluğu, panel deseni ve yönlendirme yöntemi) ise bağımsız değişken olacak şekilde üç farklı faktöryel ANOVA testi yapılmıştır.

## **Bölüm 4**

### **Bulgular ve Yorumlar**

Bu bölümde BÇAT koşullarında elde edilen sonuçlar alt problemler ve genel sonuçlar (ortalama mutlak yanlılık, RMSE ve korelasyon) bağlamında ele alınmıştır. Alt problemlere ve genel bulgulara ait bulgular ve yorumlar alt başlıklarda açıklanmıştır.

#### **Alt Problemlere İlişkin Bulgular ve Yorumlar**

##### ***Alt Problem 1'e İlişkin Bulgular ve Yorumlar***

Alt problem 1'de "Testteki çok kategorili madde oranı %10 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?" sorusuna yanıt aranmıştır.

Bu alt problemde testte bulunan çok kategorili madde oranı %10 olarak sabit tutularak, farklı test uzunluğu, panel desenleri ve yönlendirme yöntemleri koşullarında bireylerin yetenek kestirimlerinin nasıl etkilendiği incelenmiştir. Bireylerin yetenek kestirimlerinin nasıl etkilendiğini incelemek için 36 (1x3x6x2) koşul, 100 iterasyon ile elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri elde edilmiştir. Testte bulunan çok kategorili madde oranı %10 olduğunda tüm koşullarda elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri Tablo 6'da gösterilmiştir. Ayrıca çok kategorili madde oranı %10 iken, tüm koşullarda elde edilen ortalama mutlak yanlılık değerlerine ait grafik Şekil 8, RMSE değerlerine ait grafik Şekil 9 ve korelasyon değerlerinden elde edilen grafik ise Şekil 10'da sırasıyla bulgular yorumlanarak gösterilmiştir.

**Tablo 6**

*Çok Kategorili Madde Oranı %10 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları*

Yönlendirme Yöntemi	Ortalama Mutlak Yanlılık						RMSE						Korelasyon						
	MFB			DS			MFB			DS			MFB			DS			
Test Uzunluğu	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	
Panel Deseni	1-2	0.24	0.18	0.15	0.24	0.18	0.15	0.30	0.23	0.19	0.30	0.23	0.19	0.95	0.97	0.98	0.95	0.97	0.98
	1-3	0.24	0.18	0.15	0.24	0.18	0.16	0.30	0.23	0.19	0.30	0.23	0.20	0.95	0.97	0.98	0.95	0.97	0.98
	1-4	0.24	0.18	0.16	0.24	0.18	0.16	0.30	0.23	0.20	0.30	0.23	0.20	0.95	0.97	0.98	0.95	0.97	0.98
	1-2-2	0.23	0.18	0.15	0.25	0.18	0.16	0.30	0.22	0.19	0.32	0.24	0.20	0.96	0.97	0.98	0.95	0.97	0.98
	1-2-3	0.23	0.18	0.15	0.24	0.18	0.15	0.30	0.22	0.19	0.31	0.23	0.19	0.96	0.98	0.98	0.95	0.97	0.98
	1-3-3	0.23	0.18	0.15	0.24	0.18	0.16	0.30	0.22	0.19	0.31	0.23	0.21	0.96	0.97	0.98	0.95	0.97	0.98

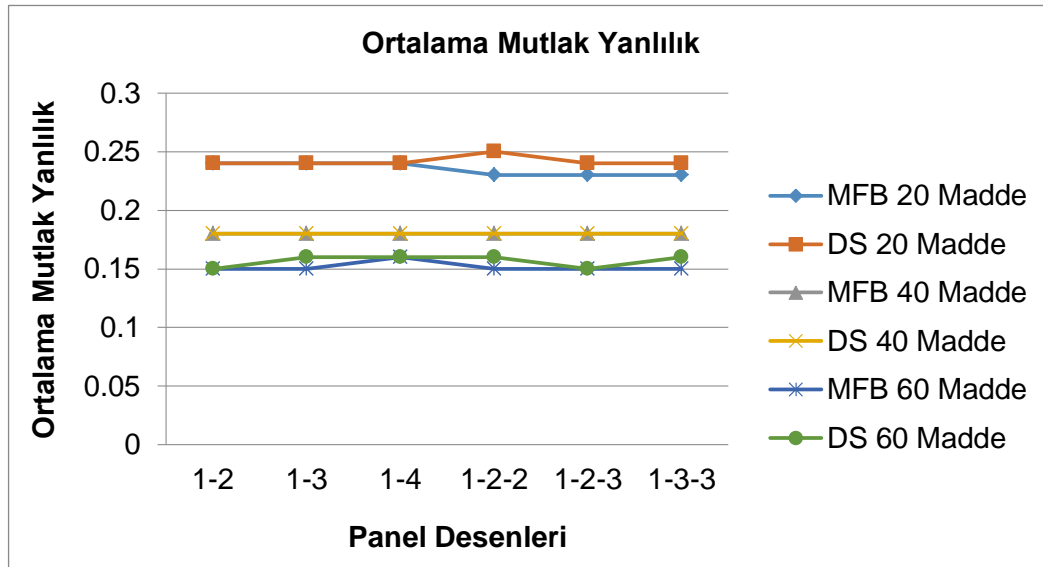
MFB= Maximum Fisher Bilgisi, DS=Doğru sayısı

Tablo 6'yı ortalama mutlak yanlılık açısından incelediğimizde değerlerin 0.15 ve 0.25 arasında değiştiği görülmektedir. MFB yönlendirme yöntemi, 60 maddelik testte, 1-4 panel deseni hariç diğer panel desenleri en düşük ortalama mutlak yanlılık değerine (0.15) sahiptir. Ayrıca DS yönlendirme yöntemi, 60 maddelik testte, 1-2 ve 1-2-3 panel desenleri en düşük değere (0.15) sahiptir. Ortalama mutlak yanlılık bağlamında elde edilen en yüksek değer (0.25) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-2-2 panel deseni koşulunda gözlenmiştir.

Testteki çok kategorili madde oranı %10 olduğunda bireylerin yetenek kestirimlerinin ortalama mutlak yanlılıkları test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 8'de gösterilmiştir.

### Şekil 8

*Testteki Çok Kategorili Madde Oranı %10 İken Ortalama Mutlak Yanlılık Değerleri*



Şekil 8'i incelediğimizde testteki çok kategorili madde oranı %10 iken, madde sayısı arttıkça ortalama mutlak yanlılık değerlerinin düştüğü görülmektedir. En yüksek ortalama mutlak yanlılık değerleri 20 maddelik koşuldakiler, en düşük değerler 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda 20 maddelik testlerde, iki aşamalı panel desenlerinde neredeyse aynı sonuçlar elde edilmiştir. Ayrıca iki aşamalı panel desenlerinden üç aşamalı desenlere gittikçe DS yönlendirme

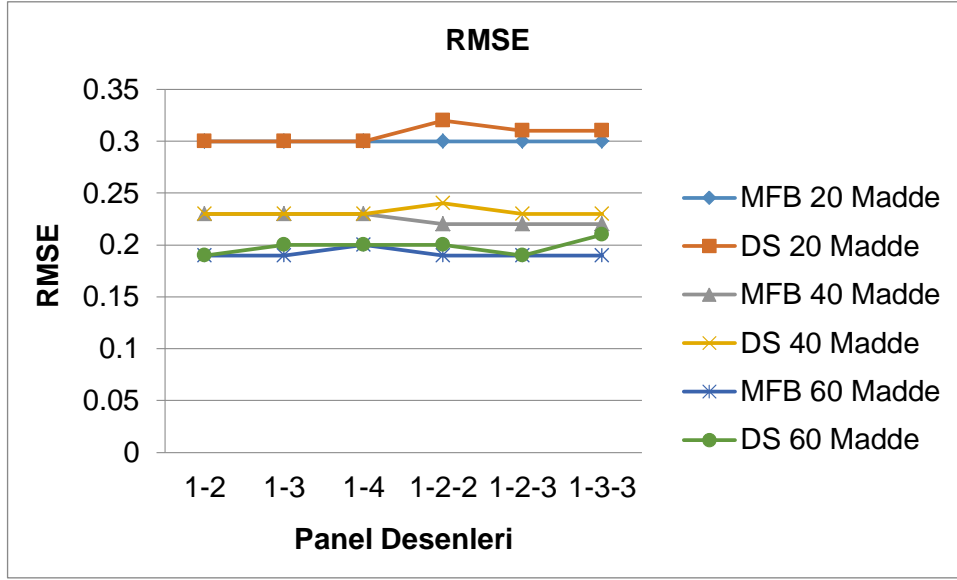
yönteminin daha yüksek ortalama mutlak yanlılık değerleri verdiği görülmüştür. MFB ve DS yönlendirme yöntemlerinin 40 maddelik koşulların tamamında aynı ortalama yanlılık değeri (0.18) elde edilmiştir. Bu da 40 maddelik test koşullarında hangi yönlendirme yönteminin kullanıldığıнын bir önemi olmadığını anlamına gelmektedir. 60 maddelik test koşullarında DS yönlendirme yönteminde nispeten daha yüksek değerler elde edilmiştir. Son olarak panel deseni açısından incelediğimizde MFB yönlendirme yöntemi, 20 maddelik koşullarda üç aşamalı panel desenlerinin (1-2-2, 1-2-3 ve 1-3-3), iki aşamalı desenlere göre daha düşük değerler aldığı görülmektedir. 40 maddelik koşullarda panel deseni sonuçları etkilememiştir. Tüm sonuçlar aynı elde edilmiştir. Diğer panel desenlerinde de benzer sonuçlar elde edilmiştir.

Tablo 6'yı RMSE değerleri açısından incelediğimizde, değerlerin 0.19 ve 0.32 arasında değiştiği görülmektedir. MFB yönlendirme yöntemi, 60 maddelik testte, 1-4 panel deseni hariç diğer panel desenleri ve DS yönlendirme yöntemi, 60 maddelik testte, 1-2 ve 1-2-3 panel desenleri en düşük RMSE değerine (0.19) sahiptir. RMSE değerleri açısından elde edilen en yüksek değer (0.32) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-2-2 panel deseni koşulunda gözlenmiştir. Gözlenen en düşük ve en yüksek değerler ortalama mutlak yanlılık değerleri ile aynı koşullarda elde edilmiştir.

Testteki çok kategorili madde oranı %10 olduğunda yetenek kestirimlerinin RMSE değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 9'da gösterilmiştir.

### Şekil 9

Testteki Çok Kategorili Madde Oranı %10 İken RMSE Değerleri



Şekil 9'u incelediğimizde testteki çok kategorili madde oranı %10 iken, madde sayısı arttıkça ortalama RMSE değerlerinin düştüğü görülmektedir. En yüksek RMSE değerleri 20 maddelik, en düşük değerler 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda 20 maddelik testlerde, iki aşamalı panel desenlerinde neredeyse aynı sonuçlar elde edilmiştir. 40 maddelik testlerde özellikle üç aşamalı panel desenlerinde, DS yönlendirme yönteminde elde edilen sonuçların daha yüksek olduğu görülmektedir. 60 maddelik test koşullarında en düşük RMSE değerleri gözlenirken, DS yönlendirme yönteminde MFB yönlendirme yöntemine göre kısmen daha yüksek değerler elde edilmiştir. Son olarak panel deseni açısından incelediğimizde MFB yönlendirme yönteminin kullanıldığı, 20 maddelik tüm panel desenlerinde aynı RMSE değeri (0.30) elde edilmiştir. Dolayısıyla testteki çok kategorili madde oranı %10 iken, 20 maddelik test koşullarında hangi panel deseninin kullanıldığının bir önemi yoktur. DS yönlendirme yönteminin 20 maddelik test koşullarında, panel desenlerindeki aşama sayısı ikiden üçe çıktığında daha yüksek ortalama RMSE değerleri verdiği görülmüştür.

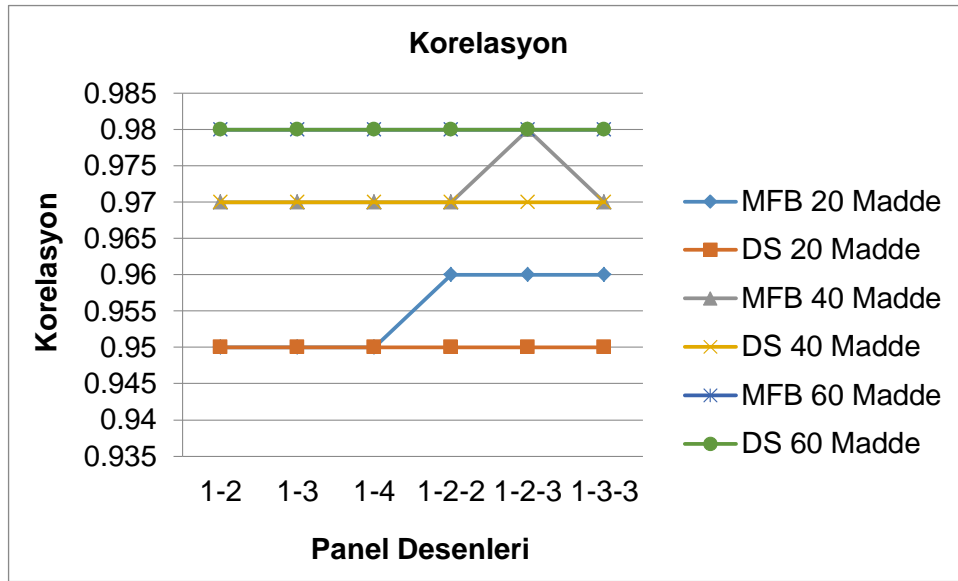
Tablo 6'yı korelasyon değerleri açısından incelediğimizde, değerlerin 0.95 ve 0.98 arasında değiştiği görülmektedir. En düşük korelasyon değeri (0.95), MFB yönlendirme

yöntemi, 20 maddelik testte, iki aşamalı panel desenleri (1-2, 1-3 ve 1-4) ve DS yönlendirme yöntemi, 20 maddelik testte, tüm panel desenlerinde görülmektedir. Elde edilen en yüksek korelasyon değeri (0.98) ise MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde ve MFB yönlendirme yöntemi, 40 maddelik testte 1-2-3 panel deseni koşulunda gözlenmiştir.

Testteki çok kategorili madde oranı %10 olduğunda yetenek düzeylerindeki korelasyon değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 10'da gösterilmiştir.

### Şekil 10

*Testteki Çok Kategorili Madde Oranı %10 İken Korelasyon Değerleri*



Şekil 10'u incelediğimizde testteki çok kategorili madde oranı %10 iken, madde sayısı arttıkça gerçek ve gözlenen yetenek düzeyleri arasındaki korelasyon değerlerinin yükseldiği görülmektedir. En yüksek korelasyon değerleri 60 maddelik, en düşük değerler 20 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda 20 maddelik testlerde, iki aşamalı panel desenlerinde aynı sonuçlar elde edilirken, üç aşamalı panel desenlerinde DS yönlendirme yönteminde daha yüksek korelasyon değerleri elde edilmiştir. 40 maddelik testlerde ise, MFB yönlendirme yöntemi, 1-2-3 panel deseni hariç

(0.98), diğer tüm koşullarda elde edilen değerlerin (0.97) aynı olduğu görülmektedir. 60 maddelik test koşullarının tamamında aynı ve en yüksek korelasyon değerleri (0.98) elde edilmiştir. Panel deseni açısından incelediğimizde, DS yönlendirme yönteminin kullanıldığı test uzunluklarının (20, 40 ve 60) kendi içerisinde tüm panel desenlerinde aynı korelasyon değerleri elde edilmiştir. Yani testteki çok kategorili madde oranı %10 iken, tüm test koşullarında hangi panel deseninin kullanıldığına bir önemi yoktur. MFB yönlendirme yöntemi ve 20 maddelik test uzunluğu koşullarında, panel desenlerindeki aşama sayısı ikiden üçe çıktığında daha yüksek korelasyon değerleri elde edilmiştir. MFB yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde sonuçları aynıdır. Ayrıca MFB ve DS yönlendirme yöntemlerinin 60 maddelik koşulların tamamında aynı korelasyon değeri elde edildiği için bu koşullarda hangi yönlendirme yönteminin kullanıldığına bir önemi bulunmamaktadır.

Özetle, ortalama mutlak yanlılık ve RMSE için en düşük değerler MFB yönlendirme yöntemi, 60 maddelik testte, 1-4 panel deseni hariç diğer panel desenleri ve DS yönlendirme yöntemi, 60 maddelik testte, 1-2 ve 1-2-3 panel desenlerinde elde edilmiştir. Ortalama mutlak yanlılık ve RMSE için en yüksek değerler ise, DS yönlendirme yöntemi, 20 maddelik testte ve 1-2-2 panel deseni koşulunda elde edilmiştir. ANOVA sonuçları incelendiğinde tüm koşullar arasında anlamlı farklılık olduğu Tablo 10 ve Tablo 12'de gösterilmiştir. Korelasyon açısından en düşük değerler MFB yönlendirme yöntemi, 20 maddelik testte, iki aşamalı panel desenleri (1-2, 1-3 ve 1-4) ve DS yönlendirme yöntemi, 20 maddelik testte, tüm panel desenlerinde elde edilmiştir. En yüksek değerler ise MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde ve MFB yönlendirme yöntemi, 40 maddelik testte 1-2-3 panel deseninde elde edilmiştir. MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde aynı çıkmıştır. ANOVA sonuçları incelendiğinde tüm koşullar arasında manidar düzeyde farklılık olduğu Tablo 14'de gösterilmiştir. En düşük ortalama mutlak yanlılık ve RMSE ve en yüksek korelasyon değerlerine sahip olan koşullar en yüksek ölçüm hassasiyetini sağlamıştır.



Bunun tam tersi değerler de en düşük ölçüm hassasiyetine sahip koşullardır. Sonuç olarak çok kategorili madde oranı %10 olduğunda, MFB yönlendirme yöntemi, 60 maddelik testler ve üç aşamalı panel desenlerinin ölçme hassasiyetleri diğer koşullara göre yüksektir. Madde sayısı arttıkça ve panel desenlerindeki aşama sayısı arttıkça daha yüksek ölçme hassasiyeti elde edilmiştir. Yönlendirme yöntemi açısından her iki yöntem de benzer düzeyde yüksek ölçme hassasiyeti göstermiştir.

### ***Alt Problem 2'ye İlişkin Bulgular ve Yorumlar***

Alt problem 2'de "Testteki çok kategorili madde oranı %30 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?" sorusuna yanıt aranmıştır.

Bu alt problemde testte bulunan çok kategorili madde oranı %30 olarak sabit tutularak, farklı test uzunluğu, panel desenleri ve yönlendirme yöntemleri koşullarında bireylerin yetenek kestirimlerinin nasıl etkilendiği incelenmiştir. Bireylerin yetenek kestirimlerinin nasıl etkilendiğini incelemek için 36 (1x3x6x2) koşul, 100 iterasyon ile elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri elde edilmiştir. Testte bulunan çok kategorili madde oranı %30 olduğunda tüm koşullarda elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri Tablo 7'de gösterilmiştir. Elde edilen verilerin grafiksel gösterimleri sırasıyla Şekil 11, Şekil 12 ve Şekil 13'te gösterilmiştir.

Tablo 7'yi ortalama mutlak yanlılık açısından incelediğimizde değerlerin 0.14 ve 0.24 arasında değiştiği görülmektedir. MFB yönlendirme yöntemi, 60 maddelik testte, 1-2-2 panel deseni en düşük ortalama mutlak yanlılık değerine (0.14) sahiptir. Testteki çok kategorili madde oranı %30 olduğunda ortalama mutlak yanlılık bağlamında elde edilen en yüksek değer (0.24) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3, 1-2-2, 1-2-3 ve 1-3-3 panel desenleri koşullarında gözlenmiştir.

**Tablo 7**

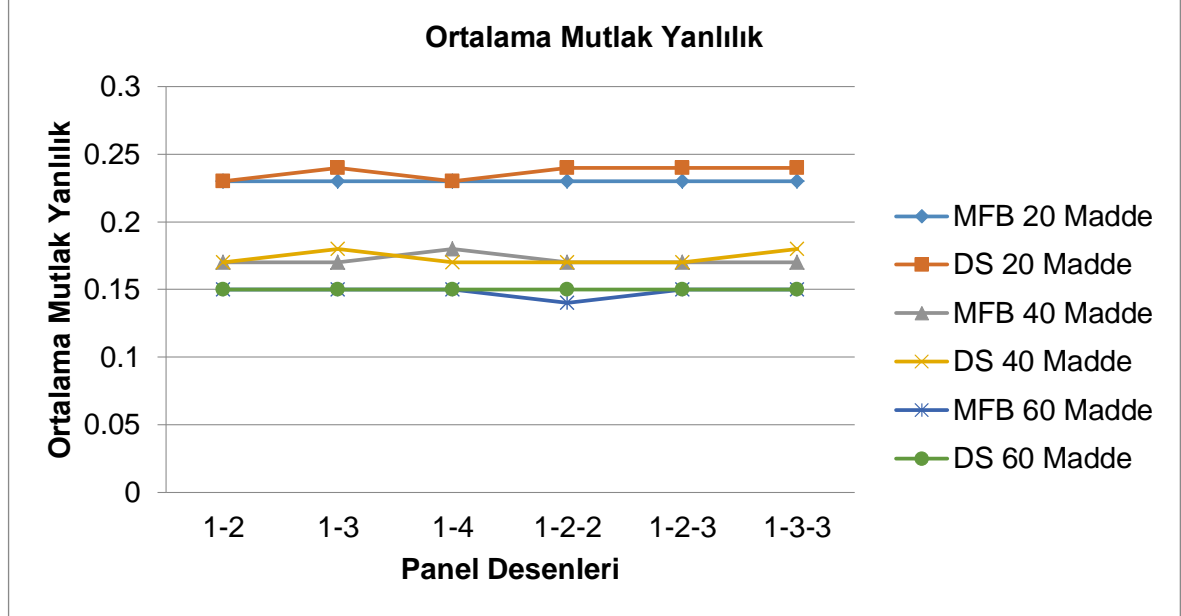
*Çok Kategorili Madde Oranı %30 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları*

Yönlendirme Yöntemi	Ortalama Mutlak Yanlılık						RMSE						Korelasyon						
	MFB			DS			MFB			DS			MFB			DS			
Test Uzunluğu	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	
Panel Deseni	1-2	0.23	0.17	0.15	0.23	0.17	0.15	0.30	0.22	0.19	0.30	0.22	0.19	0.96	0.98	0.98	0.96	0.98	0.98
	1-3	0.23	0.17	0.15	0.24	0.18	0.15	0.30	0.22	0.19	0.30	0.23	0.19	0.95	0.98	0.98	0.95	0.97	0.98
	1-4	0.23	0.18	0.15	0.23	0.17	0.15	0.30	0.22	0.19	0.30	0.22	0.19	0.95	0.97	0.98	0.96	0.98	0.98
	1-2-2	0.23	0.17	0.14	0.24	0.17	0.15	0.29	0.22	0.18	0.31	0.22	0.20	0.96	0.98	0.98	0.95	0.98	0.98
	1-2-3	0.23	0.17	0.15	0.24	0.17	0.15	0.29	0.22	0.18	0.30	0.22	0.19	0.96	0.98	0.98	0.95	0.98	0.98
	1-3-3	0.23	0.17	0.15	0.24	0.18	0.15	0.29	0.22	0.18	0.31	0.22	0.19	0.96	0.98	0.98	0.95	0.97	0.98

Testteki çok kategorili madde oranı %30 olduğunda bireylerin yetenek kestirimlerinin ortalama mutlak yanlılıkları test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 11’de gösterilmiştir.

### Şekil 11

*Testteki Çok Kategorili Madde Oranı %30 İken Ortalama Mutlak Yanlılık Değerleri*



Şekil 11’i incelediğimizde testteki çok kategorili madde oranı %30 iken, madde sayısı arttıkça ortalama mutlak yanlılık değerlerinin düştüğü görülmektedir. En yüksek ortalama mutlak yanlılık değerleri 20 maddelik koşullarda, en düşük değerler 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda DS yönlendirme yöntemi, 20 maddelik testlerde, 1-2 ve 1-4 panel desenlerinde aynı sonuçlar elde edilmiştir (0.23). Diğer panel desenlerinde ise DS yönlendirme yönteminin, MFB yönlendirme yöntemine göre daha yüksek ortalama mutlak yanlılık değerleri verdiği görülmüştür (0.24). Her iki yönlendirme yönteminde 20 maddelik testlerde 1-2 ve 1-4 panel desenlerinde aynı sonuçlar elde edilmiştir. 40 maddelik testlerde, 1-2, 1-2-2 ve 1-2-3 panel desenlerinde MFB ve DS yönlendirme yöntemlerinin her ikisinde de aynı ortalama yanlılık değeri (0.17) elde edilmiştir. Diğer panel desenlerinde ise DS yönlendirme yönteminin daha yüksek ortalama mutlak yanlılık değerleri elde edilmiştir. 60 maddelik test

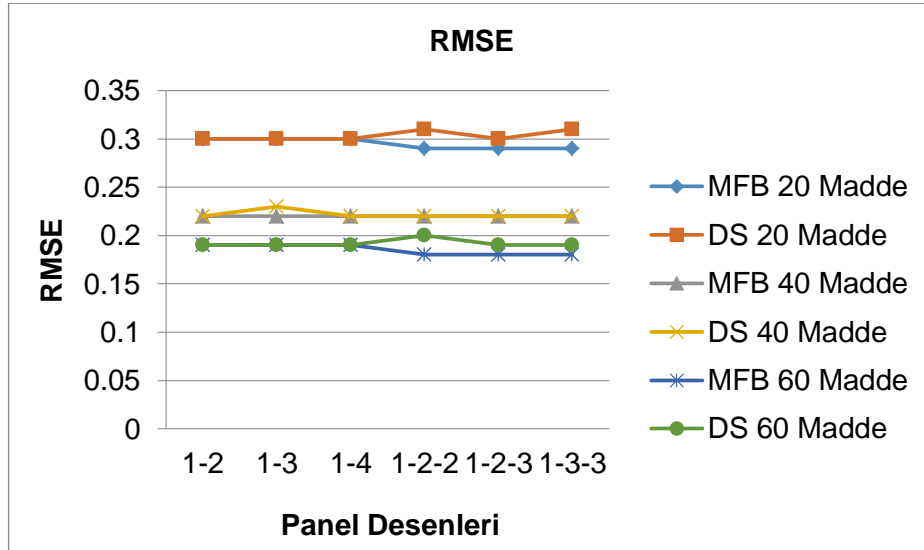
koşullarında 1-2-2 panel deseni hariç diğer tüm koşullarda aynı sonuçlar elde edilmiştir (0.15). 1-2-2 panel deseninde ise MFB yönlendirme yöntemi DS yöntemine nispeten daha iyi sonuç vermiştir (0.14). Son olarak panel deseni açısından incelediğimizde MFB yönlendirme yöntemi, 20 maddelik testte tüm panel desenlerinde aynı sonuçlar elde edilmiştir (0.23). Yani 20 maddelik %30 çok kategorili madde oranı ve MFB yönlendirme yöntemi kullanıldığında hangi panel deseninin kullanıldığının bir önemi yoktur. DS yönlendirme yönteminde 20 maddelik testte 1-2 ve 1-4 panel desenlerinde MFB yöntemi ile aynı sonuçlar elde edilmiştir (0.23). Tüm üç aşamalı 20 maddelik testlerde ve 1-3 panel deseninde ise nispeten daha yüksek ortalama mutlak yanlılık değeri elde edilmiştir (0.24). 40 maddelik testte, MFB yönlendirme yönteminde 1-4 panel deseni hariç (0.18), diğer tüm panel desenlerinin aynı ortalama mutlak yanlılık değerleri verdiği görülmüştür (0.17). 40 maddelik testte, DS yönlendirme yönteminde 1-3 ve 1-3-3 panel desenlerinde aynı değer elde edilmiştir (0.18). Diğer panel desenlerinde ise nispeten daha düşük olacak şekilde aynı ortalama mutlak yanlılık değerleri elde edilmiştir (0.17). 60 maddelik testlerde, yönlendirme yöntemi fark etmeksizin (MFB ve DS), MFB 1-2-2 panel deseni (0.14) hariç tüm koşullarda aynı sonuçlar elde edilmiştir (0.15). 60 maddelik testlerde, testteki çok kategorili madde oranı %30 iken, panel desenleri ve yönlendirme yöntemleri bir koşul hariç aynı sonuçları vermiştir. Yani çok kategorili madde oranı %30 iken herhangi bir koşul tercih edilebilir.

Tablo 7'yi RMSE değerleri açısından incelediğimizde, değerlerin 0.18 ve 0.31 arasında değiştiği görülmektedir. MFB yönlendirme yöntemi, 60 maddelik testte, üç aşamalı panel desenleri (1-2-2, 1-2-3 ve 1-3-3) en düşük RMSE değerine (0.18) sahiptir. RMSE değerleri açısından elde edilen en yüksek değer (0.31) ise DS yönlendirme yöntemi, 20 maddelik testte, 1-2-2 ve 1-3-3 panel desenleri koşullarında gözlenmiştir.

Testteki çok kategorili madde oranı %30 olduğunda yetenek kestirimlerinin RMSE değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 12’de gösterilmiştir.

### Şekil 12

Testteki Çok Kategorili Madde Oranı %30 İken RMSE Değerleri



Şekil 12’yi incelediğimizde testteki çok kategorili madde oranı %30 iken, madde sayısı arttıkça ortalama RMSE değerlerinin düştüğü görülmektedir. En yüksek RMSE değerleri 20 maddelik, en düşük değerler 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından 20 maddelik testlerde, iki aşamalı panel desenlerinde her iki yönlendirme yönteminde de aynı sonuçlar elde edilmiştir (0.30). 40 maddelik testlerde DS yönlendirme yöntemi 40 madde 1-3 panel deseni (0.23) hariç elde edilen sonuçların birbirine eşit olduğu görülmektedir (0.22). Yani 40 maddelik testlerde 1-3 panel deseni hariç hangi yönlendirme yöntemin seçildiğinin bir önemi bulunmamaktadır. 60 maddelik test koşullarında en düşük RMSE değerleri gözlenmiştir. İki aşamalı panel desenlerinde MFB ve DS yönlendirme yöntemlerinde aynı değerler elde edilmiştir (0.19). Üç aşamalı panel desenlerinde ise DS yönlendirme yönteminin daha yüksek RMSE değerleri elde edilmiştir. Panel deseni açısından incelediğimizde MFB yönlendirme yöntemi ve 20 maddelik testlerde iki aşamalı panel desenleri ve üç aşamalı panel desenleri kendi

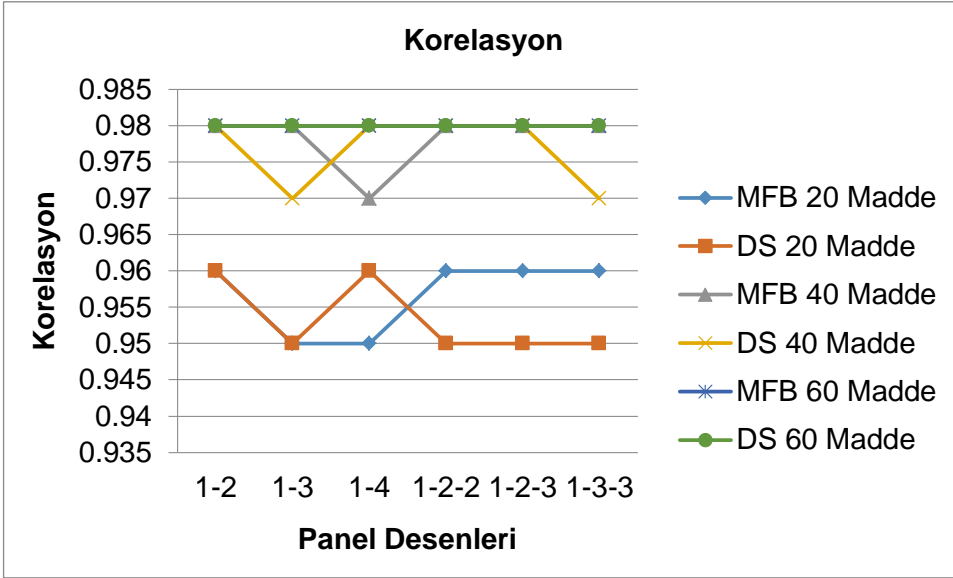
içlerinde aynı RMSE değerleri vermiştir. İki aşamalıdan üç aşamalıya geçtikçe daha düşük değerler elde edilmiştir. DS yönlendirme yöntemi ve 20 maddelik testlerde ise iki aşamalı panel desenlerinde MFB yöntemi iki aşamalı testlerle aynı sonuçlar elde edilirken (0.30), üç aşamalı panel desenlerinde daha yüksek sonuçlar elde edilmiştir. MFB yönlendirme yönteminin kullanıldığı, 40 maddelik tüm panel desenlerinde aynı RMSE değeri (0.22) elde edilmiştir. Dolayısıyla testteki çok kategorili madde oranı %30 iken, 40 maddelik test koşullarında hangi yönlendirme yönteminin ve panel deseninin kullanıldığına bir önemi yoktur. 60 maddelik testlerde ise MFB yönlendirme yönteminde üç aşamalı panel desenlerinde iki aşamalı panel desenlerine göre daha düşük RMSE değerleri elde edilmiştir. DS yönlendirme yönteminde ise 1-2-2 panel deseni (0.20) hariç diğer tüm desenlerde aynı sonuç elde edilmiştir (0.19). Yani 60 maddelik testlerde, %30 çok kategorili madde oranı ve DS yönlendirme yöntemi kullanıldığında 1-2-2 panel deseni hariç hangi panel deseninin seçildiğinin önemi bulunmamaktadır.

Tablo 7'yi korelasyon değerleri açısından incelediğimizde, değerlerin 0.95 ve 0.98 arasında değiştiği görülmektedir. En düşük korelasyon değeri (0.95), MFB yönlendirme yöntemi, 20 maddelik testte, iki aşamalı panel desenleri (1-3 ve 1-4) ve DS yönlendirme yöntemi, 20 maddelik testte, 1-3 panel deseni ve üç aşamalı tüm panel desenlerinde görülmektedir. Elde edilen en yüksek korelasyon değeri (0.98) ise MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde, MFB yönlendirme yöntemi, 40 maddelik testte 1-4 panel deseni hariç tüm koşullarda ve DS yönlendirme yöntemi, 40 maddelik testte 1-3 ve 1-3-3 panel desenleri hariç diğer tüm koşullarda gözlenmiştir.

Testteki çok kategorili madde oranı %30 olduğunda yetenek düzeylerindeki korelasyon değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 13'te gösterilmiştir.

**Şekil 13**

*Testteki Çok Kategorili Madde Oranı %30 İken Korelasyon Değerleri*



Şekil 13'ü incelediğimizde testteki çok kategorili madde oranı %30 iken, testteki madde sayısı arttıkça gerçek ve gözlenen yetenek seviyeleri arasındaki korelasyon değerlerinin yükseldiği görülmektedir. En yüksek korelasyon değerleri (0.98) 60 maddelik testlerde ve 40 maddelik testlerin bazı panel koşullarında, en düşük korelasyon değerleri (0.95) ise 20 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda 20 maddelik testlerde, MFB yönlendirme yönteminde iki aşamalı panel desenlerinden (0.95) üç aşamalıya geçtikçe korelasyonun arttığı görülmektedir (0.96). DS yönlendirme yönteminde ise iki aşamalı panel desenlerinden (0.96) üç aşamalı panel desenlerine geçtikçe korelasyonun nispeten azaldığı görülmektedir (0.95). Her iki yönlendirme yönteminde de üç aşamalı panel desenlerinde korelasyon değerleri kendi içlerinde aynıdır. 40 maddelik testlerde ise, MFB yönlendirme yönteminde, 1-4 panel deseni hariç (0.97), diğer tüm koşullarda elde edilen değerlerin (0.98) aynı olduğu görülmektedir. DS yönlendirme yönteminde ise 1-3 ve 1-3-3 panel deseni (0.97) hariç diğer tüm desenler aynı sonucu vermiştir (0.98). 40 maddelik testlerde MFB yönlendirme yönteminin nispeten daha iyi sonuçlar verdiği görülmektedir. 60 maddelik test koşullarında her iki yönlendirme yönteminde de aynı ve en yüksek korelasyon değerleri (0.98) elde

edilmiştir. Yani 60 maddelik testlerde her iki yönlendirme yöntemi de eşit düzeyde iyi çalışmıştır. Panel deseni açısından incelediğimizde, MFB yönlendirme yöntemi ve 20 maddelik test uzunluğu koşullarında, panel desenlerindeki aşama sayısı ikiden üçe çıktığında daha yüksek korelasyon değerleri elde edilmiştir. 40 ve 60 maddelik testlerde ise 40 madde 1-4 panel deseni hariç tüm panel desenlerinin sonuçları aynıdır. DS yönlendirme yöntemine baktığımızda test uzunluğu 20 olan testlerde iki aşamalı panel desenlerinden üç aşamalı panel desenlerine geçtikçe daha düşük değerler elde edilmiştir. 40 maddelik testlerde 1-3 ve 1-3-3 panel deseni hariç aynı sonuçlar elde edilmiştir. 60 maddelik testlerde ise tüm panel desenlerinde aynı korelasyon değerleri elde edilmiştir. Yani testteki çok kategorili madde oranı %30 iken, 60 maddelik testlerde DS yönlendirme yöntemi kullanıldığında hangi panel deseninin kullanıldığının bir önemi yoktur.

Özetle çok kategorili madde oranı %30 olduğunda, ortalama mutlak yanlılık açısından incelediğimizde en düşük ortalama mutlak yanlılık değeri (0.14) MFB yönlendirme yöntemi, 60 maddelik testte, 1-2-2 panel deseninde gözlenmiştir. En yüksek ortalama mutlak yanlılık değeri (0.24) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3, 1-2-2, 1-2-3 ve 1-3-3 panel desenleri koşullarında gözlenmiştir. RMSE değerleri açısından incelediğimizde, MFB yönlendirme yöntemi, 60 maddelik testte, üç aşamalı panel desenleri (1-2-2, 1-2-3 ve 1-3-3) en düşük RMSE değerine (0.18) sahiptir. en yüksek değer (0.31) ise DS yönlendirme yöntemi, 20 maddelik testte, 1-2-2 ve 1-3-3 panel desenleri koşullarında gözlenmiştir. Korelasyon değerleri açısından incelediğimizde ise, en düşük korelasyon değeri (0.95) MFB yönlendirme yöntemi, 20 maddelik testte, iki aşamalı panel desenleri (1-3 ve 1-4) ve DS yönlendirme yöntemi, 20 maddelik testte, 1-3 panel deseni ve üç aşamalı tüm panel desenlerinde görülmektedir. Elde edilen en yüksek korelasyon değeri (0.98) ise MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde, MFB yönlendirme yöntemi, 40 maddelik testte 1-4 panel deseni hariç tüm koşullarda ve DS yönlendirme yöntemi, 40 maddelik testte 1-3 ve 1-3-3 panel desenleri hariç diğer tüm koşullarda gözlenmiştir. ANOVA sonuçları incelendiğinde



ortalama mutlak yanlılık, RMSE ve korelasyon değerleri için tüm koşullar arasında anlamlı farklılık olduğu Tablo 10, Tablo 12 ve Tablo 14'te gösterilmiştir. Ortalama mutlak yanlılık ve RMSE değerleri ne kadar düşükse ve korelasyon değerleri ne kadar yüksekse, ilgili koşulların ölçüm hassasiyeti o kadar yüksek demektir. Bunun tam tersi değerler de, en düşük ölçüm hassasiyetine sahip koşullardır. Sonuç olarak %30 çok kategorili madde oranı içeren test koşullarında ortalama mutlak yanlılık ve RMSE değerlerine baktığımızda, MFB yönlendirme yöntemi, 60 maddelik testler ve 1-2-2 panel deseni başta olmak üzere üç aşamalı panel desenlerinin ölçüm hassasiyetlerinin çok yüksek olduğu görülmüştür. Korelasyon değerleri açısından baktığımızda, MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde en yüksek değerler elde edildiği için ölçüm hassasiyetleri aynıdır.

### ***Alt Problem 3'e İlişkin Bulgular ve Yorumlar***

Alt problem 3'te "Testteki çok kategorili madde oranı %50 olduğunda bireylerin yetenek kestirimleri test uzunluğuna (20, 40 ve 60), panel desenlerine ("1-2", "1-3", "1-4", "1-2-2", "1-2-3" ve "1-3-3") ve yönlendirme yöntemlerine (Maksimum Fisher Bilgisi ve doğru sayısı) göre nasıl değişmektedir?" sorusuna yanıt aranmıştır.

Bu alt problemde testte bulunan çok kategorili madde oranı %50 olarak sabit tutularak, farklı test uzunluğu, panel desenleri ve yönlendirme yöntemleri koşullarında bireylerin yetenek kestirimlerinin nasıl etkilendiği incelenmiştir. Bireylerin yetenek kestirimlerinin nasıl etkilendiğini incelemek için 36 (1x3x6x2) koşul, 100 iterasyon ile elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri elde edilmiştir. Testte bulunan çok kategorili madde oranı %50 olduğunda tüm koşullarda elde edilen ortalama mutlak yanlılık, RMSE ve korelasyon değerleri Tablo 8'de gösterilmiştir. Elde edilen verilerin grafiksel gösterimleri sırasıyla Şekil 14, Şekil 15 ve Şekil 16'da gösterilmiştir.

Tablo 8'i ortalama mutlak yanlılık açısından incelediğimizde değerlerin 0.14 ve 0.25 arasında değiştiği görülmektedir. En düşük ortalama mutlak yanlılık değeri (0.14) 60

maddelik testlerde; MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında elde edilmiştir. Testteki çok kategorili madde oranı %50 olduğunda ortalama mutlak yanlılık bağlamında elde edilen en yüksek değer (0.25) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3-3 panel deseni koşulunda gözlenmiştir.

**Tablo 8**

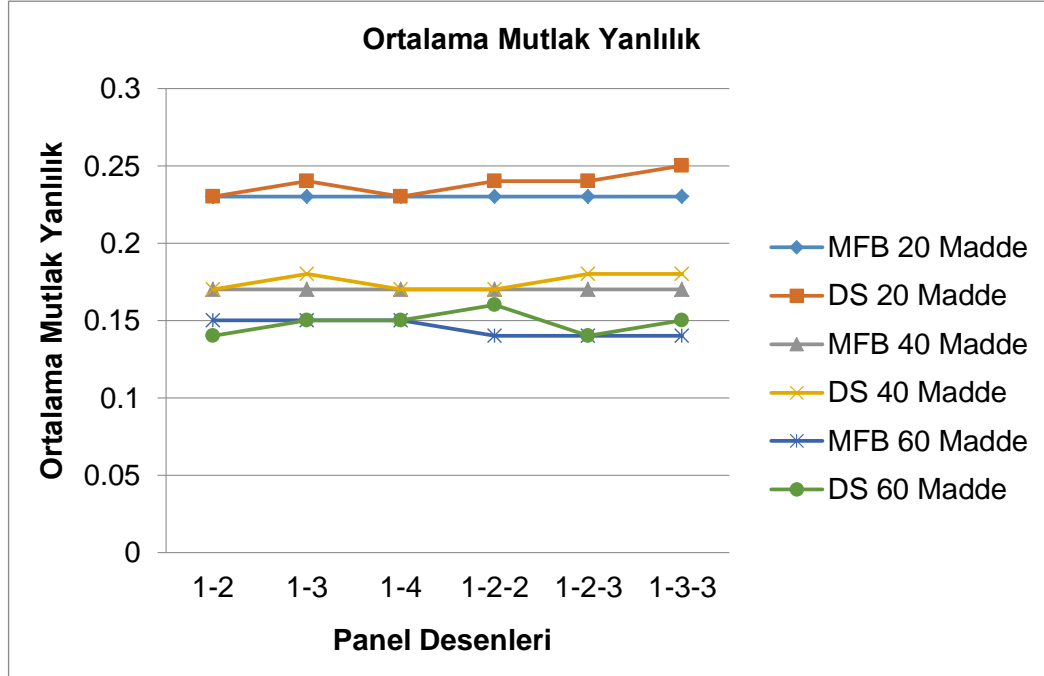
*Çok Kategorili Madde Oranı %50 Olduğunda Tüm Koşullarda Elde Edilen Ortalama Mutlak Yanlılık, RMSE ve Korelasyon Sonuçları*

Yönlendirme Yöntemi	Ortalama Mutlak Yanlılık						RMSE						Korelasyon						
	MFB			DS			MFB			DS			MFB			DS			
Test Uzunluğu	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	20	40	60	
Panel Deseni	1-2	0.23	0.17	0.15	0.23	0.17	0.14	0.29	0.22	0.19	0.29	0.22	0.18	0.96	0.98	0.98	0.96	0.98	0.98
	1-3	0.23	0.17	0.15	0.24	0.18	0.15	0.29	0.22	0.19	0.30	0.22	0.19	0.96	0.98	0.98	0.95	0.97	0.98
	1-4	0.23	0.17	0.15	0.23	0.17	0.15	0.29	0.22	0.19	0.29	0.22	0.19	0.96	0.98	0.98	0.96	0.98	0.98
	1-2-2	0.23	0.17	0.14	0.24	0.17	0.16	0.29	0.22	0.18	0.31	0.22	0.20	0.96	0.98	0.98	0.95	0.98	0.98
	1-2-3	0.23	0.17	0.14	0.24	0.18	0.14	0.29	0.22	0.18	0.30	0.22	0.18	0.96	0.98	0.98	0.95	0.97	0.98
	1-3-3	0.23	0.17	0.14	0.25	0.18	0.15	0.29	0.22	0.18	0.32	0.23	0.19	0.96	0.98	0.98	0.95	0.97	0.98

Testteki çok kategorili madde oranı %50 olduğunda yetenek kestirimlerinin ortalama mutlak yanlılıkları test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 14'te gösterilmiştir.

#### Şekil 14

*Testteki Çok Kategorili Madde Oranı %50 İken Ortalama Mutlak Yanlılık Değerleri*



Şekil 14'ü incelediğimizde testteki çok kategorili madde oranı %50 iken, madde sayısı arttıkça ortalama mutlak yanlılık değerlerinin düştüğü görülmektedir. En yüksek ortalama mutlak yanlılık değerleri 20 maddelik koşullarda, en düşük değerler ise 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından baktığımızda 20 maddelik testlerde, MFB yönlendirme yönteminin genel olarak DS yönlendirme yöntemine göre daha düşük ortalama mutlak yanlılık değerleri aldığı görülmüştür. 1-2 ve 1-4 panel desenlerinde aynı sonuçlar elde edilmiştir (0.23). Diğer panel desenlerinde ise DS yönlendirme yönteminin daha yüksek ortalama mutlak yanlılık değerleri verdiği görülmüştür. 40 maddelik testlerde, 1-3, 1-2-3 ve 1-3-3 panel desenleri (0.18) hariç, MFB ve DS yönlendirme yöntemlerinin her ikisinde de aynı ortalama yanlılık değeri (0.17) elde edilmiştir. Diğer panel desenlerinde ise DS yönlendirme yönteminde daha yüksek

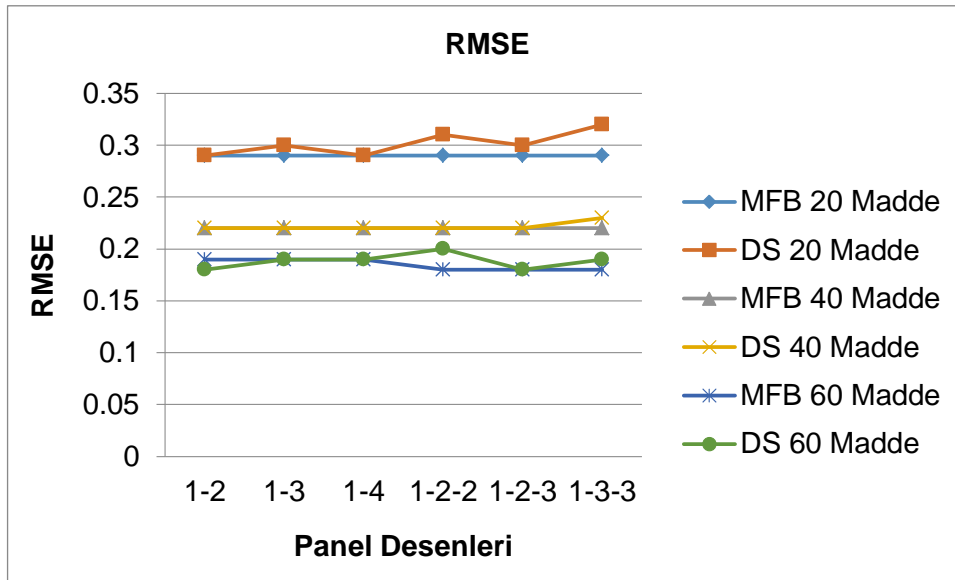
ortalama mutlak yanlılık değerleri elde edilmiştir. 60 maddelik test koşullarında MFB yönlendirme yöntemi, DS yöntemine nispeten daha iyi sonuç vermiştir. 1-3, 1-4 ve 1-2-3 panel desenlerinde her iki yönlendirme yöntemi de aynı sonuçları göstermiştir. 1-3 ve 1-4 panel deseninde ortalama mutlak yanlılık 0.15 iken 1-2-3 panel deseninde 0.14 elde edilmiştir. Panel deseni açısından incelediğimizde MFB yönlendirme yöntemi, 20 maddelik testte tüm panel desenlerinde aynı sonuçlar elde edilmiştir (0.23). Yani 20 maddelik testte, %50 çok kategorili madde oranı ve MFB yönlendirme yöntemi kullanıldığında hangi panel deseninin kullanıldığına bir önemi yoktur. DS yönlendirme yönteminde 20 maddelik testte üç aşamalı panel desenlerinde, iki aşamalı panel desenlerine nispeten daha yüksek ortalama mutlak yanlılık değerleri elde edilmiştir. 40 maddelik testte, MFB yönlendirme yönteminde tüm panel desenlerinin aynı ortalama mutlak yanlılık değerleri verdiği görülmüştür (0.17). 40 maddelik testte, DS yönlendirme yönteminde 1-3, 1-2-3 ve 1-3-3 panel desenlerinde aynı değer elde edilmiştir (0.18). Diğer panel desenlerinde ise aynı ortalama mutlak yanlılık değerleri elde edilmiştir (0.17). 60 maddelik testlerde, MFB yönlendirme yönteminde iki aşamalı panel desenlerinden üç aşamalı panel desenlerine gittikçe ortalama mutlak yanlılık değerlerinin düştüğü görülmüştür.

Tablo 8'i RMSE değerleri açısından incelediğimizde, değerlerin 0.18 ve 0.32 arasında değiştiği görülmektedir. En düşük RMSE değeri (0.18) 60 maddelik testlerde; MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında elde edilmiştir. RMSE değerleri açısından elde edilen en yüksek değer (0.32) ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3-3 panel deseni koşulunda gözlenmiştir.

Testteki çok kategorili madde oranı %50 olduğunda yetenek kestirimlerinin RMSE değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 15'te gösterilmiştir.

**Şekil 15**

*Testteki Çok Kategorili Madde Oranı %50 İken RMSE Değerleri*



Şekil 15'i incelediğimizde testteki çok kategorili madde oranı %50 iken, madde sayısı arttıkça ortalama RMSE değerlerinin düştüğü görülmektedir. En yüksek RMSE değerleri 20 maddelik, en düşük değerler 60 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından 20 maddelik testlerde, DS yönlendirme yönteminde daha yüksek değerler elde edilmiştir. Her iki yöntemde 1-2 ve 1-4 panel desenlerinde aynı sonuç elde edilmiştir (0.29). 40 maddelik testlerde her iki yöntemde, DS yönlendirme yöntemi 1-3-3 panel deseni (0.23) hariç elde edilen sonuçların birbirine eşit olduğu görülmektedir (0.22). Yani 40 maddelik testlerde 1-3-3 panel deseni hariç hangi yönlendirme yöntemin seçildiğinin bir önemi bulunmamaktadır. 60 maddelik test koşullarında DS yönlendirme yönteminde nispeten daha yüksek değerler elde edilmiştir. Panel deseni açısından incelediğimizde MFB yönlendirme yöntemi ve 20 maddelik testlerde tüm panel desenleri kendi içlerinde aynı RMSE değerleri vermiştir. DS yönlendirme yöntemi ve 20 maddelik testlerde ise 1-2 ve 1-4 panel desenlerinde MFB yönlendirme yöntemi ile aynı sonuçlar elde edilmiştir (0.29). Üç aşamalı panel desenlerinde ise MFB'ye göre daha yüksek sonuçlar elde edilmiştir. MFB yönlendirme yönteminin kullanıldığı, 40 maddelik tüm panel desenlerinde aynı RMSE değeri (0.22) elde

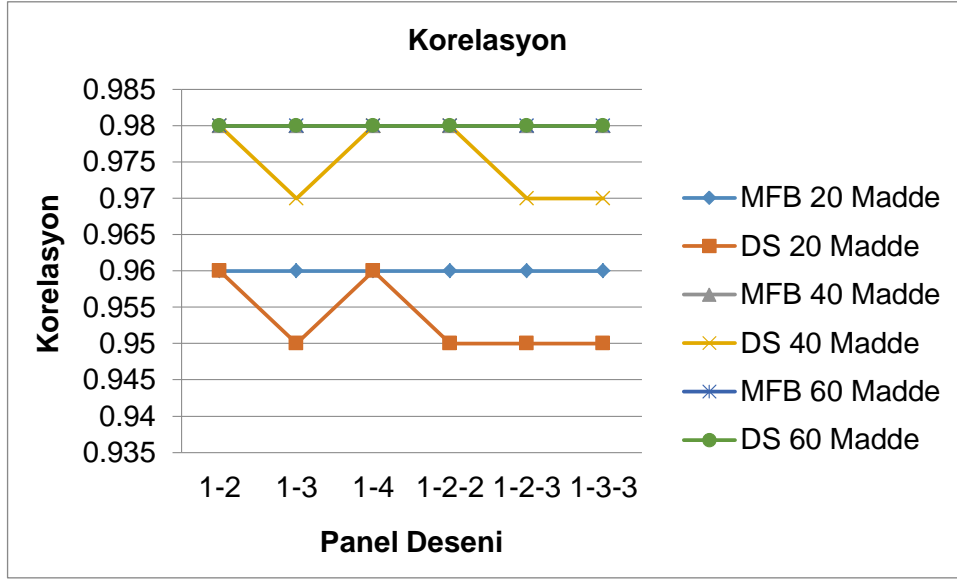
edilmiştir. Dolayısıyla testteki çok kategorili madde oranı %50 iken, 40 maddelik test koşullarında hangi yönlendirme yönteminin ve panel deseninin kullanıldığının bir önemi yoktur. DS yönlendirme yönteminin kullanıldığı 40 maddelik testlerde 1-3-3 hariç tüm panel desenlerinde aynı sonuç elde edilmiştir. 60 maddelik testlerde ise MFB yönlendirme yönteminde üç aşamalı panel desenlerinde (0.18) iki aşamalı panel desenlerine göre (0.19) daha düşük RMSE değerleri elde edilmiştir.

Tablo 8'i korelasyon değerleri açısından incelediğimizde, değerlerin 0.95 ve 0.98 arasında değiştiği görülmektedir. En düşük korelasyon değeri (0.95), DS yönlendirme yöntemi, 20 maddelik testte, 1-3 ve üç aşamalı panel desenlerinde (1-2-2, 1-2-3 ve 1-3-3) elde edilmiştir. Elde edilen en yüksek korelasyon değeri (0.98) ise MFB yönlendirme yöntemi, 40 ve 60 maddelik testlerde tüm panel desenlerinde, DS yönlendirme yöntemi, 40 maddelik testte 1-2, 1-4 ve 1-2-2 panel desenlerinde ve 60 maddelik testte tüm panel desenlerinde gözlenmiştir.

Testteki çok kategorili madde oranı %50 olduğunda yetenek kestirimlerinin korelasyon değerlerinin test uzunluğuna, panel desenlerine ve yönlendirme yöntemlerine göre nasıl değiştiğini gösteren grafik Şekil 16'da gösterilmiştir.

**Şekil 16**

*Testteki Çok Kategorili Madde Oranı %50 İken Korelasyon Değerleri*



Şekil 16'yı incelediğimizde testteki çok kategorili madde oranı %50 iken, testteki madde sayısı arttıkça gerçek ve gözlenen yetenek düzeyleri arasındaki korelasyon değerlerinin yükseldiği görülmektedir. En yüksek korelasyon değerleri (0.98) 60 maddelik testlerde ve 40 maddelik testlerin bazı panel koşullarında, en düşük korelasyon değerleri (0.95) ise 20 maddelik koşullarda gözlenmiştir. Yönlendirme yöntemi açısından incelediğimizde 20 maddelik testlerde, DS yönlendirme yönteminde 1-2 ve 1-4 panel deseni (0.96) hariç daha düşük korelasyon değerleri elde edilmiştir. 40 maddelik testlerde ise, aynı sonuçlar olmakla birlikte MFB yönlendirme yönteminde, DS yönlendirme yöntemine göre nispeten daha yüksek sonuçlar elde edilmiştir. 60 maddelik test koşullarında her iki yönlendirme yönteminde de aynı ve en yüksek korelasyon değerleri (0.98) elde edilmiştir. Yani çok kategorili madde oranı %50 olduğunda, 60 maddelik testlerde her iki yönlendirme yöntemi de eşit düzeyde iyi çalışmıştır. Panel deseni açısından incelediğimizde, MFB yönlendirme yöntemi ve 20 maddelik test uzunluğu koşullarında, tüm panel desenlerinde aynı korelasyon değerleri elde edilmiştir (0.96). 40 ve 60 maddelik testlerde de tüm panel desenlerinin sonuçları aynı ama 20 maddelik testlerden yüksek sonuçlar elde edilmiştir (0.98). DS yönlendirme yöntemine baktığımızda



test uzunluğu 20 olan testlerde iki aşamalı panel desenlerinden üç aşamalı panel desenlerine geçtikçe daha düşük değerler elde edilmiştir. 40 maddelik testlerde 1-3, 1-2-3 ve 1-3-3 panel deseni (0.97) hariç aynı sonuçlar elde edilmiştir. 60 maddelik testlerde ise tüm panel desenlerinde aynı korelasyon değerleri elde edilmiştir (0.98). Yani testteki çok kategorili madde oranı %50 iken, 60 maddelik testlerde hangi yönlendirme yönteminin herhangi panel deseninin kullanıldığına bir önemi yoktur. Ayrıca Şekil 16'yı incelediğimizde, testteki çok kategorili madde oranı %50 olduğunda, MFB yönlendirme yönteminde 40 ve 60 maddelik testlerin ve DS yönlendirme yöntemi 60 maddelik testlerin korelasyon grafiklerinin çakıştığı görülmektedir.

Özetle çok kategorili madde oranı %50 olduğunda, ortalama mutlak yanlılık açısından incelediğimizde en düşük ortalama mutlak yanlılık değeri 60 maddelik testlerde; MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında elde edilmiştir. En yüksek değer ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3-3 panel deseni koşulunda gözlenmiştir. RMSE değerleri açısından incelediğimizde, en düşük RMSE değeri 60 maddelik testlerde; MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında elde edilmiştir. En yüksek değer ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3-3 panel deseni koşulunda gözlenmiştir. Korelasyon değerleri açısından incelediğimizde ise, en düşük korelasyon değeri DS yönlendirme yöntemi, 20 maddelik testte, 1-3 ve üç aşamalı panel desenlerinde elde edilmiştir. En yüksek korelasyon değeri ise MFB yönlendirme yöntemi, 40 ve 60 maddelik testlerde tüm panel desenlerinde, DS yönlendirme yöntemi, 40 maddelik testte 1-2, 1-4 ve 1-2-2 panel desenlerinde ve 60 maddelik testte tüm panel desenlerinde gözlenmiştir. Dolayısıyla bu koşullar ölçüm hassasiyetleri en yüksek olan koşullardır. ANOVA sonuçları incelendiğinde ortalama mutlak yanlılık, RMSE ve korelasyon değerleri için tüm koşullar arasında anlamlı farklılık olduğu Tablo 10, Tablo 12 ve Tablo 14'te gösterilmiştir. Sonuç olarak %50 çok kategorili madde oranı içeren test koşullarında ortalama mutlak yanlılık ve

RMSE değerlerine baktığımızda, 60 maddelik test, MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında ölçüm hassasiyetlerinin çok yüksek olduğu görülmüştür.

## **Genel Bulgular ve Yorumlar**

### ***Ortalama Mutlak Yanlılık***

Ortalama mutlak yanlılık değerleri tüm koşullarda 0.14 ile 0.25 değerleri arasında değişmektedir. Tüm koşullar incelendiğinde en düşük ortalama mutlak yanlılık değerleri (0.14) altı farklı koşulda elde edilmiştir:

- %50 çok kategorili madde oranına sahip 60 maddelik testte, 1-2-2, 1-2-3 ve 1-3-3 panel deseninde, Maksimum Fisher Bilgisi (MFB) yönlendirme yöntemi koşulunda,
- %30 çok kategorili madde oranına sahip 60 maddelik testte, 1-2-2 panel deseninde, MFB yönlendirme yöntemi koşulunda ve
- %50 çok kategorili madde oranına sahip 60 maddelik testte, 1-2 ve 1-2-3 panel deseninde, Doğru Sayısı (DS) yönlendirme yöntemi koşulundadır.

En yüksek ortalama mutlak yanlılık değerleri (0.25) ise DS yönlendirme yöntemi, 20 maddelik testlerde; %50 çok kategorili madde oranına sahip, 1-3-3 panel deseni ve %10 çok kategorili madde oranına sahip, 1-2-2 panel deseni koşullarında elde edilmiştir. Tüm koşullardaki ortalama mutlak yanlılık bulguları Tablo 9'da gösterilmiştir. En düşük (0.14) ve en yüksek (0.25) ortalama mutlak yanlılık değerleri kalın punto ile vurgulanmıştır.

**Table 9***Tüm Koşullardaki Ortalama Mutlak Yanlılık Bulguları*

Yönlendirme Yöntemi	Panel Deseni	Testteki Çok Kategorili Madde Oranı								
		%10			%30			%50		
		20	40	60	20	40	60	20	40	60
MFB	1-2	0.24	0.18	0.15	0.23	0.17	0.15	0.23	0.17	0.15
	1-3	0.24	0.18	0.15	0.23	0.17	0.15	0.23	0.17	0.15
	1-4	0.24	0.18	0.16	0.23	0.18	0.15	0.23	0.17	0.15
	1-2-2	0.23	0.18	0.15	0.23	0.17	<b>0.14</b>	0.23	0.17	<b>0.14</b>
	1-2-3	0.23	0.18	0.15	0.23	0.17	0.15	0.23	0.17	<b>0.14</b>
	1-3-3	0.23	0.18	0.15	0.23	0.17	0.15	0.23	0.17	<b>0.14</b>
DS	1-2	0.24	0.18	0.15	0.23	0.17	0.15	0.23	0.17	<b>0.14</b>
	1-3	0.24	0.18	0.16	0.24	0.18	0.15	0.24	0.18	0.15
	1-4	0.24	0.18	0.16	0.23	0.17	0.15	0.23	0.17	0.15
	1-2-2	<b>0.25</b>	0.18	0.16	0.24	0.17	0.15	0.24	0.17	0.16
	1-2-3	0.24	0.18	0.15	0.24	0.17	0.15	0.24	0.18	<b>0.14</b>
	1-3-3	0.24	0.18	0.16	0.24	0.18	0.15	<b>0.25</b>	0.18	0.15

İstatistiksel olarak koşullardaki anlamlı farklılıkları bulmak amacıyla, ortalama mutlak yanlılık bağımlı değişken ve diğer koşullar bağımsız değişken (testteki çok kategorili madde yüzdesi, panel deseni, test uzunluğu ve yönlendirme yöntemi) olacak şekilde faktöryel ANOVA yapılmıştır. Analiz sonuçları Tablo 10'da gösterilmiştir.

**Tablo 10***Ortalama Mutlak Yanlılık İçin Faktöryel ANOVA Sonuçları*

Faktör	Sum of Squares	df	Mean Square	F	p	$\eta^2_p$
Yönlendirme	0.049	1	0.049	1594.113	0.00	0.13
Test uzunluğu	13.999	2	6.999	228391.127	0.00	0.977
Çok kategorili madde oranı	0.080	2	0.040	1297.904	0.00	0.195
Panel deseni	0.016	5	0.003	104.280	0.00	0.046
Yönlendirme*Test uzunluğu	0.007	2	0.003	112.118	0.00	0.021
Yönlendirme*Çok kategorili madde oranı	0.001	2	0.001	18.521	0.00	0.003
Yönlendirme*Panel deseni	0.042	5	0.008	273.176	0.00	0.113
Test uzunluğu*Çok kategorili madde oranı	0.001	4	0.000	10.235	0.00	0.004
Test uzunluğu*Panel deseni	0.008	10	0.001	24.606	0.00	0.022
Çok kategorili madde oranı*Panel deseni	0.004	10	0.000	12.917	0.00	0.012
Yönlendirme*Test uzunluğu*Çok kategorili madde oranı	0.001	4	0.000	12.228	0.00	0.005
Yönlendirme*Test uzunluğu*Panel deseni	0.010	10	0.001	32.315	0.00	0.029
Yönlendirme*Çok kategorili madde oranı*Panel deseni	0.002	10	0.000	8.091	0.00	0.008
Test uzunluğu*Çok kategorili madde oranı	0.007	20	0.000	11.231	0.00	0.021

kategorili madde

oranı\*Panel deseni

Yönlendirme\*Test 0.007 20 0.000 11.410 0.00 0.021

uzunluğu\*Çok kategorili

madde oranı\*Panel deseni

Hata 0.328 10692 0.000

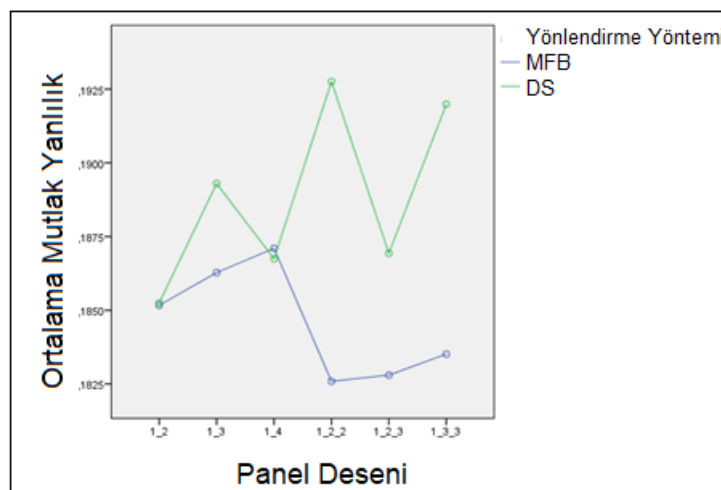
Toplam 391.019 10800

Tüm faktörler için  $p < 0.01$ .

Tablo 10'da görüldüğü üzere ana etkiler ve etkileşimlerin tamamı anlamlı çıkmıştır. Ana etkilere baktığımızda, diğer faktörler kontrol altına alındığında, sırasıyla test uzunluğu ( $\eta^2_p=.977$ ), çok kategorili madde oranı ( $\eta^2_p=.195$ ) ve yönlendirme yöntemi ( $\eta^2_p=.13$ ) ortalama mutlak yanlılığının varyansını en çok açıklayan üç faktördür. Tüm faktörler kontrol altına alındığında, yönlendirme ve panel deseni etkileşimi ortalama mutlak yanlılığının varyansını ( $\eta^2_p=.113$ ) en çok açıklayan faktördür. Ancak ana etkiler ve etkileşimlerin tamamı anlamlı çıktığı için yönlendirme ve panel deseni etkileşimine ait grafikler verilmiştir. Ortalama mutlak yanlılık üzerindeki her bir panel deseninde yönlendirme yönteminin etkileşiminin grafiği Şekil 17'de gösterilmektedir.

### Şekil 17

*Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Ortalama Mutlak Yanlılık Üzerindeki Etkileşim Etkisi*



Şekil 17’de görüldüğü üzere iki değişkene ait çizgilerin birbirini kestiği 1-2 ve 1-4 panel deseni noktasında etkileşim bulunmaktadır. 1-2 ve 1-4 panel desenlerinde, Maksimum Fisher Bilgisi (MFB) ya da Doğru Sayısı (DS) yönlendirme yöntemlerinden hangisinin kullanıldığıнын bir önemi bulunmamaktadır. 2 aşamalı panel desenlerinden (1-2, 1-3 ve 1-4) 3 aşamalı panel desenlerine (1-2-2, 1-2-3 ve 1-3-3) geçildiğinde, MFB yönlendirme yöntemi DS yönlendirme yöntemine göre belirgin düzeyde düşük ortalama mutlak yanlılık değerleri vermiştir. Bir diğer bulgu ise, özellikle MFB yönlendirme yönteminde, üç aşamalı panel desenleri iki aşamalı panel desenlerine göre daha düşük ortalama mutlak yanlılık değerleri vermektedir. Genel olarak baktığımızda MFB yönlendirme yönteminin, DS yönlendirme yöntemine göre tüm panel desenlerinde daha iyi sonuçlar verdiği görülmektedir.

### **RMSE**

Hata Kareleri Ortalamasının Karekökü (RMSE-Root Mean Square Error) değerleri tüm koşullarda 0.18 ile 0.32 değerleri arasında değişmektedir. Tüm koşullar incelendiğinde en düşük ortalama RMSE değerleri (0.18) sekiz farklı koşulda elde edilmiştir:

- %50 çok kategorili madde oranına sahip 60 maddelik testte, 1-2-2, 1-2-3 ve 1-3-3 panel deseninde, MFB yönlendirme yöntemi koşulunda,
- %50 çok kategorili madde oranına sahip 60 maddelik testte, 1-2 ve 1-2-3 panel deseninde, DS yönlendirme yöntemi koşulunda,
- %30 çok kategorili madde oranına sahip 60 maddelik testte, 1-2-2, 1-2-3 ve 1-3-3 panel deseninde, MFB yönlendirme yöntemi koşulundadır.

En yüksek ortalama RMSE değerleri (0.32) ise %50 çok kategorili madde oranına sahip 20 maddelik testte, 1-3-3 panel deseninde, DS yönlendirme yöntemi koşulunda ve %10 çok kategorili madde oranına sahip 20 maddelik testte, 1-2-2 panel deseninde, DS yönlendirme yöntemi koşulunda elde edilmiştir. Tüm koşullardaki ortalama RMSE bulguları Tablo 11’de gösterilmiştir. En düşük (0.18) ve en yüksek (0.32) RMSE değerleri kalın punto ile vurgulanmıştır.

**Table 11**

*Tüm Koşullardaki Ortalama RMSE Bulguları*

		Testteki Çok Kategorili Madde Oranı								
		%10			%30			%50		
Yönlendirme Yöntemi	Panel Deseni	20	40	60	20	40	60	20	40	60
MFB	1-2	0.30	0.23	0.19	0.30	0.22	0.19	0.29	0.22	0.19
	1-3	0.30	0.23	0.19	0.30	0.22	0.19	0.29	0.22	0.19
	1-4	0.30	0.23	0.20	0.30	0.22	0.19	0.29	0.22	0.19
	1-2-2	0.30	0.22	0.19	0.29	0.22	<b>0.18</b>	0.29	0.22	<b>0.18</b>
	1-2-3	0.30	0.22	0.19	0.29	0.22	<b>0.18</b>	0.29	0.22	<b>0.18</b>
	1-3-3	0.30	0.22	0.19	0.29	0.22	<b>0.18</b>	0.29	0.22	<b>0.18</b>
DS	1-2	0.30	0.23	0.19	0.30	0.22	0.19	0.29	0.22	<b>0.18</b>
	1-3	0.30	0.23	0.20	0.30	0.23	0.19	0.30	0.22	0.19
	1-4	0.30	0.23	0.20	0.30	0.22	0.19	0.29	0.22	0.19
	1-2-2	<b>0.32</b>	0.24	0.20	0.31	0.22	0.20	0.31	0.22	0.20
	1-2-3	0.31	0.23	0.19	0.30	0.22	0.19	0.30	0.22	<b>0.18</b>
	1-3-3	0.31	0.23	0.21	0.31	0.22	0.19	<b>0.32</b>	0.23	0.19

İstatistiksel olarak koşullardaki anlamlı farklılıkları bulmak amacıyla, ortalama RMSE bağımlı değişken ve diğer koşullar bağımsız değişken (testteki çok kategorili

madde yüzdesi, panel deseni, test uzunluğu ve yönlendirme yöntemi) olacak şekilde faktöryel ANOVA yapılmıştır. Analiz sonuçları Tablo 12’de gösterilmiştir.

**Tablo 12**

*Ortalama RMSE İçin Faktöryel ANOVA Sonuçları*

Faktör	Sum of Squares	df	Mean Square	F	p	$\eta^2_p$
Yönlendirme	0.110	1	0.110	2437,643	0.00	0.186
Test uzunluğu	22.723	2	11.362	251592.074	0.00	0.979
Çok kategorili madde oranı	0.110	2	0.055	1219.259	0.00	0.186
Panel deseni	0.024	5	0.005	107.254	0.00	0.048
Yönlendirme*Test uzunluğu	0.017	2	0.009	191.051	0.00	0.035
Yönlendirme*Çok kategorili madde oranı	0.002	2	0.001	17.598	0.00	0.003
Yönlendirme*Panel deseni	0.092	5	0.018	409.136	0.00	0.161
Test uzunluğu*Çok kategorili madde oranı	0.001	4	0.000	6.332	0.00	0.002
Test uzunluğu*Panel deseni	0.017	10	0.002	38.500	0.00	0.035
Çok kategorili madde oranı*Panel deseni	0.007	10	0.001	15.592	0.00	0.014
Yönlendirme*Test uzunluğu*Çok kategorili madde oranı	0.002	4	0.001	13.389	0.00	0.005
Yönlendirme*Test uzunluğu*Panel deseni	0.023	10	0.002	50.328	0.00	0.045
Yönlendirme*Çok kategorili madde oranı*Panel deseni	0.005	10	0.001	11.427	0.00	0.011
Test uzunluğu*Çok kategorili madde oranı*Panel deseni	0.019	20	0.001	21.157	0.00	0.038
Yönlendirme*Test uzunluğu*Çok kategorili madde oranı*Panel deseni	0.017	20	0.001	18.732	0.00	0.034



kategorili madde oranı\*Panel

deseni

Hata 0.483 10692 0.000

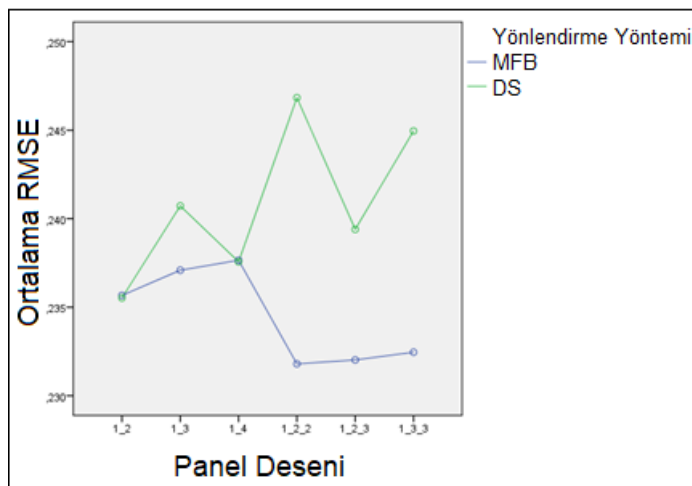
Toplam 633.563 10800

Tüm faktörler için  $p < 0.01$ .

Tablo 12'de görüldüğü gibi ana etkiler ve etkileşimlerin tümü anlamlı çıkmıştır. Tüm faktörler kontrol altına alındığında, yönlendirme ve panel deseni etkileşimi ortalama RMSE varyansını ( $\eta^2_p = .161$ ) en çok açıklayan faktördür. Ana etkilere baktığımızda, diğer faktörler kontrol altına alındığında, en çok test uzunluğu ( $\eta^2_p = .979$ ) ortalama RMSE varyansını açıklamaktadır. Çok kategorili madde oranı ( $\eta^2_p = .186$ ) ve yönlendirme yöntemi ( $\eta^2_p = .186$ ) ise ortalama RMSE varyansını eşit düzeyde, en çok açıklayan ikinci faktörlerdir. Ana etkiler ve etkileşimlerin tamamı anlamlı çıktığı için yönlendirme ve panel deseni etkileşimine ait grafik verilmiştir. Ortalama RMSE üzerindeki her bir panel deseninde yönlendirme yönteminin etkileşiminin grafiği Şekil 18'de gösterilmektedir.

### Şekil 18

*Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Ortalama RMSE Üzerindeki Etkileşim Etkisi*



Şekil 18'de görüldüğü gibi 1-2 ve 1-4 panel deseni noktasında etkileşim bulunmaktadır. Dolayısıyla 1-2 ve 1-4 panel desenlerinde, yönlendirme yöntemlerinden

hangisinin kullanıldığıнын bir önemi yoktur. 2 aşamalı panel desenlerinden (1-2, 1-3 ve 1-4) 3 aşamalı panel desenlerine (1-2-2, 1-2-3 ve 1-3-3) geçildiğinde, MFB yönlendirme yönteminde DS yönlendirme yöntemine göre daha düşük ortalama RMSE değerleri vermiştir. Bir diğer bulgu ise, özellikle MFB yönlendirme yönteminde, üç aşamalı panel desenleri iki aşamalı panel desenlerine göre daha düşük ortalama RMSE değerleri vermektedir. Genel olarak baktığımızda MFB yönlendirme yönteminin, DS yönlendirme yöntemine göre tüm panel desenlerinde daha iyi sonuçlar vermektedir.

### **Korelasyon**

Kestirilen ve gerçek yetenek parametreleri arasındaki korelasyon ( $\rho$ ) değerleri tüm koşullarda 0.95 ile 0.98 değerleri arasında değişmektedir. Tüm koşullar incelendiğinde en düşük korelasyon değerlerinin (0.95) 20 maddelik test koşullarının birçoğunda elde edilmiştir. En yüksek korelasyon değerlerinin (0.98) ise %30 ve %50'lik çok kategorili madde oranına sahip, 40 ve 60 maddelik test koşullarının neredeyse tamamında elde edildiği görülmüştür. Tüm koşullardaki korelasyon bulguları Tablo 13'te gösterilmiştir.

**Table 13**

#### *Tüm Koşullardaki Korelasyon Bulguları*

		Testteki Çok Kategorili Madde Oranı								
		%10			%30			%50		
Yönlendirme Yöntemi	Panel Deseni	20	40	60	20	40	60	20	40	60
MFB	1-2	<b>0.95</b>	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-3	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-4	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-2-2	0.96	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-2-3	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-3-3	0.96	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>

	1-2	<b>0.95</b>	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
	1-3	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>
	1-4	<b>0.95</b>	0.97	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>	0.96	<b>0.98</b>	<b>0.98</b>
DS	1-2-2	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	<b>0.98</b>	<b>0.98</b>	<b>0.95</b>	<b>0.98</b>	<b>0.98</b>
	1-2-3	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	<b>0.98</b>	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>
	1-3-3	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>	<b>0.95</b>	0.97	<b>0.98</b>

Tablo 13'te en düşük (0.95) ve en yüksek (0.98) korelasyon değerleri kalın punto ile vurgulanmıştır. Koşullardaki anlamlı farklılıkları bulmak için, korelasyon bağımlı değişken ve diğer tüm koşullar bağımsız değişken (testteki çok kategorili madde yüzdesi, panel deseni, test uzunluğu ve yönlendirme yöntemi) olacak şekilde faktöryel ANOVA analizi yapılmıştır. Analiz sonuçları Tablo 14'te gösterilmiştir.

**Tablo 14**

*Korelasyon İçin Faktöryel ANOVA Sonuçları*

Faktör	Sum of Squares	df	Mean Square	F	p	$\eta^2_p$
Yönlendirme	0.011	1	0.011	810.482	0.00	0.070
Test uzunluğu	1.340	2	0.670	51715.759	0.00	0.906
Çok kategorili madde oranı	0.010	2	0.005	389.497	0.00	0.068
Panel deseni	0.002	5	0.000	24.990	0.00	0.012
Yönlendirme*Test uzunluğu	0.006	2	0.003	244.906	0.00	0.044
Yönlendirme*Çok kategorili madde oranı	0.000	2	0.000	14.163	0.00	0.003
Yönlendirme*Panel deseni	0.007	5	0.001	106.403	0.00	0.047
Test uzunluğu*Çok kategorili madde oranı	0.005	4	0.001	99.687	0.00	0.036
Test uzunluğu*Panel deseni	0.003	10	0.000	24.699	0.00	0.023
Çok kategorili madde oranı*Panel deseni	0.001	10	9.806E-5	7.568	0.00	0.007

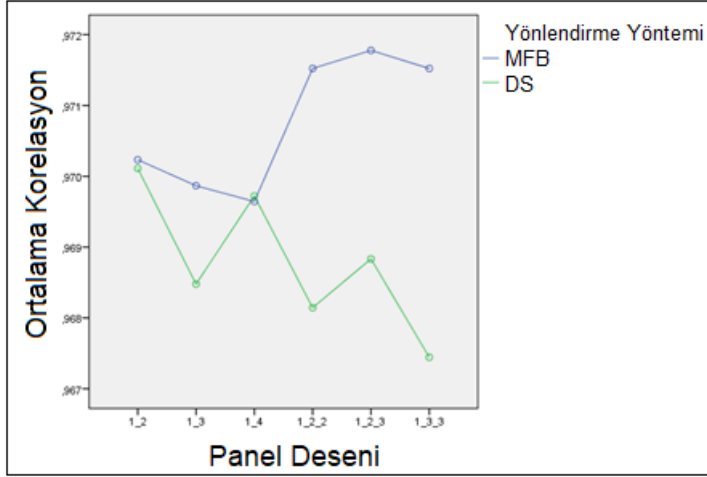
deseni							
Yönlendirme*Test uzunluğu*Çok kategorili madde oranı	0.000	4	4.589E-5	3.541	0.007	0.001	
Yönlendirme*Test uzunluğu*Panel deseni	0.004	10	0.00	34.330	0.000	0.031	
Yönlendirme*Çok kategorili madde oranı*Panel deseni	0.001	10	9.276E-5	7.159	0.000	0.007	
Test uzunluğu*Çok kategorili madde oranı*Panel deseni	0.001	20	5.999E-5	4.630	0.000	0.009	
Yönlendirme*Test uzunluğu*Çok kategorili madde oranı*Panel deseni	0.001	20	4.156E-5	3.207	0.000	0.006	
Hata	0.139	10692	0.000				
Toplam	10158.538	10800					

Tüm faktörler için  $p, < 0.01$ .

Tablo 14'te görüldüğü üzere ana etkiler ve etkileşimlerin tamamı anlamlı çıkmıştır. Ana etkilere baktığımızda, diğer faktörler kontrol altına alındığında, en çok test uzunluğu ( $\eta^2_p=.906$ ) korelasyon varyansını açıklamaktadır. Bu bulgu 60 maddelik testlerin tamamının ve 40 maddelik testlerinin çoğunun en yüksek korelasyon değerini almasını açıklamaktadır. Çok kategorili madde oranı ( $\eta^2_p=.068$ ) ve yönlendirme yöntemi ( $\eta^2_p=.070$ ) ise korelasyon varyansını yakın düzeyde açıklayan diğer faktörlerdir. Tüm faktörler kontrol altına alındığında, yönlendirme ve panel deseni etkileşimi korelasyon varyansını ( $\eta^2_p=.047$ ) en çok açıklayan faktördür. Korelasyon üzerindeki her bir panel deseni ile yönlendirme yönteminin etkileşiminin grafiği Şekil 19'da gösterilmektedir.

### Şekil 19

*Yönlendirme Yöntemi ve Panel Deseni Seviyelerinin Korelasyon Üzerindeki Etkileşim Etkisi*



Şekil 19'da görüldüğü üzere 1-2 ve 1-4 panel deseni noktasında, panel deseni ve yönlendirme yöntemi arasında etkileşim bulunmaktadır. Yani 1-2 ve 1-4 panel desenlerinde, MFB ve DS yönlendirme yöntemlerinden hangisinin kullanıldığının bir önemi olmadığı anlamına gelmektedir. Bir diğer bulgu ise, MFB yönlendirme yöntemi üç aşamalı panel desenlerinde iki aşamalı panel desenlerine göre daha yüksek korelasyon değerleri vermektedir. Genel olarak MFB yönlendirme yöntemi DS yönlendirme yöntemine göre daha yüksek korelasyon değerleri verdiği görülmektedir.

## **Bölüm 5**

### **Sonuç, Tartışma ve Öneriler**

Araştırmada, PISA 2018 yılı okuma becerileri alanında bireye uyarlanmış çok aşamalı test (BÇAT) uygulamasında kullanılan iki ve çok kategorili madde parametrelerine göre üretilen madde parametreleri kullanılarak, karma testlerdeki çok kategorili madde oranlarının, farklı yönlendirme yöntemleri, panel desenleri ve test uzunlukları koşullarında kullanıldığında yetenek kestirimlerinin kesinliği incelenmiştir. Araştırmada elde edilen sonuçların değerlendirmesinde ortalama mutlak yanlılık, RMSE ve korelasyon değerlerine bakılmıştır. Elde edilen bulgular, bulgular ve yorumlar bölümünde detaylı bir şekilde ele alınmıştır. Bu bölümde sonuçların literatürdeki çalışmalarla benzerlikleri ya da farklılıkları tartışıldıktan sonra, gelecekte yapılacak araştırmalara önerilerde bulunulmuştur.

#### **Sonuçlar**

Araştırmada karma testlerdeki çok kategorili madde oranlarının (%10, %30 ve %50) farklı koşullarda (test uzunluğu, yönlendirme yöntemi ve panel deseni) yetenek kestirimlerindeki performansları nasıl etkilediği incelenmiştir. Yetenek kestirimlerinde kesinliği değerlendirmek amacıyla ortalama mutlak yanlılık, RMSE ve korelasyon değerleri hesaplanmıştır.

Sonuçları alt problem 1 açısından incelediğimizde, yani karma testlerdeki çok kategorili madde oranları %10 olduğunda; ortalama mutlak yanlılık ve RMSE için en düşük değerler her iki yönlendirme yönteminde de 60 maddelik testlerde olduğu görülmektedir. MFB yönlendirme yönteminde 1-4 panel deseni hariç tüm panel desenleri ve DS yönlendirme yönteminde 1-2 ve 1-2-3 panel desenlerinde elde edilmiştir. Ortalama mutlak yanlılık ve RMSE için en yüksek değerler ise, DS yönlendirme yöntemi, 20 maddelik testte ve 1-2-2 panel deseni koşulunda elde edilmiştir. Korelasyon açısından en düşük değerler 20 maddelik testlerde; MFB yönlendirme yöntemi, 2 aşamalı panel desenleri ve DS yönlendirme yöntemi, tüm panel desenlerinde elde edilmiştir. En yüksek değerler ise MFB

ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde aynı çıkmıştır. Ayrıca MFB yönlendirme yöntemi, 40 maddelik testte 1-2-3 panel deseninde elde edilmiştir. Sonuç olarak çok kategorili madde oranı %10 olduğunda, MFB yönlendirme yöntemi, 60 maddelik testler ve 3 aşamalı panel desenlerinin ölçme hassasiyetleri daha iyi sonuçlar vermektedir. Yönlendirme yöntemi açısından her iki yöntem de benzer düzeyde yüksek ölçme hassasiyeti göstermiştir. ANOVA sonuçları incelendiğinde tüm koşullar arasında anlamlı farklılık bulunmaktadır.

Sonuçları alt problem 2 açısından incelediğimizde, yani karma testlerdeki çok kategorili madde oranları %30 olduğunda; en düşük ortalama mutlak yanlılık değeri MFB yönlendirme yöntemi, 60 maddelik testte, 1-2-2 panel deseninde gözlenmiştir. En yüksek ortalama mutlak yanlılık değeri ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3, 1-2-2, 1-2-3 ve 1-3-3 panel desenleri koşullarında gözlenmiştir. RMSE değerlerine baktığımızda; MFB yönlendirme yöntemi, 60 maddelik testte, 3 aşamalı panel desenleri en düşük RMSE değerine sahiptir. En yüksek değer ise DS yönlendirme yöntemi, 20 maddelik testte, 1-2-2 ve 1-3-3 panel desenleri koşullarında gözlenmiştir. Korelasyon değerlerini incelediğimizde en düşük korelasyon değeri 20 maddelik testlerde; MFB yönlendirme yöntemi, 1-3 ve 1-4 panel desenleri ve DS yönlendirme yöntemi, 1-3 panel deseni ve 3 aşamalı tüm panel desenlerinde görülmektedir. Elde edilen en yüksek korelasyon değeri ise her iki yönlendirme yönteminde, 60 maddelik testlerde tüm panel desenlerinde ve 40 maddelik testlerde, MFB yönlendirme yönteminde 1-4 panel deseni hariç tüm koşullarda ve DS yönlendirme yöntemi, 1-3 ve 1-3-3 panel desenleri hariç diğer tüm koşullarda gözlenmiştir. ANOVA sonuçları incelendiğinde ortalama mutlak yanlılık, RMSE ve korelasyon değerleri için tüm koşullar arasında anlamlı farklılık bulunmaktadır. Sonuç olarak %30 çok kategorili madde oranı içeren test koşullarında, MFB yönlendirme yöntemi, 60 maddelik testler ve 1-2-2 panel deseni başta olmak üzere 3 aşamalı panel desenlerinin ölçüm hassasiyetlerinin çok yüksek olduğu görülmüştür. Ayrıca korelasyon

değerleri açısından, MFB ve DS yönlendirme yöntemi, 60 maddelik testlerde tüm panel desenlerinde en yüksek değerler elde edildiği için ölçüm hassasiyetleri aynıdır.

Alt problem 3'te ise karma testlerdeki çok kategorili madde oranları %50 olduğundaki koşullar incelenmiştir. Sonuçları ortalama mutlak yanlılık ve RMSE açısından incelediğimizde en düşük ve en yüksek değerler her ikisinde de aynı koşullarda gözlenmiştir. 60 maddelik testlerde; MFB yönlendirme yöntemi, 3 aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında elde edilmiştir. En yüksek değer ise DS yönlendirme yöntemi, 20 maddelik testte ve 1-3-3 panel deseni koşulunda gözlenmiştir. Korelasyon değerleri açısından incelediğimizde ise, en düşük korelasyon değeri DS yönlendirme yöntemi, 20 maddelik testte, 1-3 ve 3 aşamalı panel desenlerinde elde edilmiştir. En yüksek korelasyon değeri ise MFB yönlendirme yöntemi, 40 ve 60 maddelik testlerde tüm panel desenlerinde, DS yönlendirme yöntemi, 40 maddelik testte 1-2, 1-4 ve 1-2-2 panel desenlerinde ve 60 maddelik testte tüm panel desenlerinde gözlenmiştir. ANOVA sonuçları incelendiğinde tüm koşullar arasında anlamlı farklılık bulunmaktadır. Sonuç olarak %50 çok kategorili madde oranı içeren test koşullarında 60 maddelik test, MFB yönlendirme yöntemi, üç aşamalı panel desenlerinde ve DS yönlendirme yöntemi, 1-2 ve 1-2-3 panel desenleri koşullarında ölçüm hassasiyetlerinin çok yüksek olduğu görülmüştür.

Sonuçlar ortalama mutlak yanlılık açısından incelendiğinde, en iyi sonuçlar her iki yönlendirme yönteminde ve %50 çok kategorili madde oranına sahip 60 maddelik testlerde görülmüştür. Buna ek olarak %30 çok kategorili madde oranına sahip 60 maddelik testte, 1-2-2 panel deseninde, MFB yönlendirme yöntemi koşulunda da en düşük ortalama mutlak yanlılık değerleri elde edilmiştir. En yüksek ortalama mutlak yanlılık değerleri ise 20 maddelik testte, DS yönlendirme yönteminde %10 çok kategorili madde oranına sahip, 1-3-3 panel deseninde ve %50 çok kategorili madde oranına sahip, 1-2-2 panel deseninde koşulunda gözlenmiştir. İstatistiksel olarak koşullardaki anlamlı



farklılıkları bulmak amacıyla ANOVA analizi yapılmıştır. ANOVA sonucunda tüm faktörler kontrol altına alındığında, yönlendirme ve panel deseni etkileşimi ortalama mutlak yanlılığının varyansını en çok açıklayan faktör olduğu görülmüştür. Ana etkilere baktığımızda ise sırasıyla test uzunluğu ve çok kategorili madde oranı ortalama mutlak yanlılığının varyansını en çok açıklayan iki faktördür. Çok kategorili madde oranı %10 iken 60 maddelik testlerde, MFB yönlendirme yönteminde 1-4 hariç tüm panel deseni ve DS yönlendirme yönteminde 1-2 ve 1-2-3 panel deseni koşullarında en iyi sonuçlar elde edilmiştir. Çok kategorili madde oranı %30 iken 60 maddelik testlerde, MFB yönlendirme yönteminde, 1-2-2 panel deseni, %10 çok kategorili madde oranına göre daha iyi sonuçlar elde edilmiştir. Çok kategorili madde oranı %50 olduğunda ise, 60 maddelik testlerde, MFB yönlendirme yönteminde üç aşamalı tüm panel desenleri ve DS yönlendirme yönteminde 1-2 ve 1-2-3 panel deseni koşullarında en iyi sonuçlar elde edilmiştir.

RMSE değerlerini incelediğimizde ortalama mutlak yanlılık değerlerine benzer eğilimler görülmektedir. En düşük ortalama RMSE değerleri %50 çok kategorili madde oranına sahip 60 maddelik testlerde her iki yönlendirme yönteminde ve %30 çok kategorili madde oranına sahip 60 maddelik testte, üç aşamalı panel desenlerinin tamamında elde edilmiştir. En yüksek ortalama RMSE değerleri ise DS yönlendirme yöntemi, 20 maddelik testte; %10 çok kategorili madde oranına sahip 1-2-2 panel deseni ve %50 çok kategorili madde oranına sahip, 1-3-3 panel deseni koşulunda elde edilmiştir. ANOVA analizlerinde yönlendirme ve panel deseni etkileşimi ortalama RMSE varyansını en çok açıklayan faktördür. Ana etkilere baktığımızda ise en çok test uzunluğu, daha sonra çok kategorili madde oranı ve yönlendirme yöntemi eşit düzeyde olacak şekilde ortalama RMSE varyansını açıklayan faktörlerdir. Çok kategorili madde oranı %10 iken ortalama mutlak yanlılık sonuçlarıyla aynı koşullarda düşük ve yüksek ortalama RMSE değerleri elde edilmiştir. Çok kategorili madde oranı %30 olduğu koşullarda ise en düşük RMSE değerleri 60 maddelik testlerde, MFB yönlendirme yönteminde üç aşamalı panel desenlerinde görülürken; en yüksek değerler ise 20 maddelik testlerde, DS yönlendirme

yönteminde 1-2-2 ve 1-3-3 panel desenlerinde görülmüştür. Çok kategorili madde oranı %50 olduğu koşullarda 60 maddelik testlerde; MFB yönlendirme yönteminde üç aşamalı panel desenlerinde ve DS yönlendirme yönteminde 1-2 ve 1-2-3 panel desenlerinde en düşük RMSE değerleri elde edilmiştir. En yüksek değerler ise DS yönlendirme yöntemi, 20 maddelik test, 1-3-3 panel deseni koşulunda elde edilmiştir.

Korelasyon değerlerine baktığımızda ise tüm koşullar incelendiğinde 20 maddelik test koşullarının birçoğunda en düşük korelasyon değerleri elde edilmiştir. En yüksek korelasyon değerlerinin ise %30 ve %50'lik çok kategorili madde oranına sahip, 40 ve 60 maddelik test koşullarının neredeyse tamamında elde edildiği görülmüştür. ANOVA analizlerinde 1-2 ve 1-4 panel deseni ve yönlendirme yöntemi arasında etkileşim bulunmaktadır. Dolayısıyla bu iki panel deseninde MFB ve DS yönlendirme yöntemlerinden hangisinin kullanıldığına bir önemi olmadığı şeklinde yorumlanır. Çok kategorili madde oranı %10 olduğu koşullarda en düşük korelasyon değerleri 20 maddelik testlerde; MFB yönlendirme yönteminde iki aşamalı panel desenlerinde ve DS yönlendirme yönteminde tüm panel desenlerinde elde edilmiştir. En yüksek korelasyon değerleri ise 60 maddelik testlerde; her iki yönlendirme yönteminde, tüm panel desenlerinde ve MFB yönlendirme yönteminde, 40 maddelik testte, 1-2-3 panel deseninde gözlenmiştir. Çok kategorili madde oranı %30 olduğu koşullarda en düşük korelasyon değerleri 20 maddelik testlerde; MFB yönlendirme yönteminde, 1-3 ve 1-4 panel desenlerinde ve DS yönlendirme yönteminde, 1-3, 1-2-2, 1-2-3 ve 1-3-3 panel desenlerinde elde edilmiştir. En yüksek korelasyon değerleri ise 60 maddelik testlerde tüm koşullarda ve 40 maddelik testlerde; MFB yönlendirme yönteminde 1-4 panel deseni ve DS yönlendirme yönteminde 1-3 ve 1-3-3 panel deseni koşullarında gözlenmiştir. Çok kategorili madde oranı %50 olduğunda ise en düşük korelasyon değerleri 20 maddelik testlerde, DS yönlendirme yönteminde, üç aşamalı ve 1-3 panel desenlerinde elde edilmiştir. En yüksek korelasyon değerleri ise 60 maddelik testlerde tüm koşullarda ve 40

maddelik testlerde; MFB yönlendirme yönteminde tüm koşullarda ve DS yönlendirme yönteminde 1-2, 1-4 ve 1-2-2 panel deseni koşullarında elde edilmiştir.

Tüm koşulları özetlersek en düşük ortalama mutlak yanlılık değerleri %30 ve %50 çok kategorili testlerde ve 60 maddelik testlerde elde edilmiştir. Bu da çok kategorili madde oranının artması yetenek kestirimlerinin doğruluğunu arttırdığı anlamına gelmektedir. En yüksek ortalama mutlak yanlılık değerleri ise tüm koşullarda, 20 maddelik testlerde ve DS yönlendirme yönteminde gözlenmiştir. Bu koşullardan en düşüğü ise %10 çok kategorili madde oranında, 20 maddelik testte, DS yönlendirme yöntemi 1-2-2 panel deseni koşulunda gözlenmiştir. Ayrıca testlerdeki madde sayısı arttıkça ortalama mutlak yanlılık değerleri düşmektedir. RMSE değerlerine baktığımızda en düşük %30 ve %50 çok kategorili madde oranlarında, MFB yönlendirme yöntemi, 60 maddelik testlerde, üç aşamalı panel desenlerinde ve %50 çok kategorili madde oranlarında, DS yönlendirme yöntemi, 60 maddelik testlerde elde edilmiştir. En yüksek ortalama RMSE değerleri ise %10 ve %50 çok kategorili madde oranında, 20 maddelik testlerde ve DS yönlendirme yönteminde elde edilmiştir. Bu sonuçlardan madde sayısı ve çok kategorili madde oranı arttıkça ortalama RMSE değerlerinin düştüğü yorumu yapılabilir. Ayrıca MFB yönlendirme yöntemi, DS yönlendirme yöntemine göre daha iyi sonuçlar vermektedir. Korelasyon değerlerine baktığımızda çok kategorili madde oranı ve test uzunluğu arttıkça korelasyon değerlerinin arttığı sonucuna varılmıştır.

## **Tartışma**

Araştırmada elde edilen sonuçlara göre temel bulgulardan bir tanesi, testlerde çok kategorili madde oranı arttıkça (%10'dan %50'ye gittikçe), daha doğru veya az hata ile yetenek kestirimi yapılmasıdır. Dolayısıyla testteki çok kategorili madde sayısı arttıkça daha iyi yetenek kestirimi sağlanmaktadır. Çok kategorili madde oranları %10 ile %30'a baktığımızda benzer, birbirine yakın sonuçlar verdiği görülmekteyken; %10'dan %50'ye ortalama mutlak yanlılık ve RMSE değerlerinin düştüğü, korelasyon değerlerinin ise arttığı

görülmektedir. Karma testlerde çok kategorili madde oranı arttıkça test bilgi miktarının yükselmesi beklenen bir sonuçtur. Benzer şekilde Kim, vd. (2012) sınıflandırma testi kapsamında iki, üç ve dört kategorili maddelerden oluşan karma testleri birinci modülde kullandıkları çalışma sonucunda, yönlendirme modülündeki (birinci aşamada) test bilgi fonksiyonu seviyeleri arttıkça, simülasyonlar daha fazla sayıda doğru sınıflandırma kararı vermiştir. Ancak Park, vd. (2014) iki, üç ve dört kategorili maddelerden oluşan sınıflandırma testinde, uygulanan BÇAT yeniden birleştirme koşullarının tüm koşullarda, sınıflandırma kararının kesinliği açısından eşit derecede iyi performans gösterdiği sonucunu elde etmişlerdir. Park (2015) ise çalışmasında çok kategorili madde oranı arttıkça ölçüm hassasiyetinin düştüğünü gözlemlemiştir. Bu sonucun, test oluşturma aşamasında çok kategorili maddelerin test bilgilerinin azaltılmasından kaynaklandığını belirtmiştir. Tüm bu sonuçlardan test birleştirme koşulları başta olmak üzere, uygulanan BÇAT koşullarının sonuçları ciddi şekilde etkilediği sonucuna varabiliriz.

Yönlendirme yöntemleri açısından incelediğimizde testlerdeki çok kategorili madde oranları, panel desenleri ve test uzunlukları koşulları dikkate alındığında, MFB yönlendirme yöntemi DS yönlendirme yöntemine göre daha iyi ölçüm hassasiyeti sağlamıştır. Weissman vd. (2007) çalışmalarında doğru sayısı, bilgi ve teorik miktara dayalı yeni bir yönlendirme yöntemi olan karşılıklı bilgi (mutual information) ve maksimum Fisher bilgisi yöntemlerini BÇAT simülasyonunda sınıflandırma kararları bağlamında karşılaştırmışlardır. BÇAT tasarımında kümülatif doğru sayılarına göre dört yol belirlemişlerdir. Araştırma sonucunda sınıflandırma açısından üç yönlendirme yöntemi de benzer şekilde performans göstermiştir. Bu çalışma sonuçlarına göre de MFI ve DS yönlendirme yöntemleri aynı derecede iyi performans göstermişlerdir. Bu da gösteriyor ki, testlerin kullanılış amacına uygun olacak şekilde, bireyleri gruplandırma, sınıflandırma gibi doğru sayısının önemli olduğu sınavlarda doğru sayısı yönlendirme yöntemi, maksimum Fisher gibi bilgi tabanlı yönlendirme yaklaşımları yerine tercih edilebilir (Weissman, vd., 2007). Weissman, vd., (2007) 'nin çalışmalarından farklı olarak bu araştırma koşullarında

MFB yönlendirme yönteminin DS yönlendirme yönteminden daha iyi ölçüm kesinliğine sahip olması, DS yönlendirme yönteminde kesme puanı belirleme yönteminden kaynaklanmış olabilir yorumunu yapabiliriz.

Sonuçlarda iki aşamalı panel desenlerinden üç aşamalı panel desenlerine doğru geçiş yaptıkça genel olarak mutlak yanlılık değerlerinin ve RMSE değerlerinin düştüğü görülmektedir. Literatürde de benzer şekilde aşama sayısı arttığında yetenek kestirimindeki hata miktarının azaldığı gözlenmiştir (Doğruöz, 2018; Patsula, 1999).

ANOVA testi sonuçlarına baktığımızda tüm koşullarda anlamlı farklılık bulunmaktadır. Tüm faktörler kontrol altına alındığında, ayrı ayrı ortalama mutlak yanlılık, RMSE ve korelasyon varyansını en çok açıklayan faktörün yönlendirme yöntemi ve panel deseni etkileşimi olduğu görülmüştür. Test uzunluğu ise ana etkilerden varyansları açıklayan en yüksek koşuldur. Bu sonuç çok kategorili madde oranı %50 olduğu koşullarda, test uzunluğu 40 ve 60 madde koşullarının çoğunda yüksek korelasyon değerleri elde edilmesini açıklamaktadır. Test uzunluğu koşulları açısından sonuçları incelediğimizde test uzunluğu arttıkça ortalama mutlak yanlılık ve RMSE değerlerinin düştüğü, korelasyon değerlerinin ise yükseldiği gözlenmiştir. Bu sonuç literatürdeki BÇAT çalışmaları ile paralellik göstermektedir (Erdem-Kara, 2019; Kim, vd., 2013; Patsula, 1999).

## **Öneriler**

### ***Araştırmacılara Yönelik Öneriler***

Bu çalışmada iki ve üç kategorili maddelerden oluşan karma testlerde çok kategorili madde oranı %10, %30 ve %50 olacak şekilde test birleştirme yapılmıştır. Gelecekteki çalışmalarda çok kategorili maddelerdeki kategori sayısı artırılarak farklı çok kategorili madde oranlarında sonuçlar incelenebilir.

Çalışmada çok kategorili maddeler uygulanan tüm testin %10, %30 ve %50'si olacak şekilde panel oluşturma aşamasında modüllere rastgele dağıtılmıştır. Testlerdeki çok kategorili maddeler farklı sistematik düzenler şeklinde modüllere dağıtılabilir. Çalışmada 1-2 ve 1-4 panel deseni ve yönlendirme yöntemi koşullarında etkileşimler olması açısından, farklı özelliklere sahip karma testlerden oluşan madde havuzlarının BÇAT panel desenleri ile nasıl etkileşim gösterdiğini incelemek yararlı olacaktır.

Çalışmada bireylerin yetenek seviyeleri dağılımı normal dağılım olacak şekilde simülasyon yapılmıştır. Gelecekte yapılacak çalışmalarda farklı dağılım türleri kullanılarak, çeşitli koşullarda yetenek kestirimine etkisi incelenebilir.

Karma testlerden oluşan BÇAT'larda maddeye maruz kalma kontrol yöntemlerinin ölçme hassasiyetine etkisi incelenebilir. Bu çalışmada benzer sonuçlar elde edilen koşulların, maddeye maruz kalma kontrolü ile ilişkili olup olmadığını incelemek ilginç olacaktır.

Çalışmada MFB yönlendirme yöntemi DS yönlendirme yöntemine göre daha iyi sonuçlar vermiştir. Gelecekte yapılacak çalışmalarda kullanılan yönlendirme yöntemleri sınıflama doğruluğu açısından incelenebilir.

Çalışmada otomatik test birleştirme yöntemi kullanılmıştır. Anında (on the fly) test birleştirme yöntemi ile farklı koşullarda yetenek kestirimlerinin doğruluğu incelenebilir.

Çalışmada aşamalar arasında sabit yollar izlenmiştir. Gelecekte yapılacak çalışmalarda farklı örüntüde yolların ya da rastgele yolların yetenek benzer koşullarda yetenek kestirimine etkisi araştırılabilir.

Bu çalışma bir simülasyon araştırmasıdır. Aynı koşullarda gerçek uygulamalarda elde edilen sonuçlarla karşılaştırılması, simülasyon çalışmalarının geçerliği açısından faydalı bir çalışma olacaktır. Ayrıca çalışma koşulları farklılaştırılarak yeni simülasyon çalışmaları yapmaları önerilebilir.

### ***Uygulayıcılara Yönelik Öneriler***

Genel olarak test uzunluğu ve çok kategorili madde oranı arttığında korelasyon değeri artmaktadır. Ancak çok kategorili madde oranı ve testteki madde sayısı en yüksek olduğu koşulda tüm panel desenlerinde aynı korelasyon değeri elde edilmiştir. Bu nedenle 60 test uzunluğu ve %50 çok kategorili madde oranı koşulunda panel desenlerinin ölçüm hassasiyetini etkilemediği yorumu yapılabilir. Gerçek BÇAT uygulaması söz konusu olduğunda bu sonuca dikkat edilmelidir.

Çalışmaya benzer koşullardaki uygulamalarda MFB yönlendirme yönteminin daha iyi sonuç verdiği dikkat edilmelidir.

Çalışma koşullarında 1-2 ve 1-4 panel deseni ve yönlendirme yöntemi arasında etkileşim bulunmaktadır. Yani bu iki panel deseninde MFB ve DS yönlendirme yöntemlerinden hangisinin kullanıldığının bir önemi yoktur. Gerçek uygulama koşullarında bu panel desenlerinden uygun olanı kullanılabilir.

### Kaynaklar

- Adema, J. J. (1990). The construction of customized two-stage tests. *Journal of Educational Measurement*, 27(3), 241-253. <https://doi.org/10.1111/j.1745-3984.1990.tb00746.x>
- Armstrong, R. D. (2002). Routing rules for multi-form structures (LSAC Computerized Testing Report No. 02-08). Newtown, PA: Law School Admission Council, Inc. <https://searchworks.stanford.edu/view/6794803>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In: Lord, F.M. & Novick, M.R. (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, Mass.: Addison-Wesley.
- Boztunç-Öztürk, N. (2019). How the Length and Characteristics of Routing Module Affect Ability Estimation in ca-MST?. *Universal Journal of Educational Research*, 7(1), 164-170. <https://doi.org/10.13189/ujer.2019.070121>
- Breithaupt, K., & Hare, D. R. (2007). Automated simultaneous assembly of multistage testlets for a high-stakes licensing examination. *Educational and Psychological Measurement*, 67(1), 5-20. <https://doi.org/10.1177/00131644062881>
- Cikrikci, N., Yalcin, S., Kalender, I., Emrah, G. Ü. L., Cansu, A. Y. A. N., Uyumaz, G., ... & Kamis, O. (2020). Development of a Computerized Adaptive Version of the Turkish Driving Licence Exam. *International Journal of Assessment Tools in Education*, 7(4), 570-587. <https://doi.org/10.21449/ijate.716177>
- Colvin, Kimberly F. (2014). *Effect of automatic item generation on ability estimates in a multistage test* (Doctoral dissertations). University of Massachusetts, Amherst. <https://doi.org/10.7275/5428533>
- Dallas, A. (2014). *The effects of routing and scoring within a computer adaptive multi-stage framework* (Doctoral dissertation). University of North Carolina, Greensboro.



- De Ayala, R. J. (2009). *The theory and practice of item response theory*. NY: Guilford.
- Dođruöz, E. (2018), *Bireyselleştirilmiş çok aşamalı testlerin test birleştirme yöntemlerine göre incelenmesi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Ercikan, K., Schwarz, R. D., Julian, M. W., Burket, G. R., Weber, M. M., & Link, V. (1998). Calibration and scoring of tests with multiple-choice and constructed-response item types. *Journal of Educational Measurement*, 35(2), 137-154. <https://doi.org/10.1111/j.1745-3984.1998.tb00531.x>
- Erdem-Kara, B. (2019), *Deđişen madde fonksiyonu gösteren madde oranının bireyselleştirilmiş bilgisayarlı ve çok aşamalı testler üzerindeki etkisi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36–49. <https://doi.org/10.1111/emip.12111>
- Frascati Kılavuzu, (2002). *Bilimsel ve Teknolojik Faaliyetlerin Ölçümü, Araştırma ve Deneysel Geliştirme Taramaları İçin Önerilen Standart Uygulama*, OECD.
- Georgiadou E., Triantafillous E., & Economides A. A. (2006). Evaluation Parameters for Computer-Adaptive Testing. *British Journal of Educational Technology*. 37(2), 61-278. <https://doi.org/10.1111/j.1467-8535.2005.00525.x>
- Gorin, J. S., Dodd, B. G., Fitzpatrick, S. J., & Shieh, Y. Y. (2005). Computerized adaptive testing with the partial credit model: Estimation procedures, population distributions, and item pool characteristics. *Applied Psychological Measurement*, 29(6), 433-456. <https://doi.org/10.1177/0146621605280072>

- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckease, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360. <https://doi.org/10.1111/j.1745-3984.1984.tb01039.x>
- Hambleton, R. K., & Xing, D. (2006). Optimal and nonoptimal computer-based test designs for making Pass–Fail decisions. *Applied Measurement in Education*, 19(3), 221–239. [https://doi.org/10.1207/s15324818ame1903\\_4](https://doi.org/10.1207/s15324818ame1903_4)
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications* (Vol. 7). Springer Science ve Business Media.
- Han, K. T. (2007). WinGen: Windows Software That Generates Item Response Theory Parameters and Item Responses. *Applied Psychological Measurement*, 31(5), 457–459. <https://doi.org/10.1177/0146621607299271>
- Han K. T. & Guo F. (2011). Potential impact of item parameter drift due to practice and curriculum change on item calibration in computerized adaptive testing (GMAC® Research Reports, RR-11-02). <https://pdfs.semanticscholar.org/ac5f/cb8d43575b3e1661e51a08c060e353712509.pdf>
- Han, K. T., & Hambleton, R. K. (2014). User's manual for wingen 3: windows software that generates IRT model parameters and item responses (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts.
- Harwell, M., Stone, C. A., Hsu, T., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101–125. <https://doi.org/10.1177/01466216960200020>

- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, 26(2), 44-52. <https://doi.org/10.1111/j.1745-3992.2007.00093.x>
- Kim, J., Chung, H., & Dodd, B. G. (2010). Comparing routing methods in the multistage test based on the partial credit model. In *annual meeting of the American Educational Research Association, Denver, CO*.
- Kim, J., Chung, H., Dodd, B. G., & Park, R. (2012). Panel design variations in the multistage test using the mixed-format tests. *Educational and Psychological Measurement*, 72(4), 574-588. <https://doi.org/10.1177/001316441142897>
- Kim, J., Chung, H., Park, R., & Dodd, B. G. (2013). A comparison of panel designs with routing methods in the multistage test with the partial credit model. *Behavior research methods*, 45(4), 1087-1098. <https://doi.org/10.3758/s13428-013-0316-3>
- Kim, J., & Dodd, B. G. (2014). Mixed-format multistage tests: issues and methods. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 55–67). Boca Raton, FL: CRC Press.
- Kim, H., & Plake, B. S. (1993). *Monte carlo simulation comparison of two-stage testing and computerized adaptive testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Atlanta, GA, April 13-15, 1993). <https://eric.ed.gov/?id=ED357041>
- Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement*, 14(4), 367-386. <https://doi.org/10.1177/01466216900140040>

- Lim, H. (2019). A Top-down Approach for Optimally Designing Multistage-adaptive Tests (Doctoral dissertation). University of Massachusetts, Amherst. <https://doi.org/10.7275/14218580>
- Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika*, 36(3), 227-242. <https://doi.org/10.1007/BF02297844>
- Luo, X., & Kim, D. (2018). A top-down approach to designing the computerized adaptive multistage test. *Journal of Educational Measurement*, 55(2), 243-263. <https://doi.org/10.1111/jedm.12174>
- Luecht, R. M., & Nungester, R. J. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, 35(3), 229-249. <http://www.jstor.org/stable/1435202>
- Luecht, R. M. (2003, April). Exposure control using adaptive multi-stage item bundles. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Magis, D., Yan, D., & von Davier, A. A. (2017). Overview of Adaptive Testing. In *Computerized Adaptive and Multistage Testing with R* (pp. 1-5). Springer, Cham.
- Magis, D., Yan, D., & von Davier A. A. (2018). Package 'mstR': Procedures to generate patterns under multistage testing. <https://cran.r-project.org/web/packages/mstR/mstR.pdf>
- Ostini, R., & Nering, M. L. (2006). Polytomous item response theory models. Thousand Oaks: Sage.
- Park, R. (2015). Investigating the impact of a mixed-format item pool on optimal test designs for multistage testing (Doctoral dissertation). The University of Texas, Austin. <http://hdl.handle.net/2152/31011>

- Park, R., Kim, J., Chung, H., & Dodd, B. G. (2014). Enhancing pool utilization in constructing the multistage test using mixed-format tests. *Applied Psychological Measurement, 38*(4), 268-280. <https://doi.org/10.1177/0146621613515>
- Park, R., Kim, J., Chung, H., & Dodd, B. G. (2017). The Development of MST Test Information for the Prediction of Test Performances. *Educational and Psychological Measurement, 77*(4), 570–586. <https://doi.org/10.1177/0013164416662960>
- Patsula, L. N. (1999). A comparison of computerized adaptive testing and multistage testing. (Doctoral dissertation). University of Massachusetts, Amherst.
- R Core Team. (2018). R: A language and environment for statistical computing: R foundation for statistical computing.
- Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.
- Reckase M. D. (1974). An interactive computer program for tailored testing based on the one-parameter logistic model. *Behavior Research Methods and Instrumentation, 6*, 208-212. <https://doi.org/10.3758/BF03200330>
- Rosa, K., Swygert, K. A., Nelson, L., & Thissen, D. (2001). Item response theory applied to combinations of multiple-choice and constructed-response items—scale scores for patterns of summed scores. In D. Thissen ve H. Wainer (Eds.), *Test scoring* (pp. 253–292). Routledge. <https://doi.org/10.4324/9781410604729>
- Sarı, H.İ., Yahşi Sarı, H., & Huggins Manley, A.C. (2016). Computer adaptive multistage testing: Practical issues, challenges and principles. *Journal of Measurement and Evaluation in Education and Psychology, 7*(2), 388–406. <https://doi.org/10.21031/epod.280183>

- Sari, H. I., & Raborn, A. (2018). What Information Works Best?: A Comparison of Routing Methods. *Applied psychological measurement*, 42(6), 499-515. <https://doi.org/10.1177/0146621617752990>
- Sarı, H. İ. (2019). Investigating Consequences of Using Item Pre-knowledge in Computerized Multistage Testing. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 39 (2), 1113-1134. <https://doi.org/10.17152/gefad.535376>
- Şenel, S., & Kutlu, Ö. (2018). Computerized adaptive testing design for students with visual impairment. *Eğitim ve Bilim Dergisi*, 43 (194), 261-284. <https://doi.org/10.15390/EB.2018.7515>
- Wang, K. (2017). A fair comparison of the performance of computerized adaptive testing and multistage adaptive testing (Doctoral dissertation). University of Michigan State, Michigan.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational measurement*, 24(3), 185-201. <https://doi.org/10.1111/j.1745-3984.1987.tb00274.x>
- Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6(4), 473-492. <https://doi.org/10.1177/014662168200600408>
- Weiss, D. J. (1983). Latent trait theory and adaptive testing. In Weiss D. J. (Ed.), *New horizons in testing* (pp. 5-7). New York: Academic Press.
- Weiss, D. J. & Kingsbury, G. G. (1984). Application of computer adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>

- Weissman, A., Belov, D. I., & Armstrong, R. D. (2007). Information-based versus number-correct routing in multistage classification tests, *LSAC Computerized Testing Report*, 7(5). Newtown, PA: Law School Admission Council.
- Xing, D., & Hambleton, R. K. (2004). Impact of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. *Educational and Psychological Measurement*, 64(1), 5-21.  
<https://doi.org/10.1177/0013164403258393>
- Yamamoto, K., H. Shin & L. Khorramdel (2019), Introduction of multistage adaptive testing design in PISA 2018, OECD Education Working Papers, No. 209, OECD Publishing, Paris, <https://doi.org/10.1787/b9435d4b-en>
- Yang, L. (2016). Enhancing item pool utilization when designing multistage computerized adaptive tests (Doctoral Dissertation). Michigan State University, Michigan.
- Zenisky, A. L. (2004). *Evaluating the effects of several multi-stage testing design variables on selected psychometric outcomes for certification and licensure assessment* (Doctoral dissertation). University of Massachusetts, Amherst.
- Zenisky, A. L., Hambleton, R. K., & Luecht, R. M. (2010). Multistage testing: Issues, designs, and research. In W. J. van der Linden ve C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 355-372). New York, NY: Springer.
- Zheng, Y., Nozawa, Y., Gao, X., & Chang, H. H. (2012). Multistage adaptive testing for a large-scale classification test: The designs, automated heuristic assembly, and comparison with other testing modes. ACT Research Reports 2012-6.  
[https://www.act.org/content/dam/act/unsecured/documents/ACT\\_RR2012-6.pdf](https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2012-6.pdf)

## EK-A: Testteki Çok Kategorili Madde Yüzdesi %10 Olduğunda Kesme Noktalarının Grafikleri

Madde  
sayısı

Panel deseni

1-2

1-3

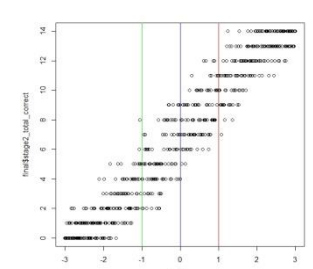
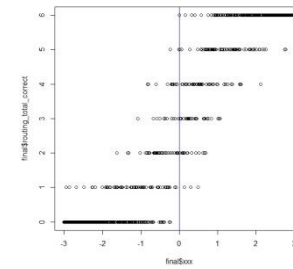
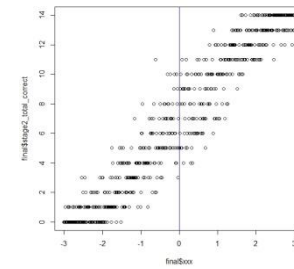
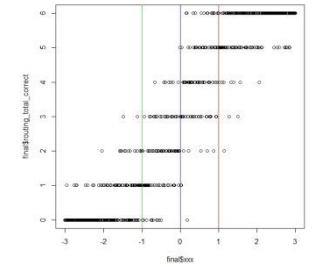
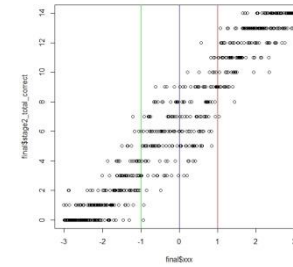
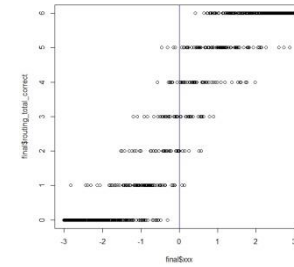
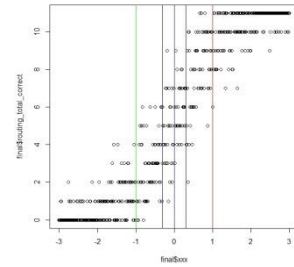
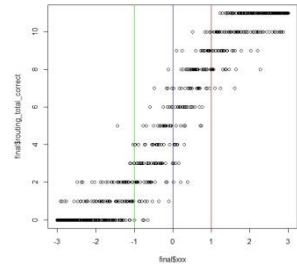
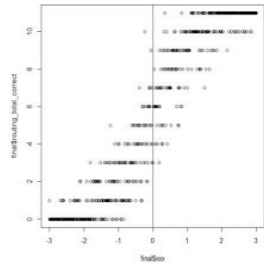
1-4

1-2-2

1-2-3

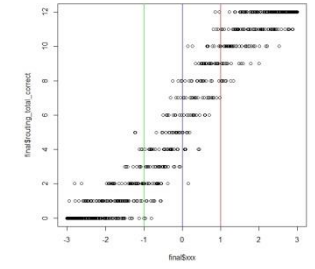
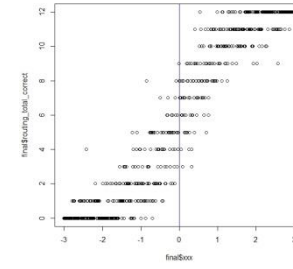
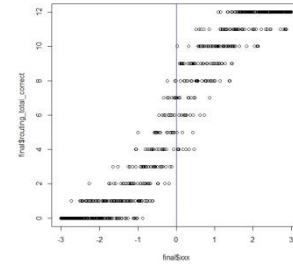
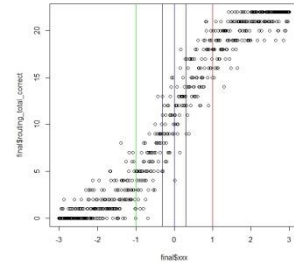
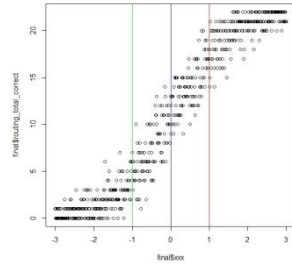
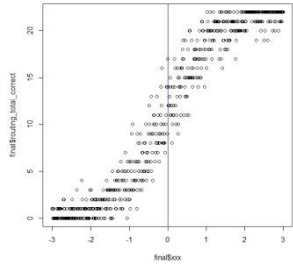
1-3-3

20

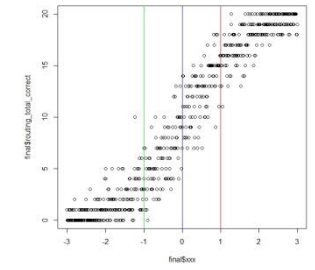
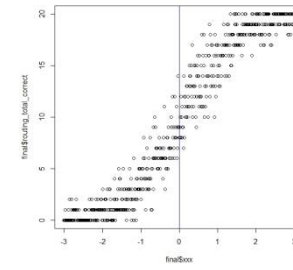
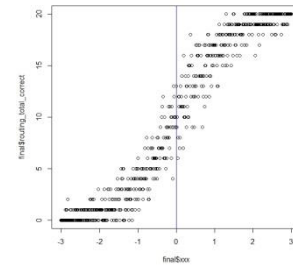
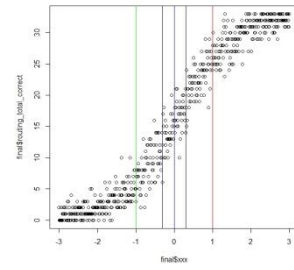
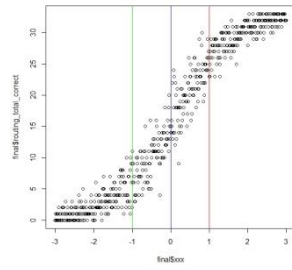
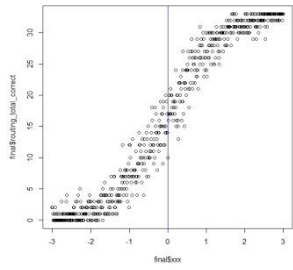


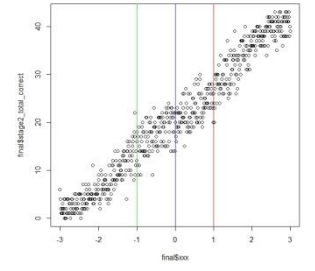
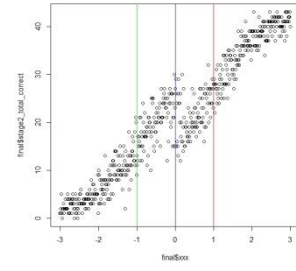
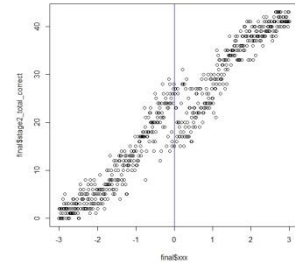


40



60





## EK-B: Testteki Çok Kategorili Madde Yüzdesi %30 Olduğunda Kesme Noktalarının Grafikleri

Madde  
sayısı

Panel deseni

1-2

1-3

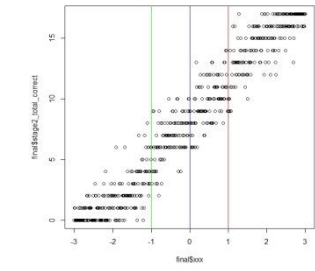
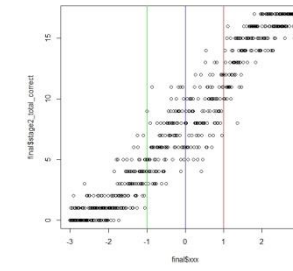
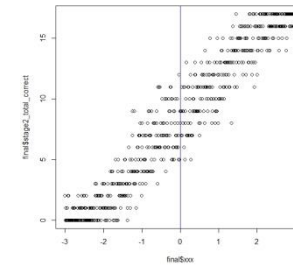
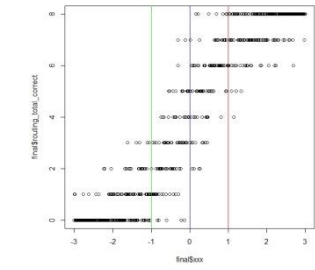
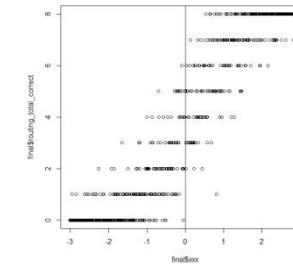
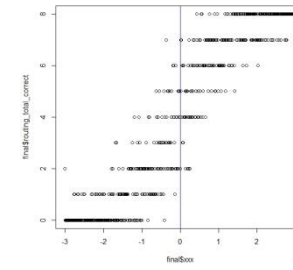
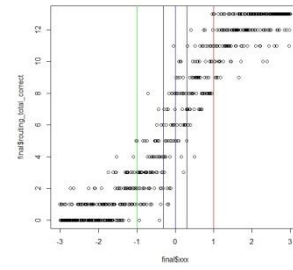
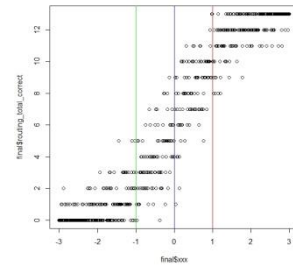
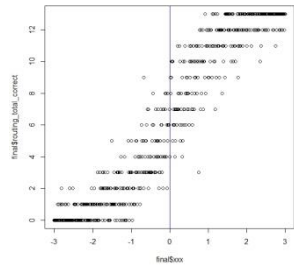
1-4

1-2-2

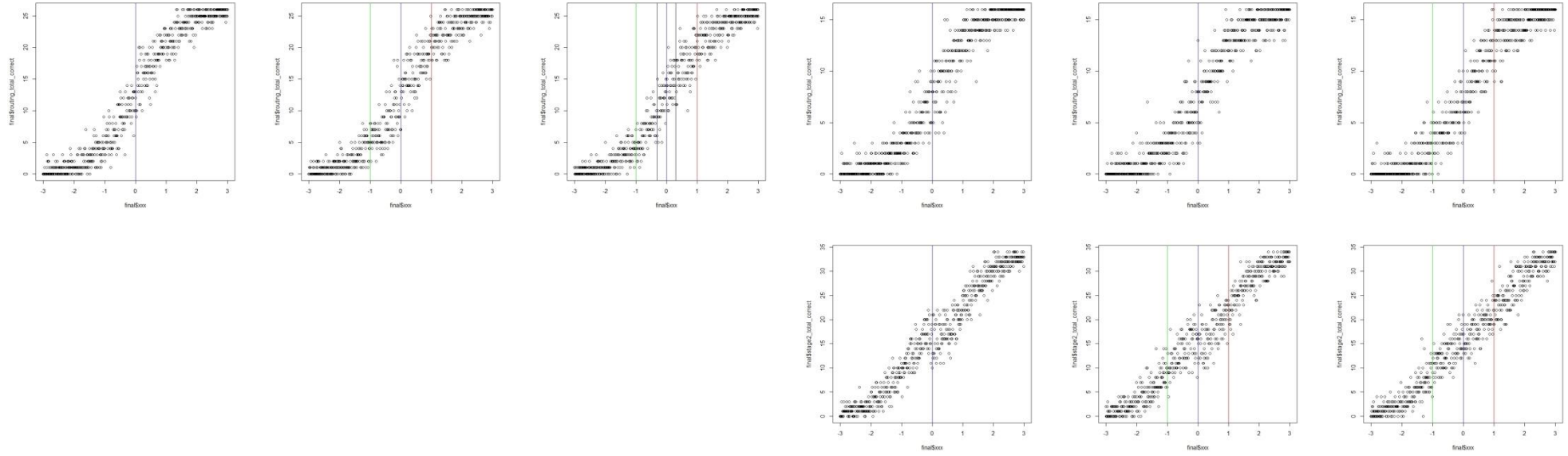
1-2-3

1-3-3

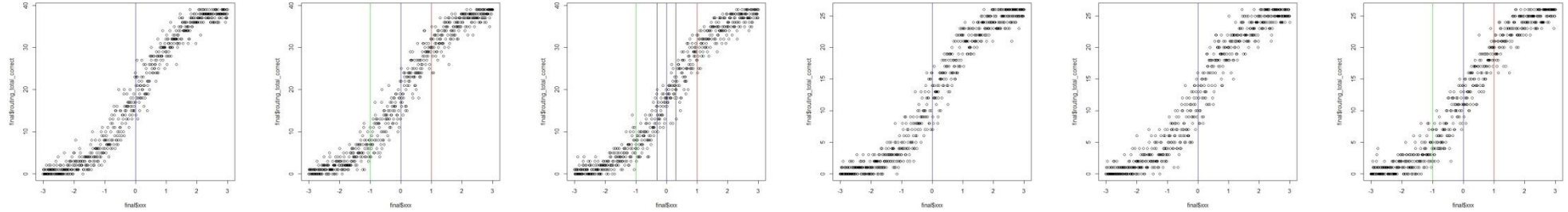
20

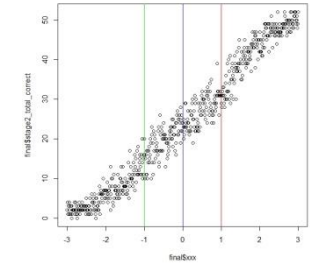
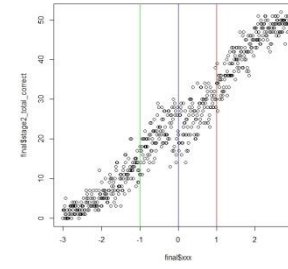
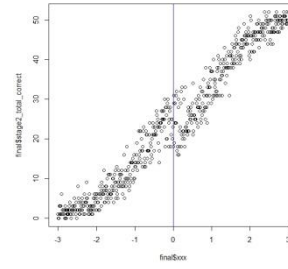


40



60





## EK-C: Testteki Çok Kategorili Madde Yüzdesi %50 Olduğunda Kesme Noktalarının Grafikleri

Madde  
sayısı

Panel deseni

1-2

1-3

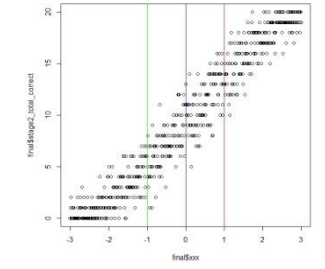
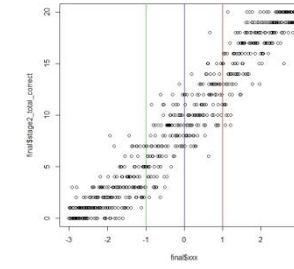
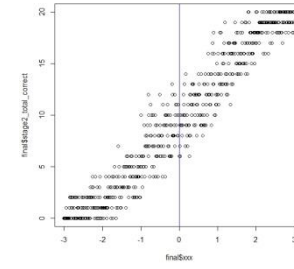
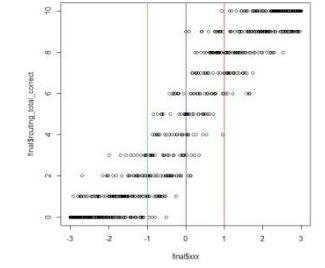
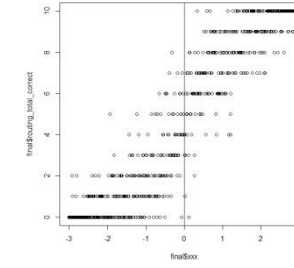
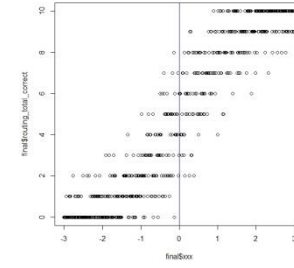
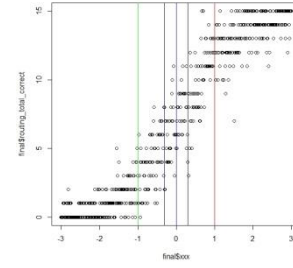
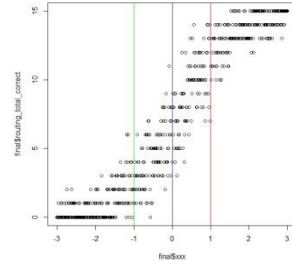
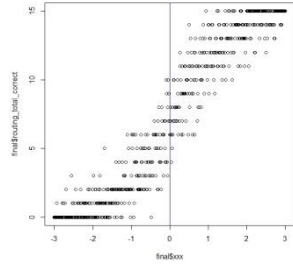
1-4

1-2-2

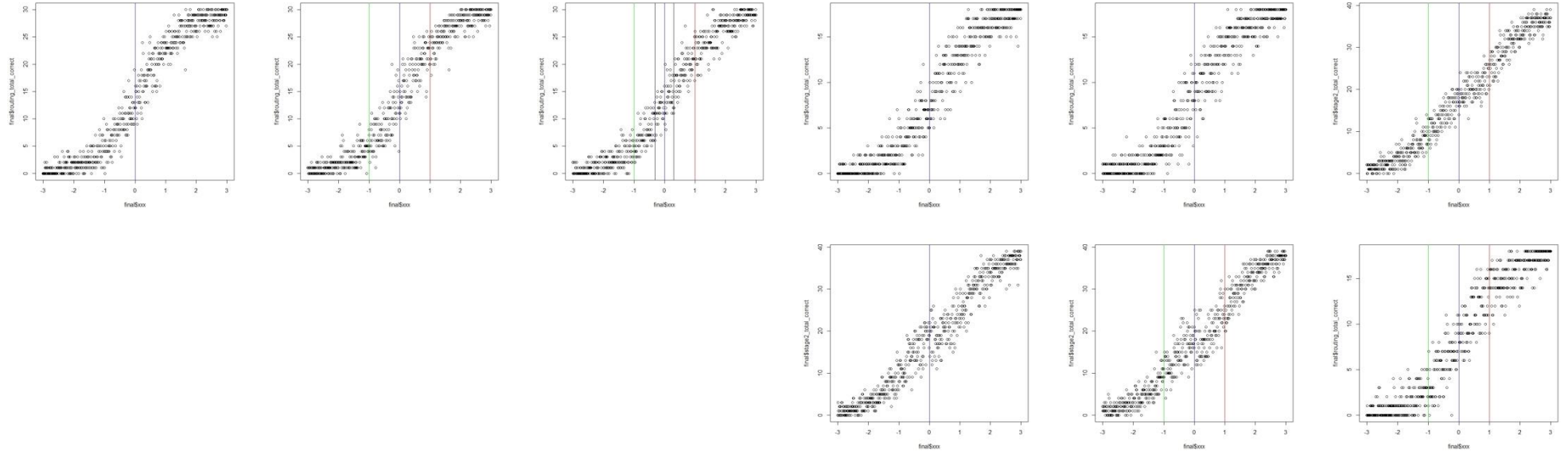
1-2-3

1-3-3

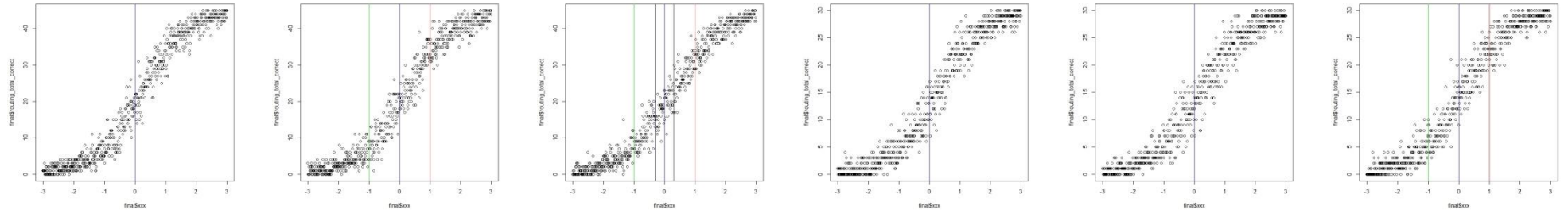
20

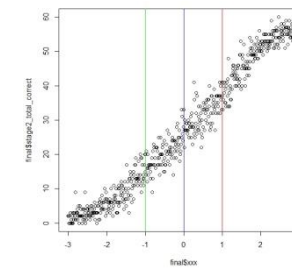
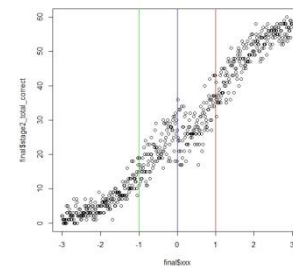
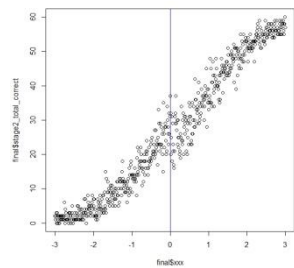


40




60







## EK-Ç: Araştırma Etik Komisyon İzin Muafiyeti Formu

	<b>Hacettepe Üniversitesi</b> <b>Eğitim Bilimleri Enstitüsü</b> <b>Tez Çalışması Etik Komisyon İzin Muafiyeti Formu</b>	<b>F46</b>
08 / 05 / 2023		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Başkanlığına		
Tez Başlığı	Bireye uyarlanmış çok aşamalı testlerde karma testlerin yetenek kestirimine etkisi	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
<ol style="list-style-type: none"> <li>1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.</li> <li>2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.</li> <li>3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.</li> <li>4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.</li> <li>5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.</li> </ol>		
Çalışmada kullanacağım veriler:		
<input checked="" type="checkbox"/> Kamusal erişime açık (buraya yazınız): OECD-PISA 2018		
<input type="checkbox"/> Özel izin ve onaya tabi (buraya yazınız): .....		
<input checked="" type="checkbox"/> Üretilmiş veri (buraya yazınız): Bilgisayar programında üretilmiş veri		
<input type="checkbox"/> Diğer (buraya yazınız): .....		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
Hasibe YAŞI SARI <small>(Araştırmacı Adı Soyadı, İmzası)</small>		
<b>Araştırmacı Bilgileri</b>		
Adı Soyadı	Hasibe YAŞI SARI	
Öğrenci ise No	N18149277	
Ana Bilim Dalı	Eğitim Bilimleri	
Programı	Eğitimde Ölçme ve Değerlendirme Bilim Dalı	
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
<b>Danışman Görüşü ve Onayı*</b>		
Prof. Dr. Hülya KELECIOĞLU <small>(İmza)</small> <small>(Danışman <b>Unvanı</b>, Adı ve Soyadı)</small>		
*Tez ve tezden üretilen yayınlarda gerekli		

**EK-D: Etik Beyanı**

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

06/ 06/ 2023

(İmza)  
Hasibe YAŞI SARI

**EK-E: Doktora Tez Çalışması Orijinallik Raporu**

05 / 06 / 2023

HACETTEPE ÜNİVERSİTESİ  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: Bireyselleştirilmiş Çok Aşamalı Testlerde Çok Kategorili Madde Oranının Yetenek Kestirimine Etkisi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
05 /06 /2023	79	113.823	29 /05 /2023	%6	2109690934

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

**Ad Soyadı:** Hasibe YAHŞI SARI

**Öğrenci No.:** N18149277

**Ana Bilim Dalı:** Eğitim Bilimleri

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

İmza

**DANIŞMAN ONAYI**

UYGUNDUR.  
(Prof. Dr., Hülya KELECİOĞLU)

## EK-F: Dissertation Originality Report

06 / 05 / 2023

HACETTEPE UNIVERSITY  
Graduate School of Educational Sciences  
To The Department of Educational Sciences

Thesis Title: The Effect of Polytomous Item Ratio on Ability Estimation in Multistage Tests

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
06 /05 /2023	79	113.823	05 /29 /2023	%6	2109690934

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

**Name Lastname:** Hasibe YAHSI SARI

**Student No.:** N18149277

**Department:** Educational Sciences

**Program:** Educational Measurement and Evaluation

**Status:**  Masters  Ph.D.  Integrated Ph.D.

Signature

### ADVISOR APPROVAL

APPROVED  
(Prof.Dr., Hülya KELECİOĞLU)

## EK-G: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/ Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

06 /06 /2023

(imza)

Hasibe YAŞI SARI

---

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir". Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlerle ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.  
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

\*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

