



**HACETTEPE ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TEST UYGULAMALARINDA MADDEYİ YENİDEN  
CEVAPLAYABİLMENİN ÖLÇME HATASINA VE YANLILIĞINA ETKİSİNİN İNCELENMESİ

Ömer Faruk ŞEN

Doktora Tezi

Ankara, 2022

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

BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TEST UYGULAMALARINDA MADDEYİ YENİDEN  
CEVAPLAYABİLMENİN ÖLÇME HATASINA VE YANLILIĞINA ETKİSİNİN İNCELENMESİ

THE EFFECT OF ITEM REVIEW IN COMPUTERIZED ADAPTIVE TESTING ON  
MEASUREMENT ERROR AND BIAS

Ömer Faruk ŞEN

Doktora Tezi

Ankara, 2022

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

mer Faruk ŐEN'in hazırladıđı "Bilgisayarda BireyselleŐtirilmiŐ Test Uygulamalarında Maddeyi Yeniden Cevaplayabilmenin lme Hatasına ve Yanlılıđına Etkisinin İncelenmesi" baŐlıklı bu alıŐma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde lme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiŐtir.

J¼ri BaŐkanı Prof. Dr. Selahattin GELBAL İmza

J¼ri Üyesi (DanıŐman) Prof. Dr. H¼lya KELECİĐLU İmza

J¼ri Üyesi Prof. Dr. Nuri DOĐAN İmza

J¼ri Üyesi Do. Dr. Murat Dođan ŐAHİN İmza

J¼ri Üyesi Dr. đr. Üyesi mer KUTLU İmza

Bu tez Hacettepe niversitesi Lisans¼st¼ Eđitim, đretim ve Sınav Ynetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 27 / 06 / 2022 tarihinde uygun gr¼lm¼Ő ve Enstit¼ Ynetim Kurulunca ..... / ..... / 2022 tarihi itibarıyla kabul edilmiŐtir.

Prof. Dr. Selahattin GELBAL  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bilgisayarda bireyselleştirilmiş test (BBT) uygulaması daha az madde ile daha hassas ve güvenilir sonuçlar vermesi sebebiyle tercih edilen bir uygulama biçimi olurken aynı zamanda bazı kısıtlamaları da içermektedir. Bu kısıtlamaların başında BBT uygulamalarında maddeyi yeniden cevaplayabilmeye izin verilmemesi gelmektedir. BBT'deki bir maddenin cevabını değiştirmek, daha önceden uygulanan maddelerin, sınava giren kişinin yeteneğini kestirmek için artık uygun olmamasına neden olabilmektedir. Cevap değişikliği sorunu için olası bir çözüm, yalnızca geçici yetenek tahminlerinin belirsizliğini hesaba katmakla kalmayıp aynı zamanda cevap değişikliğini mümkün kılan en olası optimal başlangıç değerini tanımlamaktır. Bu doğrultuda simülasyon çalışması olarak yürütülen bu araştırmada öncül bilginin tanımlandığı BBT uygulamalarında maddeyi yeniden gözden geçirmeye izin verildiğinde elde edilen yetenek kestirimlerinin standart hatasının ve yanlılığının, testin başlangıç kuralına, nihai yetenek kestirim yöntemine ve test uzunluğuna göre nasıl değiştiğini belirlemek amaçlanmıştır. Bu değerler açısından kestirimler arasında fark olup olmadığını incelemek için üç faktörlü ANOVA analizi yapılmıştır. Araştırma sonucuna göre geleneksel BBT uygulamalarında cevap değişikliğine izin verildiğinde bütün koşullarda yanlılık ve standart hata değerlerinin arttığı tespit edilmiştir. Ancak test bataryasında yer alan alt testler arasındaki korelasyondan yola çıkarak tanımlanan öncül bilgilerin hem başlama kuralında hem de nihai yetenek kestirim yönteminde kullanıldığı BBT uygulamalarında bireyler Wainer test stratejisini kullansalar bile izin verilmeyen geleneksel BBT uygulamasıyla benzer sonuçlar elde edilmektedir. Sonuç olarak öncül bilgiler BBT algoritmasını daha kullanışlı kılmakta ve test stratejilerine karşı uygulamayı dirençli hale getirmektedir.

**Anahtar sözcükler:** bilgisayarda bireyselleştirilmiş test uygulamaları, maddeyi yeniden cevaplama, yetenek kestirim yöntemleri, öncül bilgi, ölçme hassasiyeti

### **Abstract**

While computerized adaptive testing (CAT) is a preferred form of test practice because of providing more accurate and reliable results with fewer items, it also has some limitations. At the beginning of these restrictions, item review is not allowed in CAT. Changing the answer of an item in the CAT may cause previously administered items to no longer be suitable for estimating the test taker's real ability. A possible solution to the item review problem is to define the most likely optimal initial value that not only takes into account the uncertainty of temporal ability estimates, but also enables response change. In this direction, it is aimed to determine how the standard error and bias of the ability estimations obtained when prior knowledge is defined in reviewable CAT varies according to the initial rule of the test, the final ability estimation method and the test length. Three way ANOVAs were performed to examine whether there was a difference between the estimations in terms of standard error and bias. Results suggest that the bias and standard error values increased in all conditions when response changes were allowed in the CAT. However, when prior information defined based on the correlation between the subtests in the test battery is used in both the starting rule and the final ability estimation method, results similar to a traditional CAT are obtained in reviewable CAT even if individuals use the Wainer test strategy. As a result, prior knowledge makes the CAT algorithm more useful and resistant to Wainer test strategy.

**Keywords:** computerized adaptive testing, item review, ability estimation methods, prior information, measurement precision

## Teşekkür

Öncelikle danışmanlığımı üstlenerek beni onurlandıran ve gerek konu seçiminde gerekse araştırma, yazım aşamasında bana engin bilgi birikimiyle her zaman rehberlik eden saygıdeğer hocam Prof. Dr. Hülya KELEÇİOĞLU'na en derin saygılarımı ve hürmetlerimi sunarım. Tez izleme komitemde yer alan Prof. Dr. Selahattin GELBAL'a ve Dr. Öğr. Üyesi Ömer KUTLU'ya yol göstericilikleri, bana ve tezime kattıkları için çok teşekkür ederim. Yine Jürimde yer alan Prof. Dr. Nuri DOĞAN ve Doç. Dr. Murat Doğan ŞAHİN hocalarım yapıcı eleştirileri ve destekleri için ayrıca çok teşekkür ederim. Son olarak hayatımın her anında olduğu gibi doktora sürecimde de hep yanımda olan sevgili eşim Demet'e ve onu borçlu olduğum Konur aileme, her şeyden kıymetli oğlum Aras'a ve manevi desteklerini her zaman kalbimde hissettiğim kıymetli anneme, babama ve kardeşlerime teşekkürü bir borç bilirim. Sizler olmasaydınız eksik kalırdım.

**İçindekiler**

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
İçindekiler.....	vi
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	x
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	8
Araştırma Problemi.....	9
Sayıtlar.....	10
Sınırlılıklar.....	11
Tanımlar.....	11
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	12
Madde Tepki Kuramı.....	12
Bilgisayarda Bireyselleştirilmiş Test Uygulamaları.....	16
BBT Uygulamalarında Maddeyi Yeniden Cevaplamaya İzin Veren Yöntemler.....	26
BBT Uygulamalarında Test Stratejileri.....	29
BBT Uygulamalarında Öncül Bilgilerin Kullanılması.....	30
İlgili Araştırmalar.....	34
Bölüm 3 Yöntem.....	44
Araştırmanın Türü.....	44
Araştırma Deseni.....	44
Veri Kümesi.....	46



Veri Toplama Süreci.....	46
Verilerin Analizi .....	54
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	56
Birinci Alt Probleme İlişkin Bulgular ve Yorumlar .....	56
İkinci Alt Probleme İlişkin Bulgular ve Yorumlar .....	64
Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar .....	72
Tartışma.....	80
Bölüm 5 Sonuç ve Öneriler.....	83
Kaynaklar .....	86
EK-A: Araştırma Etik Komisyonu Onay Bildirimi.....	c
EK-B: Etik Beyanı.....	ci
EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu .....	cii
EK-Ç: Thesis/Dissertation Originality Report.....	ciii
EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	civ

## Tablolar Dizini

<b>Tablo 1</b> <i>İkili Puanlanan Madde Tepki Modellerine Ait Matematiksel Fonksiyonlar</i> .....	14
<b>Tablo 2</b> <i>Simülasyon Koşulları ve Düzeyleri</i> .....	45
<b>Tablo 3</b> <i>PISA Alt Testlerine Ait Madde Parametrelerine İlişkin Betimsel İstatistikler</i> .....	46
<b>Tablo 4</b> <i>Fen ve Matematik Alt Testlerine İlişkin Betimsel Değerler</i> .....	47
<b>Tablo 5</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{random}</math> ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık değerleri</i> .....	57
<b>Tablo 6</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{random}</math> ile Belirlendiğinde Elde Edilen Yanlılık ve RMSE Değerlerine Ait ANOVA Sonuçları</i> .....	62
<b>Tablo 7</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{regresyon}</math> ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık Değerleri</i> .....	65
<b>Tablo 8</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{regresyon}</math> ile Belirlendiği Koşullara İlişkin ANOVA Sonuçları</i> .....	70
<b>Tablo 9</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{tekrar}</math> ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık Değerleri</i> .....	73
<b>Tablo 10</b> <i>Başlangıç Yetenek Düzeyi Olarak <math>\theta_{tekrar}</math> ile Belirlendiği Koşullara ilişkin ANOVA Sonuçları</i> .....	78

## Şekiller Dizini

<b>Şekil 1</b> Madde Karakteristik Eğrileri.....	15
<b>Şekil 2</b> Bireylerin Gerçek ve Regresyon Eşitliği ile Belirlenen Fen Yetenek Düzeylerinin Dağılımı .....	50
<b>Şekil 3</b> Regresyon Eşitliği ile Belirlenen Fen BBT Başlangıç Değerlerinin Gerçek Yetenek Düzeyleriyle Karşılaştırılması.....	51
<b>Şekil 4</b> Aynı Testin Bir Önceki Sonucu Kullanılarak Belirlenen Fen BBT Başlangıç Değerlerinin Gerçek Yetenek Düzeyleriyle Karşılaştırılması.....	52
<b>Şekil 5</b> Başlangıç Yetenek Düzeyi $\theta_{random}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği.....	58
<b>Şekil 6</b> Başlangıç Yetenek Düzeyi $\theta_{random}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanlılık Grafiği .....	59
<b>Şekil 7</b> Başlangıç Yetenek Düzeyi $\theta_{regresyon}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği.....	67
<b>Şekil 8</b> Başlangıç Yetenek Düzeyi $\theta_{regresyon}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanlılık Grafiği .....	69
<b>Şekil 9</b> Başlangıç Yetenek Düzeyi $\theta_{tekrar}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği.....	75
<b>Şekil 10</b> Başlangıç Yetenek Düzeyi $\theta_{tekrar}$ Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanlılık Grafiği .....	77

## Simgeler ve Kısaltmalar Dizini

**BBT:** Bilgisayarda Bireyselleştirilmiş Test

**PISA:** Programme for International Student Assessment

**RMSE:** Hataların Ortalama Karekökü (Root Mean Squared Error)

**MTK:** Madde Tepki Kuramı

**ÖSYM:** Öğrenci Seçme ve Yerleştirme Merkezi

## Bölüm 1

### Giriş

Bu bölümde problem durumu, araştırmanın amacı ve önemi, problem cümlesi, alt problemler, sayılılar, sınırlılıklar ve tanımlar yer almaktadır.

#### Problem Durumu

Teknoloji, psikometri ve eğitim alanlarındaki son gelişmeler hayatımızın birçok alanını etkilemekle birlikte eğitsel değerlendirmelere de yeni bir boyut kazandırmıştır (Rezaie & Golshan, 2015). Geleneksel yöntem olan kâğıt-kalem testlerinin yerini elektronik testler almaktadır. Bilgisayarların kullanım alanlarının artması ve erişimin kolaylaşması ile bilgisayarlar bireylerin bilgi, beceri ve yeteneklerinin ölçülmesinde kâğıt-kalem testlerine nazaran birçok kolaylık sağlamaktadır.

Bilgisayar ortamında uygulanan testler; lineer testler ve bireyselleştirilmiş testler (BBT) olmak üzere iki başlıkta ele alınabilir. Kâğıt-kalem testlerinin birebir bilgisayar ortamında uygulanmasına bilgisayar tabanlı lineer test denilmektedir. Bilgisayar ortamında gerçekleştirilen lineer test uygulamaları, puanlama ve sınav formu oluşturmada kolaylık sağlamaktadır (Magis ve ark., 2017). Ülkemizde ÖSYM tarafından yapılan Yabancı Dil Sınavı (YDS) kâğıt-kalem uygulamasının yanı sıra bilgisayar ortamında da gerçekleştirilerek aynı gün puanlama imkânı sunmaktadır. Her ne kadar bilgisayar tabanlı lineer testlerin kâğıt-kalem testlerinden üstün yönleri olsa da aynı sınırlılıklara sahiptir. Bilgisayar tabanlı lineer testlerin temelinde klasik test kuramı (KTK) yer almaktadır. Kâğıt-kalem testlerinde olduğu gibi bilgisayar tabanlı lineer testlerde de bütün bireyler aynı test maddelerini cevaplarlar. KTK tabanlı bir test, evreni temsil eden iki farklı gruba uygulandığında grupların madde parametreleri farklı olabilmektedir. Bunun yanı sıra birey paralel olan iki testi yanıtladığında, bireyin yetenek düzeyinin de farklı kestirilmesi KTK'nın en göze çarpan sınırlılıkları (kısıtlılıklar) arasındadır. Bir diğer sınırlılık ise KTK ile geliştirilen bir testte güvenilirliği arttırmak amacıyla orta güçlükteki sorulara daha fazla yer

verilmesidir (Gelbal, 2013; Dolan & Burling, 2012). Bu durum, orta yetenek düzeyine sahip bireylerin yetenek kestirimlerindeki hata miktarı az iken uç yetenek düzeyine sahip bireylerin yetenek kestirimlerindeki hata miktarının yüksek olmasına neden olmaktadır (Weiss, 2011). Sonuç olarak, KTK tabanlı kâğıt-kalem testlerinde sınava giren tüm kişilerin gerçek yetenek düzeylerine yakın yetenek kestirimlerini sağlamak için sınavda tüm yeterlilik seviyelerinde yeterli sayıda madde bulunmalıdır. Bir başka deyişle, geçerli ve güvenilir geleneksel testler oluşturmak için test geliştiricilerinin teste çok sayıda madde eklemesi gerektirmektedir (Wainer, 2000). Bunların yanı sıra KTK, test eşitleme ve yanlı maddelerin belirlenmesi gibi günümüz ölçme sorunlarına çözüm sunmada yetersiz kalmaktadır. KTK bu tür sınırlılıklara rağmen uygulanabilirliğinin ve anlaşılmasının kolay olması nedeniyle günümüzde gerçekleştirilen ulusal sınavlarda sıkça kullanılmaktadır. Fakat yine bu sınırlılıklar BBT uygulamalarına yönelik araştırmaların yapılmasına öncül olmuştur (Ivие, 2007).

KTK'nın sınırlılıklarına yanıt olarak geliştirilen Madde Tepki Kuramı (MTK), gözlemlenebilir ve gözlemlenemez değişkenler arasındaki özel ilişkiyi belirtir (Hambleton & Jones, 1993). Teknolojinin gelişmesi ile birlikte MTK'nın kullanım alanları artmıştır. Bunlardan biri de Bilgisayarda Bireyselleştirilmiş Test (BBT) uygulamalarıdır. BBT bireyin daha önce uygulanan maddelere verdiği yanıtlardan yola çıkarak sonraki test maddelerinin seçildiği bir test uygulama sürecidir. Böylece, sınava girenler yalnızca yetenek seviyelerine uygun maddeleri cevaplarlar. BBT uygulamasında kullanılan madde seçim algoritması, ölçüm hassasiyetinden ödün vermeden test uzunluğunu en az %50 oranında azaltabilmektedir (Wainer, 2000; Weiss, 2011). BBT uygulamalarının lineer testlerden farklı olarak daha hassas ve güvenilir sonuçlar vermesi, uygulamada bireye özgü soruların yer alması, testte daha az madde ile bireyin yeteneğinin arzu edilen zaman diliminde belirlenmesi ve anlık değerlendirmeye olanak sağlaması gibi avantajları bulunmaktadır (Hambleton ve ark., 1991; Hendrickson, 2007; Wainer, 2000). Bu avantajlar sayesinde, BBT uygulamaları geniş çaplı test uygulamalarında hem test

düzenleyiciler için hem de testi alan bireyler için tercih edilen bir uygulama haline gelmiştir (Chang, 2015) .

BBT uygulamaları 1970'li yılların başında popülerlik kazanmış ve sonraki yıllarda eğitim alanında değerlendirmelerde kapsamlı bir şekilde araştırılmış ve uygulanmıştır (Luecht & Sireci, 2011). 1990'lı yıllardan itibaren BBT uygulamaları eğitimde yerleştirme sınavları (Computer Adaptive Placement Assessment and Support System [COMPASS]), yüksek eğitimde öğrenci seçim sınavları (Graduate Record Examination [GRE], Graduate Management Admission Test [GMAT], Scholastic Assessment Test [SAT]), askeri alanda bireyleri sınıflama (Armed Services Vocational Aptitude Battery [ASVAB]), sağlık alanında lisans programları (National Council Licensure Examination [NCLEX]) gibi birçok alanda kullanılmaktadır (Papanastasiou, 2015).

BBT uygulamalarının lineer testlere göre bazı üstünlükleri olmasına rağmen aynı şekilde bazı sınırlılıklara neden olduğu da görülmektedir. BBT sınırlılıkları, hem test geliştiricileri hem de testi alan bireyleri doğrudan etkilemektedir. BBT uygulamalarında her bir test formu uygulama sürecinde belirlendiği için test geliştiriciler test formlarını önceden gözden geçiremezler. Bu nedenle BBT uygulamalarında maddelerin sık kullanılmamasını ve tüm içeriğin eşit olarak temsil edilmesini sağlamak için büyük madde havuzlarına ihtiyaç duyulmaktadır (Gershon, 2005). Büyük bir madde havuzu da test geliştiriciler için maliyeti yükseltmektedir (Meijer & Nering, 1999).

BBT uygulamalarının testi alan bireylere yönelik en göze çarpan sınırlılığı ise testte maddeyi yeniden cevaplamaya izin verilmemesidir (Han, 2013; Papanastasiou & Reckase, 2007; Vispoel ve ark., 2000). BBT uygulamalarında maddeyi yeniden gözden geçirmeye izin verilip verilmemesi gerektiği sorunu hem test düzenleyiciler hem de testi alan bireyler için büyük önem taşımaktadır (Bowles & Pommerich, 2000). BBT uygulamalarında, testi alanların geleneksel bir testte yapabilecekleri gibi soruları cevapsız bırakmalarına veya cevaplandıktan sonra soruları gözden geçirmelerine izin verilmemektedir (Gershon, 2005; Patsula, 1999).

Kâğıt-kalem veya bilgisayar tabanlı lineer bir testte maddeyi boş bırakma ve madde cevap değişikliği yaygın olarak gözlemlenmektedir (Vispoel, Hendrickson & Bleiler, 2000). Ulusal düzeyde yapılan sınavlarda testi alanların çoğunluğunu test oturumunu erken bitirmek yerine cevapladıkları maddeleri tekrar gözden geçirme eğilimindedirler (Bezirhan ve ark., 2021). Yapılan çalışmalarda bireylerin maddelerin belirli bir oranını gözden geçirerek değiştirdikleri ve çoğunlukla yanlış cevapları düzelterek performanslarını iyileştirdikleri belirtilmektedir (Bridgeman, 2012; Kruger ve ark., 2005; Pagni ve ark., 2017).

McMorris ve ark. (1991) yüksek yetenekli adayların ilk cevaplarını değiştirme olasılıklarının daha düşük olduğunu belirtirken, değişikliği yaptıklarında çoğunlukla yanlış seçeneği doğru seçenikle değiştirdiklerini de eklemiştir. Sonuç olarak araştırmacılar dikkatsiz veya hızlı yanıtlama sonucu cevaplama hatalarının olabileceğini ve bu hataların cevap değişikliği ile düzeltileceğini belirtmekte ve cevap değişikliği davranışını desteklemektedirler (Pagni ve ark., 2017).

Öğrencilerin günlük yaşamda karşılaştıkları testlerin büyük kısmı kâğıt-kalem uygulamalarıdır. Okul hayatları boyunca, birinci sınıftan üniversiteye kadar, öğrencilere emin olmadıkları soruları atlamaları ve daha sonra zaman kalırsa geri dönerek işaretlemeleri öğretilir. Ek olarak, dikkatsizlikten kaynaklanan hataları kontrol etmek için tamamladıktan sonra tüm testi tekrar gözden geçirmeleri önerilir. Öğrenciler de test uygulamalarında cevap değişikliğine izin verilmesini ve daha doğru yetenek kestirimleri elde etmeyi arzu etmektedir (Bowles & Pommerich, 2001; Vispoel ve ark., 2000; Waddell & Blankenship, 1994).

Araştırmacılar BBT uygulamalarında cevap değişikliğine izin verilmemesinin sınav sırasında bireyin sınav kaygısını arttırabildiğini ve sınava girenlerin genel performansını etkilediğini belirtmişlerdir (Lunz ve ark., 1992; Olea ve ark., 2000). Dikkatsizlik, anlık kaygı ve hesaplama hataları gibi nedenlerden dolayı yanlış cevaplanan soruların tekrardan gözden geçirilmesine izin verilmesi ile bireyin yetenek kestirimindeki doğruluğunu



arttıracağı savunulmakta ve dolaylı olarak test puanlarının geçerliğinin de artacağı belirtilmektedir (Stylianou-Georgiou & Papanastasiou, 2017). Han'ın (2015) yaptığı araştırmada sınava girenlerin %50'sinden fazlası BBT uygulamasında cevap değişikliğine izin verilmediği için sınav kaygılarının arttığını ve %80'den fazlası ise eğer cevap değişikliğine izin verilirse sınavda daha iyi performans gösterebileceklerini belirtmiştir. Bu nedenle BBT uygulamalarında maddeyi yeniden cevaplayabilmeye izin verilmesi gerektiği belirtilmektedir (Vispoel ve ark., 2000; Wise, 1996). İzin verilmediği durumlarda yanlış yetenek kestirimlerine yol açabileceği belirtilmektedir (Lunz ve ark., 1992; Vispoel ve ark., 2000; Waddell & Blankenship, 1994). Özellikle yüksek yetenekli öğrenciler, kaygıdan ya da dikkatsizlikten dolayı doğru cevaplamaları gereken BTT uygulamasındaki ilk maddeleri yanlış cevapladığında test uzunluğu 45 madde olsa bile bireylerin yetenek kestirimi gerçek değerden daha düşük hesaplanabilmektedir (Guyer & Weiss, 2009; Rulison & Loken, 2009). Bu doğrultuda BBT uygulamalarının algoritmasına yönelik eğitim alan bireyler, yetenek seviyelerinin daha doğru kestirilmesine yardımcı olmak için ilk beş veya on maddede doğru cevabı bulmak amacıyla daha fazla zaman harcamaktırlar. Ivie (2007) yaptığı çalışmada BBT uygulamalarında ilk beş veya on maddeye daha fazla zaman ve dikkat harcayarak nihai yetenek kestiriminde daha yüksek puan elde etmeye yönelik test stratejisi eğitimi alanların daha iyi yetenek kestirimine sahip olduklarını belirlemiştir. Bu sonuçlar, test geliştirme şirketlerinin, testin başında daha fazla zaman geçirmenin testten elde edilen puanın arttıracağı iddiasını desteklemektedir. 2000 yılında Eğitsel Test Uygulama Merkezi (Educational Testing Service, [ETS]), GRE-CAT uygulamasında yaklaşık birkaç bin sınav katılımcısının puanlarının BBT algoritmasından dolayı yanlış kestirildiğini belirtmiştir (Carlson, 2000). 2002 yılında ise yaklaşık bin öğrencinin CAT-GMAT puanlarının yanlış kestirildiği belirtilmiştir (Merritt, 2003). Chang ve Ying (2008) yanlış kestirimlerin nedeni olarak madde seçim yöntemlerinden maksimum bilgi yönteminin olabileceğini belirtmişlerdir. Araştırmacılar maksimum bilgi yönteminin testin başında  $\theta$  kestiriminde büyük değişikliklere neden olabilecek en yüksek a-parametrelerine sahip maddeleri seçme eğiliminde olduğunu ifade etmişlerdir. Sonuç olarak, test yeterince

uzun değilse ve birey eğer testin başlangıcında yer alan maddeleri doğru cevaplaması gerekirken yanlış cevapladıysa, o zaman birey gerçek yetenek düzeyine yakın bir puan elde edememektedir. Bu sonuçların bireylerin en üst düzey okullara girme şansını etkilediği araştırmacılar tarafından bildirilmiştir. Çoğu BBT uygulamalarının ulusal çapta yapılan sınavlar olduğu göz önüne alındığında, bunların güvenilirliğini ve geçerliğini arttırmak gerekmektedir.

BBT uygulamalarında maddeyi yeniden cevaplamaya izin verilmemesinin nedenleri cevap değişikliğinin ölçme hassasiyetinde azalmaya neden olabileceği (Olea ve ark., 2000; Stone & Lunz, 1994) ve test stratejilerine karşı test algoritmalarının verimli çalışmayabileceğidir (Wainer, 1993). BBT'deki bir maddenin cevabını değiştirmek, daha önceden uygulanan maddelerin, sınava giren kişinin yeteneğini kestirmek için artık uygun olmamasına neden olabilir. Maddeleri yeniden cevaplamaya izin veren BBT'deki bu uygunsuz maddeler, sırayla yetenek kestiriminde yanlışlığa neden olabilir ve ölçme hassasiyetini azaltabilir. Bu durum BBT uygulamalarında zayıflama paradoksu (Lord & Novick 1968; van der Linden & Pashley, 2010) madde ve yetenek uyumsuzluğunu, daha fazla artırabilir.

Çoğu BBT uygulamalarında madde seçim algoritmaları, her bir geçici yetenek kestirimi için en uygun maddeyi seçer. Eğer birey daha önceki maddelere verdiği cevabı değiştirirse bilgisayar tarafından seçilmiş olan bir sonraki maddenin artık en uygun madde (optimal) olmayabileceği ve bu durumun yetenek kestirimini olumsuz etkileyebileceği belirtilmektedir (Stocking, 1997; Vispoel ve ark., 1999). Cevap değişikliği sonucunda hâlihazırda cevapladığı diğer maddelerden elde edilen bilgi azalabilmekte ve bu da nihai kestirimi etkileyebilmektedir. (Papanastasiou, 2005).

Test uygulayıcılar ise maddenin yeniden gözden geçirilmesine izin verildiği durumda onlara sınav süresinin uzaması, uygulanan madde miktarının artması gibi ekstra maliyetler doğuracağı ve testin geçerliğini düşüreceğini belirtmektedirler (Gershon & Bergstrom, 1995). Ayrıca yanıtlayıcıların test stratejilerini kullanarak yanlış puan elde

edebileceklerini belirtmektedirler (Zenisky ve ark., 2010; Melican ve ark., 2010). Bu test stratejileri Wainer (Wainer, 1993) ve Kingsbury (Kingsbury, 1996) olarak geçmektedir. Wainer stratejisini kullanan bir birey ilk sorudan itibaren bütün maddeleri bildiği halde yanlış cevaplayarak kolay maddelerin sorulmasını sağlamakta ve maddeyi yeniden gözden geçirmeye izin verildiğinde maddelerin hepsini doğru cevaplama amaçlamaktadır. Özellikle yetenek kestirimi olarak En Çok Olabilirlik Yöntemi (Maximum Likelihood Estimation, [MLE]) kullanıldığında bireyler testten en yüksek puanı elde edebilmektedir. Ama birey sadece bir maddeyi bile yanlış cevaplarsa bireyin puanı gerçek yetenek düzeyinden daha düşük kestirilmektedir. BBT uygulamalarında puan şişirmeye yönelik bir diğer test stratejisi olan Kingsbury yöntemini kullanan birey BBT uygulamalarında bir maddeyi cevapladığında bir sonraki maddenin güçlük düzeyini tahmin ederek önceki maddenin doğruluğu hakkında bilgi edinebilmektedir. Yapılan araştırmalarda bireylerin madde güçlük düzeylerine yakın olan maddeleri ayırt etmekte zorlandıklarından dolayı bu puan şişirme stratejilerinin küçük ve büyük çapta gerçekleştirilen sınavlarda kullanımının makul ve olası olmadığı belirtilmiştir (Lunz ve ark., 1992; Stone & Lunz, 1994; Vispoel ve ark., 2000). Bir diğer taraftan, bazı araştırmacılar maddeyi yeniden gözden geçirmeye izin veren BBT uygulamasında olası puan şişirmeye yönelik test stratejilerine rağmen yetenek kestiriminin daha sağlıklı olabileceğine vurgu yapmaktadır (Vispoel ve ark., 2002).

Yetenek kestirimdeki yanlış sonuçlar ve maddeyi yeniden cevaplamaya izin verilmemesi testi alanlar kadar test düzenleyiciler için de temel bir endişe olmuştur. Bu endişe GRE'nin BBT uygulamasından vazgeçerek çok aşamalı test modeline geçmesinin nedenlerinden biri olarak görülmektedir (ETS, 2011). Diğer yandan araştırmacılar tarafından hem madde cevap değişikliğine izin veren hem de yetenek kestiriminde yanlışlığın mümkün mertebe asgari düzeyde olmasını sağlayan BBT modelleri geliştirilmiştir. BBT uygulamasında yapılacak çok fazla cevap değişikliği, özellikle doğrudan yanlışa ya da tam tersi olduğunda, yetenek kestiriminin hassasiyetini büyük

ölçüde azalttığı için geliştirilen bu modellerde revize edilebilecek madde sayısı sınırlandırılmıştır (Stocking, 1997; Papanastasiou & Reckase, 2007; Han, 2013). Cevap sayısı değişikliğindeki kısıtlama ile ölçme hassasiyeti üzerindeki büyük ölçekli negatif etki önlenmektedir (Lin ve ark., 2021). BBT uygulamalarında sınırlı sayıda cevap değişikliğine izin verilmesi testi alan bireyler için hala bir kısıtlama olarak görülmektedir. Araştırmacılar madde cevap değişikliğinden kaynaklı sorunların giderildiği takdirde yanıt değişikliğine izin verilmesi gerektiğini belirtmektedirler (Wise, 1996).

### **Araştırmanın Amacı ve Önemi**

BBT'deki bir maddenin cevabını değiştirmek, daha önceden uygulanan maddelerin, sınava giren kişinin yeteneğini kestirmek için artık uygun olmamasına neden olabilmektedir. Cevap değişikliği sorunu için olası bir çözüm, yalnızca geçici yetenek tahminlerinin belirsizliğini hesaba katmakla kalmayıp aynı zamanda cevap değişikliğini mümkün kılan en uygun başlangıç değeri tanımlamaktır. Bu kazanım sayesinde bireylere sınırsız cevap değişikliği özgürlüğü sunulabilir. Bu doğrultuda test bataryaları maddeyi yeniden cevaplamaya izin veren BBT uygulaması olarak gerçekleştirildiğinde uygulamanın bileşenleri ile cevap değişikliklerinin yetenek kestirimlerindeki hatayı ne derece etkilediği araştırılmalıdır. Bu doğrultuda yapılan çalışmada öncül bilginin tanımlandığı BBT uygulamalarında maddeyi yeniden gözden geçirmeye izin verildiğinde elde edilen yetenek kestirimlerinin standart hatasının ve yanlılığın testin başlangıç kuralına, nihai yetenek kestirim yöntemine ve test uzunluğuna göre nasıl değiştiğini belirlemek amaçlanmıştır.

Alanyazın incelendiğinde BBT uygulamalarında madde cevap değişikliğine izin verilmediği için bu konuyla ilgili çalışmaların sınırlı sayıda olduğu görülmektedir (Wang ve ark., 2020). Mevcut araştırmalar ise çoğunlukla maddeyi gözden geçirmeye yönelik kuralların tasarlanmasına veya kaç maddenin yanıtının değiştirilmesine izin verilebileceğine yöneliktir (Stocking, 1997; Han, 2013; Papanastasiou & Reckase, 2007; Cui ve ark., 2018b). Yapılan çalışmada ise deneysel öncül bilgiler sayesinde BBT

başlangıç değerleri tanımlanarak sınırsız sayıda madde cevap değişikliğine izin verilmiştir. Bu çalışma alanyazında test uzunluğunu artırmadan sınırsız sayıda cevap değişikliğine izin veren ilk araştırma olacaktır.

Bilgisayar tabanlı lineer test (PISA-2015) verileri kullanılarak gerçekleştirilen bu çalışmadan elde edilecek bulgular, elektronik ortamda gerçekleştirilen sınavların kullanıcı dostu BBT uygulamalarına dönüşmesinde katkı sağlayacaktır.

### **Araştırma Problemi**

Bilgisayarda bireyselleştirilmiş test (BBT) uygulamalarında maddeyi yeniden gözden geçirmeye izin verildiğinde elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı testin başlangıç kuralına, nihai yetenek kestirim yöntemine ve test uzunluğuna göre nasıl değişmektedir?

### **Alt Problemler**

1. BBT uygulamalarında başlangıç yetenek düzeyi  $[-1,1]$  yetenek aralığından rastgele seçilen bir değer olarak belirlendiğinde elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı;
  - a. Maddeyi yeniden gözden geçirmeye izin verilmediğinde yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?
  - b. Maddeyi yeniden gözden geçirmeye izin verildiğinde testi alanların cevaplama stratejilerine (önce rastgele sonra bilinçli cevaplama, önce amaçlı yanlış sonra bilinçli cevaplama), yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?

2. BBT uygulamalarında başlangıç yetenek düzeyi olarak regresyon eşitliği ile belirlenen değer kullanıldığında elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı;
  - a. Maddeyi yeniden gözden geçirmeye izin verilmediğinde yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?
  - b. Maddeyi yeniden gözden geçirmeye izin verildiğinde testi alanların cevaplama stratejilerine (önce rastgele sonra bilinçli cevaplama, önce amaçlı yanlış sonra bilinçli cevaplama), yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?
3. BBT uygulamalarında başlangıç yetenek düzeyi olarak aynı testin bir önceki uygulamasından elde edilen yetenek düzeyi kullanıldığında elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı;
  - a. Maddeyi yeniden gözden geçirmeye izin verilmediğinde yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?
  - b. Maddeyi yeniden gözden geçirmeye izin verildiğinde testi alanların cevaplama stratejilerine (önce rastgele sonra bilinçli cevaplama, önce amaçlı yanlış sonra bilinçli cevaplama), yetenek kestirim yöntemlerine (EAP(0,1), EAP, MLE, MLEF) ve test uzunluğuna (5, 10, 15, 20, 25) göre nasıl değişmektedir?

### **Sayıtlar**

1. Evren madde parametrelerinin örneklem madde parametreleri ile aynı olduğu kabul edilmiştir.

2. Eksik veri matrisi Monte Carlo Simülasyonu ile tam veri matrisine dönüştürülmüştür. Bireylerin bu cevapları verdikleri varsayılmıştır.
3. Bireylerin maddeleri doğru cevaplama olasılıkları  $P$  (2 PL model ile belirlenen değer); amaçlı yanlış cevaplama olasılıkları  $(P+2)/3$  olarak kabul edilmiştir.

### **Sınırlılıklar**

1. İçerik dengeleme dikkate alınmamıştır.
2. Madde seçme yöntemi olarak Maksimum Fisher Bilgi Yöntemi kullanılmıştır.
3. Sonlandırma kuralı olarak sabit uzunluk kullanılmıştır.

### **Tanımlar**

Zayıflama paradoksu (Attenuation Paradox): BBT uygulamalarında bireyin gerçek yetenek düzeyi bilinmediğinden uygulanan madde geçici yetenek düzeyine göre belirlenir. Geçici yetenek düzeyi ile gerçek yetenek düzeyi arasında uyumsuzluk olduğunda seçilen madde gerçek yetenek düzeyinde maksimum bilgi veren madde olmayabilir. Bu çelişki zayıflama paradoksu olarak bilinmektedir (Lord & Novick, 1968).

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde tek boyutlu madde tepki kuramı, bilgisayarda bireyselleştirilmiş test uygulamalarının aşamaları ve uygulama sürecine yönelik kuramsal çerçeve çizilmiştir. Ayrıca maddeyi gözden geçirme yöntemleri hakkında bilgi verilmiş ve son olarak yurtiçi ve yurt dışında yapılan ilgili örnek çalışmalara yer verilmiştir.

#### Madde Tepki Kuramı

Madde Tepki Kuramı (MTK) ölçülmek istenen gizil değişken ile bir maddeye verilen cevap arasındaki ilişkiyi gösteren matematiksel bir ifadedir. Madde Tepki Kuramının temelleri 20. yüzyılın ortalarına dayansa da kuramın altında yatan istatistiksel hesaplamaların karmaşık olması kuramın popülerleşmesini geciktirmiştir. Tucker 1946 yılında madde karakteristik eğrisini tanımlayarak ilk adımı atan öncülerden biri olmuştur. Bir diğer öncü ise, Lord'un kuramını sistematik bir şekilde tanımlayan, gerekli bilgisayar yazılımını geliştiren ve pratik uygulamaları yapan Novick (1968)'tir (Carlson & Davier, 2013).

Başlangıçta Gizil Özellik Teorisi olarak adlandırılan MTK modelleri, test ile elde edilen değerler ile bireyin gizil değişkeni arasındaki ilişkiyi tanımlarlar (Weiss, 1983). Bir başka deyişle, MTK modelleri, gizil değişkenin testi alanların performansını nasıl etkilediğini açıklar. Veriye uyum sağlayan bir model, test kullanıcılarına bir bireyin bu özellik üzerindeki durumunu kestirmesine olanak sağlar. MTK, bireylerin bir dizi maddeye verdikleri tepkilerden yola çıkarak, bireylerin gizil özellik düzeylerini kestirir ve ortak bir ölçek üzerine yerleştirir. MTK, kişi parametrelerini madde veya test parametrelerinden ayırır. Bu, KTK'nın gerektirdiğinden çok daha güçlü varsayımları karşılayarak elde edilir.

MTK'nın KTK'ya göre en önemli avantajı parametre değişmezliğidir. Parametre değişmezliği, aynı özelliği ölçen farklı madde kümelerine dayalı yetenek kestirimlerinin doğrusal bir dönüşüm aracılığıyla karşılaştırılabilir olacağı anlamına gelir. Benzer şekilde,



farklı gruplar için de kullanıldığında, madde parametre kestirimleri doğrusal bir dönüşüm vasıtasıyla eşdeğer olacaktır.

MTK'da, testin yapısına, cevap türlerine göre birçok farklı model geliştirilmiştir. Bu çalışmada kullanılan veri seti tek boyutlu ve cevap setinin iki kategorili olması nedeniyle iki kategorili MTK modellerine yönelik bilgi yer almaktadır. Bu kapsamda en çok kullanılan modeller bir, iki ve üç parametrelili lojistik (PL) modellerdir.

### ***İki kategorili MTK modelleri***

Çoktan seçmeli test maddeleri gibi doğru ya da yanlış olarak iki kategorili puanlanan testlerde ikili puanlanan MTK modelleri kullanılır. İkili puanlanan maddeler için tipik modeller 1PL, 2PL ve 3PL'dir. Bu modeller her bir madde için doğru cevap verme olasılığı ile bireyin yeteneği arasındaki ilişkiyi gösteren fonksiyonda kullanılan madde parametre sayısına göre adlandırılır (Embretson & Reise, 2000; Hambleton ve ark., 1991). Maddeyi doğru yanıtlama olasılığı ( $P(\theta)$ ) ile madde parametreleri arasındaki ilişkiyi gösteren bu fonksiyona ise madde karakteristik fonksiyonu (MKF), bu fonksiyonun eğrisine de madde karakteristik eğrisi (MKE) denir. MKF, madde tepki kuramının temelini oluşturmaktadır. MTK'daki tüm modeller bu fonksiyon üzerine inşa edilmiştir (Baker, 2001). İkili puanlanan modeller bu fonksiyonlarda kullanılan parametre sayısına göre isimlendirilmektedir. MK fonksiyonunda madde güçlüğü (a), madde ayırt ediciliği (b) ve madde şans parametresi (c) en çok kullanılan parametrelerdir. Sadece madde güçlük parametresinin kestirildiği modele bir parametrelili lojistik model (1 PLM) denir. İkili puanlanan MTK modellerinden en basiti olan 1 PLM'de madde ayırt edicilik parametresi bütün maddeler için sabit bir değer alınırken şans parametresi ise göz ardı edilecek düzeyde olduğu varsayılır. 2 PLM'de ise b parametresinin yanı sıra a parametresi de kestirilir. 3 PLM ise madde güçlük, ayırt edicilik ve şans parametrelerinin üçü de kestirilir. Bireyin iki kategorili puanlanan bir maddeyi doğru cevaplama olasılığı ile madde parametreleri arasındaki ilişkiyi gösteren fonksiyonlar Tablo 1'de verilmiştir.

**Tablo 1***İkili Puanlanan Madde Tepki Modellerine Ait Matematiksel Fonksiyonlar*

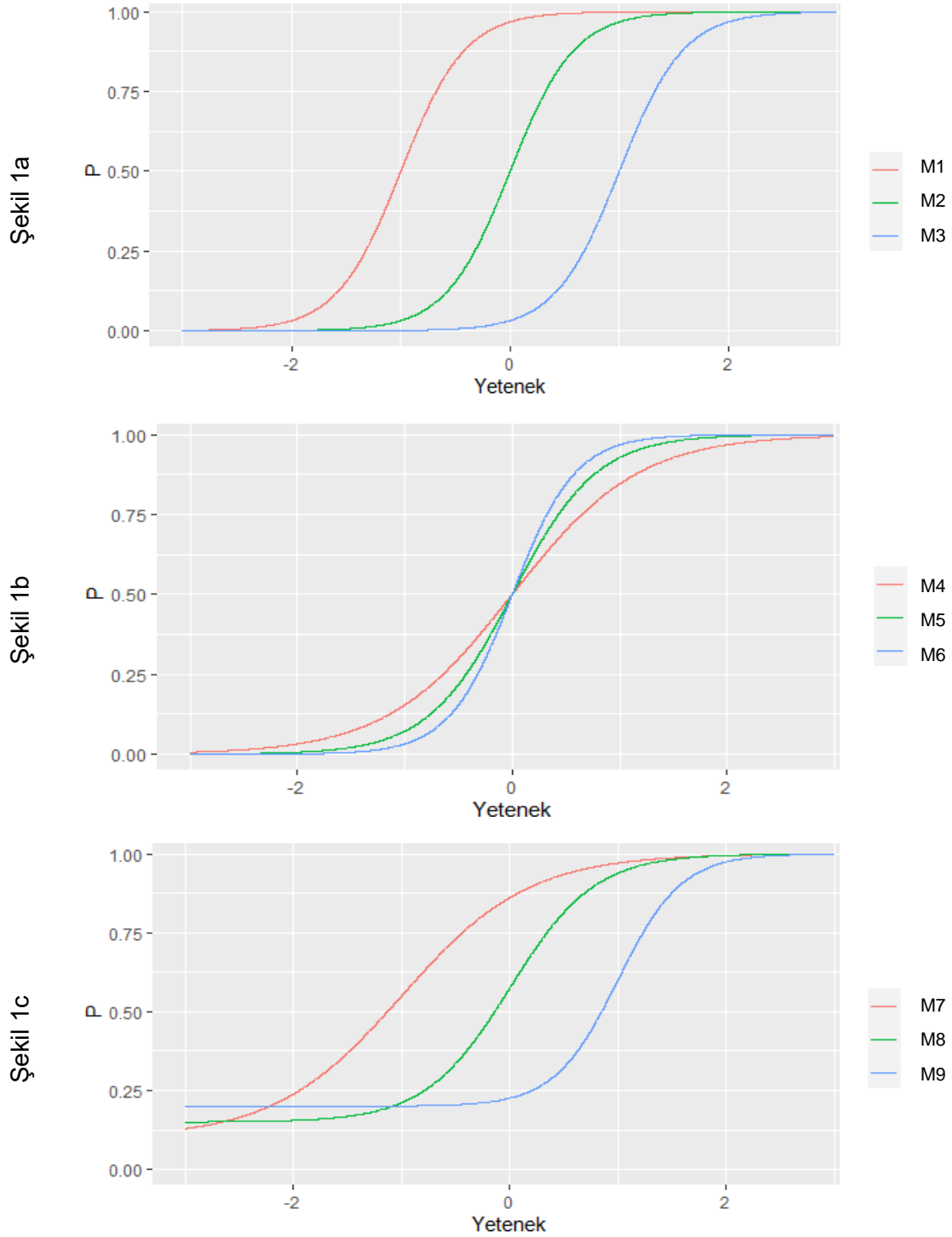
Model	Matematiksel fonksiyon
Bir Parametrelili Lojistik Model	$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}$
İki Parametrelili Lojistik Model	$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$
Üç Parametrelili Lojistik Model	$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$

1 PLM ve 2 PLM aslında 3 PLM'nin sınırlandırılmış biçimleridir. Şekil 1'de modellere ilişkin MKE'leri verilmiştir. Madde güçlük parametresi, b, maddeyi doğru yanıtlama olasılığının 0,5 olduğu yetenek düzeyidir. Ancak şans parametresinin dikkate alındığı 3PLM'de ise madde güçlük düzeyi c ile 1 arasındaki orta noktaya karşılık gelen değerdir. a parametresi ise MKE'deki eğimi ifade eder. Bir diğer parametre olan c ise şans faktörüdür. Bu parametre en düşük yetenek düzeyine sahip bireyin soruyu doğru cevaplama olasılığına eşittir. Çoktan seçmeli sorularda bu olasılık madenin çeldirici düzeylerine göre rastgele doğru cevaplama olasılığından fazla ya da eksik olabilir. Şekil 1'e göre madde 3, madde 2 ve madde 1'den daha zorken ayırt edicilik parametreleri aynıdır. Madde 4, madde 5 ve madde 6'nın b parametreleri aynı iken madde 6'nın a parametresi, madde 5 ve madde 6'dan daha fazladır. 3 PLM modele göre kalibre edilmiş Madde 9'un c parametresi madde 7 ve madde 8'den daha fazladır.

Hangi modelin en iyisi olduğunu karar verirken birçok husus göz önüne alınır. Bunlar madde ağırlıklarının eşit olup olmaması, ölçme aracının özellikleri, model veri uyumu ve madde parametre kestiriminin amacıdır (Embretson & Reise, 2000). Örneğin, çoktan seçmeli maddelerde şans faktörünü de dikkate alan 3 PLM en yaygın kullanılan modeldir. Ancak güçlü çeldiricilerin olduğu maddelerde ise daha basit model olan 2 PLM'nin de kullanılabileceği belirtilmektedir (Demars, 2010).

## Şekil 1

### Madde Karakteristik Eğrileri



Test uygulamalarında birçok avantaj sağlayan MTK modellerinin başlıca kullanım alanlarından biri Bilgisayarda Bireyselleştirilmiş Test uygulamalarıdır. MTK modelleri, farklı test formlarından elde edilen puanları aynı metrik üzerine yerleştirir. Bir başka deyişle,

madde setinden bağımsız olarak bireyin yeteneğinin kestirilmesine olanak sağlar (Hambleton ve ark., 1991). Bu özellik her bireyin bireyselleştirilmiş test formuna sahip olmasına olanak sağlar (Demars, 2010).

### **Bilgisayarda Bireyselleştirilmiş Test Uygulamaları**

Bilgisayarda Bireyselleştirilmiş Test uygulamaları ulusal ve uluslararası sınavlarda gün geçtikçe önemli bir test modu (tipi/şekli) haline gelmektedir (Chang, 2015). Geleneksel testlerde bireyler önceden belirlenen test maddelerini cevaplarırken, BBT uygulamasında ise her birey geçici yetenek düzeyine göre seçilen test maddelerini cevaplandırır. BBT'nin en büyük avantajı, geleneksel testlerde gerekli olandan daha az madde ile daha verimli yetenek kestirimleri ( $\theta$ ) sağlamasıdır. Bir başka deyişle, BBT, sınava giren kişinin geçici yetenek düzeyiyle en iyi eşleşen maddeleri seçerek bireye özgü test formu oluşturur ve geleneksel testlere göre daha kısa bir test ile istenen bir ölçme doğruluğunu elde eder.

Bilgisayarda Bireyselleştirilmiş Test uygulamaları birçok adımı içeren bir test uygulaması modülüdür. Thompson ve Weiss (2011), BBT uygulamalarını beş başlık altında incelemiştir. Bunlar madde havuzunun oluşturulması, başlama kuralı, madde seçim yöntemi, yetenek kestirimi ve sonlandırma kuralıdır. İçerik dengeleme ve madde kullanım sıklığı madde seçim yöntemi başlığı altında değerlendirilmektedir.

#### ***Madde Havuzu***

Madde havuzunun yapısı BBT uygulamalarının ölçme hassasiyeti ve işlevselliği bakımından önemli bir rol oynamaktadır (Veldkamp & van der Linden, 2000; Xing & Hambleton, 2004). BBT uygulamasının başarılı olabilmesi için yeterli sayıda kaliteli madde içeren bir madde havuzunun olması gerekir. Araştırmacılar, BBT'ler için madde havuzları geliştirilirken, sadece madde havuzu boyutunun değil, aynı zamanda madde parametrelerinin dağılımı ve içerik dağılımı gibi pratik hususların da önemli olduğuna işaret etmişlerdir (He & Reckase, 2014). Optimal test havuzu, içerik dağılımıyla uyumlu

olmalı ve hedef popülasyonun yetenek dağılımının tüm seviyelerinde yeterli bilgi sağlamalıdır (Weiss, 2004). BBT algoritmaları ideal olsa bile sınırlı bir madde havuzuyla veya düşük kaliteli maddelerle verimli çalışamaz (Flaughner, 2000). Madde havuzunun kalitesi ne kadar iyi olursa, BBT algoritması da o kadar başarılı olacaktır. BBT uygulaması geliştirme sürecinin ilk adımı, geniş bir kitleye daha önceden uygulanan ve kalibre edilmiş maddelerden madde havuzunu oluşturmaktır. Her ne kadar madde havuzu büyüklüğü hakkında bir fikir birliği olmasa da madde parametrelerini kestirmek için test maddelerini en az 1000 kişiye uygulama önerisi bugün de geçerliliğini korumaktadır (Şahin & Weiss, 2015). Madde havuzunun büyüklüğü, madde havuzunun yeterliliğinin bir başka göstergesidir. Stocking (1994) bir madde havuzunun büyüklüğünün bireye uyarlanmış test uygulamasının uzunluğunun en az altı katı olması gerektiği belirtmiştir. Way (1998) ise lisans ve sertifika programları için 6-8 katı; seçme ve yerleştirme sınavları içinse madde havuzunun, test uzunluğunun en az 12 katı olması gerektiğini belirtmiştir.

### ***Başlama Kuralı***

BBT uygulamalarının amacı olan daha az madde ile daha doğru yetenek kestirimini sağlayabilmek olduğu için ilk kullanılacak maddenin birey hakkında olabildiğince fazla bilgi vermesi gerekmektedir. BBT uygulamalarının ilk aşamalarında yeteneğin yanlış belirlenmesiyle bireyin yetenek düzeyine uygun olmayan maddelerle karşılaşması bu amaca ulaşmayı gölgelemektedir. Bu doğrultuda başlama kuralı olarak izlenen iki yol bulunmaktadır. Birincisi, bireyler hakkında öncül bilginin var olduğu durumlardır. Eğer bireyin yeteneği ( $\theta$ ) hakkında bir bilgi mevcutsa bu bilgiden yola çıkarak ilk madde belirlenebilir. Bu yöntemde, bireyin uygulanmakta olan mevcut teste benzer yapıya ait puanlarından yararlanarak başlangıç değeri tanımlanabilmektedir. (van der Linden, 1999). Böylece  $\theta$  yakınsaması daha hızlı gerçekleşecektir. Eğer bireyin yeteneği ile ilişkili herhangi bir bilgi yoksa psikometrik açıdan BBT uygulamalarında ilk madde orta zorlukta seçilir. Çünkü testi alan bireyin yeterliliği hakkında önceden bilgi bulunmadığında, yeterliliğinin en iyi tahmini ortalama yeterliliklidir (Mills & Stocking, 1996). Ölçülmek istenen

yapının normal dağıldığı kabul edilen testlerde BBT başlangıç değeri 0 olarak alınır. Ancak bireylerin hepsinin aynı güçlük düzeyindeki maddelerle teste başlaması madde kullanım sıklığı sorununu ortaya çıkarmaktadır. Bu nedenle başlama kuralı olarak sabit bir değer yerine belli bir aralıktan rastgele bir başlangıç maddesinin seçilmesi bu sorunun çözümüne yönelik yapılan uygulamalardan biridir. Thompson ve Weiss (2011) bireyin yeteneği hakkında herhangi bir bilgi bulunmadığı durumlarda başlangıç maddesinin -0,5 ile 0,5 madde güçlük düzeyi aralığının seçilebileceğini tavsiye etmektedir. Weiss (2011) testi alan bireyler hakkında ön bilgiler varsa BBT uygulamalarında en iyi başlangıç yönteminin bu bilgiler ışığında belirlenen başlangıç değerleri olabileceğini belirtmektedir. Bununla birlikte, alan yazında sınava giren bireyler hakkında ampirik bilgilerin veya daha önceden elde edilen bilgilerin kullanılmasının, sıklıkla istenmeyen, sosyal olarak önyargılı kestirimlere yol açabileceği belirtilmektedir (van der Linden, 1999; Veldkamp & Matteucci, 2013).

Başlangıç kuralının rolü üzerine yapılan araştırmalarda BBT uygulamasının performansından ziyade test güvenliği konularına odaklanılmıştır (Embretson & Reise, 2000). Belirli sayıda madde veya belirli madde zorluk aralığından ilk maddenin rastgele seçilmesi ya da ilk 5 veya 10 maddenin seçilmesi gibi yöntemler geliştirilmiştir (Kingsbury & Zara, 1989).

### ***Madde Seçim Yöntemi***

BBT uygulamalarının temel amacı olan az madde ile etkili ölçme hassasiyetini sağlamak, testte uygulanan ilk madde kadar madde seçim yöntemine de bağlıdır. Madde seçimi, testin etkililiği ve ölçme hassasiyeti ile yakından ilgilidir. Özellikle test uzunluğu kısa olan BBT uygulamalarında madde seçim yöntemi yetenek yakınsamasının hızlı gerçekleşmesinde önemli rol oynamaktadır. BBT uygulamalarını doğrusal test formlarının uygulanmasından farklı kılan unsur madde seçim algoritmasıyla bireye özgü test formlarının oluşturulabilmesidir (Veldkamp & Matteucci, 2013). Madde seçim algoritması ile testi alan bireylerin, uygulanan maddelerin yarısını doğru ve yarısını yanlış cevapladığı

bir test formu oluşturmak amaçlanır. Oluşturulan test formları sınava giren bireylere uygulanabilecek aynı sayıdaki maddelerin diğer alt kümelerine kıyasla o birey için oldukça hassas ölçümler sağlar (Weiss, 1982).

BBT uygulamalarında madde seçimine yönelik iki temel yaklaşım vardır. Bunlardan ilki olan madde bilgisi yaklaşımı, geçici yetenek kestiriminde maksimum bilgi sağlayan maddenin seçilmesini sağlarken (Birnbaum, 1968; Lord, 1977) diğer bir yaklaşım olan Bayeş yaklaşımı ise yetenek düzeyine ait öncül ve sonsal dağılımları kullanarak (Owen, 1975) madde seçimini gerçekleştirir (aktaran Meijer & Nering, 1999). Madde bilgisini kullanan başlıca yöntemler maksimum bilgi (MI), Kullaback-Leibler ve genel ağırlıklandırılmış bilgidir. Uygulama kolaylığından dolayı en çok kullanılan yöntem Lord (1980) tarafından geliştirilen Fisher En Çok Bilgi Yöntemidir (Maximum Fisher Item Information, [FMI]). Bu yöntem geçici yetenek düzeyinde en çok bilgi sağlayacak maddeyi seçer (Chang, 2015). FMI gerçek yetenek düzeyiyle uyumlu olduğu durumlarda etkili bir yöntemdir. Çünkü FMI madde seçim algoritması mevcut yetenek kestirimi ile bir sınava giren birey için gerçek yetenek seviyesi arasında bir eşleşme olduğu varsayımına dayanmaktadır. Ancak BBT uygulamasının ilk aşamalarında yetenek kestirimindeki değişkenlik oldukça büyüktür (Chang & Ying, 1999). En uygun madde ile geçici yetenek kestirimi arasındaki uyumsuzluk göz önüne alındığında, belirlenen yetenek düzeyinde maksimum bilgiyi sağlamak adına oldukça ayırt edici bir maddenin seçilmesi, gerçek  $\theta$  düzeyi hakkında çok az bilgi sağlayabilir ve bu seçim sonraki  $\theta$  kestiriminin verimsiz olmasına yol açabilmektedir. Yetenek düzeyindeki bu belirsizlikten dolayı BBT uygulamasının ilk aşamalarında kullanılan maddeler, son aşamada kullanılan maddelere nazaran potansiyel olarak daha az bilgi sunar (Chen ve ark., 2000). Diğer bir deyişle, uygulamada ilk seçilen maddeler sonrakilere nazaran daha fazla miktarda hataya (yani düşük hassasiyet) sahip olduğundan, ilk maddelere verilen yanıtlar gerçek yetenek kestirimi için kısıtlı miktarda bilgi sağlar.

BBT'nin erken aşamasında kestirilen  $\theta$  değerindeki bu tutarsızlık, ölçme hassasiyetindeki ve verimliliğindeki teorik kazancı tehlikeye atmaktadır. Ancak kestirilen  $\theta$  gerçek değere daha yakın olduğunda, MI prosedürünün verimliliği ön plana çıkar. Bireyin özelliklerinden yola çıkarak başlangıç değeri tanımlanan BBT uygulamalarında ise yakın olabileceği için FMI yönteminin kullanılması daha uygundur. Böylece  $\theta$  yakınsaması daha hızlı gerçekleşir ve en uygun test uzunluğu azalır.

Madde seçiminde başka bir yaklaşım ise Bayes'tir. Bilgisayar teknolojisinin gelişmesiyle birlikte gizil özelliğin ( $\theta$ ) sonsal dağılımını kullanabilen Bayes tabanlı madde seçim yöntemleri geliştirilmiştir. Bu yöntemler maksimum sonsal ağırlıklandırılmış bilgi (maximum posterior-weighted information), maksimum beklenen bilgi (maximum expected information), minimum beklenen sonsal varyans (minimum expected posterior variance) ve maksimum beklenen sonsal ağırlıklandırılmış bilgidir (maximum expected posterior-weighted information) (van der Linden, 1999). Bayes yaklaşımında sonsal dağılım, olabilirlik fonksiyonu ile öncül dağılımın birleşimidir. FMI ve Bayes yöntemleri, prosedürlerinin bazı benzer fonksiyonlara sahip olduğundan benzer test maddelerini seçerler. Başka bir deyişle iki seçim yöntemi arasındaki temel fark, onların varsayımıyla ilgilidir. FMI yöntemindeki geçici yetenek kestirimi ile gerçek değer uyumlu olduğu varsayımı bireyler hakkında ön bilgiler ile elde edilebilir. Bu durumda Bayes ile FMI yöntemleri benzer maddeleri seçecektir. Ayrıca Veldkamp ve Matteucci (2013) test uzunluğunun yirmi veya daha fazla madde olduğu durumlarda, madde seçim kurallarının benzer performans sergilediklerini belirtmiştir.

Parshall ve ark. (1998) madde seçim algoritmasının yalnızca ölçme hassasiyetini en üst düzeye çıkaran maddelerin seçilmesi değil, aynı zamanda havuzdaki maddeye maruz kalma seviyesini kontrol ederek test güvenliğini koruması ve her bir sınava girenin bütün içeriğe yönelik maddeler almasını sağlayarak test geçerliliğini koruması gerektiğini öne sürmüşlerdir. Madde maruz kalma seviyesinin kontrol edilmesinin amacı ölçme hassasiyetinden ödün vermeden madde havuzunda yer alan maddelerin dengeli bir



şekilde kullanılarak test güvenliğinin korunmasıdır (Pastor ve ark., 2002). Eğer bir birey testi birden çok kez alıyorsa aynı maddelerle karşılaşması olasıdır. Bu birey maddeyi hatırlayarak cevaplayabilir ya da diğer bireylerle test maddesini paylaşabilir. Bu güvenlik sorunu testin geçerliğini de düşürecektir. Diğer tarafta ise test geliştiriciler madde hazırlama maliyetleri nedeniyle büyük madde havuzu oluşturmaktan kaçınmaktadırlar. Bu sebeple madde kullanım sıklığının kontrol edilmesine yönelik çeşitli yöntemler geliştirilmiştir. Madde kullanım sıklığının kontrol edilmesine yönelik geliştirilen yöntemler koşullu, rastgele, tabakalı ve kombinasyon olmak üzere dört başlık altında ele alınmaktadır. Koşullu yöntemde her madde seçiminde daha önceden belirlenen madde kullanım oranı dikkate alınarak uygulanacak madde belirlenir. Rastgele seçim yönteminde ise, belirlenen maksimum bilgilendirici madde grubundan uygulamak için rastgele bir madde seçilir (Kingsbury & Zara, 1989). Koşullu yöntemde madde kullanım oranı bir eşik değere sahipken rastgele seçim yönteminde ise eşik değeri aşılabılır. Tabakalı yöntem ise madde ayırt ediciliklerine göre maddelerin sınıflandırılmasıdır. Buradaki amaç yüksek ayırt ediciliğe sahip maddelerin BBT uygulamasının ilk aşamalarına oranla yetenek kestirimlerin daha doğru olduğu testin son aşamalarında kullanılmasına olanak sağlamaktır (Chang & Ying, 1999). Kombinasyon yöntemi ise uygulamada birden çok madde kullanım sıklığı kontrol mekanizmasının kullanılmasıdır.

### ***Yetenek Kestirimi***

BBT'nin verimli (başarılı) bir şekilde uygulanması, az sayıda maddeye dayalı olarak  $\theta$  kestirimindeki doğruluğa ve madde seçim prosedürlerinin etkinliğine bağlıdır. Etkili bir madde seçim yönteminin yanı sıra, BBT uygulamaları sırasında  $\theta$  kestirimlerinin doğruluğunu artırmak önemlidir.

BBT uygulamaları geleneksel testlerden farklı olarak her bireyin yeteneği ile uyumlu maddeleri cevaplamasına imkân sunar. Birey ilk maddeyi cevapladıktan sonra yetenek kestirimi yöntemi ile bireyin geçici yetenek düzeyi belirlenir. Ardından en çok bilgi veren bir diğer madde uygulanır ve tekrar geçici yetenek kestirilir. Bu süreç testi

sonlandırma kuralına erişinceye kadar devam eder. Nihai yetenek kestirimi ise bütün maddelere verilen cevaplar üzerinden hesaplanır. Geçici ile nihai yetenek kestirim yöntemlerinin aynı olması gerekmez (Chang ve ark., 2000).

Yetenek kestiriminde en sık kullanılan yöntemler En Çok Olabilirlik (Maximum Likelihood, MLE), En Çok Sonsal Dağılım (Maximum A Posteriori, MAP) ve Beklenen Sonsal Dağılım (Expected A Posteriori, EAP)'dır.

MLE, olasılık fonksiyonuna dayalı olarak önceden uygulanan maddelere verilen yanıtları kullanarak yetenek parametrelerini kestirmede kullanılan bir fonksiyondur. Bu nedenle, MLE, sınava girenlerin madde tepkileri (cevapları) göz önüne alındığında en olası yetenek parametrelerini kestirir. MLE, olabilirlik fonksiyonunu maksimize eden  $\theta$ 'yı elde etmeye dayalı bir yöntemdir (Birnbaum, 1968). Olabilirlik fonksiyonu madde cevap fonksiyonlarının çarpılmasıyla elde edilir.

Olabilirlik fonksiyonunu maksimize eden  $\theta$  değerini bulmak için genellikle Newton-Raphson algoritması kullanılır. Bu algoritmada kullanılan MTK modeline dayalı log-olabilirlik fonksiyonunun birinci ve ikinci türevleri hesaplanmalıdır. Ancak fonksiyonun türevlerini hesaplayabilmek için sınava giren kişiye, gerçek gizil özellik düzeylerinin bir tahmini olarak bir değer atanır; bu değer, sınava giren hakkında ön bilgiler kullanılarak belirlenebilir veya tüm sınava girenler için eşit olabilir (Ör:  $\theta = 0$ ). Böylelikle fonksiyonun birinci ve ikinci türevleri hesaplanarak bu iki değerlerin oranı olan  $\epsilon$  bulunur. Elde edilen  $\epsilon$  değeri bir önceki değerden çıkarılarak bir sonraki olabilirlik fonksiyonun başlangıç değeri elde edilir. Bu süreç,  $\epsilon$  belli bir değerden küçük olana kadar tekrarlanır (Embretson & Reise, 2000; Wang & Vispoel, 1998).

Bayes yöntemleri olan EAP ve MAP, popülasyon yetenek dağılımı varsayımının ya da ön bilgilerin yetenek kestirim sürecine dahil edilmesiyle MLE'den farklılaşır (Wang & Vispoel, 1998). Temel olarak, Bayes temelli yöntemler, bir maddenin doğru veya yanlış cevaplanma olasılığı ile öncül dağılımının bir kombinasyonu sonucunda sonsal dağılım elde edilir (Embretson & Reise, 2000). Bu sonsal dağılım bir sonraki madde için öncül

dağılım olarak değerlendirilir. Ön bilgi ile bireyin tahmin edilen yetenek düzeyinin hangi aralıkta olabileceği belirlenerek öncül dağılım elde edilir. Bu ön bilgi, hem testin etkinliğini hem de yetenek kestiriminin hassasiyetini artırarak, sınava giren hakkında daha fazla bilgi verir ve tüm olası yanıt modelleri için yetenek kestirimi sağlayabilir. Ancak, uygun olmayan bir ön dağılım seçilirse, özellikle aşırı uçtaki yetenek düzeylerindeki kestirimler yanlı olabilmektedir (Embretson & Reise, 2000). Ön bilgilerin olmadığı durumlarda ise öncül dağılımın ortalama 0 ve standart sapma 1 olan normal dağılıma sahip olduğu varsayılır.

Bayes Model Kestirim olarak da bilinen Maksimum Sonsal Dağılım'ın (Maximum a posteriori, MAP) amacı, sonsal dağılımı veya modu maksimize eden  $\theta$  değerini belirlemektir (Bock & Aitkin, 1981). Beklenen Sonsal Dağılım (Expected a Posteriori, EAP) sonsal dağılımın ortalamasını bulur (Bock & Aitkin, 1981). Bir başka deyişle, MAP yetenek kestiriminde sonsal dağılım modu hesaplanırken, EAP'da ise  $\theta$  kestirimi olarak sonsal dağılımın ortalaması kullanılır. EAP, belirlenen  $\theta$  değerleri için öncül bilgilerden yola çıkarak olabirlik fonksiyonunu ağırlıklandırır.

Bayes yöntemleriyle tek bir değer elde edilmez. Öncül bilgiler kullanılarak kestirilen parametrenin olasılık dağılımını yani sonsal dağılımı elde edilir. Sonsal dağılıma ilişkin denklem aşağıda verilmiştir.

$$P(\theta | \text{öncül}) = \frac{P(\text{öncül} | \theta) * P(\theta)}{\int P(\text{öncül} | \theta) * P(\theta) d\theta}$$

Sözel olarak ifade edersek;

$$\text{Sonsal} = \frac{\text{maksimum\_olabilirlik} * \text{öncül}}{\text{ortalama\_olabilirlik}}$$

EAP yönteminin hesaplama kolaylığı sürekli bir fonksiyon olan olabirlik fonksiyonunun süreksiz hale getirilmesidir. Fonksiyonun altında kalan alanı bulmak yerine sadece belirlenen noktalardaki yükseklik bulunur (Baker & Kim, 2004). Gauss-Hermite Quadratures tekniği (Stround & Sechrest, 1966) olarak adlandırılan bu yöntemde theta

ölçeğinde belli sayıda nokta belirlenir. Bu noktalardaki bireyin öncül bilgi ve maksimum olabilirlik fonksiyonundaki değerlerin çarpımı bulunur. Bu çarpım değerlerinin, tüm noktalardaki çarpım değerlerinin toplamına bölünmesiyle bireyin yeteneğine ilişkin sonsal dağılım elde edilir (de Abreu e Lima, 2019).

BBT uygulamalarının doğası gereği birey yetenek seviyesine uygun maddelerle karşılaşacağı için MLE yönteminde yanlılık değerinin çok düşük olması beklenir. Ancak bu öngörü, madde havuzunun tüm yetenek seviyelerinde yeterli sayıda kabul edilebilir madde içerdiği varsayımına dayanmaktadır (Wang & Vispoel, 1998). Özellikle BBT'de test uzunluğu kısa ve sabit olduğunda, yetenek kestirimleri tutarsız olabilmektedir (Embretson & Reise, 2000). Uygulanan madde sayısı 15 ve daha fazla olduğunda kullanılan yetenek kestirim yöntemleri benzer sonuçlar vermektedir (Babcock & Weiss, 2009).

MLE düşük yetenek düzeyine sahip bireylerin yeteneklerini daha düşük kestirirken yüksek yetenek düzeyine sahip bireylerinkini daha fazla kestirmektedir (Meijer & Nering, 1999). Ayrıca düşük yetenek düzeylerindeki yanlılık miktarı daha fazladır (Lord, 1980). MLE yönteminin bir diğer sınırlılığı ise tek tip cevap türünde yetenek kestirimini gerçekleştirememesidir. Bayes yöntemleri tek bir cevap türünde bile düşük standart hata ile yetenek kestirimi yapılmasını sağlar. Ancak Bayes yöntemleri yetenek kestirimlerini ortalamaya yaklaştırma eğilimindedir. Yani, EAP kestirimlerinin dağılımı, gerçek  $\theta$  dağılımından daha küçük bir varyansa sahiptir (Lord, 1980). Ayrıca, Bayes yöntemlerinde özellikle kısa testlerde uygun olmayan ya da yanlış ön bilgiler, yanlı kestirimlere sebep olmaktadır (Embretson & Reise, 2000). Öncül bilgilerin kestirim üzerindeki etkisi test uzunluğu arttıkça azalmaktadır (Wainer, 2000).

Hambleton ve ark. (1991), MLE'nin artık hata puanlarının EAP'den daha yüksek olduğunu ve EAP kestiriminin doğruluğunun MLE'den daha iyi olduğunu belirtmektedirler. Swaminathan ve ark. (2003) Bayes prosedürünün genellikle MLE'ye kıyasla daha küçük varyansa neden olabileceğini belirtmişlerdir.

Bayes kestirimlerini kullanmanın bazı avantajları (kestirimlerde daha iyi kararlılık ve test sırasında mevcut bilgilerin daha iyi kullanımı) olmasına rağmen, MLE yönteminden ve Bayes yönteminden elde edilen nihai kestirimler nadiren önemli ölçüde farklılık gösterir. Bunun özellikle en az 20 madde uygulandıktan sonra doğru olduğu bulunmuştur (Thissen & Mislavy, 2000).

MLE yönteminin tek tip cevap setine karşı olan sınırlılığına yönelik Han (2016) tarafından Sınırlandırılmış En Çok Olabilirlik (MLEF) alternatif bir yetenek kestirim yöntemi olarak tanımlanmıştır. MLE yönteminde sabit üst ve alt sınırlar belirlenirken MLEF yönteminde ise hayali ayırt ediciliği yüksek çok kolay ve zor iki maddeyi daha bireylerin cevapladığı varsayılır. Bireylerin hepsinin kolay maddeyi doğru, zor maddeyi ise yanlış cevapladığı kabul edilir. Böylece MLEF yöntemi, tipik MLE için sorun teşkil eden tek tip yanıtlara karşı çözüm sunmakla birlikte, yansız kestirim gibi MLE'nin olumlu özelliklerini korur.

### ***Sonlandırma Kuralı***

BBT uygulamalarının son bileşeni ise sonlandırma kuralıdır. Değişken ve sabit uzunluk olmak üzere iki temel kural vardır. BBT uygulamasının doğası gereği ne kadar fazla madde uygulanırsa yetenek o kadar doğru olacaktır. Özellikle de kısa testlerde uzunluk arttıkça bireyin yeteneği daha yansız kestirilir. Ancak belli bir uzunluk sonrasında ise daha fazla madde uygulamak da önemli bir kazanç sağlamayabilir (Babcock & Weiss, 2012).

Sabit uzunluk sonlandırma kuralında, önceden belirlenmiş sayıda madde uygulanınca test sonlanmaktadır. Tüm bireylere testin sonunda ulaşılan ölçme hassasiyetinin derecesi göz ardı edilerek aynı sayıda madde uygulanmaktadır. En önemli avantajı uygulamasının kolay olmasıdır. Yalnızca daha önce uygulanan madde sayısının izlenmesini gerektirir ve madde havuzu kullanımının daha iyi korunmasını sağlayabilir (Thissen & Mislavy, 2000). Ayrıca testi alan adaylar sabit uzunlukta testi almanın daha adil olduğunu düşünmektedirler (Babcock & Weiss, 2012). Sabit uzunlukta durdurma kuralının en büyük dezavantajı, tüm bireylerin farklı hassasiyet derecelerinde ölçülüyor

olması ve genellikle uç yetenek seviyelerinde daha fazla ölçme hatasına neden olmasıdır. Bu duruma ek olarak bireyin yetenek düzeyinde daha az bilgiye sahip maddelerin gereksiz yere uygulanması BBT'nin etkililiğini sınırlandırabilmektedir (Choi ve ark., 2011). Blaise ve Raiche (2002) çalışmalarında madde sayısının sonlandırma kuralı olarak temel alındığı çalışmalarda en az 13 maddenin uygulanmasını önermişlerdir.

Değişken uzunlukta durdurma kuralıyla, araştırmacının amacına özgü belirlenen eşik standart hata değerine ulaşıncaya test sonlandırılır (Weiss, 2004). Tipik olarak, yetenek kestiriminin standart hatası veya sonsal  $\theta$  dağılımının varyansı, ölçme hassasiyet katsayısı olarak kullanılabilir (Thissen & Mislevy, 2000). Örneğin, bireyin yetenek kestirimindeki theta değişimi de sonlandırma kuralı olarak kullanılmaktadır. BBT uygulamasında daha fazla madde uygulandıkça bireyin ardışık maddeler sonucunda kestirilen theta değerleri birbirine yaklaşır. Yetenek düzeyindeki bu yakınsama belli bir değere ulaşıncaya test sonlandırılır. Değişken uzunluklu BBT uygulamalarında sonlandırma kriteri her test için değişebilir ve testin amacına bağlıdır. Test geliştiriciler, sonlandırma kuralının seçiminin genel olarak; testin amacına, madde havuzunun özelliğine ve işlevsel kısıtlılıklara dayalı olarak yapılması gerektiğini belirtmişlerdir (Segall, 2004).

Değişken uzunluklu bir BBT uygulamasının başlıca avantajı, tüm sınava girenler için ölçme hassasiyetinin aynı seviyede olmasıdır (Thissen & Mislevy, 2000). Genel olarak, sınava giren kişinin  $\theta$  kestirimi ile test sırasında madde güçlük parametresi arasında mükemmel bir eşleşme varsa, sınava giren kişi daha kısa bir test alabilir, bu da madde havuzunun en uygun düzeyde kullanımını sağlar. Bununla birlikte, sınava girenler, test sırasında farklı sayıda madde alırlarsa, değişken uzunluklu bir testin hakkaniyetini anlamakta zorluk çekebilirler (Bergstrom & Lunz, 1999).

### **BBT Uygulamalarında Maddeyi Yeniden Cevaplamaya İzin Veren Yöntemler**

Araştırmacılar hem maddeyi yeniden cevaplamaya izin veren BBT uygulamalarının ölçme hassasiyetini hem de bu uygulamaların test puanı şişirme stratejilerine karşı

dayanıklılığını artırmaya yönelik farklı yöntemler geliştirmektedirler. Çok sayıda cevap değişikliği, BBT'yi daha fazla uyumsuz madde seçmeye zorlayacak ve özellikle bu değişikliklerin tümü yanlıştan doğruya veya tam tersi olduğunda, kestirim hassasiyetini büyük ölçüde azaltacaktır (Stone & Lunz, 1994). Bu nedenle, madde incelemesinin aşırı kullanımını sınırlamak (Han, 2013; Stocking, 1997) veya kestirim prosedürünü iyileştirmek için çeşitli yöntemler önerilmiştir (Papanastasiou & Reckase, 2007). Bu yöntemlerden Stocking (1997) tarafından geliştirilen Model 1, Papanastasiou ve Reckase (2007) tarafından geliştirilen madde dizisini yeniden düzenleme (rearrangement procedure) ve Han'ın (2013) geliştirdiği maddeyi askıya alma (item pocket) ve Cui ve ark. (2018b) tarafından geliştirilen Harmanlanmış BBT uygulamaları (Computerized Adaptive Testing With Salt) aşağıda sırasıyla açıklanmıştır.

**Stocking Modelleri:** Stocking (1997) maddeyi yeniden gözden geçirmeye yönelik farklı modeller geliştirmiştir. Modellerin hepsinde sınırlı madde incelemesine ve yanıt değişikliğine izin verilmektedir. Model 1'de testteki bütün maddeleri yanıtladıktan sonra belirli bir sayıda madde yanıtının değiştirilmesine izin verilmektedir. Model 2'de ise test farazi olarak bölümlere ayrılmakta ve bölümü tamamladıktan sonra tekrar cevap değişikliğine izin verilmemektedir. Son modelde ise ortak soru köküne sahip maddeler içerisinde cevap değişikliğine izin verilmektedir.

**Madde Dizisini Yeniden Düzenleme (Papanastasiou & Reckase, 2007):** Bireylerin testi tamamladıktan sonra maddeleri yeniden gözden geçirmesine izin verilmektedir. Bu yöntemde belirli bir sayıda madde yanıtı değişikliğine izin verilmektedir. Bireylerin maddeleri yeniden cevaplamaları üç modelle tasvir edilmiştir. Birinci model, bireyin daha önce yanlış yanıtladığı bir maddeyi yine yanlış yanıtlamasıdır. Bu modelde madde dizisinde değişiklik yapılmaz ve yetenek kestiriminde var olan madde cevapları kullanılır. İkinci model ise bireyin daha önce yanlış yanıtladığı maddeyi bu sefer doğru cevaplamasıdır. Bireyin daha önceden kestirilen yeteneği  $\theta_i$  iken bu maddeye verdiği yanlış yanıtı düzelttiği zaman bireyin yeteneği  $\theta'_i$  ( $\theta'_i > \theta_i$ ) olarak belirlenir. Bu nedenle

bireyin bir sonraki karşılaştığı maddenin daha zor olması gerekir. Bu maddeden sonraki ilk dört madde arasında, zor maddeye kadarki maddelere ilişkin yanıtlar yetenek kestiriminde kullanılmaz. Eğer ilk dört madde arasında yanlış yanıt yer almazsa dördüncü madde yetenek kestiriminde kullanılır. Üçüncü model ise bireyin daha önce doğru yanıtladığı maddeyi bu sefer yanlış cevaplamasıdır. Bireyin daha önceden kestirilen yeteneği  $\theta_i$  iken bu maddeyi düzelttiği zaman bireyin yeteneği  $\theta'_i$  ( $\theta'_i < \theta_i$ ) olarak belirlenir. Bu nedenle bireyin bir sonraki karşılaştığı maddenin daha kolay olması gerekir. Bu maddeden sonraki ilk dört madde arasında kolay maddeye kadarki maddelere ilişkin yanıtlar yetenek kestiriminde kullanılmaz. Eğer ilk dört madde arasında doğru yanıt yer almazsa dördüncü madde yetenek kestiriminde kullanılır.

Maddeyi Askıya Alma Yöntemi (item pocket): Stocking modelleri test tamamlandıktan sonra maddeyi yeniden gözden geçirmeye izin verirken maddeyi askıya alma yöntemi ise bu kısıtlamayı ortadan kaldırmaktadır. Yanıtlayıcı cevabından emin olmadığı soruları askıya alabilmekte ve daha sonra cevaplayabilmektedir. Birey istediği zaman askıya aldığı maddeyi yeniden gözden geçirebilmekte ve maddeye cevap verme zorunluluğu bulunmamaktadır (Han, 2013). Ek olarak, test süresi dolduğunda askıya alınan maddeler cevaplandırılmamışsa yanlış sayılır. Han, literatürdeki cevap değiştirmeye yönelik çalışmaları inceleyerek hangi maddenin askıya alınacağına yönelik olasılık modelleri oluşturmuş ve bu verilerle simülasyon çalışması gerçekleştirmiştir. Bireyin yetenek düzeyine göre zor bir maddenin askıya alınacağı kabul edilmiştir. Eğer madde kolay ise bireyin maddeyi yanıtlayacağı düşünülmüştür. Askıya alma özelliğinin tercih edilebileceği maddelerin  $b$  parametrelerinin bireyin theta düzeyinden yüksek olduğu maddeler için simüle edilmiştir. Eğer  $b$  parametresi theta değerlerinin yarısından fazla ise zor bir madde olarak kabul edilmekte ve %70 olasılıkla madde askıya alınmaktadır. Eğer  $b$  parametresi theta değerinin yarısından küçükse %50 olasılıkla askıya alınmaktadır. Eğer daha önceden belirlenen sayıda madde askıya alındıysa, askıya alınması planlanan maddenin  $b$  parametresi ile askıdaki en düşük  $b$  parametresine sahip madde karşılaştırılır. Eğer



askıdaki maddenin  $b$  parametresi bireyin maddeyi cevaplama olasılığının logaritmik değerinin yarısından fazla ise o madde %70 olasılıkla en kolay madde olarak belirlenmektedir. Eğer düşükse de %50 olasılıkla en kolay madde olarak belirlenmektedir.

Harmanlanmış BBT Uygulamaları (Computerized Adaptive Testing With Salt; Cui ve ark., 2018b): Bu yöntemin kullanıldığı BBT uygulamalarında bireylere uyarlanmış test maddelerinin yanı sıra daha önceden belirlenmiş test maddeleri uygulanır. Bir başka deyişle, bireyler kestirilen geçici yetenek düzeyinden bağımsız maddelerle karşılaşır. Böylece birey bir önceki madde ile bir sonraki maddenin günlük düzeylerini tahmin edebilse bile bu çıkarım bir önceki maddenin doğru ya da yanlış olduğu hakkında ipucu vermemektedir. Her ne kadar harmanlanmış BBT uygulamaları sınırsız sayıda madde cevap değişikliğine izin verse de sabit test maddelerinin uygulamada yer alması test uzunluğunu artırmakta ve uygulamanın verimliliğini düşürmektedir.

4 Parametrelili Lojistik Model (Barton & Lord, 1981): BBT uygulamalarında yüksek yetenekli bir aday bir testteki ilk maddeleri dikkatsizce yanlış cevapladığında, sonraki maddeler onun gerçek yetenek düzeyinden çok daha kolay maddeler olacaktır. Barton ve Lord (1981) bu problemle başa çıkmak için, yüksek yetenekli bir öğrencinin, yeteneğini daha doğru kestirebilmek için aynı şans parametresi gibi bir başka parametreyi daha modele eklemişlerdir. Oluşturulan 4PL modelde bireylerin dikkatsizlik sonucu doğru cevaplama gereken maddeyi yanlış cevaplayabileceğini de dikkate alan 4PLM geliştirilmiştir. Bu model gözden geçirilebilir BBT'deki kestirim yanlılığını önemli ölçüde azaltmaya yönelik olarak geliştirilmiştir. Test puanı şişirme stratejilerine karşı diğer cevap değişikliği modelleriyle birlikte kullanılmaktadır (Yen ve ark., 2012).

### **BBT Uygulamalarında Test Stratejileri**

BBT uygulamalarında bireylerin kullanabileceği Wainer ve Kingsbury olmak üzere iki temel test puanı artırma stratejisi bulunmaktadır. Wainer stratejisinin kullanıma yönelik gerçekleştirilen simülasyon çalışmasında yetenek kestirimi olarak EAP yöntemi

kullanıldığında bireylerin sistematik puan kazanımlarının yalnızca düşük ila orta yetenekli adaylar için gerçekleştiğini ve kazanımlarının çok küçük olduğu tespit edilmiştir (Vispoel ve ark., 1999) Bununla birlikte, MLE yöntemi kullanıldığında yüksek yetenek düzeyine sahip bireyler için Wainer stratejisini bütün test maddelerinde uygulayarak maddelerin hepsini doğru cevapladığında kazanım elde edilebilmektedir. Wainer stratejisinin başarılı bir şekilde uygulanması çok zordur. Gershon ve Bergstrom (1995) bir maddede bile test stratejisi uygulanmadığında yeteneğin büyük ölçüde küçümsenmesine yol açtığını belirtmektedirler.

### **BBT Uygulamalarında Öncül Bilgilerin Kullanılması**

BBT uygulamalarında ilk aşamada geçici yetenek değeri belirlenir. Bu sabit, rastgele veya öncül bilgilerle belirlenen değer olabilir. Başlangıç yetenek düzeyi olarak, sabit ya da rastgele bir değer belirlendiği durumlarda gerçek yetenek seviyesinden çok uzak başlangıç değerleri ortaya çıkabilir (Matteucci & Veldkamp, 2013). Öncül bilgi ise Bayes yetenek kestirim yöntemlerini avantajlı kılmaktadır. Öncül bilgi BBT uygulamalarının ilk maddelerinde bile kişinin gerçek yeteneğine daha yakın maddelerin seçilmesini sağlayarak theta yakınsamasını hızlandırır (van der Linden, 1999). Özellikle test uzunluğu kısa olan BBT uygulamalarında öncül bilginin kullanılması yetenek kestirimindeki yanlılığı azaltabilmektedir. Ayrıca sınava giren tüm kişiler için geçici yetenek düzeylerinin aynı belirlenmesi, bu yetenek düzeyine yakın maddelerin daha sık kullanılmasına bu da test güvenliğinin sınırlandırılmasına yol açmaktadır. Öncül bilgilerin kullanıldığı durumlarda ise bireylerin başlangıç yetenek düzeyleri farklılık gösterecek ve her düzeyde yer alan maddelerin daha eşit bir şekilde kullanılmasını sağlayacaktır. Öncül bilgiler testin tasarımını iyileştirmek için kullanılabilir gibi izin verildiğinde nihai yetenek kestiriminde de kullanılabilirler (van der Linden & Pashley, 2010).

van der Linden (1999), Hollanda Genel Yetenek Testi Bataryasının BBT uygulamasında Karşılaştırma testi ile Kelime Bilgisi testi arasında bir ilişki kurarak öncül

bilgiyi tanımlamıştır. Araştırmacı bu ilişkiden ve bir testteki önceki performansından (İsim Karşılaştırma) yola çıkarak Kelime Testi'ne ilişkin bireylerin geçici yetenek düzeylerini belirlemiştir. Veldkamp ve Matteucci (2013) de ilişkili iki alt testten yola çıkarak gerçekleştirdikleri simülasyon çalışmasında öncül bilgilerin kullanılarak test uzunluğunun önemli bir ölçüde azalacağını belirtmişlerdir.

BBT uygulamalarında öncül bilgilerin varlığı, başlangıç kuralı ve madde seçim yöntemini belirlemeye yardımcı olmaktadır. Belli bir uzunluktaki BBT uygulamalarında bütün yetenek kestirim yöntemleri ile elde edilen sonuçlar gerçek değere yakınsamaktadır (Thissen & Mislevy, 2000). Algoritmanın yakınsama hızı, algoritmanın başlatılmasına, başlama kuralına bağlıdır. İlk yetenek tahmini bireyin gerçek yeteneğinden ne kadar uzaksa, yakınsama o kadar yavaş gerçekleşecektir (van der Linden, 1999). Wainer'e (2000) göre BBT uygulamalarında anahtar sorulardan bir tanesi de ilk maddenin nasıl belirleneceğidir. Xu ve ark. (2016) BBT uygulamalarında kötü bir başlangıcın nihai yetenek kestirimlerini olumsuz etkilemediği konusunda alan yazında fikir birliği olduğunu ancak bu çıkarımın kısa bir test için geçerli olmadığını, 20'den fazla maddenin uygulandığı durumlarda geçerli olduğunu belirtmişlerdir. Alan yazında ilk maddenin seçimine yönelik farklı yöntemler kullanılmaktadır. Sınava girenlerle ilgili ön bilgiler (yani önceki sınav puanları, notları vb.) kullanılabilen veya sınava girenlerin nihai puanlarını etkilemeyen bir dizi madde, ilk maddeyi belirlemek için tüm sınav katılımcılarına uygulanabilmektedir. (Sireci, 2003).

Geleneksel testlerde ön bilgilerin kullanılması yaygın değildir. Bireyselleştirilmiş testlerde bireye ait ön bilginin kullanımı geleneksel testlere nazaran daha kolaydır. BBT uygulamalarında ön bilgilerin kullanılmasına olanak sağlayan Bayes yöntemlerinin kullanılması uygulamanın verimliliğini artırmaktadır. Bu yaklaşım, bir yetenek kestirimine yakınsama için gereken madde sayısını azaltabilecek, madde kullanım sıklığını düşürebilecek ve ölçme hassasiyetini artırabilecektir (van der Linden, 1999; Veldkamp & Matteucci, 2013). Örneğin MLE yönteminin kullanıldığı BBT uygulamalarında genellikle

başlangıç değeri olarak popülasyon için beklenen dağılımın orta değeri olan 0 (sıfır) tayin edilir. Ancak yararlı ön bilgilerin kullanıldığı Bayes yöntemlerine nazaran MLE yönteminin yakınsama hızı daha yavaş olacaktır. Guyes (2008) BBT uygulamalarında kötü bir başlangıcın çok yavaş bir theta yakınsamasına neden olabileceğini göstererek, daha doğru başlangıç yetenek tahminlerinin önemini vurgulamaktadır. Test sırasında adayların yetenekleriyle ilgili çeşitli sosyo-ekonomik veya demografik bilgileri veya daha önceki testlerden elde edilen puanları da ön bilgi olarak değerlendirilebilmektedir. Alanyazını ön bilginin kullanılıp kullanılmaması gerektiğine dair fikir birliğinin olmadığını göstermektedir. Bireyler hakkında elde edilen her ön bilginin yetenek kestirimine olumlu bir katkısı olmayacağı, yanlış sonuçlar doğurabileceği ve etik açıdan uygun olmayacağı belirtilmektedir (van der Linden, 1999, van der Linden & Pashley, 2010). Yararlı ön bilgilerin olmadığı ve Bayes yöntemlerinin kullanıldığı BBT uygulamalarında genellikle öncül dağılım olarak ya tek düze bir dağılım ya da beklenen yetenek dağılımının ortalamasına yönelik büyük varyanslı normal bir dağılım seçilir. Tek düze dağılım sınava giren birey hakkında mevcut olası bilgileri yok sayar ve büyük varyanslı öncül dağılım ise sınava girenler hakkında arzu edilen ön bilgiyi sağlayamayabilir (van der Linden, 1999). Daha küçük bir varyansa sahip bilgilendirici ön bilgilerle daha öznel sonuçlar elde edilebilir (van de Schoot & Depaoli, 2014).

Daha önce uygulanan bir testten elde edilen bilgi, özellikle bir test bataryasındaki tüm testler için ölçüm hassasiyetini iyileştirmenin basit bir yoludur (Weiss, 2004). Aynı sınava girenlere uygulanan birkaç alt testten oluşan bir test grubu, test bataryası olarak tanımlanır. Test bataryasının alt testlerinin ölçtüğü değişkenler birbiriyle yüksek düzeyde ilişkilidir (Weiss, 2004). Günümüzde uygulanan test bataryalarına örnek olarak ASVAB, GMAT, TIMSS ve PISA gösterilebilir. Örneğin ASVAB aritmetik muhakeme, kelime bilgisi, okuduğunu anlama ve matematik becerileri gibi birçok alt testten oluşmaktadır. Bu boyutlar arasındaki korelasyonlar 0,5 ile 0,7 arasında değişmektedir (Yao, 2013).

Günümüzde BBT olarak uygulanan test bataryaları genellikle tek boyutlu olarak ele alınmaktadır. Ayrıca alt testlere ait yetenek düzeyleri herhangi bir diğer alt testten elde edilen ön bilgi kullanılmadan bağımsız olarak kestirilir. Hâlbuki birbiriyle ilişkili olan bu alt testler birbirine bilgi sağlayıp yetenek kestirimini iyileştirebilmektedirler (van der Linden, 1999). Alt testler arasındaki korelasyon ne kadar yüksek olursa, elde edilen ön bilgi diğer alt testlerdeki performansını tahmin etmede iyi bir yordayıcı olacaktır (de la Torre, 2009). Böylece aynı ölçme hassasiyeti daha az madde ile elde edilebilir. Weiss (2004) yaptığı çalışmada ölçme hassasiyetinde azalma olmadan test bataryasının uzunluğunun %80'i veya daha fazlası kadar azaltılabileceğini göstermiştir.

BBT uygulamalarında ön bilginin kullanılmasının bir diğer avantajı da madde havuzunun daha verimli kullanılmasıdır. Ön bilginin kullanıldığı BBT uygulamalarında her birey yetenek düzeyine daha uygun olan ilk madde ile teste başlar. Ortalama zorlukta bir maddeden ziyade düşük yetenek düzeyine sahip bireyler daha kolay maddeleri cevaplar. Yüksek yetenek düzeyine sahip bireyler daha zor maddeleri cevaplar. Böylece orta güçlükteki madde kullanım oranı azalacaktır (van der Linden, 1999).

Sonuç olarak test bataryalarında ön bilginin kullanılması, test uzunluğu ve madde kullanım oranlarını azaltabilmektedir. Dolaylı olarak da testin daha kısa zamanda gerçekleştirilmesini ve testin daha ekonomik olmasını sağlar (Luecht & Sireci, 2011).

Alan yazını incelendiğinde ön bilginin kullanıldığı BBT uygulamaları sınırlı sayıda. Veldkamp ve Matteucci'nin (2013) ön bilginin başlatma kuralında kullanılması ile nihai sonucu etkileyecek şekilde kullanılmasını karşılaştırdıkları çalışmalarında nihai puanın hesaplanmasındaki performansı daha iyi görülmektedir. Ancak, aynı araştırmacılar ve diğerleri ön bilgilerin sadece uygulamanın başında ve sürecinde kullanılmasını, nihai kestirimde ise yalnızca madde cevaplarının kullanılmasını önermektedir (Veldkamp & Matteucci, 2013). Fakat bu görüş testin amacına göre yorumlanabilir. Eğer testin amacı değerlendirme içeriyorsa bu görüş ağır basmaktadır. Ancak eğer izleme ya da durum tespiti amaçlanıyorsa ölçme hassasiyetinden minimal düzeyde taviz vererek ön bilgiler

uygulamayı daha ekonomik yapacaktır. Veldkamp ve Matteucci (2013) BBT maliyetlerini azaltmak için ön bilgilerin dikkate alındığı Bayes tabanlı BBT uygulamalarının kullanılmasını tavsiye etmektedir. Böylece uç yetenek düzeylerinde yer alan adaylar için test uzunluğunda önemli bir azalma elde edilebilecektir.

### **İlgili Araştırmalar**

Psikometrideki gelişmelerle birlikte BBT uygulamaları test düzenleyicileri tarafından geleneksel testlere göre daha fazla tercih edilmektedir. BBT uygulamalarını geleneksel testlerden üstün kılan özellik ise daha az sayıda madde ile ölçme hassasiyetinden ödün vermeden bireyin yeteneğinin kestirilmesine olanak sağlamasıdır. Araştırmacıların ortak hedefi, bu özelliği en iyi seviyeye taşıırken testi alan bireylere olabildiğince kâğıt-kalem uygulamalarında sunulan olanakları BBT uygulamalarında da sunmaktır. Bu doğrultuda araştırmacılar daha az madde ile ölçme hassasiyetini artırmak için ön bilgilerin kullanılabilmesine işaret etmişlerdir. Ayrıca, BBT uygulamalarını testi alan bireyler açısından da değerlendiren araştırmacılar, cevap değişikliğine yönelik çalışmalar da gerçekleştirmişlerdir. Ancak, alanyazında yer alan çalışmalar incelendiğinde her iki durumu birlikte ele alan çalışma bulunmamaktadır. Alan yazında yer alan ilgili çalışmalar ön bilginin BBT uygulamalarında kullanıldığı ve madde cevap değişikliğine izin verildiği durumlar olmak üzere iki başlık altında ele alınmıştır.

### ***Cevap Değişikliğine İzin Verilen BBT Uygulamaları***

Lunz ve ark. (1992) sertifika almaya hak kazanan bireyleri belirlemek amacıyla BBT uygulaması gerçekleştirmişlerdir. Bu çalışmada maddeyi yeniden cevaplayabilmeye izin veren BBT uygulamalarının ölçme hassasiyetine etkisi incelenmiştir. Deneysel desen olarak tasarlanan çalışmada gruplar rastgele oluşturulmuştur. Kontrol grubuna BBT uygulamasında maddeyi yeniden gözden geçirmeye izin verilmezken deney grubunda ise bireylerin testi sonlandırma kriteri sağlandığında maddeleri yeniden gözden geçirmelerine ve yanıt değiştirmelerine izin verilmiştir. Deney grubundaki bireylerin yanıt değişikliği

öncesi ve sonrası cevapları kaydedilmiştir. Araştırma sonucunda testin etkililiği %1 azalmıştır. Yanıt değişikliğine izin verilmeden önce ve verildikten sonra kesme noktasına yakın sadece üç kişinin sınıflama grubunda değişiklik meydana gelmiştir. Araştırmacılar iki eşdeğer grup arasındaki ortalama yetenek farkının, bireylerin dikkatsizlik sonucu yaptıkları hataları düzeltmelerinden kaynaklandığı sonucuna varmıştır. Bireylerin çoğunun belirlenen sınıflama düzeylerinin değişmediğini ve ölçme hassasiyetindeki kaybın, madde cevap değişikliğini kısıtlayacak düzeyde olmadığını belirtmişlerdir.

Stone ve Lunz (1994) yaptıkları deneysel çalışmada BBT uygulamalarındaki cevap değişikliğinin sınıflama doğruluğu üzerindeki etkisini incelemişlerdir. BBT uygulamalarında değişken uzunluklu sonlandırma yöntemi kullanılmıştır. 50 madde uygulandıktan sonra geçme puanına yönelik %5 güven aralığı göz önüne alınarak sonlandırma gerçekleştirilmiştir. Bütün maddeleri cevapladıktan sonra bireylere sınırlandırma olmaksızın cevap değişikliği hakkı tanınmıştır. Bireyler gerçek yetenek düzeylerine göre düşük, normal ve yüksek olmak üzere üç gruba ayrılarak kestirilen yetenek düzeyleri incelenmiştir. Çalışma sonucunda düşük ve yüksek yetenek düzeyi gruplarındaki bireylerin sınıflama doğruluğu değişmezken orta yetenek düzeyine sahip bireylerin sınıflama grubu kısmen değişmiştir. Cevap değişikliği öncesi ve sonrası bireylerin %6'sının sınıflaması değişmiştir. Cevap değişikliği sonucunda bireylerin çoğunlukla puanlarını artırdıkları ancak sistematik bir test stratejisi (Örneğin Wainer stratejisi) kullanmadıkları belirtilmiştir. Cevap değişikliği sonucu oluşan bilgi kaybının ekstra bir madde daha uygulanarak telafi edilebileceğini belirtmişlerdir. Ayrıca, cevap değişikliğinin ölçüm hatası, yetenek kestirimleri ve BBT'nin verimliliği üzerindeki etkisinin minimum düzeyde olduğunu ve BBT uygulamalarında cevap değişikliğine izin verilmesi gerektiğini belirtmişlerdir.

Stocking (1997) maddeyi yeniden gözden geçirmeye yönelik yaptığı simülasyon çalışması ile kendi önerdiği üç modeli sınamıştır. Modellerin hepsinde yetenek kestirimi için MLE, madde seçimi için MFI ve sabit uzunlukta sonlandırma kriteri kullanılmıştır.

Modellerin test stratejilerine karşı dayanıklılığı test edilmiştir. Model 1’de 28 maddelik bir testte 2 madde, 7 madde, 14 madde ve bütün maddelere verilen yanıtların değiştirilmesine izin veren 4 farklı test modülüyle BBT uygulaması simüle edilmiştir. 2 madde değişikliğine izin veren koşul haricinde diğer bütün koşullarda özellikle yüksek yetenek düzeyine sahip bireylerin yetenek kestirimlerinde yüksek düzeyde standart hataya neden olmuştur. Model 2’de ise BBT bölümlere ayrılarak her bölümün kendi içinde maddeyi yeniden gözden geçirmesine izin verilmektedir. 28 maddelik BBT uygulaması, 4 maddelik yedi bölüm, 7 maddelik dört bölüm, 14 maddelik iki bölüm ve 28 maddelik bir bölüm oluşturulmuştur. Araştırma sonucunda 4 ve 7 maddelik yöntemlerde kestirilen yeteneklerin yanlılığının azaldığı gözlemlenmiştir. Model 3’te ise birbiriyle ilişkili madde takımları kullanılarak BBT uygulaması gerçekleştirilmiştir. Madde takımları kendi içerisinde maddeyi yeniden cevaplamaya izin vermektedir. Bu model sonucunda maddeyi yeniden gözden geçirmeye izin veren ve vermeyen BBT uygulamalarından elde edilen yetenek kestirimi standart hatasının benzer olduğu belirtilmiştir. Ama maddelerin birbiri ile ilişkili olma gerekliliği modelin kullanımını sınırlamaktadır.

Olea ve ark. (2000) yaptıkları deneysel çalışmada ise maddeyi yeniden gözden geçirmenin hem BBT uygulamalarının hem de geleneksel testlerin performansları üzerindeki etkilerini ve bunun sınav kaygısı ile ilişkisini incelemişlerdir. Araştırmacılar, bireyleri dört koşuldaki birine (maddeyi yeniden cevaplamaya izin veren BBT ve lineer test uygulamaları, izin verilmeyen BBT ve lineer uygulamaları) rasgele atadılar ve sınava girenlerin kaygı durumlarını uygulama öncesi ve sonrası değerlendirdiler. Kaygı düzeyi ile maddeyi yeniden gözden geçirme koşulları arasında anlamlı bir olumsuz ilişki olduğunu, cevaplarını gözden geçirmelerine izin verilen sınava girenlerin kaygı düzeylerinde bir azalma olduğunu ve gözden geçirmesine izin verilmeyenlerin kaygı düzeylerinde bir artış olduğunu tespit ettiler. Araştırmacılar ayrıca cevap değişikliğinin BBT’nin psikometrik özellikleri üzerindeki etkilerini de incelediler. Cevap değişikliğinin doğru yanıtların sayısını önemli ölçüde artırmasına ve sınava giren kişilerin yetenek düzeylerinin daha yüksek



kestirilmesini sağlamasına rağmen, ölçme hassasiyetini etkilemediği sonucuna ulaşılmıştır. Araştırmacılar, maddeleri gözden geçirmenin yetenek kestirimine katkı sağlayacağını, madde incelemesine izin vermemenin ise sınava girenlerin yeteneklerinin potansiyel olarak yanlış yetenek kestirimlerine yol açabileceğini belirtmişlerdir.

Vispoel ve ark. (2002) test stratejisini kullanan bireylerin yetenek kestirimlerinde kazanç elde edip edemeyeceklerine yönelik deneysel bir çalışma yapmışlardır. Bireyler rastgele olarak iki gruba ayrılıp bir gruba test stratejilerine yönelik eğitim verilmiştir. Her iki gruptaki bireyler rastlantısal olarak test sonunda ya da bölmeler sonunda cevap değişikliğine izin verilen BBT uygulamasını almışlardır. Test stratejisi eğitimi almayan bireylerin cevap değişikliği sonrası yetenek kestirimlerinde olumlu bir değişim olurken eğitim alan bireyler ise test maddelerinin zorluk düzeylerini belirleyemedikleri için cevap değişikliği sonrası daha düşük yetenek kestirimi elde etmişlerdir. Araştırmacılar test stratejisinin muhtemelen sınava girenlere herhangi bir avantaj sağlamayacağını belirtmişlerdir. Ayrıca, bu stratejileri kullanan bireylerin test sürelerinin %50'ye kadar arttığını ve bu süre artışının büyük çaplı uygulamalarda stratejilerin kullanımını zorlaştırabileceğini belirtmişlerdir.

Papanastasiou ve Reckase (2007) yaptıkları simülasyon çalışmasında BBT uygulamalarında maddeleri yeniden gözden geçirmeye izin veren madde dizisini yeniden düzenleme yönteminin (Rearrangement Procedure) etkililiğini belirlemeyi amaçlamışlardır. 30 maddelik BBT uygulaması sonucunda bireylere sadece 5 maddenin cevabının değiştirilmesine izin verilmiştir. Yetenek kestirim yöntemi olarak ML kullanılmıştır. Yeniden düzenleme prosedürünün doğruluğunu ve etkinliğini belirlemek için gözden geçirmeden önce, gözden geçirmeden sonra ve yeniden düzenleme prosedüründen sonra olmak üzere ML ile kestirilen yetenek düzeyleri karşılaştırılmıştır. Cevap değişikliği öncesi yanlışlık -0,1374; sonrasında 0,0718 ve yeniden düzenleme prosedüründen sonra yanlışlık 0.061 olarak belirlenmiştir. Ancak, uç yetenek düzeylerinde, özellikle de düşük yetenek düzeylerinde, cevap değişikliği sonrası yanlışlık değeri artmıştır. Araştırmacılar yeniden

düzenleme yönteminin güvenilirliği olumsuz etkilediği ancak bunun ihmal edilebilir düzeyde olduğunu belirtmişlerdir.

Yen ve ark. (2012) yaptıkları simülasyon çalışmasında gözden geçirilebilir BBT uygulamasında dikkatsizlik sonucunda uygun olmayan maddelerin neden olduğu yanlılığı azaltmak için 4PL modelini kullanmışlardır. Araştırmacılar yüksek yetenek düzeyine sahip bireylerin BBT uygulamasının ilk maddelerini dikkatsizlikten dolayı yanlış cevaplayabileceklerini ve bu durumun bireyin yeteneğinin gerçek değerden daha düşük kestirilmesine sebep olacağını ifade etmektedir. Araştırmacılar çalışmanın sonucunda 4PL modelinin gözden geçirilebilir BBT için yanlılığı önemli ölçüde azaltabileceğini ve yeniden düzenleme prosedürü ile birlikte kullanıldığında daha doğru yetenek kestirimi elde edilebileceğini belirtmişlerdir.

Han (2013) BBT uygulamasını alan bireye maddeyi askıya almaya (item pocket) olanak sunan esnek bir yöntem geliştirmiştir. Araştırmacı maddeyi askıya almaya yöntemini sınıadığı çalışmada sabit uzunlukta bir BBT uygulamasında farklı sayılarda maddeyi askıya almaya izin veren BBT uygulamalarını simüle etmiştir. Yapılan çalışma sonucunda 2 ve 4 maddeyi askıya almaya izin veren uygulamalarda referans modele (maddeyi yeniden gözden geçirmeye izin vermeyen BBT uygulaması) göre benzer sonuçlar elde edilmiştir. Birey madde cevap değişikliğinden ziyade madde cevaplamayı sonraya bıraktığı için test stratejilerinin kullanımını olanaksız kılmaktadır.

Jensen (2017) çoklu puanlanan test maddeleriyle gerçekleştirdiği simülasyon çalışmasında maddeyi askıya alma yönteminin kullanıldığı BBT uygulamasının verimliliğini araştırmıştır. Simülasyonda maddeyi askıya alma miktarı (2, 3 ve 4), sabit ve değişken uzunluklu sonlandırma kurallarıyla oluşturulan koşullarla BBT uygulamasının performansları değerlendirilmiştir. Çalışma sonucunda her iki sonlandırma kuralının kullanıldığı uygulamalarda referans modele göre askıya alınan madde miktarı arttıkça gerçek ve kestirilen yetenek düzeyleri arasında daha yüksek korelasyon, daha hassas ölçüm ve daha az yanlılık elde edildiği tespit edilmiştir.

Cui ve ark. (2018a) BBT uygulamalarında madde cevap değişikliğine izin veren “Madde dizisini yeniden oluşturma” ve “Stocking Model 2 (bölümlere ayırma)” yöntemleri ile sınırsız sayıda cevap değişikliğine izin verildiğinde elde edilen ölçme hassasiyeti ile bu yöntemlerin test stratejisine karşı dirençlerini araştırmışlardır. Cevap değişikliğine izin verilmeyen durum referans model olarak belirlenmiştir. Araştırmada bireylerin test stratejilerinden Wainer, Kingsbury ve rastgele cevaplama gerçekleştirdikleri üç farklı cevap seti simüle edilmiştir. Çalışma sonucunda herhangi bir kısıtlama olmaksızın madde cevap değişikliğine izin veren BBT uygulanırsa, test manipülasyon stratejilerinin sınava girenlerin yeteneklerini ölçme üzerinde etkili olduğunu tespit etmişlerdir. Sınava girenlerin madde cevap değişikliğine izin veren BBT uygulamalarının nasıl uygulandığından bağımsız olarak Kingsbury stratejisini kullanarak çok az kazanım elde edilebildiğini belirtmişlerdir. Wainer stratejisi, madde cevap değişikliğinde herhangi bir kısıtlama olmaksızın uygulandığında, üst yetenek düzeyine sahip bireylerin puanlarını şişirmede etkili olduğu bulunurken hem madde dizisini yeniden oluşturma yönteminin hem de blok inceleme yönteminin bu stratejiye karşı dirençli olduğunu tespit etmişlerdir. Blok büyüklüğü küçüldükçe ölçme hassasiyetinin arttığı tespit edilmiştir. Rastgele cevaplama tekniğinin de Wainer stratejisi kadar etkili olabileceğini vurgulamışlardır. Araştırmacılar bölümlere ayırma yönteminin hem madde cevap değişikliğine izin verdiği hem de test stratejilerinin etkisini kontrol ettiği için test uygulayıcılar ve test alan bireyler için ideal bir BBT uygulaması olabileceğini belirtmişlerdir.

Cui ve ark. (2018b) yaptıkları çalışmada BBT uygulamasına adaptif olmayan maddeler ekleyerek uygulamanın test stratejilerine karşı ne kadar dirençli hale getirmeyi amaçlamışlardır. Çalışma sonucunda araştırmacılar BBT uygulamasında adaptif olmayan madde sayısı arttıkça, uygulamanın daha çok lineer forma benzer hale geldiğini belirtmişlerdir. Araştırmacılar bu olumsuz etkiyi ortadan kaldırmak için test formlarında az sayıda adaptif olmayan ama zor maddelere yer verilebileceğini ya da adaptif olmayan

maddelerin sadece test manipülasyon stratejileri tespit edilen (örneğin kişi-uyumunun zayıf olması vb.) durumlarda kullanılmasını önermektedir.

Lin ve ark. (2021) tek boyutlu BBT uygulaması olarak gerçekleştirdikleri Monte Carlo simülasyon çalışmasında maddeyi askıya yöntemi (Han, 2013) ile Stocking Model 2'yi (Stocking, 1997) bir arada kullanarak uygulamanın performansını test etmişlerdir. Çalışmada test birkaç büyük bloğa bölünmüş (Model 2) ve her blok içinde belirli sayıda maddenin askıya alınmasına izin verilmiş ve testi alanların bir sonraki bloğa geçtiklerinde önceki bloklardaki maddeleri gözden geçirmelerine izin verilmemiştir. Testi alanların Wainer test stratejisini kullandıkları varsayılmıştır. Çalışma sonucunda blok ve maddeyi askıya alma yöntemlerinin kombinasyonu, askıya alınabilen madde sayısı fazla olduğu durumlarda bile kabul edilebilir ölçme hassasiyetine sahip olduğu belirlenmiştir. Elde edilen sonuçlar, bu kombinasyonun gözden geçirilebilir BBT uygulaması için güvenilir bir çözüm olduğunu göstermiştir.

### ***Öncül Bilgilerin Kullanıldığı BBT Uygulamaları***

van der Linden (1999) BBT uygulamasından önce elde edilen ampirik değişkenler arasındaki istatistiksel ilişkiye dayanan BBT başlangıç prosedürü ile test bataryaları için gereken sürenin ölçme hassasiyetinden ödün vermeden azaltılabileceğini önermiştir. Araştırmacı yaptığı deneysel çalışmada Hollanda Genel Yetenek Testi Bataryası'ndan elde edilen verileri kullanmıştır. Daha önceden elde edilen verilerden yola çıkarak test bataryasının ilk uygulanan alt testi cevaplama süresi ile ikinci testin yetenek düzeyi arasındaki istatistiksel ilişki irdelenmiştir. Bu istatistiksel ilişkiden yola çıkarak ikinci testin başlangıç yetenek tahmini ile güven aralığı belirlenmiştir.

de la Torre (2009) test uygulamalarında elde edilen verilerin birbiriyle ilişkili olması durumunda bu verilerin yetenek kestiriminde kullanımını araştıran doktora tez çalışmasını gerçekleştirmiştir. Çalışmada bu verilerin de yetenek kestiriminde kullanıldığı bir model önermiştir. Araştırmacı oluşturduğu modelin, çeşitli madde tepki modellerini kullanarak sınava girenlerin puanlanmasında ön bilgilerin de dâhil edilmesini sağlayan genel bir

formülasyonu olduğunu ve EAP yönteminin özel bir durumu olduğunu belirtmiştir. Çalışma sonucunda yetenek kestirimlerinde ön bilgilerin kullanılmasının ölçme hassasiyetini iyileştirdiğini belirtmiştir. Özellikle yüksek düzeyde korelasyon gösteren yetenekleri ölçen kısa testler kullanıldığında ön bilgilerden elde edilen kazancın en yüksek olduğunu vurgulamaktadır.

Wang ve ark. (2012) test bataryalarında ön bilgilerin kullanımının test bataryasının ölçme hassasiyetine etkisini farklı yetenek kestirim yöntemlerini kullanarak test etmişlerdir. Üç alt testten oluşan 60 maddelik test bataryasında alt testler arasındaki korelasyondan yola çıkarak bir sonraki alt testin BBT başlangıç değerleri tanınmıştır. Simülasyon çalışmasından elde edilen sonuçlar alt testler ve test bataryasının ön bilgilerinin kullanılmasının sınıflandırma doğruluğunu iyileştirdiğini göstermektedir. Önceki alt testten kestirilen değeri bir sonraki test için başlangıç değeri olarak kullanmanın en güvenilir ve en basit yöntem olduğunu belirtmişlerdir. Ayrıca, nihai yetenek kestirimi olarak kullanılan MLE ve Bayes yöntemlerinin her ikisinin de ön bilgilerin kullanıldığı BBT-test bataryası uygulamalarında oldukça iyi performans gösterdiğini vurgulamışlardır.

Matteucci ve Veldkamp (2013) sabit uzunluklu ve değişken uzunluklu BBT uygulamalarında ön bilginin kullanımını inceleyen simülasyon ve deneysel çalışmalar gerçekleştirmişlerdir. Simülasyon çalışmasında ön bilgiye yönelik üç farklı uygulama gerçekleştirilmiştir. Ön bilginin  $N(0,1)$  dağılıma sahip olduğu, ön bilginin sadece başlama kuralı olarak kullanıldığı ve nihai yetenek kestiriminde de ön bilginin kullanıldığı üç farklı simülasyon gerçekleştirilmiştir. Çalışma sonucuna göre ön bilgilerin nihai yetenek kestiriminde kullanılması standart modele göre madde sayısını ve madde kullanım sıklığını azaltmaktadır. Ayrıca hem değişken uzunluklu hem de sabit uzunluklu uygulamalarda standart modele göre daha hassas yetenek kestirimi gerçekleştirilmiştir. Ön bilginin sadece başlama kuralında kullanılması durumunda standart modele göre daha iyi ölçme hassasiyeti sağlarken nihai yetenek kestiriminin kullanıldığı modele göre ise daha az ölçme hassasiyeti sağladığı görülmüştür. Deneysel uygulamada ise ön bilgilerin

nihai yetenek kestiriminde kullanıldığında özellikle uç yetenek düzeylerinde ölçme hassasiyetinin iyileştiği belirlenmiştir.

Wang ve ark. (2014) ön bilgilerin yetenek kestiriminde kullanımının sınıflama doğruluğuna etkisini araştırmışlardır. Çalışmada test bataryasında üç alt test arasındaki korelasyonlardan yola çıkarak bir sonraki alt test için başlangıç değeri tanımlanmıştır. Ayrıca araştırmacılar alt testler arasındaki korelasyondan yola çıkarak bir diğer alt teste ilişkin BBT uygulamasının başlangıç değeri belirlendiğinde uygulamanın daha küçük standart hata değerine sahip olabildiğini vurgulamışlardır. Çalışma sonucunda ön bilginin kullanımının sınıflandırma doğruluğunu artırdığını ve Bayes yöntemlerinin (EAP ve MAP), doğru sınıflandırma oranı açısından MLE yönteminden biraz daha iyi performans gösterdiği belirlenmiştir.

Xie (2019) yaptığı doktora çalışmasında tek ve çok boyutlu BBT uygulamaları çerçevesi altında bireylerin yetenek kestiriminde ön bilgilerin farklı açılardan kullanımını karşılaştırarak alt testler arasındaki korelasyonun yetenek kestirimini nasıl etkilediğini belirlemeyi amaçlamıştır. Araştırmacı tek boyutlu BBT uygulamasında ilk alt testten elde edilen puan ile alt testler arasındaki korelasyondan yola çıkarak bir diğer test için yordanan değeri, ikinci testin başlangıç değeri olarak kullanmıştır. Araştırmacı gerçekleştirdiği simülasyonlar sonucunda özellikle uç değerlerde kabul edilebilir bir seviyede ölçme hassasiyetinden kayıp olduğunu ancak ölçme verimliliğinin iyileştiğini tespit etmiştir. Ön bilginin alt testler arasında yüksek korelasyon olduğunda daha faydalı olduğunu belirtmiştir. Yetenek kestiriminde ön bilginin kullanımının madde sayısını azaltarak güvenilir ölçmelerin yapılabileceğine vurgulamıştır.

### ***İlgili Çalışmaların Özeti***

Alanyazın incelendiğinde BBT uygulamalarında madde cevap değişikliğine izin verilip verilmemesi sorununa yönelik çalışmaların 1990'lı yıllardan günümüze kadar devam ettiği görülmektedir. Cevap değişikliğine yönelik birçok model geliştirilse hiçbirinde test uzunluğunu artırmadan sınırsız sayıda cevap değişikliğine izin verilmemektedir. İzin

verilmemesinin nedeni olarak test stratejileri gösterilmektedir. Yapılan çalışmalarda bireylerin test stratejilerini kullanarak yanlı puanlar elde edebilecekleri ifade edilmektedir. Çalışmalarda oluşturulan modellerin test stratejilerine karşı dirençli olmadığı için madde cevap değışikliđi sayısı genellikle sınırlandırılmıştır.

Alanyazında BBT uygulamalarında öncül bilginin kullanımına yönelik sadece simülasyon çalışmaları yer almaktadır. Öncül bilgiler yapılan çalışmalarda daha çok test uzunluđunu azaltmak amacıyla kullanılmaktadır.

## Bölüm 3

### Yöntem

Bu bölüm araştırmanın türü, araştırma deseni, verilerin elde edilmesi, simülasyon süreçleri ve veri analizi olmak üzere beş kısımdan oluşmaktadır

#### Araştırmanın Türü

Bu araştırma öncül bilginin tanımlandığı bilgisayarda bireyselleştirilmiş test uygulamalarında madde cevap değişikliğinin farklı koşullar altında ölçme hatası ve yanlılığı üzerindeki etkisini belirlemeyi amaçlayan simülatif bir araştırma niteliğindedir. Simülasyon çalışmaları gerçek veri setlerinden ya da dağılımlarından yola çıkarak belirlenen hipotezleri test etmek amacıyla veri üretilmesi ve veri analiz aşamalarından oluşur. Simülasyon çalışmaları gerçek yaşam uygulamasının zor olduğu, çalışmanın koşullarına ilişkin gerçek verinin olmadığı ya da koşulların çok karmaşık olduğu durumlarda sıkça tercih edilir (Ranganathan & Foster, 2003).

Çalışmada ayrıca, test bataryasında yer alan alt testler arasındaki korelasyondan yola çıkarak öncül bilgiler tanımlanmıştır. Çalışma bu yönüyle korelasyonel araştırma türlerinden biri olan yordayıcı korelasyonel araştırma deseni kapsamına girmektedir. Yordayıcı korelasyonel araştırmalarda belirlenen bir ya da daha fazla bağımsız değişkenden yola çıkarak bağımlı değişkendeki değişimler açıklanır (Creswell, 2012).

#### Araştırma Deseni

Çalışmada bilgisayarda bireyselleştirilmiş test (BBT) uygulamalarında maddeyi yeniden gözden geçirmeye izin verildiğinde elde edilen yetenek kestirimlerinin standart hatası ve yanlılığının testin başlangıç kuralına, nihai yetenek kestirim yöntemine ve test uzunluğuna göre nasıl değiştiği incelenmiştir. Ön bilgilerin ve maddeyi yeniden yanıtlamanın etkilerini gözlemlemek için test bataryasında yer alan alt testler arasındaki korelasyon değerinden yararlanarak başlama kuralları ile cevap değişikliğinde



gerçekleşebilecek amaçlı yanlış cevaplama ve rastgele cevaplama verileriyle simülasyon çalışması gerçekleştirilmiştir. Alt testler arasındaki korelasyondan yola çıkarak belirlenen başlama kuralının BBT uygulamasındaki etkisini incelemek için üç model kurulmuştur. Bunlardan ilki herhangi bir ön bilginin olmadığı durumlarda ilk maddenin  $[-1,1]$  aralığından seçilmesidir. İkincisi ise alt testler arasındaki korelasyondan yararlanarak BBT uygulamasının başlatılmasıdır. Sonuncusu ise aynı testin bir önceki uygulamasından elde edilen yetenek kestirimleri kullanılarak BBT uygulamasının başlatılmasıdır.

Çalışma kapsamında BBT uygulamasında bireye yanıt değişikliğine izin verilip/verilmeme durumunun, başlangıç kuralı, nihai yetenek kestirim yöntemi ve test uzunluğu koşullarına göre etkililiği incelenmiştir. Bu doğrultuda bireylerin daha önce uygulanan maddelere verdikleri yanıtları değiştirmeye izin verilip vermeme durumuna göre üç farklı cevaplama stili (bilinçli, rastgele ve amaçlı yanlış cevaplama), bireyler hakkında herhangi bir öncül bilginin kesinliği durumuna göre 3 farklı başlama kuralı (öncül bilginin olmadığı, öncül bilginin regresyon ve aynı testin daha önceki sonucu ile belirlenmesi), 4 farklı nihai yetenek kestirim yöntemi (EAP, EAP(0,1), MLE, MLEF) ve 5 farklı test uzunluğu (5, 10, 15, 20, 25) olmak üzere toplam  $(3*3*4*5)$  180 koşul belirlenmiştir. Araştırmanın simülasyon koşulları ve düzeyleri Tablo 2’de yer almaktadır.

**Tablo 2**

*Simülasyon Koşulları ve Düzeyleri*

Simülasyon koşulları	Düzye Sayısı	Düzeyler
Cevaplama stratejisi	3	Bilinçli, rastgele, amaçlı
Başlama kuralı	3	Rastgele, regresyon, tekrar
Yetenek kestirimi	4	EAP, EAP(0,1), MLE, MLEF
Test uzunluğu	5	5, 10, 15, 20, 25

## Veri Kümesi

Araştırmada kullanılan veri seti 2015 yılında uygulanan PISA verilerinden oluşmaktadır. Çalışma grubu için bilgisayar ortamında testi alanlar ile fen ve matematik testlerinden en az on madde cevaplayan bireyler belirlenmiştir. Demografik değişken şartı gözetmeksizin bu şartları taşıyan bireylerin hepsi çalışmaya dâhil edilmiştir. Çalışma grubu 13605 kişiden oluşmaktadır.

## Veri Toplama Süreci

Veri toplama süreci test maddelerinin belirlenmesi, fen ve matematik yetenek düzeylerinin kestirilmesi, cevap matrislerinin oluşturulması ve koşulları olmak üzere dört alt başlıktan oluşmaktadır.

### *Test Maddelerinin Belirlenmesi*

Araştırmaya PISA 2015 yılında sadece ikili puanlanan 171 fen ve 73 matematik maddeleri dâhil edilmiştir. Madde parametreleri olarak PISA tarafından 2 PL model ile kestirilen uluslararası/ortak madde parametreleri kullanılmıştır (OECD, n.d.). Madde parametrelerine ilişkin betimsel istatistikler Tablo 3'te verilmiştir. Matematik alt testindeki maddelerin ortalama güçlük düzeyi -0,093, fen maddelerinin ortalama güçlük düzeyi ise 0,025'tir. Matematik maddelerinin madde ayırt edicilikleri [0,220;2,303], fen testinin madde ayırt edicilikleri ise [0,402;2,475] aralığındadır. Madde sayısının az olması nedeniyle düşük ayırt ediciliğe sahip maddeler, madde havuzundan çıkarılmamıştır.

**Tablo 3**

*PISA Alt Testlerine Ait Madde Parametrelerine İlişkin Betimsel İstatistikler*

Alt Testler	Parametre	N	Ortalama	Standart Sapma	Minimum	Maksimum
Matematik	a	73	1,139	0,389	0,220	2,303
	b	73	-0,093	0,771	-1,810	1,710
Fen	a	171	1,171	0,392	0,402	2,475
	b	171	0,025	0,557	-1,412	1,953

### ***Fen ve Matematik Yetenek Düzeylerinin Kestirilmesi***

Çalışma grubu ve maddelerin belirlenme aşamasından sonra elde edilen eksik cevap matrisi ile bireylerin matematik ve fen yetenek düzeyleri CATSim programında hibrit simülasyon özelliği ile EAP yöntemi kullanılarak 2PL modele göre kestirilmiştir. Bireylerin eksik cevap matrisi ile kestirilen yetenek düzeylerine ilişkin psikometrik değerler Tablo 4'te verilmiştir. Bireylerin ayrıca oluşturulan tam cevap matrisleri ile de yetenek düzeyleri kestirilmiştir. Eksik ve tam cevap matrisleriyle elde edilen yetenek kestirimleri arasındaki korelasyon ise 0,99 ( $p < 0,01$ ) olarak hesaplanmıştır. Tam cevap matrisleriyle kestirilen yetenek düzeyleri bireylerin gerçek yetenek düzeyi olarak kabul edilmiştir. Bu değerler alt problemlerde referans değerler olarak kullanılmıştır. Bireylerin yetenek düzeylerine ilişkin betimsel değerler Tablo 4'te verilmiştir. Bireylerin matematik yetenek düzeylerinin ortalaması -0,262; fen yetenek düzeylerinin ortalaması ise -0,070'dir. Yetenek dağılımlarının basıklık ve çarpıklık değerleri incelendiğinde normal dağılım gösterdiği söylenebilir (George & Mallery, 2010).

**Tablo 4**

#### ***Fen ve Matematik Alt Testlerine İlişkin Betimsel Değerler***

Alt testler	N	Ortalama	Standart Sapma	Minimum	Maksimum	Basıklık	Çarpıklık
Matematik	13605	-0,262	0,805	-4,000	4,000	-0,136(0,042)	0,018(0,021)
Fen	13605	-0,070	0,694	2,700	2,674	0,235(0,042)	0,099(0,021)

#### ***Cevap Matrislerinin Oluşturulması***

Çalışmada eksik cevap matrisinden tam cevap (bilinçli) matrisi ile rastgele ve amaçlı yanlış cevap matrisleri türetilmiştir. Bireyler belirlenen maddelerin tamamını almadıkları için madde ve çalışma grubu belirleme süreci sonunda eksik cevap matrisi elde edilmiştir. Bireylerin kestirilen yetenek düzeylerinden yola çıkarak CATSim programında gerçekleştirilen hibrit simülasyonla tam cevap matrisi elde edilmiştir.

Bireyin almadığı maddeye verdiği cevabı simüle etmek için öncelikle madde parametrelerine ve yetenek düzeyine dayanarak, sınava giren kişinin ilgili MTK modeline göre maddeyi doğru cevaplama olasılığı ( $p$ -değeri) hesaplanmıştır. Bu  $p$  değeri daha sonra tek biçimli bir dağılımdan  $U(0,1)$  rastgele oluşturulmuş bir sayı ( $x$ ) ile karşılaştırılmıştır.  $p$ -değeri, tek tip rastgele sayıdan büyük veya ona eşitse, doğru; aksi halde madde için yanlış cevapladığı kabul edilmiştir.

Rastgele cevap matrisi ise bireylerin önce maddeleri rastgele olarak cevapladıkları, cevap değişikliğine izin verildiğinde ise bilinçli bir şekilde cevaplayacakları varsayımı ile oluşturulmuştur. Bu nedenle bireylerin maddeye doğru cevap verme olasılığı (%25) dikkate alınarak rastgele cevap matrisi oluşturulmuştur.

Amaçlı yanlış cevap matrisi de bireylerin bütün maddeleri amaçlı olarak yanlış cevaplamaya çalışacakları ve kolay maddelerle karşılaşacakları durumu dikkate alınarak oluşturulmuştur. Bireylere sınırsız sayıda cevap değişikliği verildiğinde olabildiğince cevap değişikliği yaptıkları farz edilmiştir. Yani BBT uygulaması için en kötü senaryo olan bütün maddelere verilen cevapların değişmesidir. Bireylerin maddeleri amaçlı olarak yanlış cevaplama olasılıklarını hesaplarken Cui ve arkadaşları (2018b) tarafından belirlenen amaçlı yanlış cevaplama denklemi kullanılmıştır. Sınava giren kişi doğru cevabı biliyorsa, kasıtlı olarak yanlış cevap verme olasılığı 1 olmalıdır. Sınava giren kişi doğru cevabı bilmiyorsa, o zaman sınava giren kişi önce doğru olduğunu düşünerek yanlış bir seçeneği ortadan kaldıracaktır. Sınava giren kişi daha sonra kalan üç seçenektan birini (dört seçenekli bir madde varsayarak) seçer ve bu da üçte iki oranında yanlış bir yanıt alma şansına neden olur. Sınava giren kişinin amaçlı yanlış cevaplama olasılığı ( $q$  olarak gösterilir)  $q = [p + (1 - p)*2/3] = (p + 2)/3$ . Burada  $p$ , sınava giren kişinin, madde parametre değerleri ve simüle edilmiş yetenek değerinden 2PL modeli kullanılarak hesaplanan doğru cevaplama olasılığıdır. En düşük yetenek düzeyine sahip bireyin dört seçenekli çoktan seçmeli test maddesini amaçlı yanlış cevaplama olasılığı 0,75 olarak ele alınmıştır. En

yüksek yetenek düzeyine sahip bireyin de dört seçenekli çoktan seçmeli test maddesini amaçlı yanlış cevaplama olasılığı 1 olarak ele alınmıştır.

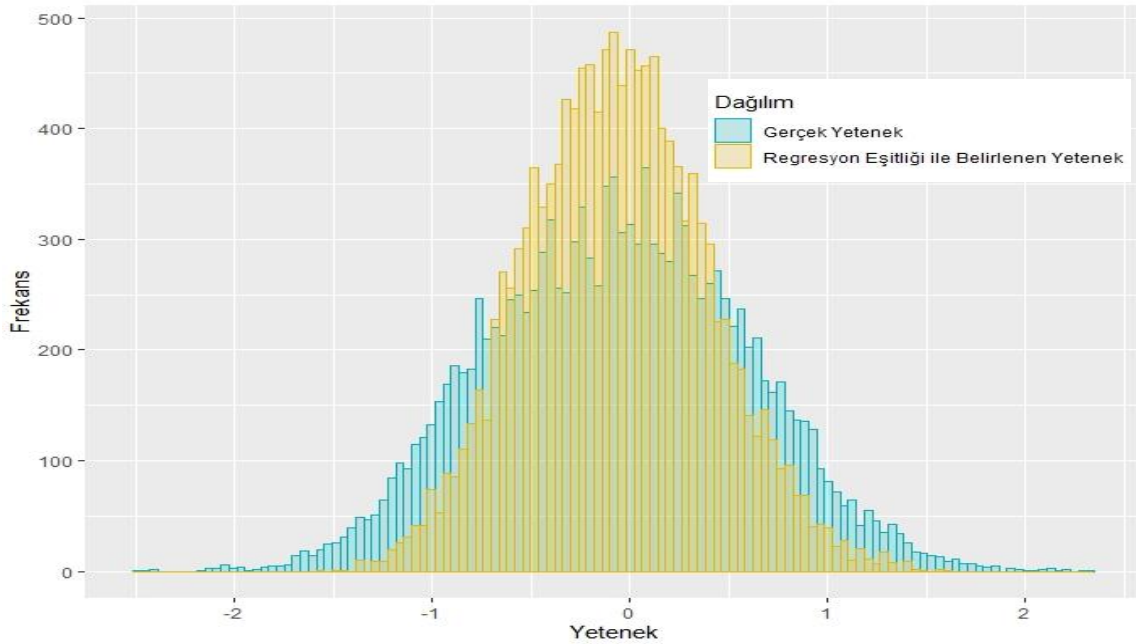
Her replikasyon için birer tane olmak üzere toplam 30 tane rastgele ve amaçlı yanlış cevaplama cevap matrisleri oluşturulmuştur. Araştırmada kullanılan cevap setlerine yönelik kısaltmalar kullanılmıştır. Bu kısaltmalar için izin verilmeyen BBT uygulamalarındaki gerçek cevaplar için  $p\_cevaplama$ ; cevap değişikliğine izin verildiği uygulamalarda ise rastgele cevaplama için  $r\_cevaplama$  ve amaçlı cevaplamanın yapıldığı durumlarda  $q\_cevaplama$  ifadeleri kullanılmıştır.

### ***BBT Uygulamalarında Tanımlanan Öncül Bilgilerin Elde Edilmesi***

Çalışmada öncül bilgiler ikinci ve üçüncü alt problemlerde kullanılmıştır. İkinci alt problem kapsamında bireylerin öncelikle PISA matematik testini sonrasında ise fen testini BBT uygulaması olarak aldıkları varsayılmıştır. Simülasyonlar CATSim programı ile gerçekleştirilmiştir. Eğer alt testler arasında daha önceden bir regresyon modeli kurulmuşsa bireylerin matematik yetenek düzeylerinden yola çıkarak fen BBT uygulamasında başlangıç değerleri tanımlanabilmektedir. Bu doğrultuda ilk aşamada PISA test bataryasının matematik alt test testinden elde edilen yetenek düzeyinin bağımsız değişken, fen alt testinden elde edilen yetenek düzeyinin ise bağımlı değişken olduğu bir regresyon modeli kurulmuştur. İkinci aşamada ise matematik BBT uygulaması simüle edilmiştir. Simülasyonda ilk madde [-1,1] aralığından rastgele seçilmiş, geçici ve nihai yetenek kestirimi EAP, madde seçim yöntemi olarak MFI ve sonlandırma kuralı değişken uzunluklu sonlandırma kuralı kullanılmıştır. Kestirimlerin standart hatası 0,35'ten küçük olduğunda ve en az 10 madde ya da 50 madde uygulandığında test sonlandırılmıştır. Elde edilen matematik yetenek düzeyi ve varyansı (standart hatanın karesi,  $SEM^2$ ) ile regresyon modeli kullanılarak fen BBT uygulamasındaki başlangıç yetenek düzeyleri ve bu düzeylere ait varyans değerleri hesaplanmıştır. Bireylere ait gerçek ve regresyon modeliyle yordanan fen yetenek düzeylerinin dağılım grafiği Şekil 2'de verilmiştir.

## Şekil 2

### Bireylerin Gerçek ve Regresyon Eşitliği ile Belirlenen Fen Yetenek Düzeylerinin Dağılımı



Bireylerin öncelikle Matematik-BBT uygulaması ile matematik yetenek düzeyi ve bu düzeyin standart hatası (SEM) belirlenir. Standart hatanın karesi kestirimin varyansına eşittir. Regresyon eşitliği ile elde edilen Fen-BBT başlangıç değerleri ve bu değerlere ilişkin varyanslar bireylerin öncül bilgilerinin dağılımını ifade etmektedir. Öncül bilgilerin dağılımını daha geniş bir yetenek düzeyine yaymak için Bayes kestirimlerinde öncül bilginin varyansı olarak belirlenen değerlerin iki katı olarak ele alınmıştır. Böylelikle bireylerin öncül bilgilerinin varyansları  $[0,49;0,73]$  aralığında olmaktadır. Fen BBT uygulamasında tanımlanan öncül bilgilerin belirlenme sürecine ilişkin denklemler aşağıda verilmiştir.

Matematik-BBT uygulaması ile kestirilen  $j$ . bireyin matematik yetenek düzeyi ve varyansı:

$$\theta_{mat_j} \sim N(\hat{\theta}_{mat_j}, \sigma_{mat_j}^2)$$

Fen-BBT uygulaması için tanımlanan  $j$ . bireyin fen öncül yetenek düzeyi ve varyansı:

$$\theta_{fen_j} \sim N(\hat{\theta}_{fen_j}, \sigma_{fen_j}^2)$$

Alt testler arasındaki regresyon eşitliği:

$$\hat{\theta}_{fen,j} = \beta_0 + \beta_1 * \hat{\theta}_{mat,j} + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

$$\hat{\theta}_{fen,j} = 0,085 + 0,595 * \hat{\theta}_{mat,j} + \varepsilon, \varepsilon \sim N(0, \sigma^2 = 0,227)$$

Fen-BBT uygulaması için  $j$ . bireyin fen yetenek öncül bilgisinin dağılımının ortalaması:

$$\hat{\theta}_{fen,j} = 0,085 + 0,595 * \hat{\theta}_{mat,j}$$

Fen-BBT uygulaması için  $j$ . bireyin fen yetenek öncül bilgisinin dağılımının varyansı:

$$\sigma_{fen,j}^2 = \beta_1 * \sigma_{mat,j}^2 + \sigma_{reg}^2$$

Öncül bilginin dağılımı daha geniş yetenek düzeyine yaymak için eşitliğin iki katı alınmıştır.

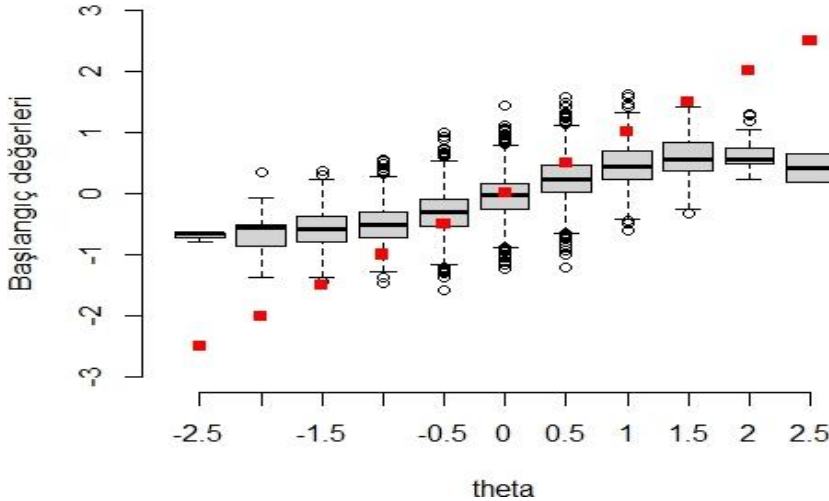
$$\sigma_{fen,j}^2 = 2(\beta_1 * \sigma_{mat,j}^2 + \sigma_{reg}^2)$$

$$\sigma_{fen,j}^2 = 2(0,595 * \sigma_{mat,j}^2 + 0,227)$$

Regresyon modeliyle başlatılan fen BBT uygulamasına ilişkin başlangıç değerleri ile gerçek yetenek düzeyinin karşılaştırılması Şekil 3'te verilmiştir.

### Şekil 3

*Regresyon Eşitliği ile Belirlenen Fen BBT Başlangıç Değerlerinin Gerçek Yetenek Düzeyleriyle Karşılaştırılması*

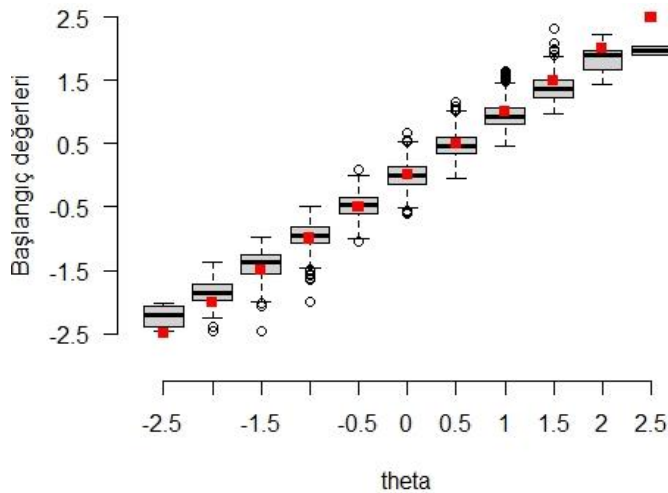


Üçüncü alt problem kapsamında ise bireylerin fen testini iki kez aldıkları varsayılmış ve ilk testten elde edilen yetenek düzeyleri ve standart hata değerleri (öncül bilgilerin varyansı =  $SEM^2$ ) ikinci uygulamada başlangıç değerleri olarak kullanılmıştır. İlk

uygulamada kestirilen yetenek düzeylerinin standart hatası (SEM) ikinci uygulamada bireylerin öncül bilgilerinin dağılımın standart sapması olarak alınmıştır. Birinci Fen BBT simülasyonunda da ilk madde  $[-1,1]$  aralığından rastgele seçilmiş, geçici ve nihai yetenek kestirimi EAP, madde seçim yöntemi olarak MFI ve sonlandırma kuralı değişken uzunluklu sonlandırma kuralı kullanılmıştır. Kestirimlerin standart hatası 0,35'ten küçük olduğunda ve en az 10 madde ya da 50 madde uygulandığında test sonlandırılmıştır. Regresyon modeliyle başlatılan fen BBT uygulamasına ilişkin başlangıç değerleri ile gerçek yetenek düzeyinin karşılaştırılması Şekil 4'te verilmiştir.

#### Şekil 4

*Aynı Testin Bir Önceki Sonucu Kullanılarak Belirlenen Fen BBT Başlangıç Değerlerinin Gerçek Yetenek Düzeyleriyle Karşılaştırılması*



#### **BBT Simülasyon Koşulları**

Bir hibrit simülasyon çalışması olan bu araştırma, PISA'yı bireye uyarlanmış bir test bataryası olarak ele almaktadır. PISA test bataryasında öncelikle bireylerin Matematik BBT uygulamalarını aldıkları ve elde edilen sonuçlara dayanarak başlangıç değeri tanımlandığı ve maddeyi yeniden gözden geçirmeye izin verilen fen BBT uygulamasının performansının nasıl değiştiği araştırılmıştır.



Çalışmada 3 farklı başlama kuralı, maddeyi yeniden cevaplamaya yönelik 3 farklı cevap seti ve 4 farklı son yetenek kestirimi olmak üzere 36 farklı koşul oluşturulmuştur. Koşulların nitelikleri her bir BBT aşamasında açıklanmıştır.

BBT uygulamasının ilk aşaması olan başlama kuralı olarak üç farklı yöntem izlenmiştir. Birinci yöntem, bireyler hakkında herhangi bir öncül bilginin olmadığı durumlarda ilk madde -1 ile 1 güçlük düzeyi arasından rastgele seçilmiştir. İkinci başlama kuralı ise alt testler arasındaki korelasyondan yola çıkarak öncül bilgilerin oluşturulmasıdır. “BBT Uygulamalarında Tanımlanan Öncül Bilgilerin Elde Edilmesi” başlığı altında belirtilen regresyon eşitliği ile öncül bilgiler tanımlanmıştır. Üçüncü başlama kuralı ise aynı testin bir önceki sonucunun kullanılmasıdır. Bu durumda fen BBT uygulaması yapılmış ve elde edilen sonuçlar ikinci fen BBT uygulamasında öncül bilgi olarak kullanılmıştır.

BBT uygulamasının ikinci aşaması olan yetenek kestirim yönteminde ise geçici yetenek kestirimi olarak bütün koşullarda EAP yöntemi tercih edilmiştir. EAP yöntemi yararlı öncül bilgilerin olduğu durumlarda ölçme hassasiyetini iyileştirmektedir (Wang & Vispoel, 1998). Öncül bilgilerin olmadığı ve EAP yönteminin kullanıldığı durumlarda ise alan yazında da kabul gören popülasyon ortalaması 0 ve standart sapması 1 olan standart normal dağılım kullanılmıştır (Yao, 2013). Son yetenek kestirimi olarak çalışmada deneysel öncül dağılımın tanımlandığı EAP, öncül bilginin olmadığı ve ön bilginin standart normal dağılıma  $N(0,1)$  sahip olduğu EAP(0,1), ML ile MLEF olmak üzere dört farklı yöntem kullanılmıştır.

Madde seçim yöntemi olarak Lord (1970) tarafından tanımlanan MFI kullanılmıştır. Bu yöntem madde havuzundan geçici yetenek düzeyinde en çok bilgiyi sağlayan maddeyi seçer. Simülasyonlarda ilk madde olarak her bir koşulda en çok bilgiyi veren beş maddeden biri rastgele seçilmiştir.

Sonlandırma kuralı olarak ise sabit uzunluk seçilmiştir. Öncül bilgilerin kısa testlere de etkisini gözlemleyebilmek amacıyla 5, 10, 15, 20 ve 25 olmak üzere beş farklı test

uzunluęu belirlenmiřtir. Madde havuzunun sınırlı olması nedeniyle en uzun testte 25 madde yer almaktadır.

Psikometrik özelliklere ek olarak, operasyonel test yönetimlerini yansıtmak amacıyla BBT simülasyonları için madde kullanım sıklığı kontrolü gibi pratik kısıtlamalar da uygulanmıştır. Madde kullanım sıklığı 0,25 olarak belirlenmiştir.

### **Verilerin Analizi**

Bir simülasyon çalışması olan bu arařtırmada CATSim programı (Weiss & Guyer, 2010) ile R kullanılmıştır. CATSim programı öncül bilgilerin tanımlamasına olanak sağladığı için tercih edilmiştir. CATSim replikasyon olanağı sunmadığı için 45 koşul 30 kez manuel olarak tekrarlanmıştır. Yanlılık ve RMSE deęerleri her replikasyon için hesaplanıp sonrasında ortalamaları alınmıştır.

CATSim programında maddeyi yeniden gözden geçirmeye yönelik bir simülasyon sekmesi yer almamaktadır. Maddeyi yeniden geçirmeye izin verilen uygulamalarda bireylerin öncelikle maddeleri ya rastgele cevapladıkları ya da amaçlı olarak yanlış cevapladıkları varsayılmıştır. Bu süreç CATSim programında simüle edilirken rastgele ve amaçlı yanlış cevap matrisleri kullanılmıştır. Bireylerin testi tamamladıklarında cevap deęişikliğine izin verildiği varsayılmış ve bu süreç R programında simüle edilmiştir. Bireylerin aldıkları maddeler belirlenmiş ve bu maddelere verdikleri cevapları gerçek cevap matrisinden elde edilerek belirlenen nihai yetenek kestirim yöntemleri ile arařtırmacı tarafından oluşturulan R kodları ile simüle edilmiştir.

BBT uygulamalarında başlama kuralı, maddeyi yeniden gözden geçirme durumları ve yetenek yöntemleri kullanılarak oluşturulan her bir koşul sonucu kestirilen yetenek düzeyi ile hibrit simülasyon sonucu kestirilen yetenek düzeyi arasındaki yanlılık ve RMSE deęerleri hesaplanmıştır.

RMSE ve yanlılık deęerlerini hesaplamaya yönelik eşitlikler aşağıda verilmiştir.

$$Yanlilik[\hat{\theta}] = \frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)}{N} \quad RMSE[\hat{\theta}] = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{N}}$$

Her bir koşula ait yanlilik ve RMSE deęerleri 30 replikasyonun ortalaması alınarak hesaplanmıřtır. Hesaplanan deęerlerden yola çıkılarak, hangi koşulun dięerlerine göre daha yüksek ölçüm hassasiyetine sahip olduğuna yönelik betimsel deęerlendirme yapılmıřtır. Ayrıca ele alınan deęişkenlerin etkisini ve etkileşimlerini görmek için her bir koşula ait yanlilik ve RMSE bağımlı deęişkenleri ile ANOVA gerçekleştirilmiştir. Bağımsız deęişkenler ise cevapları gözden geçirme davranış türü, nihai yetenek kestirim yöntem ve önsel bilginin dağılımıdır. Varyans analizi, çeşitli koşullar ve bunların etkileşimleri arasında manidar düzeyde farklılık olup olmadığını test etmek amacıyla uygulanmıştır. Varyans analizi sonucu elde edilen etki büyüklükleri deęerlendirmek için Cohen (1988) tarafından önerilen ařağıdaki kriterler kullanılmıştır.

- Etki yok ( $\eta^2 < 0,0099 \approx 0,01$ )
- Küçük etki ( $0,01 < \eta^2 < 0,0588 \approx 0,06$ ),
- Orta etki ( $0,06 < \eta^2 < 0,1379 \approx 0,14$ )
- Büyük etki ( $\eta^2 > 0,14$ )

## Bölüm 4

### Bulgular, Yorumlar ve Tartışma

Bu bölümde her bir araştırma probleminde koşullara yönelik RMSE ve yanlılık değerleri ile üç faktörlü ANOVA'dan elde edilen sonuçlara yer verilmiştir. ANOVA analizleri 30 tekrardan elde edilen RMSE ve yanlılık değerleri ile gerçekleştirilmiştir.

#### Birinci Alt Probleme İlişkin Bulgular ve Yorumlar

BBT uygulamalarında başlangıç yetenek düzeyi  $[-1,1]$  yetenek aralığından rastgele seçilen bir değer olarak belirlendiğinde ( $\theta_{\text{random}}$ ) elde edilen yetenek kestirimlerinin RMSE değeri ve yanlılığı maddeyi yeniden gözden geçirmeye izin verildiğinde bireylerin kullandığı madde cevaplama stiline, yetenek kestirim yöntemlerine ve test uzunluğuna göre nasıl değişmektedir?

İlgili alt problem kapsamında başlangıç yetenek düzeyi  $[-1,1]$  yetenek aralığından rastgele seçilen bir değer olarak belirlenmiş ve maddeyi yeniden cevaplama izin verilme durumunun uygulama üzerindeki etkisinin test uzunluğuna ve nihai yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. Üç cevaplama stili ile gerçekleştirilen simülasyonlar sonucu elde edilen her koşula ilişkin ortalama RMSE ve yanlılık değerleri Tablo 5'te verilmiştir. Öncül bilginin kullanılmadığı bu alt problemde EAP ile EAP(0,1) yetenek kestirim yöntemlerine ilişkin RMSE ve yanlılık değerleri aynı olacağı için EAP yöntemine ait sonuçlar verilmemiştir.

Simülasyonlarda herhangi bir öncül bilginin olmadığı durumlarda başlangıç yetenek düzeyi  $\theta_{\text{random}}$  yöntemi ile belirlenen tüm koşullar için RMSE değerleri 0,184 ile 1,660 arasında değişmektedir. Bütün test uzunluklarında Bayes yöntemi en az RMSE değerlerine, MLE ise en yüksek RMSE değerlerine sahiptir. Ayrıca bütün simülasyon koşullarında elde edilen RMSE değerleri test uzunluğu arttıkça azalmaktadır. Cevap değişikliğine izin verildiğinde ise RMSE değerlerinin arttığı görülmektedir. Maddelerin

amaçlı olarak yanlış cevaplandığı ve sonrasında cevap değişikliğinin yapıldığı (Q-BBT) test uygulamaları rastgele cevaplama (R-BBT) göre daha fazla RMSE değerine sahiptir.

**Tablo 5**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{random}$  ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık Değerleri*

Yetenek Kestirimi		RMSE					Yanlılık				
		5	10	15	20	25	5	10	15	20	25
P-BBT	EAP(0,1)	0,343	0,259	0,220	0,198	0,184	0,007	0,003	0,002	0,002	<0,001
	MLE	0,527	0,318	0,253	0,222	0,204	-0,013	-0,004	-0,002	-0,002	-0,002
	MLEF	0,465	0,308	0,250	0,221	0,203	-0,007	-0,003	-0,002	-0,002	-0,002
R-BBT	EAP(0,1)	0,433	0,350	0,296	0,257	0,229	0,047	0,039	0,029	0,020	0,018
	MLE	1,433	0,912	0,614	0,436	0,355	0,311	0,172	0,088	0,045	0,034
	MLEF	0,885	0,643	0,478	0,369	0,311	0,161	0,106	0,060	0,033	0,026
Q-BBT	EAP(0,1)	0,444	0,379	0,332	0,295	0,265	0,059	0,053	0,045	0,035	0,027
	MLE	1,660	1,180	0,834	0,600	0,446	0,490	0,300	0,156	0,081	0,045
	MLEF	0,951	0,796	0,623	0,487	0,389	0,246	0,192	0,116	0,069	0,043

P-BBT: Maddeyi yeniden cevaplama izin verilmeyen uygulama

R-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce rastgele cevaplama, sonra bilinçli cevaplama

Q-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce amaçlı yanlış cevaplama, sonra bilinçli cevaplama

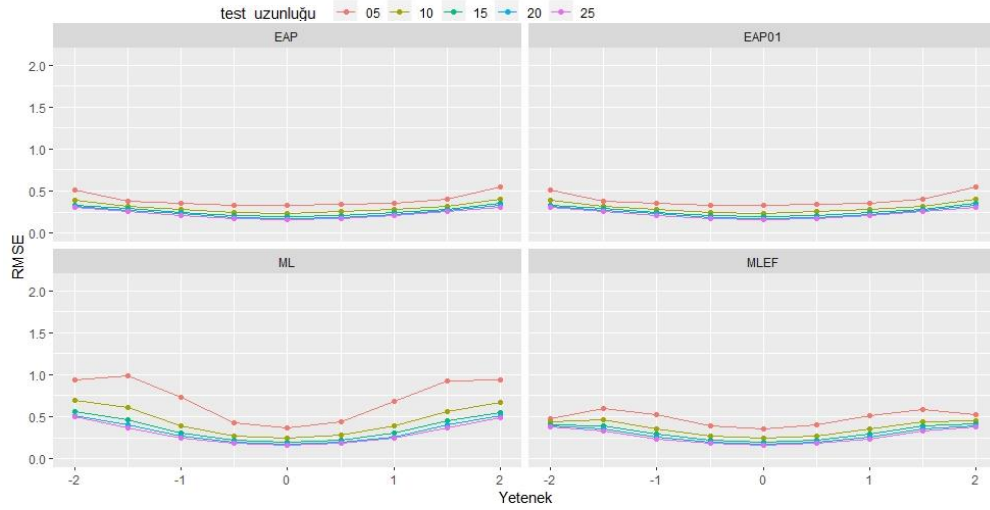
Tablo 5'te başlangıç yetenek düzeyinin  $\theta_{random}$  ile belirlendiği koşullara ilişkin yanlılık değerlerinin -0,013 ile 0,490 arasında değiştiği görülmektedir. Yanlılık değerleri test uzunluğu arttıkça genellikle azalmaktadır. Birçok koşulda kestirilen yetenek düzeyleri gerçek değerden fazla iken maddeyi yeniden cevaplama izin verilmeyen BBT uygulamalarında MLE ve MLEF yetenek kestirim yöntemleriyle kestirilen yetenek düzeyleri gerçek değerlerden daha düşüktür. Cevap değişikliğine izin verilen R-BBT ve Q-BBT uygulamalarındaki yanlılık değeri, izin verilmeyen uygulamaya nazaran daha fazla ve hepsi pozitifdir. Ayrıca amaçlı olarak yanlış cevaplanan BBT uygulaması, rastgele cevaplama simüle edilen BBT uygulamasına göre daha fazla yanlılık göstermektedir.

Şekil 5'te başlangıç yetenek düzeyi olarak [-1;1] aralığından rastgele bir değer seçilen BBT uygulamalarında kullanılan nihai yetenek kestirimlerine yönelik koşullu RMSE değerleri yer almaktadır.

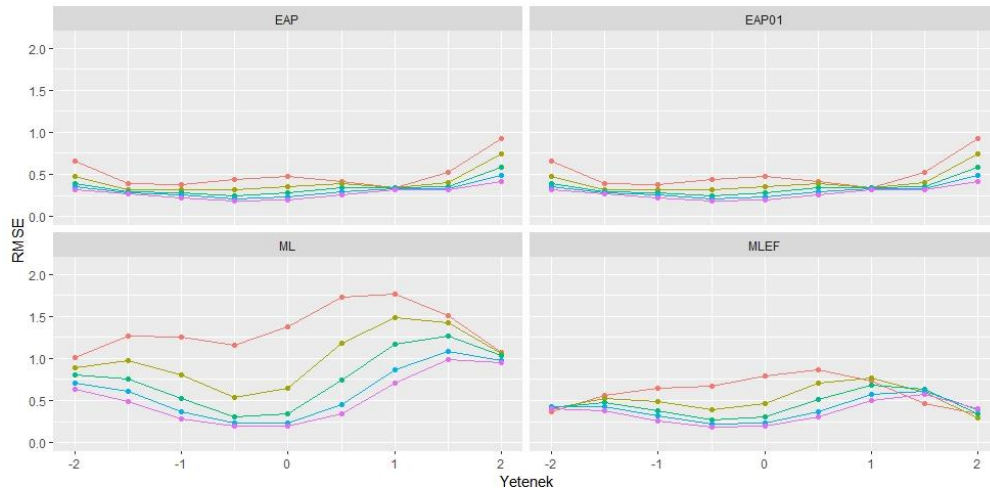
## Şekil 5

Başlangıç Yetenek Düzeyi  $\theta_{random}$  Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği

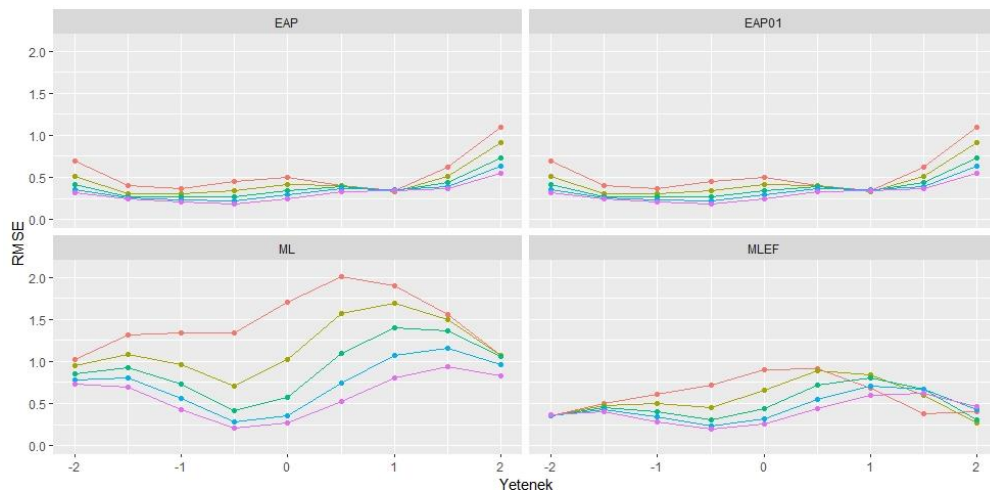
Şekil 5a



Şekil 5b



Şekil 5c



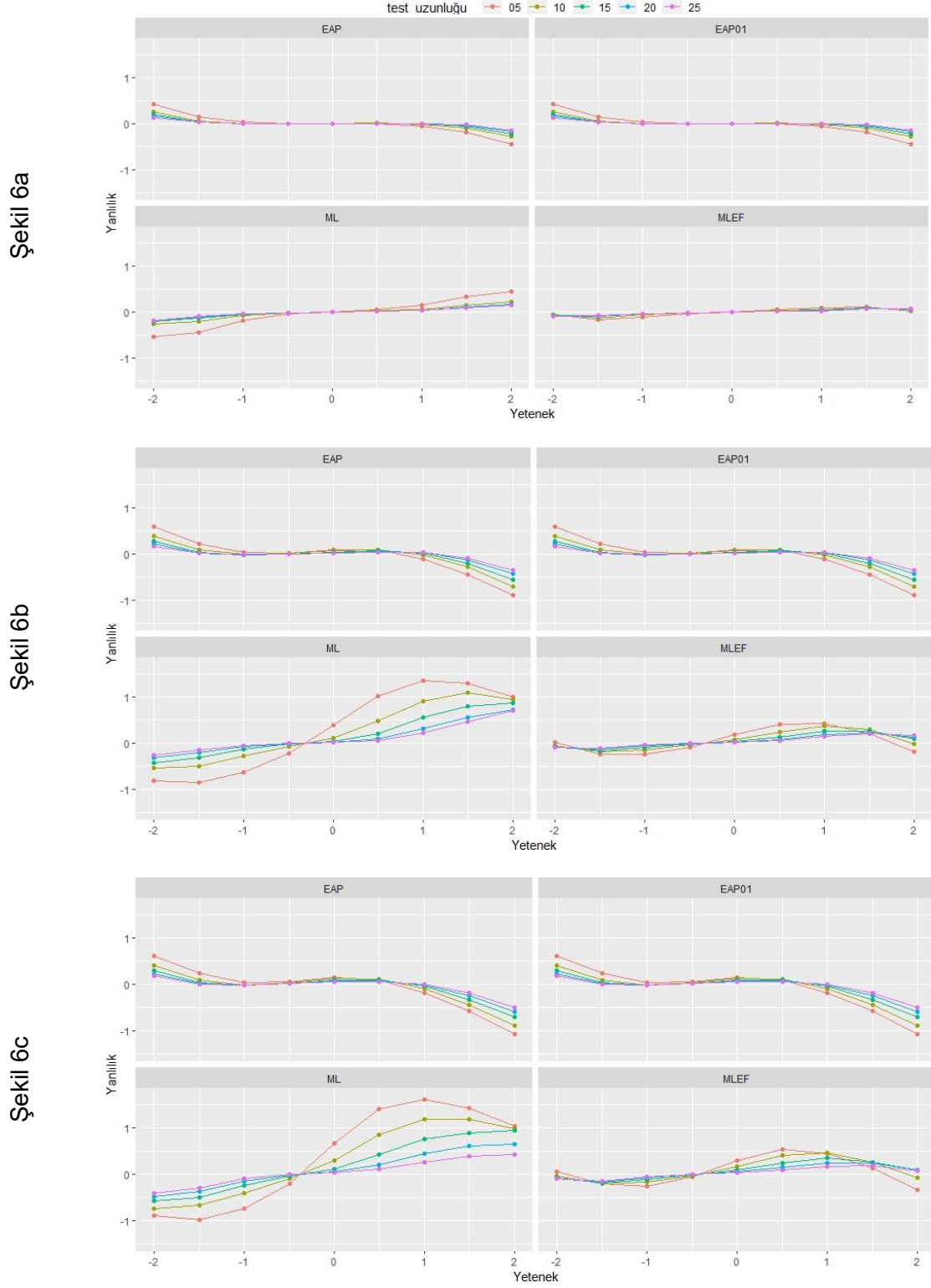
$\theta_{\text{random}}$  yöntemi ile başlangıç yetenek düzeyi belirleyerek oluşturulan bütün test desenlerinde (P-BBT, R-BBT, Q-BBT) elde edilen RMSE değerleri bütün yetenek düzeylerinde test uzunluğu arttıkça azalmaktadır. Cevap değişikliğine izin verilmeyen P-BBT uygulamasında test uzunluğu 5 olduğunda bütün yetenek düzeylerinde Bayes yöntemlerine ilişkin RMSE değerleri MLE ve MLEF'e göre daha azdır. Ayrıca bu test deseninde yer alan daha uzun testlerde ise sadece uç yetenek düzeylerindeki RMSE değerleri farklılaşırken orta yetenek düzeylerinde  $[-1,1]$  RMSE değerleri benzeşmektedir. Şekil 5b ve 5c'de ise  $\theta_{\text{random}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen ve madde cevap değişikliğine izin verilen diğer test desenlerine ilişkin RMSE değerleri yer almaktadır.

Madde cevap değişikliğine izin verilen uzun test uygulamalarında Bayes yöntemlerine ilişkin RMSE değeri izin verilmeyen duruma göre uç noktalarda artarken, orta yetenek düzeyinde yöntemler benzer değerlere sahiptir. Şekil 5'te MLE ve MLEF yöntemlerinin kullanıldığı durumlarda ise RMSE değerinin cevap değişikliğine izin verilmeyen uygulamaya nazaran bütün koşullarda arttığı görülmektedir. Bütün test desenlerinin RMSE değerleri, yetenek düzeylerinin hepsinde test uzunluğu arttıkça azalmaktadır. Ayrıca Bayes kestirim yöntemine ilişkin RMSE değerleri bütün yetenek düzeylerin MLE ve MLEF'e göre daha az olduğu görülmektedir. MLE ve MLEF yöntemlerine ilişkin RMSE değerleri arasındaki farklılık test uzunluğu arttıkça azalmaktadır.

RMSE değerlerinin yanı sıra başlangıç yetenek düzeyi olarak  $[-1,1]$  aralığından rastgele bir değer seçilerek simüle edilen her bir koşula ilişkin yanlılık değerlerinin yetenek düzeylerine göre değişim grafikleri Şekil 6'da yer almaktadır. Şekil 6'da bütün koşullardaki uç yetenek düzeylerinin RMSE değerlerinin orta yetenek düzeylerine nazaran daha fazla olduğu görülmektedir. Bu farklılığın cevap değişikliğine izin verildiğinde daha belirgin olduğu Şekil 6b ve 6c'de görülmektedir.

## Şekil 6

Başlangıç Yetenek Düzeyi  $\theta_{random}$  Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanıllık Grafiği





Şekil 6'da bütün test desenlerinde (P-BBT, R-BBT, Q-BBT) yetenek kestirim yöntemi olarak Bayes ve MLE kullanıldığında yanlılık değerlerinin her bir yetenek düzeylerinde test uzunluğu arttıkça azaldığı görülmektedir. Bayes yönteminin kullanıldığı test desenlerinde düşük yetenek düzeylerine ait yanlılık değeri negatifken yüksek yetenek düzeylerinde ise pozitifdir. Ancak MLE ve MLEF yöntemlerine ilişkin yanlılık değerleri düşük yetenek düzeyinde negatif, yüksek yetenek düzeylerinde genellikle pozitif değere sahiptir. Bayes yönteminin kullanıldığı madde cevap değişikliğine izin verilen ve verilmeyen koşullarda yanlılık değeri orta yetenek düzeylerinde benzerlik gösterirken uç yetenek düzeylerinde madde cevap değişikliğine izin verildiğinde yanlılık değeri artmaktadır. Şekil 6'da MLE yöntemleriyle gerçekleştirilen ve cevap değişikliğine izin verilen durumlarda orta yetenek düzeyinde en yüksek yanlılık değerine sahip olduğu görülmektedir.

$\theta_{\text{random}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen simülasyon koşullarına yönelik elde edilen RMSE ve yanlılık değerleriyle çizilen grafiklerin yanı sıra bu değerlerin cevaplama stiline, yetenek kestirim yöntemine ve test uzunluklarına göre istatistiksel olarak manidar düzeyde farklılaşıp farklılaşmadığını test etmek için üç faktörlü ANOVA testi gerçekleştirilmiştir. ANOVA test sonuçları Tablo 6'da verilmiştir. Başlama kuralı  $\theta_{\text{random}}$  yöntemi ile belirlendiği için uygulamalarda öncül bilgi yoktur. Bu nedenle öncül bilginin kullanıldığı EAP yöntemine ait sonuçlara bu alt problemde yer verilmemiştir.

Tablo 6'da  $\theta_{\text{random}}$  yöntemi ile başlangıç yetenek düzeyi belirlendiğinde elde edilen ortalama yanlılık ve RMSE değerleri üzerindeki temel etkilerin, ikili ve üçlü etkileşimlerin etkililiğinin istatistiksel olarak manidar düzeyde olduğu görülmektedir. Üç yönlü bir etkileşim, bir veya daha fazla yönlü etkileşimin üçüncü bir değişkenin seviyeleri arasında farklılık gösterdiği anlamına gelmektedir (UCLA, n.d.). Bu nedenle üçlü etkileşim olduğunda temel etkileri ve ikili etkileşimleri yorumlamak yerine araştırmanın amacına uygun olarak CS\*YKY etkileşimin yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. İki yönlü etkileşim analizi yaparken, daha önce tüm veri seti kullanılarak elde

edilen üç yönlü ANOVA sonucundaki genel hata terimi (artıkları) kullanılmaktadır (Kirk, 1995; Keppel & Wickens, 2004). O nedenle iki yönlü etkileşimin F değeri hesaplanırken asıl model olan üç yönlü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

**Tablo 6**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{random}$  ile Belirlendiğinde Elde Edilen Yanlılık ve RMSE Değerlerine Ait ANOVA Sonuçları*

Varyansın Kaynağı	Sd	Yanlılık		RMSE	
		F	$\eta^2$	F	$\eta^2$
Cevaplama stili (CS)	2	116330*	0,29	762849	0,22
Yetenek kestirim yöntemi (YKY)	2	50083*	0,12	728457	0,21
Test Uzunluğu(TU)	4	40711*	0,20	532526	0,31
CS*YKY	4	17548*	0,09	136221	0,08
CS*TU	8	12048*	0,12	37283	0,04
YKY*TU	8	10455*	0,10	85108	0,10
CS*YKY*TU	16	3540*	0,07	13454	0,03
CS*YKY@ TU:5	4	22746*	0,11	108633*	0,03
CS@ TU:5, @EAP(0,1)	2	1280*	<0,01	4365*	<0,01
CS@ TU:5, @MLE	2	113380*	0,28	516451*	0,15
CS@ TU:5, @MLEF	2	28796*	0,07	100144*	0,03
CS*YKY@ TU:10	4	7097*	0,04	51607*	0,03
CS@ TU:10, @EAP(0,1)	2	1163*	<0,01	5659*	<0,01
CS@ TU:10, @MLE	2	40547*	0,10	280096*	0,08
CS@ TU:10, @MLEF	2	16754*	0,04	89688*	0,03
CS*YKY@ TU:15	4	1513*	0,01	20144*	0,01
CS@ TU:15, @EAP(0,1)	2	814*	<0,01	4700*	<0,01
CS@ TU:15, @MLE	2	11053*	0,03	123645*	0,04
CS@ TU:15, @MLEF	2	6108*	0,02	50695*	0,01
CS*YKY@ TU: 20	4	291*	<0,01	7194*	<0,01
CS@ TU:20, @EAP(0,1)	2	465*	<0,01	3405*	<0,01
CS@ TU:20, @MLE	2	2947*	0,01	51511*	0,02
CS@ TU:20, @MLEF	2	2211*	0,01	25612*	0,01
CS*YKY@ TU:25	4	58*	<0,01	2494*	<0,01
CS@ TU:25, @EAP(0,1)	2	349*	<0,01	2350*	<0,01
CS@ TU:25, @MLE	2	1047*	<0,01	213285*	0,01
CS@ TU:25, @MLEF	2	931*	<0,01	12566*	<0,01

\*p<0.05

Tablo 6'da her bir test uzunluğunda CS\*YKY ikili etkileşimin elde edilen ortalama yanlılık ve RMSE değeri üzerindeki istatistiksel olarak anlamlı olduğu görülmektedir. Başka bir deyişle, bütün test uzunluklarında yetenek kestirim yönteminin ortalama yanlılık ve RMSE üzerindeki etkisi cevaplama stiline göre farklılaşmaktadır. İkili etkileşimlerin anlamlı olması nedeniyle her bir test uzunluğunda ve yetenek kestiriminde cevaplama stiline ortalama yanlılık ve RMSE üzerindeki etkililiği incelemek için basit etkiler (simple simple effect) hesaplanmıştır. Basit etkilerin hesaplanmasında asıl model olan üç yönlü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

Basit etkiler incelendiğinde cevaplama stiline ortalama yanlılık değerleri üzerindeki etkililiği, test uzunluğuna ve yetenek kestirim yöntemine göre farklılaşmaktadır. Etki büyüklükleri incelendiğinde test uzunluğu arttıkça bütün yetenek kestirim yöntemlerinde cevaplama stiline ortalama yanlılık üzerindeki etkililiğinin azaldığı görülmektedir. Test uzunluğu 5 olduğunda cevaplama stiline ortalama yanlılık üzerindeki etki düzeyi MLE yönteminde büyük, MLEF yönteminde orta ve EAP(0,1) yönteminde ise küçüktür. Test uzunluğu 10 olduğunda ise MLE yöntemi kullanıldığında cevaplama stiline ortalama yanlılık üzerindeki etki düzeyi orta diğer yöntemlerde ise küçük düzeydedir. Test uzunluğunun daha fazla olduğu durumlarda ise bütün yetenek kestirim yöntemlerinde cevaplama stiline ortalama yanlılık üzerindeki etki düzeyi küçüktür. Yapılan bütün Post-Hoc testlerde cevap değişikliğine izin verildiğinde yanlılık değerlerinin manidar düzeyde arttığı tespit edilmiştir.

Tablo 6'daki RMSE değerlerine yönelik basit etkiler incelendiğinde ise cevaplama stiline ortalama RMSE üzerindeki etkililiğinin test uzunluğuna ve yetenek kestirim yöntemine göre farklılaştığı görülmektedir. MLE yönteminde cevaplama stiline ortalama RMSE üzerindeki etki düzeyi test uzunluğu 5 olduğunda büyük, 10 olduğunda ise orta düzeydedir. Diğer uzunluklarda ve yetenek kestirim yöntemlerinde ise cevaplama stiline ortalama RMSE üzerindeki etki düzeyleri ise küçüktür. Yapılan Post-Hoc ikili

karşılaştırmaları ile cevap değişikliğine izin verildiğinde bütün yetenek kestirim yöntemlerinde ve test uzunluklarında amaçlı cevaplama ile elde edilen yetenek düzeylerine ilişkin ortalama RMSE değerlerinin istatistiksel olarak daha fazla olduğu belirlenmiştir.

Sonuç olarak başlangıç yetenek düzeyi  $\theta_{\text{random}}$  yöntemi belirlendiğinde ve cevap değişikliğine izin verilen bütün koşullarda yanlılık ve RMSE değerleri artmaktadır. Bu artış MLE yönteminde en fazla olurken EAP(0,1) yönteminde en azdır. Test uzunluğu arttıkça bütün yetenek kestirimlerinde cevap değişikliğine izin verildiğinde elde edilen yanlılık ve RMSE değerlerindeki artış azalmaktadır.

### **İkinci Alt Probleme İlişkin Bulgular ve Yorumlar**

BBT uygulamalarında başlangıç yetenek düzeyi olarak regresyon eşitliği ile belirlenen değer kullanıldığında elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı maddeyi yeniden gözden geçirmeye izin verildiğinde bireylerin kullandığı cevaplama stiline, yetenek kestirim yöntemlerine ve test uzunluğuna göre nasıl değişmektedir?

İlgili alt problem kapsamında başlangıç yetenek düzeyi, test bataryasında yer alan alt testler arasındaki korelasyon katsayısından ve uygulanan bir önceki alt testten elde edilen değerden yola çıkarak belirlenmiştir. Bu başlama kuralı ile birlikte maddeyi yeniden cevaplama izin verilmesi durumunun uygulama üzerindeki etkisinin test uzunluğuna ve nihai yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. Üç cevaplama stili ile gerçekleştirilen simülasyonlar sonucu elde edilen her koşula ilişkin gerçekleştirilen replikasyonlara ait ortalama RMSE ve yanlılık değerleri Tablo 7’de verilmiştir.

Tablo 7’de başlangıç yetenek düzeyi regresyon eşitliği ile belirlenen bütün koşullarda RMSE değeri 0,175 ile 1,451 arasında değişmektedir. RMSE değerlerinin test uzunluğu arttıkça azaldığı görülmektedir. Nihai yetenek kestiriminde öncül bilginin kullanıldığı üç cevaplama stilinde de diğer yetenek kestirim yöntemlerine nazaran daha

az RMSE değerine sahip olduğu görülmektedir. En yüksek RMSE değerinin ise üç cevaplama stilinde de MLE yöntemine ilişkin koşullarda olduğu görülmektedir. Cevap değişikliğine izin verilmeyen koşullarda kısa testlerde MLEF, MLE'ye göre daha az RMSE değerine sahipken, uzun testlerde ise her iki yöntem benzer RMSE değerlerine sahiptir. Cevap değişikliğine izin verildiği durumlarda ise RMSE değerleri bütün koşullarda artmaktadır. R-BBT ve Q-BBT uygulamalarında bütün test uzunluklarında MLE yöntemine ilişkin ortalama RMSE değeri, MLEF yönteminden daha fazladır. Maddelerin amaçlı olarak yanlış cevaplandığı ve sonrasında cevap değişikliğinin yapıldığı (Q-BBT) test uygulamaları, rastgele cevaplama (R-BBT) göre daha fazla RMSE değerine sahiptir.

**Tablo 7**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{regresyon}$  ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık Değerleri*

	Yetenek Kestirimi	RMSE					Yanlılık				
		5	10	15	20	25	5	10	15	20	25
P-BBT	EAP(0,1)	0,344	0,256	0,218	0,197	0,189	0,008	0,003	0,002	0,001	<0,001
	EAP	0,279	0,227	0,200	0,183	0,175	0,001	0,001	0,001	0,001	<0,001
	MLE	0,637	0,320	0,253	0,222	0,232	-0,001	-0,004	-0,003	-0,002	-0,003
	MLEF	0,501	0,305	0,249	0,220	0,221	-0,002	-0,004	-0,002	-0,002	-0,002
R-BBT	EAP(0,1)	0,409	0,309	0,259	0,230	0,208	0,032	0,027	0,021	0,016	0,012
	EAP	0,293	0,244	0,216	0,198	0,184	0,003	0,004	0,003	0,002	0,002
	MLE	1,229	0,718	0,499	0,380	0,304	0,128	0,080	0,050	0,032	0,020
	MLEF	0,776	0,510	0,386	0,316	0,268	0,064	0,046	0,031	0,021	0,014
Q-BBT	EAP(0,1)	0,420	0,350	0,307	0,276	0,250	0,075	0,072	0,060	0,050	0,040
	EAP	0,297	0,260	0,238	0,221	0,206	0,004	0,003	0,003	0,003	0,002
	MLE	1,451	1,104	0,870	0,677	0,522	0,406	0,348	0,244	0,159	0,099
	MLEF	0,838	0,682	0,578	0,489	0,408	0,194	0,185	0,144	0,104	0,072

P-BBT: Maddeyi yeniden cevaplama izin verilmeyen uygulama

R-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce rastgele cevaplama, sonra bilinçli cevaplama

Q-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce amaçlı yanlış cevaplama, sonra bilinçli cevaplama

Başlangıç yetenek düzeyinin  $\theta_{regresyon}$  ile belirlendiği ve cevap değişikliğine izin verilmeyen koşullarında kullanılan yetenek kestirim yöntemlerinin yanlılık değerleri -0,004 ile 0,406 arasında değişmektedir. Cevap değişikliğine izin verildiğinde ise yanlılık değerleri

artmaktadır. Tablo 7’de cevap deęişikliğine izin verilmeyen durumlarda farklı test uzunlukları benzer yanlılık deęerlerine sahipken izin verilen durumlarda ise yanlılık deęerleri artmaktadır. Cevap deęişikliğine izin verilen uygulamalarda da test uzunluğu arttıkça yanlılık deęerleri azalmaktadır. Amaçlı yanlış cevaplanan test uygulamalarının yanlılık deęerlerinin rastgele cevaplama göre daha fazla olduęu görölmektedir. Bütün koşullarda öncül bilginin kullanıldığı EAP yöntemine ilişkin yanlılık deęerleri en az iken MLE yöntemine ilişkin yanlılık deęeri ise en fazladır. Cevap deęişikliğine izin verilmeyen uygulamalarda MLE ve MLEF yöntemlerine ilişkin ortalama yanlılık deęerleri birbirine benzerken cevap deęişikliğine izin verildiğinde ise MLE yönteminin bütün koşullarda MLEF yönteminden daha fazla yanlılığa sahip olduęu görölmektedir.

$\theta_{\text{regresyon}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen simülasyon koşullarına yönelik hesaplanan ortalama RMSE ve yanlılık deęerlerine ek olarak bu deęişkenlerin yetenek düzeylerine göre deęişimlerinin incelenmesi amacıyla koşullu RMSE ve yanlılık deęerlerini içeren grafikler çizilmiştir.

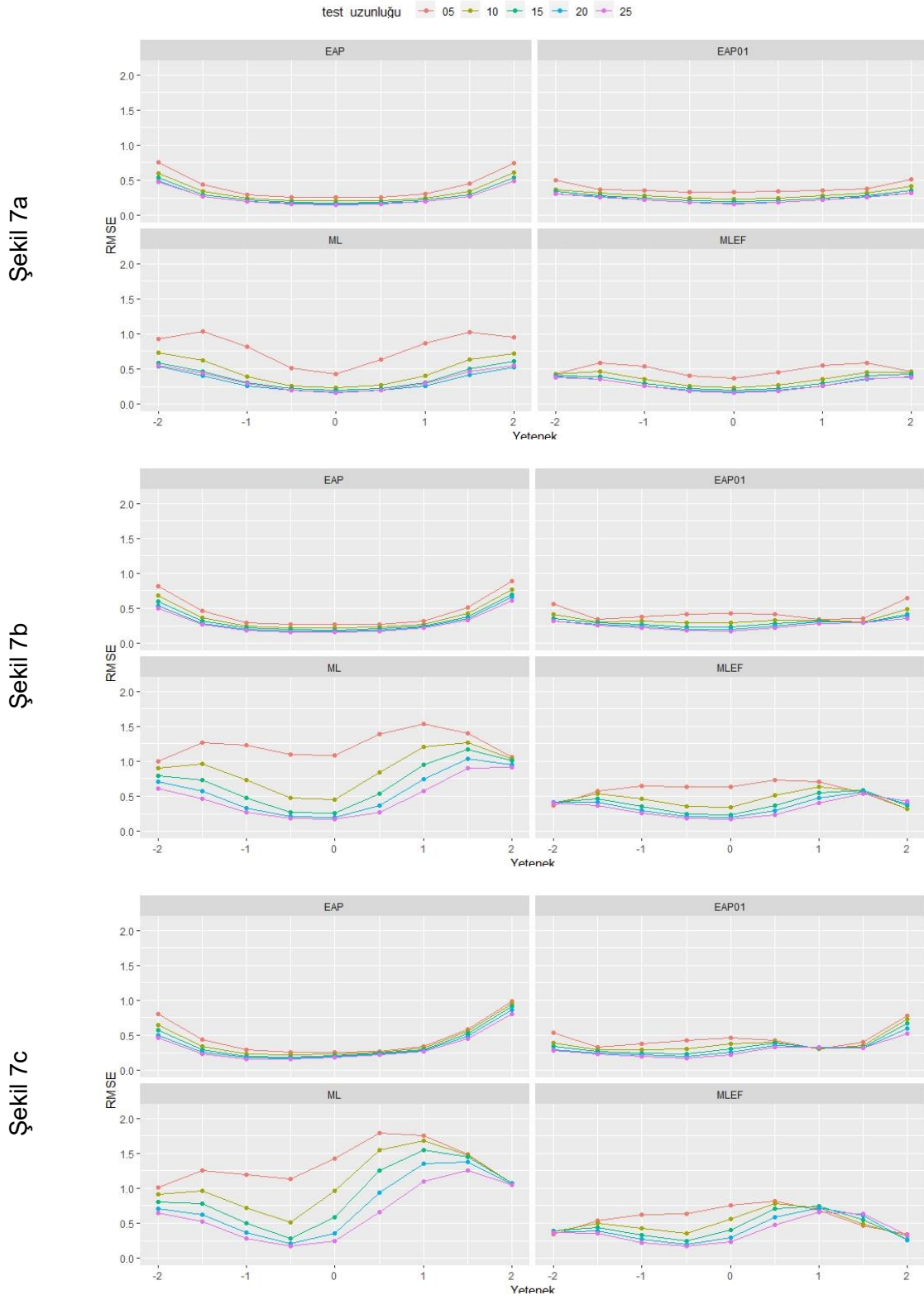
Şekil 7’de başlangıç yetenek düzeyinin regresyon eşitliği ile belirlendięi BBT uygulamalarında kullanılan nihai yetenek kestirimlerine yönelik koşullu RMSE deęerleri yer almaktadır. Başlama kuralı olarak  $\theta_{\text{regresyon}}$  yöntemi kullanılan BBT uygulamalarında elde edilen RMSE deęerleri bütün yetenek düzeylerinde genellikle test uzunluğu arttıkça azalmaktadır.

Cevap deęişikliğine izin verilmeyen P-BBT uygulamalarındaki bütün test uzunluklarında orta yetenek düzeyindeki bireylerin (yetenek düzeyi  $[-1,1]$  aralığında olan bireyler) nihai yetenek kestirim yöntemlerine ilişkin RMSE deęerleri benzerlik göstermektedir. Ancak test uzunluğu 5 ve 10 olduęunda ise uç yetenek düzeylerinde MLE ve MLEF yöntemlerinin Bayes kestirim yöntemlerine nazaran daha fazla RMSE deęerlerine sahip olduęu görölmektedir.

## Şekil 7

Başlangıç Yetenek Düzeyi  $\theta_{\text{regresyon}}$  Yöntemi ile Belirlenen BBT Uygulamalarında

Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği



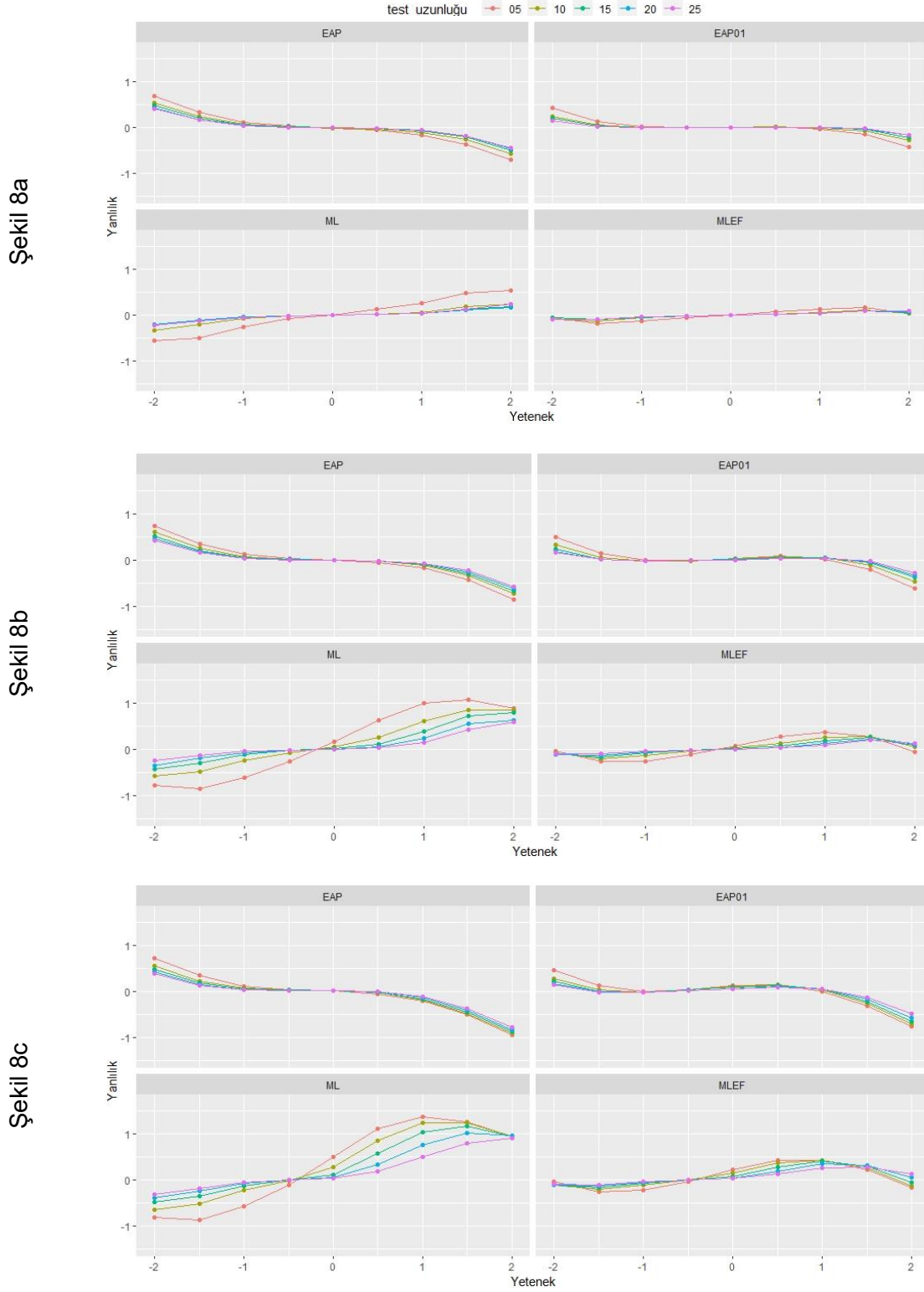
Şekil 7a'da cevap değişikliğine izin verilmediğinde EAP(0,1) yönteminin uç yetenek düzeylerinde öncül bilginin kullanıldığı EAP yöntemine nazaran daha az RMSE değerlerine sahiptir. Bu durum öncül bilginin madde bilgilerini baskılamasından kaynaklanmaktadır. Şekil 7b ve 7c'de ise  $\theta_{\text{regresyon}}$  yöntemi ile başlama kuralı tanımlanan ve madde cevap değişikliğine izin verilen test desenlerine ilişkin RMSE değerleri yer almaktadır. Madde cevap değişikliğine izin verilen uzun test uygulamalarındaki Bayes yöntemlerine ilişkin RMSE değeri izin verilmeyen duruma göre uç noktalarda artmakta, orta yetenek düzeyinde ise benzerlik göstermektedir. MLE yönteminin kullanıldığı test desenlerinde ise RMSE değerinin cevap değişikliğine izin verilmeyen uygulamaya nazaran bütün koşullarda arttığı görülmektedir. MLEF yönteminde ise cevap değişikliğine izin verilen ve verilmeyen uygulamalarda uç değerlere ilişkin RMSE değeri benzerlik gösterirken orta yetenek düzeylerinde RMSE artışı gözlemlenmiştir. Cevap değişikliğine izin verildiğinde en fazla RMSE artışı, amaçlı yanlış cevaplamanın yapıldığı ve nihai yetenek kestirimi olarak MLE yönteminin kullanıldığı orta yetenek düzeyindeki bireylere aittir. Bu artış test uzunluğu arttıkça azalmaktadır.

Başlama kuralının regresyonla belirlendiği her bir koşula ilişkin yanlılık değerlerinin yetenek düzeylerine göre değişimi Şekil 8'de yer almaktadır. Bütün test desenlerinde Bayes kestirim yöntemlerine ilişkin yanlılık değerleri düşük yetenek düzeylerinde pozitif, yüksek yetenek düzeylerinde ise negatiftir. MLE kestirim yönteminde ise düşük yetenek düzeylerinde negatif, yüksek yetenek düzeylerinde ise pozitifdir. Bütün test desenlerinde yetenek kestirim yöntemi olarak Bayes ve MLE kullanıldığında yanlılık değerlerinin test uzunluğu arttıkça azaldığı görülmektedir. Madde cevap değişikliğine izin verildiğinde Bayes yöntemlerindeki yanlılık değerlerindeki artış uç yetenek düzeylerinde iken MLE ve MLEF yöntemlerine ilişkin yanlılık değerlerindeki artışın çoğunun yetenek düzeyinde olduğu görülmektedir. Madde cevap değişikliğine izin verilen koşullarda yanlılık değerlerindeki değişim Bayes ve MLEF yetenek kestirim yöntemlerinde çok düşük iken, MLE yönteminde ise yanlılık değerleri [0,020;0,406] aralığında değişmektedir.



## Şekil 8

Başlangıç Yetenek Düzeyi  $\theta_{\text{regresyon}}$  Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanlılık Grafiği



$\theta_{\text{regresyon}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen simülasyon koşullarına yönelik elde edilen RMSE ve yanlılık değerleriyle çizilen grafiklerin yanı sıra bu değerlerin cevaplama stiline, yetenek kestirim yöntemine ve test uzunluklarına göre istatistiksel olarak manidar düzeyde farklılaşıp farklılaşmadığını test etmek için ANOVA testi gerçekleştirilmiştir. Üç faktörlü ANOVA test sonuçları Tablo 8'de verilmiştir.

**Tablo 8**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{\text{regresyon}}$  ile Belirlendiği Koşullara İlişkin ANOVA*

*Sonuçları*

Varyansın Kaynağı	Sd	Yanlılık		RMSE	
		F	$\eta^2$	F	$\eta^2$
Cevaplama stili (CS)	2	222792*	0,34	138447*	0,15
Yetenek kestirim yöntemi (YKY)	3	90968*	0,21	215717*	0,35
Test Uzunluğu(TU)	4	19158*	0,06	107799*	0,23
CS*YKY	6	52642*	0,24	35985*	0,12
CS*TU	8	7727*	0,05	3848*	0,02
YKY*TU	12	5938*	0,05	18003*	0,12
CS*YKY*TU	24	2820*	0,05	1536*	0,02
CS*YKY@ TU:5	6	26815*	0,12	15465*	0,05
CS@ TU:5, @EAP	2	8*	0,00	44*	0,00
CS@ TU:5, @EAP(0,1)	2	3647*	0,01	749*	0,00
CS@ TU:5, @MLE	2	139440*	0,21	78000*	0,08
CS@ TU:5, @MLEF	2	32178*	0,05	14187*	0,02
CS*YKY@ TU:10	6	21023*	0,10	12909*	0,04
CS@ TU:10, @EAP	2	8*	0,00	117*	0,00
CS@ TU:10, @EAP(0,1)	2	3969*	0,01	984*	0,00
CS@ TU:10, @MLE	2	108871*	0,17	67690*	0,07
CS@ TU:10, @MLEF	2	30677*	0,05	15656*	0,02
CS*YKY@ TU:15	6	20916*	0,05	7828*	0,03
CS@ TU:15, @EAP	2	8*	0,00	162*	0,00
CS@ TU:15, @EAP(0,1)	2	3969*	0,00	867*	0,00
CS@ TU:15, @MLE	2	108871*	0,08	42444*	0,07
CS@ TU:15, @MLEF	2	18771*	0,02	12087*	0,02
CS*YKY@ TU: 20	6	4290*	0,02	4200*	0,01
CS@ TU:20, @EAP	2	3	0,00	147*	0,00
CS@ TU:20, @EAP(0,1)	2	2038*	0,00	676*	0,00
CS@ TU:20, @MLE	2	23276*	0,04	23513*	0,03
CS@ TU:20, @MLEF	2	10083*	0,02	8180*	0,01
CS*YKY@ TU:25	6	1716*	0,01	1718*	0,01
CS@ TU:25, @EAP	2	3	0,00	103*	0,00
CS@ TU:25, @EAP(0,1)	2	2682*	0,00	426*	0,00
CS@ TU:25, @MLE	2	18234*	0,04	10001*	0,01
CS@ TU:25, @MLEF	2	9761*	0,02	4171*	0,00

\*p<0.05

Tablo 8'de başlama kuralı olarak regresyon kullanıldığında elde edilen ortalama yanlılık ve RMSE değerleri üzerindeki temel etkilerin, ikili ve üçlü etkileşimlerin etkililiğinin istatistiksel olarak manidar düzeyde olduğu görülmektedir. Üç yönlü bir etkileşim, bir veya daha fazla iki yönlü etkileşimin üçüncü bir değişkenin seviyeleri arasında farklılık gösterdiği anlamına gelmektedir (UCLA, n.d.). Bu nedenle üçlü etkileşim olduğunda temel etkileri ve ikili etkileşimleri yorumlamak yerine araştırmanın amacına uygun olarak CS\*YKY etkileşimin yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. İki yönlü etkileşim analizi yaparken, daha önce tüm veri seti kullanılarak elde edilen üç yönlü ANOVA sonucundaki genel hata terimi (veya artıkları) kullanılmaktadır (Kirk, 1995; Keppel & Wickens, 2004). O nedenle iki yönlü etkileşimin F değeri hesaplanırken asıl model olan üç yönlü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

Tablo 8'de her bir test uzunluğunda CS\*YKY ikili etkileşimin elde edilen ortalama yanlılık ve RMSE değerleri üzerindeki istatistiksel olarak anlamlı olduğu görülmektedir. Başka bir deyişle, bütün test uzunluklarında yetenek kestirim yönteminin ortalama yanlılık ve RMSE üzerindeki etkisi cevaplama stiline göre farklılaşmaktadır. İkili etkileşimlerin anlamlı olması nedeniyle her bir test uzunluğunda ve yetenek kestiriminde cevaplama stiline ortalama yanlılık ve RMSE değerleri üzerindeki etkililiğini incelemek için basit etkiler (simple simple effect) hesaplanmıştır. Basit etkilerin hesaplanmasında asıl model olan üç yönlü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

Basit etkiler incelendiğinde cevaplama stiline ortalama yanlılık değerleri üzerindeki etkililiği test uzunluğuna ve yetenek kestirim yöntemine göre farklılaşmaktadır. Yetenek kestirimi olarak EAP yöntemi kullanıldığında ve test uzunluğu 5, 10 ve 15 olduğunda cevaplama stiline ortalama yanlılık üzerindeki etkililiği istatistiksel olarak anlamlı farklılık gösterirken, test uzunluğu 20 ve 25 olduğunda ise cevaplama stiline ortalama yanlılık üzerindeki etkililiği istatistiksel olarak anlamlı değildir. Yapılan Post-Hoc ikili

karşılaştırmalar sonucunda test uzunluğu 5, 10 ve 15 olduğunda ve cevap değişikliğine izin verildiğinde bütün yetenek kestirim yöntemlerinde amaçlı cevaplama ile elde edilen yetenek düzeylerinin ortalamasının daha yüksek olduğu tespit edilmiştir. Diğer yetenek kestirim yöntemlerinde (EAP(0,1), MLE ve MLEF) ise cevaplama stilinin ortalama yanlılık üzerindeki etkililiği bütün test uzunluklarında anlamlı farklılık göstermektedir. Yapılan Post-Hoc analiz sonuçlarına göre EAP(0,1), MLE ve MLEF yöntemlerinin kullanıldığı bütün test uzunluklarında cevap değişikliğine izin verildiğinde yanlılık değerlerinin arttığı tespit edilmiştir. Bireylerin amaçlı cevaplama ile daha fazla puan elde ettikleri belirlenmiştir.

Tablo 8'deki RMSE değerlerine yönelik basit etkiler incelendiğinde ise cevaplama stilinin ortalama RMSE üzerindeki etkililiğinin test uzunluğuna ve yetenek kestirim yöntemine göre farklılaştığı görülmektedir. Yapılan Post-Hoc ikili karşılaştırmaları ile cevap değişikliğine izin verildiğinde bütün yetenek kestirim yöntemlerinde ve test uzunluklarında amaçlı cevaplama ile elde edilen yetenek düzeylerine ilişkin ortalama RMSE değerlerinin istatistiksel olarak daha fazla olduğu belirlenmiştir.

Sonuç olarak başlangıç kuralında, nihai yetenek kestirimlerinde regresyon eşitliği ile belirlenen öncül bilgiler kullanıldığında ve test uzunluğu 20 ya da 25 olduğunda bireyin amaçlı cevap değişikliği yapsa bile cevap değişikliğine izin verilmeyen durumla istatistiksel olarak benzer yanlılık değerleri elde edilmektedir. BBT uygulamalarında bireylere cevap değişikliğine izin verildiğinde en kötü senaryo durumunda bile öncül bilginin kullanıldığı nihai yetenek kestirim yöntemlerinde elde edilen yanlılık değerleri [0,002;0,004], RMSE değerleri ise [0,206;0,297] aralığındadır. Cevap değişikliğine izin verilmeyen modelde ise yanlılık değerleri 0,001 civarında ve RMSE değerleri [0,175;0,279] aralığındadır.

### **Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar**

BBT uygulamalarında başlangıç yetenek düzeyi olarak aynı testin bir önceki uygulama sonucu elde edilen yetenek düzeyi kullanıldığında elde edilen yetenek kestirimlerinin standart hatası ve yanlılığı maddeyi yeniden gözden geçirmeye izin

verildiğinde bireylerin kullandığı cevaplama stiline, yetenek kestirim yöntemlerine ve test uzunluğuna göre nasıl değişmektedir?

İlgili alt problem kapsamında başlangıç yetenek düzeyi olarak aynı testin bir önceki uygulama sonucu kullanılarak koşullara ilişkin RMSE ve yanlılık değerleri belirlenmiştir. Bu başlama kuralı ile birlikte BBT uygulamalarında madde cevap değişikliğinin uygulama üzerindeki etkisinin test uzunluğuna ve nihai yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. Üç cevaplama stili ile gerçekleştirilen simülasyonlar sonucu elde edilen bulgulara ait ortalama RMSE ve yanlılık değerleri Tablo 9'da verilmiştir.

**Tablo 9**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{tekrar}$  ile Belirlendiği Koşullara İlişkin RMSE ve Yanlılık Değerleri*

	Yetenek Kestirimi	RMSE					Yanlılık				
		5	10	15	20	25	5	10	15	20	25
P-BBT	EAP(0,1)	0,375	0,263	0,222	0,200	0,194	0,006	0,003	0,002	0,001	<0,001
	EAP	0,121	0,121	0,121	0,120	0,120	0,002	0,002	0,002	0,002	0,002
	MLE	0,701	0,328	0,255	0,223	0,245	-0,009	-0,004	-0,002	-0,003	-0,004
	MLEF	0,543	0,310	0,250	0,221	0,230	-0,007	-0,004	-0,002	-0,003	-0,003
R-BBT	EAP(0,1)	0,406	0,282	0,232	0,205	0,187	0,010	0,006	0,004	0,003	0,002
	EAP	0,121	0,121	0,121	0,120	0,120	0,002	0,002	0,002	0,002	0,002
	MLE	0,902	0,412	0,282	0,234	0,208	0,002	-0,001	-0,001	-0,002	-0,002
	MLEF	0,645	0,357	0,268	0,229	0,206	0,000	-0,001	-0,001	-0,002	-0,002
Q-BBT	EAP(0,1)	0,364	0,240	0,193	0,168	0,152	0,015	0,010	0,011	0,009	0,004
	EAP	0,121	0,122	0,121	0,121	0,120	0,002	0,002	0,002	0,002	0,002
	MLE	0,704	0,285	0,211	0,180	0,161	0,019	0,003	0,006	0,005	0,001
	MLEF	0,522	0,271	0,210	0,180	0,161	0,011	0,002	0,006	0,005	0,002

P-BBT: Maddeyi yeniden cevaplama izin verilmeyen uygulama

R-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce rastgele cevaplama, sonra bilinçli cevaplama

Q-BBT: Maddeyi yeniden cevaplama izin verilen uygulamalarda önce amaçlı yanlış cevaplama, sonra bilinçli cevaplama

Tablo 9'da tüm koşullar için RMSE değerlerinin 0,120 ile 0,704 arasında değiştiği görülmektedir. BBT uygulamalarında başlangıç kuralı, aynı testin bir önceki uygulama sonucu kullanılarak belirlendiğinde test uzunluğu arttıkça MLE, MLEF ve EAP(0,1) nihai yetenek kestirim yöntemlerine ilişkin RMSE değeri azalırken öncül bilginin kullanıldığı Bayes yönteminde ise RMSE değerleri yaklaşık aynıdır. Cevap değişikliğine izin verilen ve verilmeyen uygulamalarda öncül bilginin kullanıldığı Bayes yetenek kestirim yöntemine

ilişkin RMSE değerleri en az iken MLE yöntemi ise en fazla RMSE değerlerine sahiptir. Tablo 9'da nihai yetenek kestirim yöntemleri olarak MLE, MLEF ve EAP(0,1) kullanılan koşullarda cevap değişikliğine izin verildiğinde, izin verilmeyen duruma göre RMSE değerlerinin arttığı görülmektedir.

Tablo 9'da tüm koşullar için yanlılık değerlerinin -0,009 ile 0,019 arasında değiştiği görülmektedir. Yanlılık değerlerinin test uzunluğu arttıkça genellikle azaldığı görülmektedir. Madde cevap değişikliğine izin verildiğinde ise yanlılık değerleri MLE, MLEF ve EAP(0,1) yetenek kestirim yöntemlerinin kullanıldığı durumlarda artarken, öncül bilginin kullanıldığı koşullarda ise değişmemektedir. Amaçlı yanlış cevaplanan test uygulamalarının yanlılık değerlerinin rastgele cevaplama göre daha fazla olduğu görülmektedir.

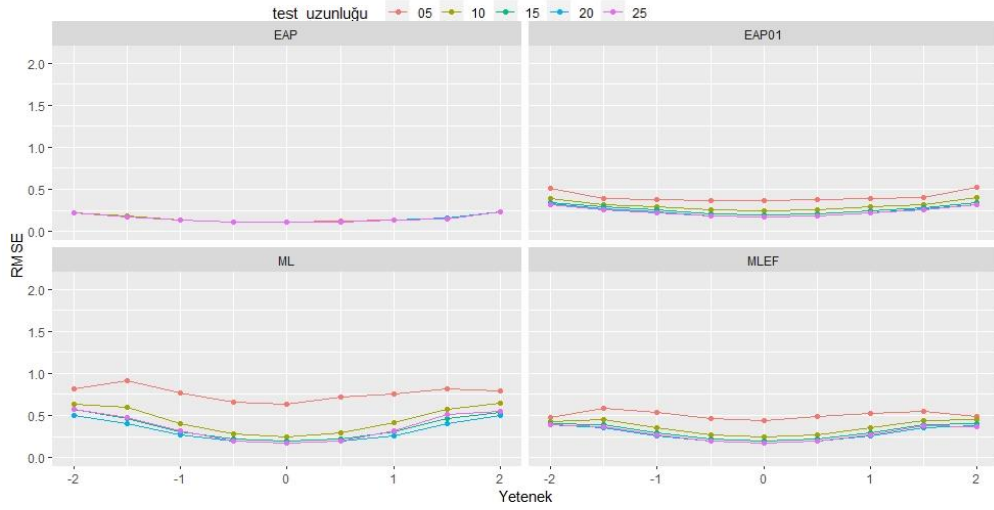
$\theta_{\text{tekrar}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen uygulamalarda elde edilen ortalama RMSE ve yanlılık değerlerine ek olarak bu değişkenlerin yetenek düzeylerine göre değişimlerinin incelenmesi amacıyla koşullu RMSE ve yanlılık değerlerini içeren grafikler çizilmiştir (Şekil 9).

Şekil 9'da başlangıç kuralı olarak aynı testin bir önceki sonucunun kullanıldığı BBT uygulamalarında nihai yetenek kestirimlerine yönelik koşullu RMSE değerleri yer almaktadır. Nihai yetenek kestiriminde öncül bilginin kullanıldığı (EAP) koşullarda yetenek düzeylerine ait RMSE değerleri kısa ve uzun testlerde benzerlik göstermektedir. Ancak öncül bilginin kullanılmadığı EAP(0,1), MLE ve MLEF yöntemleriyle kestirilen bütün yetenek düzeylerinin RMSE değeri, test uzunluğu arttıkça azalmaktadır. Şekil 9a'da madde cevap değişikliğine izin verilmeyen uygulamalarda öncül bilginin kullanıldığı EAP modelinde bütün test uzunluklarında benzer RMSE değerleri gözlemlenirken diğer yöntemlerde (EAP(0,1), MLE ve MLEF) ise test uzunluğu arttıkça bütün yetenek düzeylerinde RMSE değeri azalmaktadır.

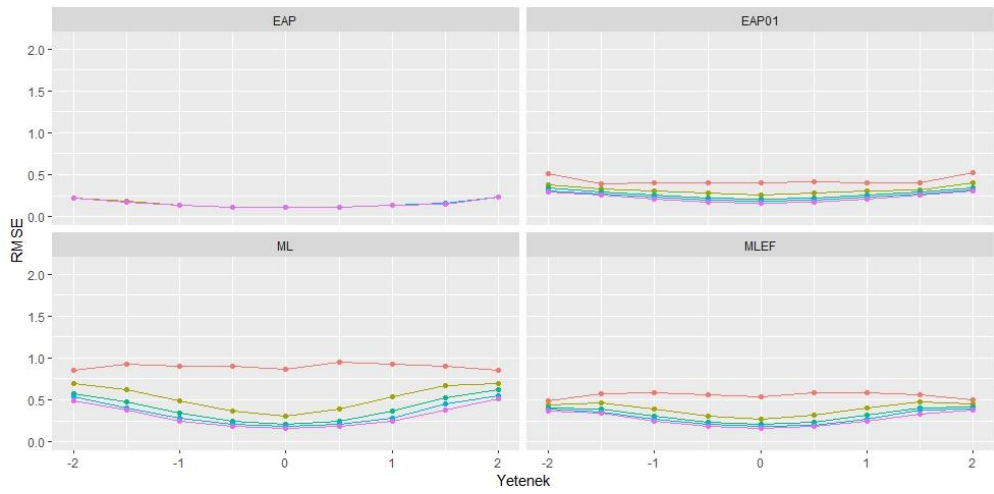
## Şekil 9

Başlangıç Yetenek Düzeyi  $\theta_{\text{tekrar}}$  Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu RMSE Grafiği

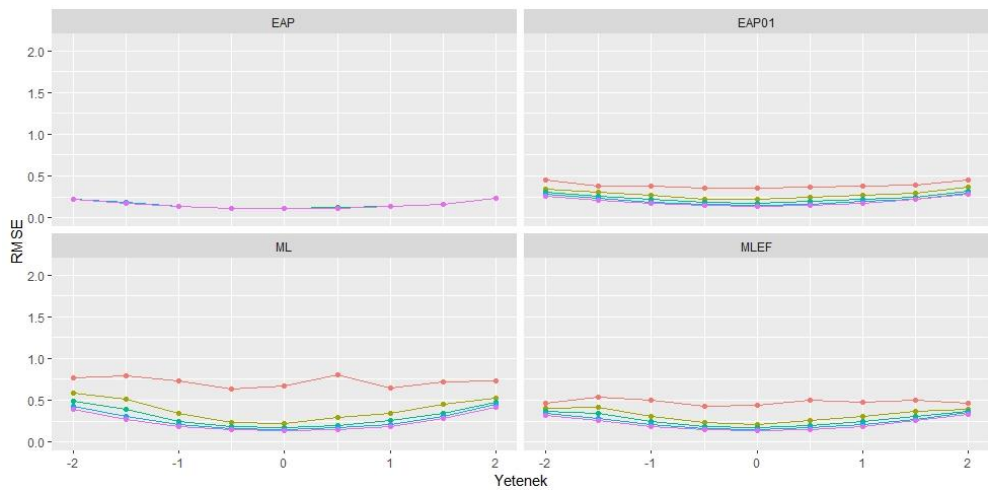
Şekil 9a



Şekil 9b



Şekil 9c



Cevap deęişikliğine izin verilmeyen uygulamalarda en küçük RMSE deęeri öncül bilginin kullanıldığı EAP; en yüksek ise MLE yetenek kestirim yöntemi ile elde edilmiştir. Şekil 9b ve 9c'de ise  $\theta_{\text{tekrar}}$  yöntemi ile başlama kuralı tanımlanan ve madde cevap deęişikliğine izin verilen test desenlerine ilişkin RMSE deęerleri yer almaktadır. Cevap deęişikliğine izin verildiğinde EAP yöntemine ait RMSE deęerleri cevap deęişkenliğine izin verilmeyen koşulla benzerlik gösterirken dięer yetenek kestirim yöntemlerine ilişkin RMSE deęerleri artmıştır. Amaçlı yanlış cevaplanan BBT uygulamalarında RMSE deęeri rastgele cevaplama göre Bayes yetenek kestirim yöntemlerinde farklılık göstermezken, MLE ve MLEF yetenek kestirim yöntemleri kullanıldığında ise orta yetenek düzeyinde artmaktadır. Cevap deęişikliği sonucu RMSE deęerindeki artış ise MLE yönteminde daha fazladır.

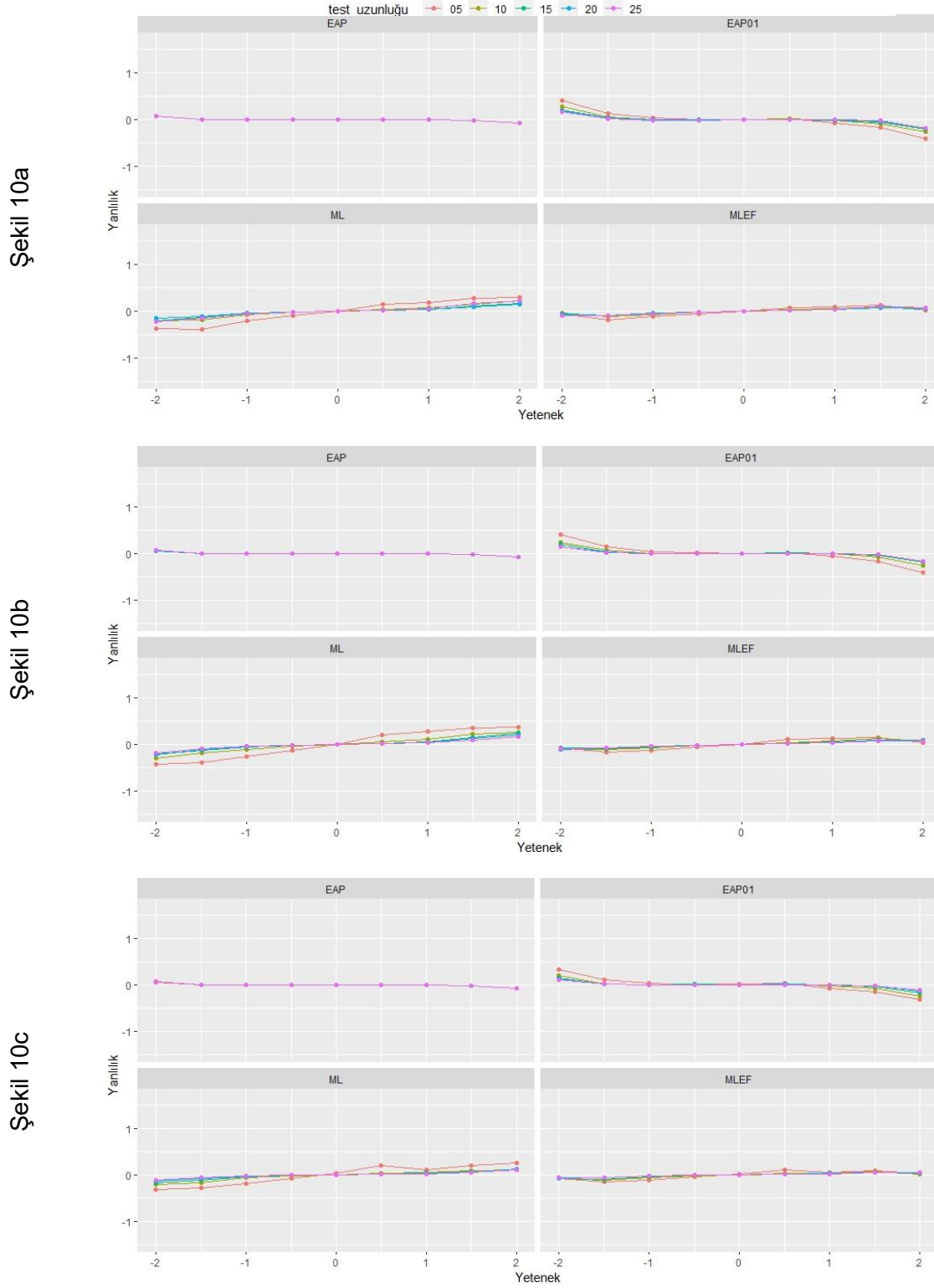
Başlangıç yetenek düzeyi  $\theta_{\text{tekrar}}$  yöntemi ile belirlenen uygulamalarda kullanılan nihai yetenek kestirimlerine yönelik koşullu yanlılık deęerlerinin yetenek düzeylerine göre deęişim grafikleri Şekil 10'da yer almaktadır. Öncül bilginin kullanıldığı bütün koşullarda yanlılık deęerleri sifıra yakındır. Cevap deęişikliğine izin verilen ve verilmeyen uygulamaların hepsinde EAP yöntemi en az yanlılık deęerine sahipken MLE en yüksek deęere sahiptir.

Uygulamalarda cevap deęişikliğine izin verilen ve verilmeyen durumlarda yetenek kestirim yöntemlerinin yanlılık deęerleri uç noktalarda ve kısa testlerde farklılık göstermektedir. Cevap deęişikliğine izin verilen ve verilmeyen bütün koşullarda EAP(0,1) yetenek kestirim yöntemi kullanıldığında düşük yetenek düzeyinde yanlılık deęeri pozitifken, yüksek yetenek düzeyinde ise negatiftir. MLE yöntemi kullanıldığında ise düşük yetenek düzeyinde yanlılık deęeri negatif, yüksek yetenek düzeyinde ise pozitifdir. MLEF yönteminde ise uç deęerlerde yanlılık deęerlerinin MLE yöntemi sonucunda elde edilen deęerlerden daha az olduđu görülmektedir.



### Şekil 10

Başlangıç Yetenek Düzeyi  $\theta_{tekrar}$  Yöntemi ile Belirlenen BBT Uygulamalarında Kullanılan Nihai Yetenek Kestirimlerine Yönelik Koşullu Yanlılık Grafiği



$\theta_{\text{tekrar}}$  yöntemi ile başlangıç yetenek düzeyi belirlenen simülasyon koşullarına yönelik elde edilen RMSE ve yanlılık değerleriyle çizilen grafiklerin yanı sıra bu değerlerin uygulama koşullarına göre istatistiksel olarak manidar düzeyde farklılaşp farklılaşmadığını test etmek için ANOVA testleri gerçekleştirilmiştir. Her bir koşula ait ANOVA test sonuçları Tablo 10'da yer almaktadır.

**Tablo 10**

*Başlangıç Yetenek Düzeyi Olarak  $\theta_{\text{tekrar}}$  ile Belirlendiği Koşullara İlişkin ANOVA Sonuçları*

Varyansın Kaynağı	sd	Yanlılık		RMSE	
		F	$\eta^2$	F	$\eta^2$
Cevaplama stili (CS)	2	222792*	0,34	138447*	0,15
Yetenek kestirim yöntemi (YKY)	3	90968*	0,21	215717*	0,35
Test Uzunluğu(TU)	4	19158*	0,06	107799*	0,23
CS*YKY	6	52642*	0,24	35985*	0,12
CS*TU	8	7727*	0,05	3848*	0,02
YKY*TU	12	5938*	0,05	18003*	0,12
CS*YKY*TU	24	2820*	0,05	1536*	0,02
CS*YKY@ TU:5	6	201*	0.14	1062	0.01
CS@ TU:5, @EAP	2	<1	<0,01	<1	<0,01
CS@ TU:5, @EAP(0,1)	2	201*	0,02	188	<0,01
CS@ TU:5, @MLE	2	1207*	0,24	5348	<0,01
CS@ TU:5, @MLEF	2	402*	0,10	1747	0.02
CS*YKY@ TU:10	6	17*	0.01	322	0.01
CS@ TU:10, @EAP	2	<1	<0,01	<1	<0,01
CS@ TU:10, @EAP(0,1)	2	72*	0.02	188	<0,01
CS@ TU:10, @MLE	2	68*	0.02	1666	<0,01
CS@ TU:10, @MLEF	2	55*	0.02	752	0,01
CS*YKY@ TU:15	6	33*	0.02	107	<0,01
CS@ TU:15, @EAP	2	<1	<0,01	<1	<0,01
CS@ TU:15, @EAP(0,1)	2	145*	0.02	161	<0,01
CS@ TU:15, @MLE	2	126*	0.02	511	<0,01
CS@ TU:15, @MLEF	2	125*	0.02	363	<0,01
CS*YKY@ TU: 20	6	26*	0.02	67	<0,01
CS@ TU:20, @EAP	2	<1	<0,01	<1	<0,01
CS@ TU:20, @EAP(0,1)	2	105*	0.02	161	<0,01
CS@ TU:20, @MLE	2	104*	0.02	322	<0,01
CS@ TU:20, @MLEF	2	103*	0.02	282	<0,01
CS*YKY@ TU:25	6	9*	0.01	148	<0,01
CS@ TU:25, @EAP	2	<1	<0,01	<1	<0,01
CS@ TU:25, @EAP(0,1)	2	28*	0.01	202	<0,01
CS@ TU:25, @MLE	2	396*	0.01	712	<0,01
CS@ TU:25, @MLEF	2	35*	0.01	497	<0,01

\*p<0.05

Tablo 10'da başlangıç yetenek düzeyi  $\theta_{\text{tekrar}}$  yöntemi ile belirlendiğinde ortalama yanlılık ve RMSE değerleri üzerindeki temel etkilerin, ikili ve üçlü etkileşimlerin etkililiği istatistiksel olarak manidar düzeyde olduğu görülmektedir. Üç yönlü bir etkileşim, bir veya daha fazla iki yönlü etkileşimin üçüncü bir değişkenin seviyeleri arasında farklılık gösterdiği anlamına gelmektedir (UCLA, n.d.). Bu nedenle üçlü etkileşim olduğunda temel etkileri ve ikili etkileşimleri yorumlamak yerine araştırmanın amacına uygun olarak CS\*YKY etkileşimin yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. İki yönlü etkileşim analizi yaparken, daha önce tüm veri seti kullanılarak elde edilen üç yönlü ANOVA sonucundaki genel hata terimi (veya artıkları) kullanılmaktadır (Kirk, 1995; Keppel & Wickens, 2004). O nedenle iki yönlü etkileşimin F değeri hesaplanırken asıl model olan üç yönü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

Tablo 10'da her bir test uzunluğunda CS\*YKY ikili etkileşimin elde edilen ortalama yanlılık ve RMSE değerleri üzerindeki istatistiksel olarak anlamlı olduğu görülmektedir. Başka bir deyişle, bütün test uzunluklarında yetenek kestirim yönteminin ortalama yanlılık ve RMSE üzerindeki etkisi cevaplama stiline göre farklılaşmaktadır. İkili etkileşimlerin anlamlı olması nedeniyle her bir test uzunluğunda ve yetenek kestiriminde cevaplama stilinin ortalama yanlılık ve RMSE değerleri üzerindeki etkililiğini incelemek için basit etkiler (simple simple effect) hesaplanmıştır. Basit etkiler hesaplanmasında asıl model olan üç yönlü ANOVA analizinin artık değeri ve serbestlik derecesi kullanılmıştır. Hesaplanan F değeri anlamlılığı, Bonferroni düzeltmesi yapılarak test edilmiştir.

Tablo 10'daki basit etkiler incelendiğinde cevaplama stilinin ortalama yanlılık ve RMSE değerleri üzerindeki etkililiği test uzunluğuna ve yetenek kestirim yöntemine göre farklılaşmaktadır. Yetenek kestirimi olarak EAP yönteminin kullanıldığı bütün test uzunluklarında cevaplama stillerine ait yanlılık ve RMSE değerleri arasında manidar farklılık bulunmadığı görülmektedir. Bir başka deyişle, öncül bilginin nihai yetenek kestiriminde kullanıldığı koşullarda cevap değişikliğine izin verildiğinde yanlılık ve RMSE

değerleri istatistiksel olarak değişmemektedir. Ancak yetenek kestirim yöntemi olarak EAP(0,1), MLE ve MLEF'nin kullanıldığı uygulamalarda ise cevaplama stillerine ait yanlılık ve RMSE değerleri arasında manidar farklılık bulunmaktadır. Bu farklılık test uzunluğu arttıkça azalmaktadır. Bu durum öncül bilgilerin varyansının çok küçük olmasından kaynaklanabilmektedir. Öncül bilgilerin varyansının küçük olduğu durumlarda maddelerden elde edilen bilgi öncül bilgi tarafından baskılanabilmektedir.

Sonuç olarak yetenek kestirim yönteminde en uygun öncül bilgilerin tanımlandığı ve nihai yetenek kestirimlerinde kullanıldığında bireyin amaçlı cevap değişikliği yapsa bile cevap değişikliğine izin verilmeyen durumla benzer yanlılık ve RMSE değerleri elde edilmektedir.

## **Tartışma**

Başlangıç yeteneğinin belli bir aralıktan rastgele seçildiği BBT uygulamalarında bireylere cevap değişikliği izni verilince yetenek kestirimlerinin yanlılık ve standart hata değerleri artmaktadır. Amaçlı yanlış cevaplamanın rastgele cevaplamaya nazaran yanlılık ve standart hata değerindeki artış daha fazladır. Chang ve Ying (2008) test stratejilerini kullanan bireylerin BBT sürecini olumsuz etkileyebileceğini ve gerçek yetenek düzeyinden daha fazla puan elde edebileceklerini belirtmiştir. ANOVA test sonuçlarına göre başlangıç yetenek düzeyi  $\theta_{random}$  yöntemi belirlendiğinde ve cevap değişikliğine izin verildiği bütün koşullarda yanlılık ve RMSE değerleri artmaktadır. Bu artış MLE yönteminde en fazla olurken EAP(0,1) yönteminde en azdır. Özellikle amaçlı yanlış cevaplamanın yapıldığı BBT uygulamalarında nihai yetenek kestirimi olarak MLE yönteminin kullanıldığı durumlarda cevap değişikliği sonucu RMSE değerleri test uzunluğu 25 olduğunda bile [0,500;1,660] aralığındadır. Blaise ve Raiche (2010) çalışmalarında [-3,00;+3,00] aralığında değişen yetenek düzeylerinde standart hatanın 0,40'a eşit veya daha düşük olmasının ölçme hassasiyeti için uygun olduğunu önermişlerdir. Alanyazında bu probleme karşılık geliştirilen çözümler sınırlı sayıda cevap değişikliğine izin verilmesidir (Stocking,

1997; Papanastasiou & Reckase, 2007; Cui ve ark., 2018b). Ancak bu çalışmada öncül bilgilerin, başlama kuralında ve nihai yetenek kestirimlerinde kullanılmasıyla bütün madde cevaplarının değiştirilmesine olanak sağlanmıştır. Başlama kuralı ile nihai yetenek kestirimlerinde regresyon eşitliği ile belirlenen öncül bilgiler kullanıldığında ve test uzunluğu 20 ve daha fazla olduğunda bireyin amaçlı cevap değişikliği yapsa bile cevap değişikliğine izin verilmeyen durumla istatistiksel olarak benzer yanlılık değerleri elde edilmektedir. Wang ve Vispoel (1998) bireyin bilgilerinden yola çıkarak başlatılan BBT uygulamasında Bayes yetenek kestirim yöntemlerinin yanlılık değerlerinin azaldığını belirtmektedirler. BBT uygulamalarında bireylere cevap değişikliğine izin verildiğinde en kötü senaryo durumunda bile öncül bilginin kullanıldığı nihai yetenek kestirim yöntemlerinde elde edilen yanlılık değerleri [0,002;0,004], RMSE değerleri ise [0,206;0,297] aralığındadır. Cevap değişikliğine izin verilmeyen modelde ise yanlılık değerleri 0,001 civarında ve RMSE değerleri [0,175;0,279] aralığındadır.

BBT uygulamalarında başlangıç yetenek düzeyi olarak aynı testin bir önceki uygulama sonucu elde edilen yetenek düzeyi kullanıldığında ise EAP kestirim yöntemlerinde yanlılık değerleri bakımından test uzunlukları arasında manidar farklılık varken bireyin başlangıç yetenek düzeyi regresyonla belirlenen uygulamada ise manidar farklılık bulunmamıştır. Bireyin başlangıç yeteneğinin en uygun tanımlanmasına rağmen öncül bilginin varyansının çok küçük tanımlanması test uzunluğunun etkisinden daha baskın olmuştur. van der Linden (1998) öncül bilgi bireyin yeteneğine tam olarak uygun değilse çok yanlı sonuçlar elde edilebileceğini belirtmiştir. Bu nedenle alanyazında yetenek tahmininde öncül bilgilerin kullanılmasına yönelik etik kaygılar mevcuttur (van der Linden, 1999; Veldkamp & Matteucci, 2013). Özellikle eğitim alanında sonuç odaklı değerlendirmelerde kullanılması uygun karşılanmamaktadır. Sonuç odaklı değerlendirmelerde BBT uygulaması sırasında ampirik bilgilerin kullanılması, nihai puanların yalnızca yanıt modellerine dayalı olarak rapor edilmesi önerilmektedir. BBT uygulamaları öğrenme sonuçlarının ölçülmesi yerine öğrenme sürecini optimize etmek için

de kullanılmaktadır (Veldkamp & Matteucci, 2013). Ülkemizde de akademik başarıların incelenmesi ve değerlendirilmesi amacıyla yapılan ABİDE eğer BBT olarak uygulanacak olursa alt testler arasındaki ilişkiden yola çıkarak öncül bilgiler tanımlanabilir. Böylece testi alan bireylere cevap değişikliği hakkı tanınarak daha kullanıcı dostu bir test modülü oluşturulabilir. Öncül bilgilerin kullanım alanları sadece eğitim amaçlı olarak değil sağlık taramalarında, personel alımında da kullanılabilir. Buradaki temel sorun BBT uygulamasının performansını değerlendirmede birincil önemin hangi değişken olduğudur. Amaç, sınava girenleri ölçülen özelliğe göre sıralamaksa, standart hata birincil; yanlışlık ise ikincil öneme sahip olur. Çünkü yanlışlık yetenek kestirimleriyle monoton olarak ilişkili olduğu sürece yetenek kestirimlerinin sırasını etkilemez (Wang & Vispoel, 1998). Bayes yöntemleri, özellikle 30 maddeden daha kısa test uzunluklarında, sıralama amacıyla MLE'den daha iyi sonuçlar sağlamaktadır (Wang & Vispoel, 1998). Bu nedenle eğer standart hatanın az olması isteniyorsa öncül bilginin kullanılması gerekirken, yanlışlık değerinin az olması isteniyorsa da tekdüze öncül dağılıma sahip Bayes yetenek kestirim yöntemi kullanılmalıdır (Morrison, 2017).

BBT uygulamalarında cevap değişikliğine izin verilmemesi birçok testin BBT formatına dönüşmesinde engel olabilir. Wise (2021) sadece düzey belirleme amacıyla gerçekleştirilen PISA gibi testlerde testi alan bireylerde düşük test motivasyonu olduğunu ve durumun testin geçerliğini olumsuz etkilediğini belirtmiştir (Shuttleworth, 2009). Bu tarz testler cevap değişikliğine izin verilmeyen BBT formatında uygulandığında, düşük olan test motivasyonları daha da düşebilir. Bu nedenle öncül bilginin uygun olduğu durumlarda bu bilgiyi kullanarak hem uygulama etkililiği artırılabilir hem de test stratejilerine karşı uygulama direnç kazanabilir. Unutulmamalıdır ki “*Test uygulamasını iyileştireceksek, sonraki nesil testlerin öncekilerden daha iyi olmasına izin vermeliyiz*” (Wainer, 1993).

## Bölüm 5

### Sonuç ve Öneriler

Araştırma kapsamında cevap değişikliğine izin veren BBT uygulamalarının etkililiği farklı test koşulları altında incelenmiştir. BBT uygulamalarının performansını değerlendirmek için RMSE ve yanlılık değerleri incelenmiştir. Çalışma sonucunda elde edilen sonuçlar;

1. Başlama kuralında herhangi bir öncül bilginin kullanılmadığı, ilk maddenin rastgele belirlendiği BBT uygulamalarında cevap değişikliğine izin verildiğinde yanlılık ve RMSE değerleri artmaktadır. Test uzunluğu arttıkça yanlılık ve RMSE değerleri azalmaktadır. Cevap değişikliğine izin verildiğinde yanlılık değerlerinin hepsinin pozitif olduğu görülmektedir yani bireylerin cevap değişikliği sonrası puanları artmaktadır. Amaçlı yanlış cevaplama, rastgele cevaplama göre daha yüksek standart hata ve yanlılık değerlerine sahiptir. MLE ve MLEF yöntemlerine ilişkin yanlılık ve RMSE değerleri arasındaki farklılık test uzunluğu arttıkça azalmaktadır.
2. Başlangıç kuralında, yetenek düzeyi regresyon eşitliği ile belirlenen değer kullanıldığı ve test uzunluğunun 20 ile 25 olduğu BBT uygulamalarında nihai yetenek kestirimi olarak öncül bilginin kullanılırsa yanlılık değerleri cevap değişikliğine izin verilme durumuna göre farklılık göstermezken standart normal dağılıma sahip  $EAP(0,1)$ , MLE ve MLEF yöntemlerinin kullanıldığı uygulamalarda cevaplama stilleri arasında anlamlı farklılık bulunmaktadır. Başka bir deyişle, nihai yetenek kestiriminde öncül bilginin kullanıldığı ve 20 ile 25 maddelik testlerde bireyler bütün maddelerin cevaplarını değiştirse bile cevap değişikliğine izin verilmeyen durumla benzer yanlılık değerleri elde edilmektedir.
3. Aynı testin bir önceki sonucu ile başlama kuralının belirlendiği BBT uygulamalarında öncül bilginin kullanıldığı nihai yetenek kestirim yöntemi olan

EAP kullanıldığında kısa testlerde de uzun testlerde de cevap deęişikliğine izin verilmediğinde ve verildiğinde yanlılık ve RMSE deęerleri bakımından benzer sonuçlar elde edilmektedir. EAP(0,1), MLE ve MLEF yetenek kestirim yöntemlerinin kullanıldığı uygulamalarda cevap deęişikliğine izin verilme durumunda yanlılık ve RMSE deęerleri artmaktadır.

Özetle madde cevap deęişikliğine izin verilen BBT uygulamalarında yanlılık ve standart hata deęerleri artmaktadır. Bu deęerlerdeki artış büyüklükleri sırasıyla MLE, MLEF ve EAP yöntemlerindedir. Bu deęerler test uzunluğu arttıkça manidar düzeyde azalmaktadır. Cevap deęişikliğine izin verilmeyen ve başlama kuralı olarak bireye rastgele bir deęer tanımlanan BBT uygulamaları sonucu elde edilen yanlılık ve RMSE deęerleri ile cevap deęişikliğine izin verilen, başlama kuralında ve nihai yetenek kestiriminde öncül bilginin kullanıldığı BBT uygulamalarında elde edilen deęerler birbirine paralellik göstermektedir.

Araştırma sonucunda elde edilen bulgulara yönelik öneriler test uygulayıcılarına ve araştırmacılara olmak üzere iki kısımdan oluşmaktadır.

Test uygulayıcılara yönelik öneriler;

1. Kısa BBT uygulamalarında öncül bilgi kullanımının etkisini gözlemlemek adına farklı test uzunlukları belirlenmiştir. Çalışmada en kısa test uzunluğu 5 olarak belirlenmiştir. Konu kapsam dengelenmesini de dikkate alarak daha uzun testler tercih edilebilir.
2. Cevap deęişikliğine izin veren BBT uygulamalarında öncül bilgi sadece başlama kuralında kullanılacaksa yetenek kestirim yöntemi olarak MLEF kullanılabilir.
3. Öncül bilgi her ne kadar uygun olsa da başlama kuralında tanımlanan yeteneęe ait varyansın küçük olmamasına dikkat edilmelidir.

Araştırmacılara yönelik öneriler;



1. Arařtırmada kullanılan cevaplama stilleri simülasyon verisi ile sınırlıdır. Gerçek cevap deęişikliklerini içeren veri setiyle çalışılması önerilir.
2. Çalışmadaki madde havuzu PISA 2015'te ikili puanlanan 171 fen maddesinden oluşmaktadır. Madde havuzu büyüklüğünün öncül bilgiye etkisi araştırılabilir.
3. Madde cevap deęişikliği sayısı sınırlandırılarak ölçme hassasiyetine etkisi araştırılabilir.
4. Çalışmada sabit uzunluk sonlandırma kuralı kullanılmıştır. Farklı sonlandırma kurallarının ve cevap deęişiklięinin ölçme hassasiyetinin üzerindeki etkisi araştırılabilir.
5. Başlama kuralında kullanılan öncül bilginin yardımıyla belirlenen yetenek tayinine ilişkin farklı varyans deęerlerinin ölçme hassasiyetine etkisi araştırılabilir.
6. Çalışmada konu kapsam dengelenmesine yer verilmemiştir. Konu kapsam dengelenmesini de dikkate alacak şekilde test uzunluğu belirlenebilir.

## Kaynaklar

- Babcock, B., & Weiss, D.J., (2012). Termination criteria in computerized adaptive tests: Do variable-length CAT's provide efficient and effective measurement? *International Association for Computerized Adaptive Testing*, 1, 1-18. <https://doi.org/10.7333/1212-0101001>
- Baker, F. B. (2001). The basics of item response theory (2nd. ed.). <https://eric.ed.gov/?id=ED458219>.
- Baker, F. B., & Kim, S. H. (2004). Bayesian Parameter Estimation Procedures. In: F. B. Baker & S. H. Kim (Eds.), *Item response theory parameter estimation techniques*, (p. 171-201). Marcel Dekker.
- Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, 1981(1), 1-8. <https://doi.org/10.1002/j.2333-8504.1981.tb01255.x>
- Bergstrom, B. A., & Lunz, M. E. (1999). CAT for certification and licensure. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 67-91). Lawrence Erlbaum Associates Publishers.
- Bezirhan, U., von Davier, M., & Grabovsky, I. (2021). Modeling item revisit behavior: The hierarchical speed-accuracy-revisits model. *Educational and Psychological Measurement*, 81(2), 363-387. <https://doi.org/10.1177/0013164420950556>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. F. M. Lord & M. R. Novick (Ed), *Statistical theories of mental test scores* (pp. 397-472). Reading MA: Addison-Wesley.
- Blais, J. G., & Raiche, G. (2010). Features of the sampling distribution of the ability estimate in computerized adaptive testing according to two stopping rules. *Journal of Applied Measurement*, 11(4), 424-431.

- Bock, R. D., & Aitkin, M. (1981) Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459. <https://doi.org/10.1007/BF02293801>
- Bowles, R., & Pommerich, M. (2001, April). *An examination of item review on a CAT using the specific information item selection algorithm*. National Council of Measurement in Education, Seattle, WA, United States. <https://eric.ed.gov/?id=ED454275>
- Bridgeman B. (2012). A simple answer to a simple question on answer changing. *Journal of Educational Measurement*, 49(4), 467-468. <https://doi.org/10.1111/j.1745-3984.2012.00189.x>
- Carlson, J. E., & von Davier, M. (2013). Item response theory. *ETS Research Report Series*, 2013(2), 1-69. <https://doi.org/10.1002/J.2333-8504.2013.TB02335.X>
- Carlson, S. (2000). ETS finds flaws in the way online GRE rates some students. *Chronicle of Higher Education*, 47(8), A47.
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1-20. <https://doi.org/10.1007/s11336-014-9401-5>
- Chang, H. H., & Ying, Z. (1999). a-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 211-222. <https://doi.org/10.1177/01466219922031338>
- Chang, H. H., & Ying, Z. (2008). To weight or not to weight? Balancing influence of initial items in adaptive testing. *Psychometrika*, 73(3), 441-450. <https://doi.org/10.1007/s11336-007-9047-7>
- Chen, S. Y., Ankenmann, R. D., & Chang, H. H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement*, 24(3), 241–255. <https://doi.org/10.1177/01466210022031705>

- Choi, S. W., Grady, M. W., & Dodd, B. G. (2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71(1), 37-53. <https://doi.org/10.1177/0013164410387338>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavior Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Cui, Z., Liu, C., He, Y., & Chen, H. (2018a). *Comparison of algorithms that allow item review in computerized adaptive testing* (ACT Research Report No:17009).
- Cui, Z., Liu, C., He, Y., & Chen, H. (2018b). Evaluation of a new method for providing full review opportunities in computerized adaptive testing—Computerized adaptive testing with salt. *Journal of Educational Measurement*, 55(4), 582-594. <https://doi.org/10.1111/jedm.12193>
- de Abreu e Lima, F (2019). Bayesian models in R. *Poissonisfish*. Retrieved June 20, 2022. <https://poissonisfish.com/2019/05/01/bayesian-models-in-r/>
- de la Torre, J. (2009). Improving the quality of ability estimates through multidimensional scoring and incorporation of ancillary variables. *Applied Psychological Measurement*, 33(6), 465-485. <https://doi.org/10.1177/0146621608329890>
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- Dolan, R. P., & Burling, K. S. (2012). Computer-based testing in higher education. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education* (p. 312-335). Routledge. <https://doi.org/10.4324/9780203142189>
- Educational Testing Service [ETC] (2011). GRE Information Bulletin 2015-16. [www.ets.org/s/gre/pdf/gre\\_info\\_reg\\_bulletin.pdf](http://www.ets.org/s/gre/pdf/gre_info_reg_bulletin.pdf)

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Erlbaum.
- Flaugher, R. (2000). Item pools. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 37-59). Lawrence Erlbaum.
- Gelbal, S. (2013). Ölçme ve değerlendirme. Eskişehir: Açıköğretim Fakültesi Yayını
- George, D. & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference* (10th ed.) Pearson.
- Gershon, R. (2005). Computerized adaptive testing. *Journal of Applied Measurements*, 6(1), 109-127.
- Gershon, R., & Bergstrom, B. (1995, April). *Does cheating on CAT pay: NOT!* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC document reproduction service No. ED 392 844).
- Guyer, R. D., & Weiss, D. J. (2009). Effect of early misfit in computerized adaptive testing on the recovery of theta. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. [www.psych.umn.edu/psylabs/CAT Central/](http://www.psych.umn.edu/psylabs/CATCentral/)
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Han, K. T. (2013). Item pocket method to allow response review and change in computerized adaptive testing. *Applied Psychological Measurement*, 37(4), 259-275. <https://doi.org/10.1177/0146621612473638>

- Han, K. T. (2016). Maximum Likelihood Score Estimation Method with Fences for Short Length Tests a Computerized Adaptive Tests. *Applied Psychological Measurement, 40*(4), 289-301. <https://doi.org/10.1177/01466216166631317>
- He, W., & Reckase, M. D. (2014). Item pool design for an operational variable length computerized adaptive test. *Educational and Psychological Measurement, 74*(3), 473-494. <https://doi.org/10.1177/0013164413509629>
- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice, Summer 2007*, 44-52. <https://doi.org/10.1111/j.1745-3992.2007.00093.x>
- Ivie, J. L. (2007). *Test taking strategies in computer adaptive testing that will improve your score: fact or fiction?* (Doctoral Dissertation). University of Michigan, United State of America.
- Jensen, M. G. (2017). *Extension of the item pocket method allowing for response review and revision to a computerized adaptive test using the generalized partial credit model* (Doctoral dissertation). The University of Texas at Austin, United State of America.
- Keppel, G., & Wickens, T. D. (2004). *Design and Analysis. A Researcher's Handbook* (4<sup>th</sup> ed.). Pearson.
- Kingsbury, G. (1996). *Item review and adaptive testing*. National Council on Measurement in Education, New York.
- Kingsbury, G. G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education, 2*(4), 359-375. [https://doi.org/10.1207/s15324818ame0204\\_6](https://doi.org/10.1207/s15324818ame0204_6)
- Kirk, Roger E. (1995) *Experimental Design: Procedures for the Behavioral Sciences* (3<sup>th</sup> ed.). Brooks/Cole Publishing.

- Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology*, 88(5), 725-735. <https://doi.org/10.1037/0022-3514.88.5.725>
- Lin, Z., Chen, P., & Xin, T. (2021). The block item pocket method for reviewable multidimensional computerized adaptive testing. *Applied psychological measurement*, 45(1), 22-36. <https://doi.org/10.1177/0146621620947177>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luecht, R. M., & Sireci, S. G. (2011). A Review of Models for Computer-Based Testing. Research Report 2011-12. *College Board*.
- Lunz, M. E., Bergstrom, B. A. & Wright, B. D. (1992). The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement*, 16, 41-51. <https://doi.org/10.1177/014662169201600103>
- Magis, D., Yan, D., & von-Davier, A. (Eds.). (2017). *Computerized adaptive and multistage testing with R: Using packages catR and mstR*. Springer.
- Matteucci, M. & Veldkamp, B. P. (2013). On the use of MCMC computerized adaptive testing with empirical prior information to improve efficiency. *Statistical Methods & Applications*, 22, 243-267. <https://doi.org/10.1007/s10260-012-0216-1>
- McMorris, R. F. (1991). *Why do young students change answers on tests?* <https://eric.ed.gov/?id=ED342803>
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied psychological measurement*, 23(3), 187-194. <https://doi.org/10.1177/01466219922031310>

- Melican, G. J., Breithaupt, K., & Zhang, Y. (2010). Designing and implementing a multistage adaptive test: The Uniform CPA Examination. In W. J. van der Linden & C. E. W. Glas (Eds.), *Elements of adaptive testing* (pp. 167–189). Springer.
- Merritt, J. (2003, December 29). Why the folks at ETS flunked the course—a tech-savvy service will soon be giving B-school applicants their GMATs. *Business Week*.
- Mills C. N., & Stocking, M. L. (1996) Practical issues in Large-Scale Computerized Adaptive Testing, *Applied Measurement in Education*, 9(4), 287-304, [https://doi.org/10.1207/s15324818ame0904\\_1](https://doi.org/10.1207/s15324818ame0904_1)
- Morrison, K. M. (2017). *Impact of composite population priors on computer adaptive test proficiency estimates* (Doctoral dissertation). Georgia Institute of Technology, United States of America.
- OECD (n.d.). PISA 2015 Technical Report: Annex G: Common Unique Item Parameters in Each Domain, by Countries and Languages. <https://www.oecd.org/pisa/data/2015-technical-report/>
- Olea, J., V., Revuelta, J., Ximenez, M. C., & Abad, F. J. (2000). Psychometric and psychological effects of review on computerized fixed and adaptive tests. *Psicología*, 21, 157-173.
- Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive testing. *Journal of the American Statistical Association*, 70, 351-356. <https://doi.org/10.1007/BF02294775>
- Pagni, S., Bak, A., Eisen, S., Murphy, J., Finkelman, M., & Kugel, G. (2017). The benefit of a switch: answer-changing on multiple-choice exams by first-year dental students. *Journal of Dental Education*, 81(1), 110–115. <https://doi.org/10.1002/j.0022-0337.2017.81.1.tb06253.x>



- Papanastasiou, E. C. (2005). Item review and the rearrangement procedure: Its process and its results. *Educational Research and Evaluation*, 11(4), 303–321.  
<https://doi.org/10.1080/13803610500110521>
- Papanastasiou, E. C. (2015). Psychometric changes on item difficulty due to item review by examinees. *Practical Assessment, Research & Evaluation*, 20(3).  
<https://doi.org/10.7275/jcyv-k456>
- Papanastasiou, E. C., & Reckase, M. D. (2007) A “rearrangement procedure” for scoring adaptive tests with review options. *International Journal of Testing*, 7(4), 387-407.  
<https://doi.org/10.1080/15305050701632262>
- Parshall, C. G., Davey, T. & Nering, M. L. (1998). *Test development exposure control for adaptive testing*. National Council on Measurement in Education, San Diego.  
<https://eric.ed.gov/?id=ED421526>
- Pastor, D. A., Dodd, B. G., & Chang, H. H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement*, 26(2), 147–163.  
<https://doi.org/10.1177/01421602026002003>
- Patsula, L. N. (1999). *A comparison of computerized adaptive testing and multistage testing*. (Doctoral dissertation). University of Massachusetts Amherst, United States of America.
- Ranganathan, K., & Foster, I. (2003). Simulation studies of computation and data scheduling algorithms for data grids. *Journal of Grid Computing*, 1, 53-62.  
<https://doi.org/10.1023/A:1024035627870>
- Rezaei, M., & Golshan, M. (2015). Computer Adaptive Test (CAT). Advantages and Limitations. *International Journal of Educational Investigations*, 2(5), 128-137.

- Rulison, K. L., & Loken, E. (2009). I've fallen and I can't get up: Can high-ability students recover from early mistakes in CAT? *Applied Psychological Measurement*, 33(2), 83-101. <https://doi.org/10.1177/0146621608324023>
- Şahin, A., & Weiss, D. J. (2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory & Practice*, 15(6). <https://doi.org/10.12738/estp.2015.6.0102>
- Segall, D. O. (2004). A sharing item response theory model for computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 29(4), 439-460. <https://doi.org/10.3102/10769986029004439>
- Shuttleworth, M. (2009). *Repeated measures design*. *Experiment Resources*.
- Sireci, S. (2003). Computerized adaptive testing: An introduction. In Wall, & Walz (Eds), *Measuring up: Assessment issues for teachers, counselors and administrators* (pp. 684-694). CAPS Press.
- Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools* (ETS Research Rep. No. 94-5). Princeton: Educational Testing Service.
- Stocking, M. L. (1997). Revising item responses in computerized adaptive tests: A comparison of three models. *Applied Psychological Measurement*, 21(2), 129-142. <https://doi.org/10.1177/01466216970212003>
- Stone, G. E., & Lunz, M. E. (1994). The effect of review on the psychometric characteristics of computerized adaptive tests. *Applied Measurement in Education*, 7(3), 211-222. [https://doi.org/10.1207/s15324818ame0703\\_4](https://doi.org/10.1207/s15324818ame0703_4)
- Stroud, A., & Secrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs. Prentice-Hall.
- Stylianou-Georgiou, A., & Papanastasiou, E. C. (2017). Answer changing in testing situations: The role of metacognition in deciding which answers to

review. *Educational Research and Evaluation*, 23(3-4), 102-118.

<https://doi.org/10.1080/13803611.2017.1390479>

Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small Sample Estimation in Dichotomous Item Response Models: Effect of Priors Based on Judgmental Information on the Accuracy of Item Parameter Estimates. *Appl. Psychol. Meas.* 27(1), 27-51. <https://doi.org/10.1177/0146621602239475>

Thissen, D. & Mislevy, R.J. (2000). *Testing algorithms*. In H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (2<sup>nd</sup> ed.) 101-133. Lawrence Erlbaum Associates.

Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1). 1-9. <https://doi.org/10.7275/wqzt-9427>

UCLA (n.d.). FAQ How can I understand a three-way interaction in Anova? *Advance Research Computing Statistical Methods and Data Analysis*. Retrieved June 20, 2022. <https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faqhow-can-i-understand-a-three-way-interaction-in-anova/>

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, 16(2), 75-84.

van der Linden, W. J. (1999). Empirical initialization of the trait estimator in adaptive testing. *Applied Psychological Measurement*, 23(1), 21-29. <https://doi.org/10.1177/01466219922031149>

van der Linden, W. J., & Pashley, P. J. (2010). Item Selection and Ability Estimation in Adaptive Testing. In van der Linden, W. J. & Glas, C. A. W. (Eds.), *Elements of adaptive testing* (pp. 3–30). Springer.

- van der Linden, W. J., Klein Entink, R. H., & Fox, J.-P. (2010). IRT Parameter Estimation with Response Times as Collateral Information. *Applied Psychological Measurement, 34*(5), 327–347. <https://doi.org/10.1177/0146621609349800>
- Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerized adaptive testing. *Ensaio: Avaliação e Políticas Públicas em Educação, 21*(78), 57-81.
- Veldkamp, B. P., & van der Linden, W. J. (2000). Designing item pools for computerized adaptive testing. In van der Linden, WJ, & Glas, CAW (Eds.), *Computerized adaptive testing: Theory and practice* (p. 149-162). Kluwer Academic Publishers.
- Vispoel, W. P., Clough, S. J., Bleiler, T., Hendrickson, A. B., & Ihrig, D. (2002). Can examinees use judgments of item difficulty to improve proficiency estimates on computerized adaptive vocabulary tests? *Journal of Educational Measurement, 39*(4), 311–330.
- Vispoel, W. P., Henderickson, A. B., & Bleiler, T. (2000). Limiting answer review and change on computerized adaptive vocabulary test: Psychometric and attitudinal results. *Journal of Educational Measurement, 37*(1), 21-38.
- Vispoel, W. P., Rocklin, T. R., Wang, R., & Bleiler, T. (1999). Can examinee use a review option to obtain positively biased ability estimates on a computerized adaptive test? *Journal of Educational Measurement, 36*(2), 141-157. <https://doi.org/10.1111/j.1745-3984.1999.tb00551.x>
- Waddell, D. L., & Blankenship, J. C. (1994). Answer changing: A meta-analysis of the prevalence and patterns. *The Journal of Continuing Education in Nursing, 25*(4), 155-158. <https://doi.org/10.3928/0022-0124-19940701-06>
- Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practice, 12*(1), 15–20. <https://doi.org/10.1111/j.1745-3992.1993.tb00519.x>

- Wainer, H. (2000). Introduction and history. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (p. 1-22). Lawrence Erlbaum Associates.
- Wang, S., Xiao, H., & Cohen, A. (2021). Adaptive weight estimation of latent ability: application to computerized adaptive testing with response revision. *Journal of Educational and Behavioral Statistics*, 46(5), 560-591.  
<https://doi.org/10.3102/1076998620972800>
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135. <https://doi.org/10.1111/j.1745-3984.1998.tb00530.x>
- Wang, X., Nozawa, Y., Gao, X., & Lin, H. (2014). *A study on adaptive test battery: the impact of collateral information and subtest sequencing on classification accuracy*. American Educational Research Association, Philadelphia, PA.
- Wang, X., Nozawa, Y., & Gao, X. (2012). A Study on Adaptive Test Battery: The Impact of Collateral Information and Subtest Sequencing on Classification Accuracy. *ACT Research Report Series 2012*.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27.  
<https://doi.org/10.1111/j.1745-3992.1998.tb00632.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. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.  
<https://doi.org/10.1080/07481756.2004.11909751>

- Weiss, D. J. (2011) Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Science*, 2(1), 1-27. <https://doi.org/10.2458/v2i1.12351>
- Weiss, D. J., & Guyer, R. (2010). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing*. St. Paul MN: Assessment Systems Corporation. <http://iacat.org/sites/default/files/biblio/CATSIM%20Manual.pdf>
- Wise, S. L. (1996, April). *A critical analysis of the arguments for and against item review in computerized adaptive testing*. National Council on Measurement in Education, New York.
- Xie, Q. (2019). *The impact of collateral information on ability estimation in an adaptive test battery* (Doctoral dissertation), The University of Iowa. United States of America.
- 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>
- Xu, G., Wang, C., & Shang, Z. (2016). On initial item selection in cognitive diagnostic computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 69(3), 291-315. <https://doi.org/10.1111/bmsp.12072>
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, 37(1), 3-23. <https://doi.org/10.1177/0146621612455687>
- Yen, Y. C., Ho, R. G., Laio, W. W., Chen, L. J., & Kuo, C. C. (2012). An empirical evaluation of the slip correction in the four parameter logistic models with computerized adaptive testing. *Applied Psychological Measurement*, 36(2), 75-87. <https://doi.org/10.1177/0146621611432862>

Zenisky, A., Hambleton, R. J., & Luecht, R. M. (2010). Multistage testing: Issues, designs, and research. In W. J. van der Linden & C. E. W. Glas (Eds.), *Elements of adaptive testing* (pp. 355-372). Springer.

**EK-A: Arařtırma Etik Komisyonu Onay Bildirimi**

T.C.  
HACETTEPE ÜNİVERSİTESİ  
Rektörlük

Tarih: 12.04.2019 17:07  
Sayı: 35853172-300-E.00000543149  
  
E.00000543149

Sayı : 35853172-300  
Konu : Ömer Faruk ŐEN Hk.

**EĐİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĐÜNE**

İlgi : 27.03.2019 tarihli ve 51944218-300/00000522641 sayılı yazı.

Enstitünüz Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Deđerlendirme Bilim Dalı doktora programı öğrencilerinden **Ömer Faruk ŐEN**'in **Prof. Dr. Hülya KELECİOĐLU** danışmanlığında yürüttüğü "**Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamalarında Maddeyi Yeniden Cevaplayabilme Yöntemlerinin İncelenmesi**" başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonunun **02 Nisan 2019** tarihinde yapmış olduđu toplantıda incelenmiş olup,etik açıdan uygun bulunmuştur.

Bilgilerinizi ve geređini saygılarımla rica ederim.

e-izmalıdır  
Prof. Dr. Rahime Meral NOHUTCU  
Rektör Yardımcısı

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 26845677-6004-46101-0234022-050078670 kodu ile erişebilirsiniz.  
Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara  
Telefon:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta:yazimd@hacettepe.edu.tr İnternet  
Adresi: www.hacettepe.edu.tr

Duygu Didem ILFPI





**EK-B: Etik Beyanı**

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

- \* 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.

20/07/2022

(İmza)

Ömer Faruk ŞEN

**EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu**

05/07/2022

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

Tez Başlığı: Bilgisayarda Bireyselleştirilmiş Test Uygulamalarında Maddeyi Yeniden Cevaplayabilmenin Ölçme Hatasına ve Yanlılığına Etkisinin İncelenmesi  
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
05/07/2022	82	133840	27/06/2022	2	1866849056

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 Esaslarını 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:** Ömer Faruk ŞEN

**Öğrenci No.:** N14246116

**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-Ç: Thesis/Dissertation Originality Report

05/07/2022

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

Thesis Title: The Effect of Item Review in Computerized Adaptive Testing on Measurement Error and Bias

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
05/07/2022	82	133840	27/06/2022	2	1866849056

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:** Ömer Faruk ŞEN

**Student No.:** N14246116

**Department:** Educational Sciences

**Program:** Educational Measurement and Evaluation

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

Signature

### ADVISOR APPROVAL

APPROVED  
Prof. Dr. Hülya KELECİOĞLU

## EK-D: Yayınlama 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. <sup>(1)</sup>
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir. <sup>(3)</sup>

20/07/2022

Ömer Faruk ŞEN

---

"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 internetten 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.

