



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

ÇOK AŞAMALI BİREYE UYARLANMIŞ TESTLERDE FARKLI KOŞULLARDAN
ELDE EDİLEN YETENEK KESTİRİMLERİNİN KARŞILAŞTIRILMASI

Fatma Güneş ERTAŞ POLAT

Doktora Tezi

Ankara, 2022



Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

ÇOK AŞAMALI BİREYE UYARLANMIŞ TESTLERDE FARKLI KOŞULLARDAN
ELDE EDİLEN YETENEK KESTİRİMLERİNİN KARŞILAŞTIRILMASI

COMPARISON OF ABILITY ESTIMATIONS OBTAINED FROM DIFFERENT
CONDITIONS IN MULTI STAGE ADAPTIVE TESTING

Fatma Güneş ERTAŞ POLAT

Doktora Tezi

Ankara, 2022

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Güneş ERTAŞ'ın hazırladığı “Çok Aşamalı Bireye Uyarlanmış Testlerde Farklı Koşullardan Elde Edilen Yetenek Kestirimlerinin Karşılaştırılması” başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Nükhet ÇIKRIKCI	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Duygu ANIL	İmza
Jüri Üyesi	Prof. Dr. Nilüfer KAHRAMAN	İmza
Jüri Üyesi	Doç. Dr. Ergül DEMİR	İmza
Jüri Üyesi	Doç. Dr. Burcu ATAR	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 08 / 04 / 2022 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. Selahattin GELBAL
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu çalışmanın amacı, farklı tasarımlardaki çok aşamalı bireye uyarlanmış testlerin etkinliğini, farklı koşullar altında incelemektir. Çalışmada madde güçlük parametreleri normal ve uniform dağılan iki ayrı madde havuzu oluşturulmuştur. Araştırmanın koşulları olan modül uzunluğu, yönlendirme yöntemi ve kestirim yöntemi, bu iki ayrı madde havuzu üzerinden gerçekleştirilen simülasyonlar ile ayrı ayrı sınanmıştır. Simülasyonlar R yazılımında yer alan mstR paketi yardımıyla gerçekleştirilmiştir. Artan, azalan, eşit, ikinci aşamanın en uzun ilk aşamanın en kısa ve ikinci aşamanın en uzun son aşamanın en kısa olduğu beş farklı modül uzunluğu tasarımı, dinamik ve statik yönlendirme kuralları ve EAP ve MLE son kestirim yöntemleri koşulları altında ortalama hata ve yanlılık değerlerine göre karşılaştırılmıştır. Araştırmanın bulgularından elde edilen sonuçlar, değişen modül uzunluklarının ortalama hata ve yanlılık üzerinde etkisi olduğunu göstermektedir. Ancak madde güçlük dağılımlarına göre bu etkinin farklı koşullar altında farklılaştığı görülmüştür. Yönlendirme yönteminin statik veya dinamik olması, madde güçlüklerinin normal dağılımdan türetildiği koşulda yetenek kestiriminde bir fark yaratmazken, uniform dağılımdan türetildiği koşulunda düşük de olsa bir fark oluşturmaktadır. Alan yazınında altı çizildiği gibi, bu çalışmanın sonuçları da her durumda en iyi olan tek bir tasarımın olmadığını göstermektedir. Bu nedenle, tüm gereklilikleri ve amaçları karşılayan en uygun çok aşamalı bireye uyarlanmış test tasarımının belirlenmesi için benzer çalışmaların yapılması önerilmektedir. Çok aşamalı bireye uyarlanmış testler, testi hazırlayanların, uygulayıcıların ve testi alan bireylerin tamamen kabul edebileceği ve kolay anlaşılır bir test olma avantajına sahiptir. Bu avantajı ve diğer özellikleri de göz önünde bulundurulduğunda, Türkiye’de uygulanan merkezi sınavların çok aşamalı bireye uyarlanmış test olarak uygulanmasına yönelik çalışmaların yapılmasının önemli olduğu düşünülmektedir.

Anahtar sözcükler: çok aşamalı bireye uyarlanmış testler, güçlük dağılımı, modül uzunluğu, yönlendirme yöntemi, kestirim yöntemi, geniş ölçekli testler

Abstract

The purpose of this study is to examine the effectiveness of multi-stage adaptive tests of different designs under different conditions. In the study, two separate item pools were generated, with item difficulty parameters that were either normally or uniformly distributed. The study factors were module length, routing method and estimation method, and they were tested separately using data simulated to represent the two item pools. The simulation was conducted with the help of the mstR package in the R software. Five different module length designs were considered: increasing, decreasing, equal, second stage longest (first stage shortest), and second stage longest (last stage shortest). These designs were compared according to mean error and bias computed under the dynamic & static routing rule and EAP & ML final estimation methods conditions. The findings show that varying module lengths has an effect on the mean error and bias. However, it was observed that this effect differed under the two item difficulty distribution conditions studied. Whether the routing method is static or dynamic did not make a difference in the ability estimations when the item difficulties were generated using a normal distribution, but there was a slight difference when the difficulties were generated using a uniform distribution. Consistent with the literature, the results of this study also show that there is no single design that is best in all cases. Therefore, future studies are recommended to formulate alternative MST designs meeting various test requirements and objectives. MST designs are seen as user friendly and widely accepted by the test developers, test administrators and the examinees around the world. Given its advantages, the MST can be a useful alternative for the existing central exam designs in Turkey.

Keywords: multistage adaptive testing, difficulty distribution, module length, routing rules, estimation methods, large scale tests

Teşekkür

Doktora eğitimim süresince, desteğini ve şefkatini esirgemeyen kıymetli danışman hocam Prof. Dr. Duygu Anıl'a bana sundukları için çok teşekkür ediyorum.

Hacettepe Üniversitesi, Eğitimde Ölçme ve Değerlendirme Bilim Dalında doktora eğitimim boyunca ders aldığım, bilgi ve deneyimlerinden faydalandığım tüm hocalarıma ve birlikte çalıştığım tüm arkadaşlarıma tek tek teşekkür ederim.

Doktora tezimin ilerlemesini titizlikle takip eden ve değerli katkılarını sunan Doç. Dr. Burcu Atar ve Doç. Dr. Ergül Demir; tez jürimde yer alan ve önemli katkılar sunan Prof. Dr. Nükhet Çıkrıkçı ve Prof. Dr. Nilüfer Kahraman hocalarıma en içten teşekkürlerimi sunarım.

Doktora yolculuğuna birlikte başladığımız, desteğini her zaman hissettiğim ve fikir danıştığım sevgili arkadaşım Melike Özer'e bu uzun yolculukta bana yoldaş olduğu için çok teşekkür ederim.

Doktora eğitimim boyunca, görev yaptığım Boğaziçi Üniversitesi, Matematik ve Fen Bilimleri Eğitimi Bölümündeki tüm hocalarım ve araştırma görevlisi arkadaşlarıma süreç boyunca bana gösterdikleri akademik katkı, manevi destek ve hoşgörü için sonsuz teşekkürlerimi sunarım.

Tüm eğitim hayatım boyunca bana inanan ve desteklerini esirgemeyen, tezimin son aşamaya gelmesinde önemli katkıları olan canım annem İnci Ertaş, canım babam Recep Ertaş'a; beni her zaman rahatlatan ve yanımda olduklarını her zaman hissettiğim kardeşlerim Bulut Ertaş ve İlke Cambazoğlu'na çok teşekkür ediyorum. İyi ki varsınız.

Bu süreç boyunca beni hep destekleyen, çıkmaza düştüğümde beni ayağa kaldıran sevgili eşim, can yoldaşım Haluk Polat'a tüm şefkat ve fedakârlıkları için teşekkür ederim. Yaşama sevincim, motivasyon kaynağım biricik oğlum Can Deniz, iyi ki varsın.

Bu zorlu yolculuğun en başından beri birbirimize destek olduğumuz dostlarım Nezih'e ve bu yolculukta ruhumu besleyen AN Vokal'e ayrıca teşekkürlerimi sunarım.

İçindekiler

Öz	ii
Abstract	iii
Teşekkür	iv
Tablolar Dizini	vii
Şekiller Dizini	viii
Simgeler ve Kısaltmalar Dizini	ix
Bölüm 1 Giriş	1
Problem Durumu	1
Araştırmanın Amacı ve Önemi	8
Araştırma Problemi	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar	10
Madde Tepki Kuramı	10
Bilgisayar Ortamında Bireye Uyarlanmış Testler (Computer Adaptive Testing)	16
Çok Aşamalı Bireye Uyarlanmış Testler (Multi Stage Testing)	22
İlgili Araştırmalar	32
Bölüm 3 Yöntem	39
Araştırma Yöntemi	39
Çok Aşamalı Bireye Uyarlanmış Test Simülasyonu	39
Verilerin Analizi	50
Bölüm 4 Bulgular ve Yorum	52
Birinci Alt Probleme İlişkin Bulgular ve Yorum	52
İkinci Alt Probleme İlişkin Bulgular ve Yorum	57
Bölüm 5 Sonuç, Tartışma ve Öneriler	63
Sonuçlar	63
Tartışma	66
Öneriler	70

Kaynaklar.....	73
EK-A1: Modül Bilgi Fonksiyonları (Normal Dağılım)	86
EK-A2: Modül Bilgi Fonksiyonları (Uniform Dağılım)	89
EK-B1: Madde Parametreleri (Normal Dağılım).....	92
EK-B2: Madde Parametreleri (Uniform Dağılım).....	97
EK-C: R Kodları	102
EK-D: Etik Komisyonu Onay Bildirimi	108
EK-E: Etik Beyanı	109
EK-F: Doktora Tez Çalışması Orijinallik Raporu.....	110
EK-G: Dissertation Originality Report	111
EK-Ğ: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	112

Tablolar Dizini

Tablo 1 <i>Simülasyon Koşulları</i>	40
Tablo 2 <i>Modül Madde Güçlük Dağılımları</i>	49
Tablo 3 <i>Madde Güçlükleri Normal Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerleri</i>	53
Tablo 4 <i>Madde Güçlükleri Normal Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları</i>	56
Tablo 5 <i>Madde Güçlükleri Uniform Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerleri</i>	58
Tablo 6 <i>Madde Güçlükleri Uniform Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları</i>	61

Şekiller Dizini

Şekil 1. Madde karakteristik eğrisi	11
Şekil 2. Test bilgi fonksiyonu ve standart hata grafiği (DeMars, 2010, ss.92).....	15
Şekil 3. Madde bilgi fonksiyonları (De Ayala, 2008, ss.102)	16
Şekil 4. Bilgisayar ortamında bireye uyarlanmış test döngüsü.....	17
Şekil 5. İki aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-3)	22
Şekil 6. Üç aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-2-3)	23
Şekil 7. Üç aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-3-3)	23
Şekil 8. 1-2-3 tasarımı çoklu paneller	25
Şekil 9. 1-2-3 tasarımı modül bilgi fonksiyonları	28
Şekil 10. 1-2-3 tasarımında izlenebilecek yollar.....	29
Şekil 11. Birey odaklı statik yönlendirme kuralı.....	30
Şekil 12. Grup odaklı statik yönlendirme kuralı.....	31
Şekil 13. Dinamik yönlendirme kuralı.....	32
Şekil 14. 1-2-3 panel tasarımı	44
Şekil 15. Statik yönlendirme hesaplama örneği	47
Şekil 17. Normal dağılımdan türetilmiş modül tasarımları için RMSE değerleri	54
Şekil 18. Normal dağılımdan türetilmiş modül tasarımları için yanlılık değerleri	55
Şekil 19. Uniform dağılımdan türetilmiş modül tasarımları için RMSE değerleri	59
Şekil 20. Uniform dağılımdan türetilmiş modül tasarımları için yanlılık değerleri	59

Simgeler ve Kısaltmalar Dizini

BOBUT: Bilgisayar Ortamında Bireye Uyarlanmış Test

CPA: The Uniform Certified Public Accounting Examinations (Yeminli Mali Müşavirlik Sınavı)

ETS: Educational Testing Service (Eğitsel Test Hizmetleri)

GRE: Graduate Record Examination (Lisansüstü Kayıt Sınavı)

LSAT: Law School Admission Test (Hukuk Fakültelerine Kabul Sınavı)

MEB: Milli Eğitim Bakanlığı

MST: Multi Stage Testing

MTK: Madde Tepki Kuramı

OECD: Organization of Economic Cooperation and Development (Ekonomik İşbirliği ve Kalkınma Örgütü)

ÖSYM: Ölçme Seçme ve Yerleştirme Merkezi

PISA: The Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)

PIAAC: The Programme for the International Assessment of Adult Competencie (Yetişkin Becerileri Araştırması)

RMSE: Root Mean Squared Error (Hata Kareleri Ortalamasının Karekökü)

TOEFL: Test of English as a Foreign Language (Yabancı bir Dil olarak İngilizce Testi)

Bölüm 1

Giriş

Eğitimde ölçme ve değerlendirme amacıyla kullanılan testler, bireyler ve gruplar hakkında nesnel veriler toplayarak öznel izlenimlerle beraber daha iyi, daha güvenilir ve hesap verilebilir kararlar almak için kullanılan araçlar olarak tanımlanmaktadır (AERA, APA, & NCME, 2014; Kubiszyn & Borich, 2013; Woolfolk, 2011). Günümüzde testler, eğitim alanında birçok farklı amaç için kullanılabilir. Bireylerin genel başarısını ve gelişimin değerlendirmek, güçlü ve zayıf yönlerini teşhis etmek, onları uygun bir eğitim programına yerleştirmek, sertifika veya diploma vermek veya bireylerin eğitimi ile ilgili müdahalede bulunmak için de farklı testler uygulanabilmektedir (AERA, APA, & NCME, 2014; Kubiszyn & Borich, 2013).

Uzun yıllar boyunca testler çoğunlukla kağıt-kalem testi olarak uygulansa da kişisel bilgisayarların hayatın bir parçası olmasıyla beraber testler de bilgisayar ortamına taşınmıştır. Bilgisayar yazılımlarının istatistiksel analiz ve performans değerlendirme süreçlerini kolaylaştırması sayesinde klasik testler de yerini bilgisayar ortamında bireye uyarlanmış testlere bırakmaktadır (van der Liden ve Glass, 2010). Günümüzde birçok geniş ölçekli test uygulaması bilgisayar ortamında bireye uyarlanmış çok aşamalı test olarak uygulanmaktadır (Breirhaupt, Zhang ve Hare, 2014; Educational Testing Service, 2019; Kirsch & Lennon, 2017; Robin, Stefan ve Liang, 2014). Bireye uyarlanmış çok aşamalı testlerin ölçme, değerlendirme ve uygulamadaki avantajları düşünüldüğünde test uygulayıcılarının test modellerini çok aşamalı bireye uyarlanmış test olarak yapılandırması beklenen bir durumdur (Wang, 2017). Türkiye’de bu alanda yapılacak çalışmaların, özellikle merkezi olarak uygulanan sınavların yeniden yapılandırılmasına katkı sağlayabileceği düşünülmektedir.

Problem Durumu

Tüm dünyada olduğu gibi Türkiye’de de bazı önemli kararların alınması için testler birer gereklilik ve ön koşul olarak görülmektedir. Bu amaçla yapılan merkezi değerlendirme, seçme ve yerleştirme odaklı sınavlar, Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) ve Milli Eğitim Bakanlığı (MEB) tarafından yılın belli zamanlarında uygulanmaktadır. Özellikle ortaöğretim ve yükseköğretime geçiş için yapılan sınavlar ülke gündeminde de her zaman önemli yer tutmuş ve zaman zaman eğitim politikalarına yön vermiştir. Bu sınavlarla ilgili yapılan değişiklikler sadece sınava

girecek öğrencileri değil onlarla beraber aileleri, öğretmenleri, idarecileri ve eğitim politikaları geliştiricileri de kapsayan milyonlarca insanı etkilemektedir. Milyonlarca kişi bu sınavlardan aldıkları sonuçlara göre değerlendirilmekte ve hayatlarına yön vermektedir. “Bireysel düzeyde, bir öğrencinin eğitim süreci ya da seçimleri testten aldığı performansa göre doğrudan etkileniyorsa; örneğin bir öğrencinin sınıf atlaması veya kalması, mezun olması ya da istenilen bir programa kabul edilip edilmemesi gibi durumlarda kullanılan testler, *high stakes – yüksek beklentili* olarak tanımlanmaktadır” (AERA, APA, & NCME, 1999, p.139). Türkiye’de de ÖSYM ve MEB tarafından yapılan ortaöğretime ve yükseköğretime geçiş sınavları da yüksek beklentili testler olarak tanımlanabilir. Yükseköğretime geçiş sınavları 1974’te ÖSYM’nin kurulması ile beraber merkezi olarak yapılmaya başlanmıştır. Bu sınavlarda amaç, başvurulacak programda başarılı olma olasılığı en yüksek öğrencileri seçmek ve yerleştirmektir. Okullara yerleştirilecek öğrencilerin sayısı başvurulacak kontenjanlardan çok daha fazla olduğu için seçme ve yerleştirme yapılması kaçınılmaz olmaktadır. Seçme ve yerleştirme sınavlarının ölçmek istediği becerilerin ne olduğuna ve ne olması gerektiğine ilişkin tartışmalar ile beraber adalet ve güvenilirlik ile ilgili tartışmalar da geçmişten bugüne devam etmektedir (Atkinson, 2001; Zwick, 2006). Bireylerin eğitim hayatlarını etkileyen, önemli kararların alındığı bu yüksek beklentili testlerin adil ve güvenilir olmasının yanında yüksek standartlara sahip olması da gerekmektedir (AERA vd., 1999). Her ne kadar bireyin performansını kusursuz bir şekilde doğru belirlemek mümkün olmasa da burada yapılan ölçmede amaç bireyin puanını belirlerken ve bireyle alakalı önemli kararları verirken hatayı en aza indirmektir.

Uzun yıllardır eğitim alanında yapılan sınavlar çoğunlukla kâğıt kalem testlerine ve performans değerlendirmesine dayanmaktadır (van der Linden & Glas, 2010). Teknolojinin gelişmesi ve kişisel bilgisayarların hayatımıza girmesi ile beraber eğitim alanında da testlerin gelişiminde bilgisayarlar önemli yer tutmaya başlamıştır. Testlerin bilgisayar ortamına taşınması öncelikle kâğıt israfını ortadan kaldıran, puanlamayı ve değerlendirmeyi kolaylaştıran hızlı bir yöntem olsa da bilgisayar ortamında uygulanan testlerin kâğıt-kalem testleri karşısında etkinliğinin araştırıldığı çalışmalar (Akdemir & Oguz, 2008; Aybek & Demirtaşlı, 2014; Clariana & Wallace, 2002; Kim & Huynh, 2007; McClelland & Cuevas, 2020; Piaw Chua, 2012; Prisacari & Danielson, 2017; Rotou, Patsula, Steffen & Rizavi, 2007; Reckase, Ju & Kim, 2019) halen yapılmaktadır. Kâğıt-kalem yerine aynı testin bilgisayar ortamına taşınması ilk aşamadır ve burada sadece

testi uygulama şekli değişmiştir. Bireylerin hepsi yine aynı testi alırlar, ancak maddeler ve değerlendirme bilgisayar ortamına taşınmıştır ve uygulama bilgisayar ortamında yapılmaktadır. Testlerin bilgisayar ortamına taşınması sonrasında ikinci aşamada paradigma değişir ve her bireye aynı maddelerin uygulandığı klasik *lineer testler (linear tests)* yerine her bireyin yeteneğine göre uyarlanmış maddelerin uygulandığı *bireye uyarlanmış testler (adaptive tests)* ortaya çıkar (Schaeffer, Bridgeman, Gloub-Smith, Lewis, Potenza & Steffen, 1998). 1980'ler sonrasında kullanılmaya başlayan ve hala gelişmekte olan Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT) eğitim alanındaki önemli dönüm noktalarındandır (Murphy & Davidshofer, 2005; van der Linden & Glas, 2010). Günümüzde hala kâğıt-kalem testleri geçerliliğini korumakta ve birçok yüksek beklentili test bu şekilde uygulanmaktadır. Kâğıt-kalem testleri için de hala bilinmeyen araştırılmaya devam edilen konular olsa da çok kez denenmiş ve doğruluğu kabul edilen madde seçimi ve yetenek kestirimi yöntemleri olduğu bilinmektedir (van der Linden & Pashley, 2010). Ancak kâğıt-kalem testlerinin de bazı sınırlılıkları vardır ve bu sınırlılıklar bireye uyarlanmış testler ile ortadan kalkmaktadır. Tüm dünyada geçerliliği olan ve birçok ülkede uygulanan GRE (Graduate Record Examination) ilk olarak 1991 de bilgisayar ortamına taşınmış, 1993'te BOBUT olarak uygulanmaya başlanmıştır. Günümüzde ise GRE çok aşamalı bireye uyarlanmış test olarak bilgisayar ortamında uygulanmaktadır.

Türkiye'de ÖSYM ve MEB tarafından yapılan sınavlar kâğıt-kalem testi şeklinde uygulanmakta sadece ÖSYM tarafından uygulanan YDS (Yabancı Dil Bilgisi Seviye Tespit Sınavı) hem kâğıt-kalem testi olarak hem de bilgisayar ortamında uygulanmaktadır. Klasik lineer test formatında uygulanan bu sınavlarla ilgili yapılan eleştiriler kâğıt kalem testlerinin sınırlılıklarından kaynaklanmaktadır. En çok dikkat çekilen konulardan biri sınavın zorluk düzeyi üzerinedir. Madde güçlüğü ile öğrencilerin yetenek düzeylerinin eşleşmiyor oluşu ölçmeyi sınırlandırmaktadır. Sınava giren tüm bireyler yetenek düzeylerinden bağımsız olarak aynı testi alırlar. Bu yüzden test geliştiricilerin tercihi de kolay, orta ve zor güçlükte maddeleri içeren her yetenek düzeyini kapsayabilecek bir test hazırlamak yönünde olmaktadır. Dolayısı ile bu durum testlerdeki madde sayısını arttırır. Farklı yetenek düzeylerindeki öğrencilerin tek bir test ile ölçülmesi test puanlarının güvenilirliğine bir tehdit oluşturabilir. Bireyin yetenek düzeyine uygun olmayan maddelerle yapılan bir ölçme bireyin sıkılmasına ve yorulması da yol açarken yetenek düzeyine ilişkin de daha az bilgi verir. Bu sınırlılıkları

ortadan kaldırmak için bireye uyarlanmış testler kullanılmaktadır. Bireye uyarlanmış testlerde yapılmaya çalışılan test maddelerinin güçlüklerini bireyin yetenek düzeyleri ile eşleştirmektir (Hambleton & Swaminathan, 1985). Geleneksel lineer testler ile bireye uyarlanmış testler arasında test düzeni, test maddelerinin güçlük düzeyi, test uzunluğu, test zamanı, test organizasyonu ve test puanı hesaplama açısından farklılıklar vardır. Geleneksel testlerde her birey ortalama güçlükte aynı test maddeleri ile karşılaşırken, bireye uyarlanmış testlerde her birey yetenek düzeyine göre farklı güçlüklerde farklı test maddeleri ile karşılaşır. Bu sayede bireye uyarlanmış testler, geleneksel testlere göre hem daha az madde ile sonuca ulaşabilir hem de her bir bireye herhangi bir zamanda uygulanabilir. Geleneksel testlerde ölçme kesinliği en yüksek istenilen ölçme grubunun ortalama performans gösteren bireylerinin puanlarında görülür ve gruba bağımlılık söz konusudur. Öte yandan bireye uyarlanmış testlerde ölçme herkes için aynı kesinlikte ve gruptan bağımsızdır (Hendrickson, 2007).

Bir BOBUT uygulamasında bilgisayarlar bireyin yeteneğini kestirmek, bireyin kestirilen yeteneğine göre madde seçmek ve seçilen maddeleri bireylere atamak için kullanılır. Böylelikle testler bilgisayar ortamında uygulanarak bireye uyarlanmış olur. Testleri bilgisayar ortamında uygulamak çoğu zaman sınav güvenliğini, test verimliliğini arttıran ve puanlamayı iyileştiren de bir yöntemdir (Sireci, 2007). Test organizasyonu geleneksel testlere göre bireye uyarlanmış testlerde daha karmaşık olsa da BOBUT ile testin uygulanması için harcanan zaman azalır ve puanlama anında yapılabilir. BOBUT uygulamasında madde seçimi ve yetenek kestirimi döngüsel olarak aynı anda gerçekleşmektedir. Kişinin maddeye verdiği cevaba göre yetenek kestirilir, kestirilen yetenek düzeyine göre de yeni madde seçilir. Bu yüzden yetenek kestiriminin verimliliği ve kesinliği birey için doğru maddenin seçilmiş olmasıyla ilişkili iken, maddenin bireye ne kadar uygun olduğu da yetenek kestiriminin kalitesine bağlıdır (van der Linden & Pashley, 2010). Madde seçimi ve yetenek kestirimi arasındaki bu döngüsel ilişki, madde seçim ve yetenek kestirim yöntemlerinin iyi ve doğru seçilmesini gerektirir. Test geliştiriciler bir BOBUT uygulaması tasarlarlarken, ilk ve test boyunca yapılan yetenek kestirimlerinin nasıl hesaplanacağına ve maddelerin nasıl seçileceğine en baştan karar vermek zorundadırlar (van der Linden & Pashley, 2010).

BOBUT uygulamasının kâğıt-kalem uygulamalarına göre avantajları olsa da bazı sınırlılıkları ve dezavantajları da bulunmaktadır. BOBUT uygulamasında maddeler tek tek bireyin bir önceki maddeye verdiği cevaba göre otomatik olarak gelir ve test

bütünü uygulama anında otomatik olarak oluşur. Bu yüzden test geliştiricilerinin testin bütününe önceden oluşturma ve kontrol etme şansı yoktur. Bu durum ölçülmek istenen içerik alanının kontrolünü engelleyen bir sınırlılıktır. Madde düzeyinde otomatik olarak sırasıyla gelen maddeler yazılımsal açıklar ve tasarım kusurlarından da kaynaklanan sınırlılıklarla da içerik bütünlüğü anlamında her zaman en iyi sonuçları vermeyebilir (Mead, 2006). Bunun yanı sıra maddelerin hangi sıra ile geldiği kontrol edilemediğinden bir uygulama içinde birbirine ipucu olabilecek maddelerin kontrolü de yapılamamaktadır. Madde düzeyinde bireye uyarlanmış testlerde testi alan bireyleri etkileyen bir diğer sınırlılık da bir sonraki maddeyi görmek için uygulanan maddenin cevaplanması zorunluluğudur. Klasik bir BOBUT uygulamasında bireyin bir maddeyi boş bırakıp sonra geri dönüp cevaplaması ya da emin olmadığı bir sorunun cevabını değiştirmesi gibi bir durum söz konusu değildir. Bir sonraki maddeyi görebilmesi için önce bireyin gelen soruyu cevaplaması gerekmektedir. Otomatik olarak gelen maddelerin sebep olabileceği bir diğer sorun da madde ifşasıdır. Test güvenliğinin sağlanması için madde ifşasının kontrolüne ihtiyaç vardır (Hendrickson, 2007).

Yukarıda söz edilen BOBUT uygulamasının sınırlılıklarını ortadan kaldırmak veya iyileştirmeye yönelik Çok Aşamalı Testler (Multistage Testing-MST) bireye uyarlanmış olarak kullanılmaya başlanmıştır. Çok aşamalı test fikri esasında bireye uyarlanmış test fikrinin de öncesine dayanmaktadır ancak BOBUT ile beraber de başka bir noktaya taşınmıştır (Mead, 2006; Yan, Lewis, & von Davier, 2014). Çok aşamalı bireye uyarlanmış testlerin uygulaması ve bu konuda yapılan çalışmalar günden güne artmaktadır (Zheng, Nozawa, Gao, & Chang, 2012). Çok aşamalı testlerin popülerliğinin artmasında hem geleneksel testlerin hem de bireye uyarlanmış testlerin avantajlarına sahip olması önemli bir etken olabilir. Çok aşamalı bireye uyarlanmış test uygulamalarında bireyler, yetenek düzeylerine göre maddelerle karşılaşır ancak burada uyarılma madde düzeyinde değil, bireyin yetenek düzeyine uygun birçok maddenin yer aldığı madde blokları düzeyinde olur. Bu sayede de içeriği kontrol edememe, birbirine ipucu olabilecek maddelerle karşılaşma, soruya geri dönememe gibi maddelerin otomatik birbiri ardına gelmesinden kaynaklanan sınırlılıklar ortadan kalkmış olur.

Çok aşamalı bireye uyarlanmış testlerde madde blokları aşamalı olarak uygulanmak üzere önceden hazırlanır. Böylece bir bireyin geçeceği olası aşamalar, bir başka deyişle bireyin izleyeceği olası yollar test başlamadan önce de bellidir. Böylece

bireyin tamamlayacağı testin bütünü önceden kontrol edilebilmektedir. Bu özelliği ile çok aşamalı bireye uyarlanmış testler madde düzeyinde BOBUT'un içerik dengesi ile ilgili sınırlılığını ortadan kaldırmaktadır. Test uygulanmadan önce uzmanlar ve test geliştiriciler tarafından test bütünü'nün kontrolleri yapılabilir. Ayrıca maddeler otomatik olarak gelmeyeceğinden birbirine ipucu oluşturabilecek maddelerin olup olmadığı da böylelikle kontrol edilebilir. Çok aşamalı testlerde aşamalar durağan olduğu için bireylerin aşama içinde bir maddeye geri dönme, cevabını kontrol etme şansı da vardır. Bu da bireyin boş bırakıp soruya geri dönme şansını da korumaktadır.

Çok aşamalı bireye uyarlanmış testler madde düzeyinde BOBUT'un sınırlılıklarını ortadan kaldırarak, özellikle uygulama sürecinde madde düzeyinde BOBUT'a göre birçok avantaj sağlamaktadır. Bu yüzden çok aşamalı bireye uyarlanmış test uygulamalarının daha kolay uygulanabilir ve yüksek beklentili testler için daha uygun olduğu da söylenmektedir (Wang, 2017). Ayrıca bireyler, geleneksel test formuna yakın, aşına oldukları bir tarzda maddeleri test blokları halinde görürler ve o test bloğu içerisinde boş bıraktıkları maddelere geri dönme, cevaplarını tekrar düzenleme gibi olanaklara sahiptirler. Bu yüzden çok aşamalı bireye uyarlanmış test uygulamaları, sınavı alan bireylere daha aşına, daha doğal ve daha uygun geldiği için test boyunca bireylerin kaygı ve stresi de madde daha az olabilmektedir (Zheng, Nozawa, Gao, & Chang, 2012). Çok aşamalı bireye uyarlanmış test uygulamaları tüm bu özellikleri ile bireylerin yadırgamadıkları ve tümüyle kabul edebildikleri uygulamalardır (Melican, Breithaupt, & Zhang, 2010).

Geleneksel testlerde bireylerin performansının en doğru kestirildiği nokta grubun ortalama performansının olduğu nokta olurken bireye uyarlanmış testlerde bireylerin yetenek kestirimi her birey için aynı kesinlikte olmaktadır (Hendrickson, 2007). Madde düzeyinde bireye uyarlanmış test uygulamalarının uygulama öncesi, uygulama sırası ve sonrasındaki sınırlılıklarını ortadan kaldıran avantajları düşünüldüğünde çok aşamalı bireye uyarlanmış testlerin yüksek beklentili testlerin uygulamalarında kullanılması hem test geliştiriciler hem de sınavı alan bireyler için olumlu sonuçlar doğuracağı açıkça görülebilmektedir.

Bireye uyarlanmış testlerin kâğıt kalem testlerine göre birçok avantajının bulunması Türkiye'de de araştırmacıları bu alanda çalışmalar yapmaya yönlendirmiştir. Bu alanda yapılan çalışmaların bir kısmı kâğıt-kalem testlerinin BOBUT uygulanabilirliğini araştırmaktadır (Aybek, 2016; Bulut & Kan, 2012; Özbaşı &

Demirtaşlı, 2015) bir kısmı da farklı BOBUT uygulama yöntemlerini karşılaştıran çalışmalarıdır (Bozunç Öztürk, 2014; Eroğlu, 2013; Kalender 2011; Kezer 2013; Özdemir, 2015). Bu çalışmalar ayrı ayrı BOBUT uygulanabilirliğini ortaya koymaktadır.

Aynı bilgisayar ortamında bireye uyarlanmış testler gibi, çok aşamalı testler de uygulama anında bireyin yeteneğine göre testin uyarlandığı karmaşık bir test tasarımını gerektirmektedir. Çok aşamalı bireye uyarlanmış testler (Multistage Testing) hem geleneksel kağıt-kalem testlerinin hem de bilgisayar ortamında bireye uyarlanmış testlerin bazı test tasarım özelliklerini içerdiğinden, iki uygulamanın bir birleşimi olarak görülebilir. Bu sayede çok aşamalı bireye uyarlanmış test uygulaması her iki uygulamanın da avantajlarından faydalanmaktadır. Bu özellikleriyle son yıllarda teknolojinin de gelişmesi ile beraber test endüstrisinde çok aşamalı bireye uyarlanmış testlere olan ilgi de giderek artmaktadır (Magis, Yan & von Davier, 2017).

Uygulama esnasında bireyin performansına göre karşısına yeni maddelerin geliyor olması çok aşamalı bireye uyarlanmış testler ve BOBUT için benzerdir. Ancak, BOBUT uygulamasında birey, her bir maddeye verilen cevaba göre yeni bir madde ile karşılaşırken çok aşamalı bireye uyarlanmış test uygulamasında bireyden, modül olarak adlandırılan birçok maddeden oluşan madde bloğunu cevaplaması istenir. Bireyin, bir modülde yer alan maddelere verdiği cevaplara ilişkin performansı bir sonraki aşamada karşılaşacağı yeni madde bloğunu yani yeni modülü belirler. Test uygulaması, yönlendirme modülü olarak adlandırılan bir grup maddenin tüm bireylere uygulanması ile başlar. Sonra her bir bireyin yönlendirme modülündeki maddelere verdiği cevaplara göre performansı hesaplanır ve yönlendirme yöntemine göre belirlenen eşik değer (kesme puanı ya da teta düzeyi) ile karşılaştırılır. Eğer bireyin performansı belirlenen eşik değerden daha yüksek ise bireye daha zor modül uygulanır; eğer düşükse bireye daha kolay modül uygulanır.

Çok aşamalı bireye uyarlanmış testlerin bu işleyişi göz önüne alındığında, uygulamanın esnekliği, bireye uyarlanmış yapısı ve pratikliği ile geleneksel kağıt-kalem testleri ve madde düzeyinde BOBUT arasında dengeli model olduğu söylenebilir. Bir çok aşamalı bireye uyarlanmış test uygulaması, test uzunluğu açısından geleneksel kağıt kalem testinden daha kısadır ve bireyin yeteneğinin ölçülmesi açısından da bir BOBUT uygulaması kadar etkilidir (Magis, Yan & von Davier, 2017).

Araştırmanın Amacı ve Önemi

Bu çalışma, yüksek beklentili olarak tanımlayabileceğimiz yüksek öğretime giriş sınavının bir alt testi olabilecek çok aşamalı bireye uyarlanmış test tasarımlarından elde edilen yetenek kestirimlerini farklı modül uzunlukları, farklı yönlendirme yöntemleri ve farklı son yetenek kestirim yöntemleri ile farklı madde güçlük dağılımları kullanılarak simülatif olarak karşılaştırmayı amaçlamaktadır.

Türkiye’de bireye uyarlanmış test uygulamaları, yüksek beklentili testlerde ne madde düzeyinde ne de çok aşamalı test olarak uygulanmamıştır. Ölçmelerin daha geçerli ve daha güvenilir olabilmesi için, madde düzeyinde BOBUT’un uygulamadaki sınırlılıklarını da ortadan kaldırarak daha kullanışlı olmasını sağlayan çok aşamalı bireye uyarlanmış testlerin farklı yapıdaki uygulamalarının sonuçlarının değerlendirilmesi ve Türkiye’de uygulanabilirliğinin tartışılmasının ölçme ve değerlendirme sürecine önemli bir katkı sağlayacağı düşünülmektedir.

Çok aşamalı test uygulaması genel olarak oldukça esnek gibi görünse de testin verimliliği için testin uygulama amacına da bağlı olarak hangi yapıda olacağı, uygulama boyunca nasıl bir yöntem izleneceği sonuçları da düşünülerek karara bağlanmalıdır (Zenisky, Hambleton, & Luecht, 2010). Çok aşamalı testlerin uygulama öncesi hazırlık süreci bu anlamda oldukça önemlidir. Panel tasarımlarının, modül uzunlukları ile modüllerde yer alacak maddelerin ve yönlendirme yöntemlerinin önceden belirlenip kararlaştırılması gerekmektedir. Testin amacı ve uygulama koşulları düşünülerek çok aşamalı test yapısı oluşturulmalıdır.

Farklı koşullar altında incelenen çeşitli modellerin değerlendirilmesinde, madde tepki kuramı bağlamında yapılan çalışmalarda, simülasyonlar en etkili biçimde en iyi cevabı verebilmektedir (Bulut & Sünbül, 2017; Feinberg & Rubright, 2016; Harwell, Stone, Hsu ve Kirisci 1996). Bu anlamda da farklı ülkelerde yüksek beklenti testler için kullanılan ve alan yazınında güncel olarak yer alan çok aşamalı bireye uyarlanmış testlerin farklı koşullar altında uygulamalarını simülatif veri üreterek incelemeyi ve elde edilen sonuçlar ışığında Türkiye’de uygulanabilirliğini kuramsal olarak tartışmayı amaçlayan bu çalışmanın alana katkı sağlayacağı düşünülmektedir.

Çok aşamalı bireye uyarlanmış testler, BOBUT’un en çok eleştirilen, içerik dengesini kontrol edememe, bir maddeyi boş bırakamama ve bir maddeye geri dönememe sınırlılıklarını ortadan kaldırmaktadır. Özellikle test uzunluğu ile ilgili

kaygıların olmadığı durumlarda BOBUT yerine çok aşamalı test kullanmak kolay uygulanabilir ve yüksek beklentili test uygulamalarına daha uygun olmaktadır (Wang, 2017). Alan yazınında çok aşamalı bireye uyarlanmış testlerle ilgili birçok kuramsal çalışma ve uygulama varken, Sarı, Yahsi-Sarı ve Huggins-Manley (2016) Türkiye’de çok aşamalı bireye uyarlanmış testler ile ilgili henüz yeterli çalışma bulunmadığını ve araştırmacıların bu alanda çalışmalarının önemli olduğunu vurgulamaktadır. Sarı ve diğerlerinin iddialarının üzerinden geçen zamana rağmen, halen bu alanda yapılan çalışmaların sınırlı olması sebebiyle çok aşamalı bireye uyarlanmış test modellerinin incelenmesinin alana katkı sağlayacağı düşünülmektedir. Ayrıca çok aşamalı bireye uyarlanmış test uygulamalarının Türkiye’de uygulanan yüksek beklenti testler ekseninde farklı tasarımlarının karşılaştırılması üzerine henüz çalışılmamış olması da bu çalışmanın özgünlüğünü de ortaya koymaktadır.

Araştırma Problemi

Çok aşamalı bireye uyarlanmış test uygulamasında farklı güçlük dağılımlarına göre oluşturulan üç aşamalı 1-2-3 panel tasarımında; farklı modül uzunlukları (artan, azalan, eşit, ikinci aşama en uzun ilk aşama en kısa, ikinci aşama en uzun son aşama en kısa), farklı yönlendirme (statik ve dinamik) ve son kestirim (MLE ve EAP) yöntemlerine göre elde edilen yetenek kestirimlerinin yanlılık (bias) ve RMSE değerleri nasıl değişmektedir?

Alt problemler.

1. Çok aşamalı bireye uyarlanmış test uygulamasında, madde güçlülüklerinin normal dağılımdan türetildiği 1-2-3 panel tasarımından elde edilen yetenek kestirimlerinin yanlılık (bias) ve RMSE değerleri; yönlendirme yöntemine (statik ve dinamik), modül uzunluklarına (artan, azalan, eşit, ikinci aşama en uzun ilk aşama en kısa ve ikinci aşama en uzun son aşama en kısa) ve son kestirim yöntemine göre (MLE ve EAP) nasıl değişmektedir?
2. Çok aşamalı bireye uyarlanmış test uygulamasında, madde güçlülüklerinin uniform dağılımdan türetildiği 1-2-3 panel tasarımından elde edilen yetenek kestirimlerinin yanlılık (bias) ve RMSE değerleri; yönlendirme yöntemine (statik ve dinamik), modül uzunluklarına (artan, azalan, eşit, ikinci aşama en uzun ilk aşama en kısa ve ikinci aşama en uzun son aşama en kısa) ve son kestirim yöntemine göre (MLE ve EAP) nasıl değişmektedir?

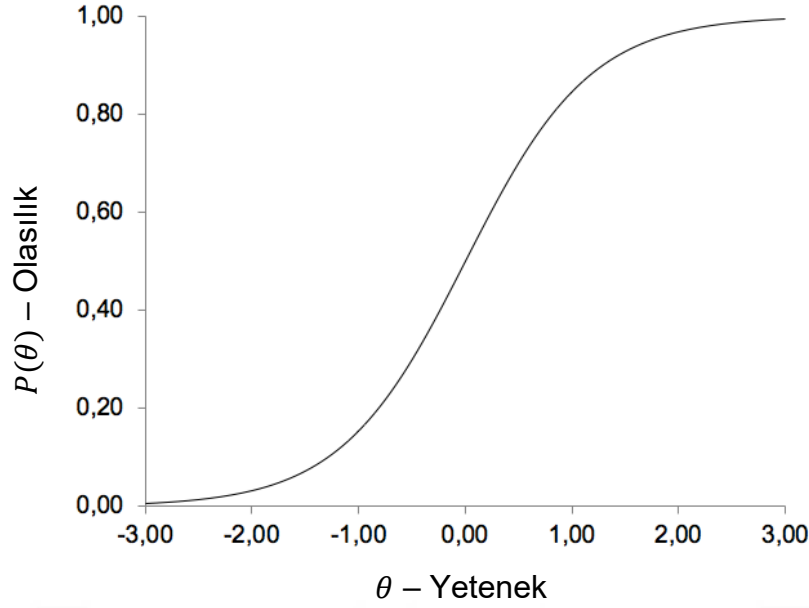
Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu çalışma, çok aşamalı bireye uyarlanmış testlerin farklı tasarım ve yöntemlerden elde edilen yetenek kestirimlerine ait sonuçların simülatif veriler yardımıyla karşılaştırılmasını ve çok aşamalı bireye uyarlanmış testlerin geniş ölçekli ve yüksek beklenti testler ekseninde tartışmayı amaçlamaktadır. Bireye uyarlanmış testlerde bireyin yetenek düzeyine uygun güçlükte madde ile karşılaşması, madde tepki kuramının bireyin yeteneği ile madde güçlüğü aynı ölçekte ele alıyor olması ile sağlanmaktadır. Bu çalışma kapsamında da madde tepki kuramı (MTK) modelleri kullanılarak, çok aşamalı bireye uyarlanmış test tasarımları oluşturulmuştur. Bu yüzden bu bölümde, çalışmanın kuramsal temelini oluşturan MTK hakkında kuramsal bilgiler ile birlikte bilgisayar ortamında bireye uyarlanmış testler (BOBUT) ve çok aşamalı bireye uyarlanmış testlerin geliştirilmesi ve uygulanması ile ilgili bilgiler ve ilgili araştırmalar sunulacaktır.

Madde Tepki Kuramı

Testler yoluyla bireylerin yetenek düzeyleri hakkında çıkarım yapabilmek için bireyin yeteneğinin, bir maddeye verdiği cevabı nasıl etkilediğinin bilinmesi gerekmektedir (Lord, 1980). Madde tepki kuramı (MTK) da bireyin yeteneği (gözlenemeyen özellik) ile bir maddeye verdiği cevap (gözlenen özellik) arasında nasıl bir ilişki olduğunu gösteren matematiksel bir model ortaya koyar (De Ayala, 2008; Hambleton & Swaminathan, 1985; Lord, 1980). Bu ilişki madde karakteristik eğrisi (madde tepki fonksiyonu) olarak isimlendirilen bireyin bir maddeye doğru cevap verme olasılığının bireyin yetenek düzeyine göre bir fonksiyonu olarak tanımlanır (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Lord, 1980). Bu fonksiyon monoton artan bir fonksiyondur. Bireyin yetenek düzeyi olarak isimlendirilen örtük özellik θ ile, ve olasılık da $P(\theta)$ ile sembolize edilir. θ artarken $P(\theta)$ da artar. İki kategorili (doğru yanlış olarak ikili puanlanan) bir maddeye ait madde karakteristik eğrisi Şekil 1'dekine benzer bir yapıda olacaktır. Madde karakteristik eğrisinin şeklini ve konumunu uyumlu olduğu MTK modeline göre madde parametreleri (güçlük, ayırt edicilik ve şans) belirler.



Şekil 1. Madde karakteristik eğrisi

MTK modelleri ile bireyin örtük özelliği - yetenek düzeyi (θ) ve madde parametreleri, n tane bireyin k tane maddeye verdiği cevaplar üzerinden kestirilir ve böylece bir ölçme aracı olarak testin kalitesi değerlendirilerek, farklı koşullarda (farklı bireyler, farklı test maddeleri, farklı ortamlar) yapılacak gelecek uygulamalar için tahminler yapılabilir (Molenaar, 1995). Çünkü bu yetenek ve madde parametresi kestirimleri, uygulanan test maddelerinden, uygulamanın yapıldığı grubun özelliğinden ve sayısından bağımsızdır. θ yetenek düzeyindeki bir bireyin i maddesine doğru cevap verme olasılığı, $P_i(\theta)$, θ yetenek düzeyindeki kişi sayısından ve diğer yetenek düzeylerindeki kişi sayılarından etkilenmez (Lord, 1980). Bu, yetenek parametresinin (θ) değişmezliği olarak tanımlanır (Hambleton, Swaminathan & Rogers, 1991). Aynı yetenek düzeyine sahip fakat farklı gruplardan (ör.: alt grup, üst grup) gelen iki bireyin aynı test maddelerine doğru cevap verme olasılıkları da benzer olacaktır. Bu durum da madde parametrelerinin değişmezliği olarak tanımlanır (Hambleton, Swaminathan & Rogers, 1991; Lord, 1980).

MTK modelleri ile yetenek ve madde parametre kestirimlerinin yapılabilmesi için iki temel varsayımın sağlanması gerekmektedir; *tek boyutluluk* ve *yerel bağımsızlık*. *Tek boyutluluk*, test maddelerinin tek bir örtük özelliği ölçtüğünü söyleyen varsayımdır. Bu varsayımın kontrol edilmesi, her zaman aynı özelliğin ölçüldüğünden emin olmak için önemlidir (Hambleton & Swaminathan, 1985). Pratikte, test performansını etkileyecek başka faktörler de olduğundan bu varsayım kusursuz olarak

sağlanamayabilir ancak test maddeleri tarafından ölçülen baskın bir bileşenin ya da faktörün olması bu varsayımın sağlanması için yeterli olur ve bu faktör test maddelerinin ölçtüğü yetenek olarak tanımlanır (Hambleton & Swaminathan, 1985). *Yerel bağımsızlık* ise aynı özelliği ölçen test maddelerinin birbirinden bağımsız olması varsayımdır. Bireyin yeteneğine göre bir testteki her bir maddeye vereceği cevabın istatistiksel olarak diğer maddelere nasıl cevap verdiğinden bağımsız olmasıdır (De Ayala, 2008; Hambleton, Swaminathan & Rogers, 1991).

Tek boyutluluk varsayımı güçlü bir varsayımdır ve bu varsayımın sağlanmadığı koşullar için çok boyutlu MTK modelleri de geliştirilmiştir (Reckase, 2009). MTK modelleri, iki kategorili ve çok kategorili modeller olarak da sınıflandırılmaktadır. Bu çalışma kapsamında ise iki kategorili maddeler için kullanılan tek boyutlu modeller ele alınacaktır. En yaygın olarak kullanılan iki kategorili maddeler için tek boyutlu modeller bir, iki ve üç parametrelili lojistik modellerdir. Bu çalışmada da iki parametrelili lojistik model kullanıldığından sadece bu lojistik modellere yer verilmiştir.

$j = 1, 2, 3, \dots, n$ bireyinin $i = 1, 2, 3, \dots, k$ maddesine doğru cevap verme ($u_{ji} = 1$ ise doğru $u_{ji} = 0$ ise yanlış cevap) olasılığı her bir MTK modeli için aşağıdaki eşitliklerde gösterilmiştir.

Bir parametrelili lojistik model – 1PLM (Lord & Novick, 1968; Rasch 1960).

$$P_i(\theta_j) = P_i(u_{ji} = 1 | \theta_j, b_i) = \frac{1}{1 + \exp\{-(\theta_j - b_i)\}}$$

İki parametrelili lojistik model – 2PLM (Birnbbaum 1968).

$$P_i(\theta_j) = P_i(u_{ji} = 1 | \theta_j, b_i, a_i) = \frac{1}{1 + \exp\{-a_i(\theta_j - b_i)\}}$$

Üç parametrelili lojistik model – 3PLM (Birnbbaum 1968).

$$P_i(\theta_j) = P_i(u_{ji} = 1 | \theta_j, b_i, a_i) = c_i + (1 - c_i) \frac{1}{1 + \exp\{-a_i(\theta_j - b_i)\}}$$

Bu üç modeli birbirinden ayıran madde parametre sayılarıdır. i maddesine ait a_i , b_i ve c_i parametreleri modeller için belirleyicidir. Sadece 3PLM modelde yer alan c_i , şans parametresi bireyin yetenek düzeyinden bağımsız, en düşük yetenekteki bireylerin maddeye doğru cevap verme olasılığını gösterir. Madde karakteristik eğrisinin başlangıç noktasıdır. 0 ile 1 arasında bir değer alır.

Madde güçlüğü parametresi, b , madde karakteristik eğrisinin yetenek ölçeği boyunca konumunu belirlediği için konum parametresi olarak da isimlendirilir (Lord, 1980). 1PLM (Rach Model) ve 2PLM için b , doğru cevap verme olasılığının 0.5 olduğu θ değerine karşılık gelir. 3PLM'de ise bu değer doğru cevap verme olasılığının $\frac{c}{2} + 0.5$ olduğu θ değerine karşılık gelir. $\theta = b$ noktası da eğrinin dönüm noktası olarak tanımlanır. b parametresi teorik olarak $-\infty$ ile $+\infty$ arasında yer alır, ancak genellikle -3 ile $+3$ arasında değişmektedir. b değeri arttıkça maddenin güçlüğü de artar ve grafik sağa doğru kayar. b değeri azaldıkça maddenin güçlüğü azalır ve grafik sola doğru kayar. $b = 0$ olduğunda ise maddenin orta güçlükte olduğu kabul edilir.

Maddenin ayırt edicilik gücünü gösteren a parametresi, 2PL ve 3PL modellerde yer alır ve eğrinin yetenek ölçeği boyunca nasıl değiştiğini (daha dik / daha düz) gösterir. a parametresi, eğrinin dönüm noktasındaki eğimi ile orantılıdır. Madde ayırt edicilik, a , parametresi, teorik olarak $-\infty < a < +\infty$ aralığında bir değer alabilse de pratikte pozitif değerler alır ve genellikle 2.5'ten küçüktür (Baker & Kim, 2004). a değeri arttıkça eğri dikleşir ve maddenin ayırt edicilik gücü artar, a değeri azaldıkça eğri düzleşir ve maddenin ayırt edicilik gücü azalır.

MTK modellerinde, bir, iki, üç parametrelili lojistik modellerde bireyin bir maddeye doğru cevap verme olasılığı, bireyin yeteneği ve madde parametrelerine bağlıdır. Burada hem bireyin yeteneği hem de madde parametreleri bilinmeyen değerlerdir. Bilinen değer, bireyin test maddelerine verdiği cevapları, test performansıdır. Madde parametreleri bilindiğinde bireyin yetenek düzeyi, test performansına göre kestirilebilir.

n tane birey arasından rastgele seçilmiş bir j bireyinin, k tane maddeden oluşan bir testte i maddesine verdiği cevap doğru ise $u_{ij} = 1$, yanlış ise $u_{ij} = 0$ olarak ifade edilebilir. Böylece j bireyinin testteki tüm maddelere verdiği cevaplar $U_j = (u_{1j}, u_{2j}, \dots, u_{kj} | \theta_j)$ ile gösterilir. Yerel bağımsızlık varsayımı bir bireyin her bir maddeye verdiği cevaplarının, yani u_{ij} değerlerinin istatistiksel olarak bağımsız olduğunu söyler. Bu yüzden bireyin test performansına ait ortak olasılık değeri, olabilirlik fonksiyonundan aşağıdaki eşitlikte gösterildiği gibi olur.

$$P(U_j | \theta_j) = \prod_{i=1}^k P_i(u_{ij} | \theta_j)^{u_{ij}} (1 - P_i(u_{ij} | \theta_j))^{1-u_{ij}}$$

Daha basit ifade edebilmek için $P_{ij} = P_i(u_{ij}|\theta_j)$ ve $Q_{ij} = 1 - P_i(u_{ij}|\theta_j)$ olsun. Bu durumda ortak olasılık değeri aşağıdaki eşitlikte gösterildiği gibi bir olabilirlik fonksiyonu (likelihood) olacaktır.

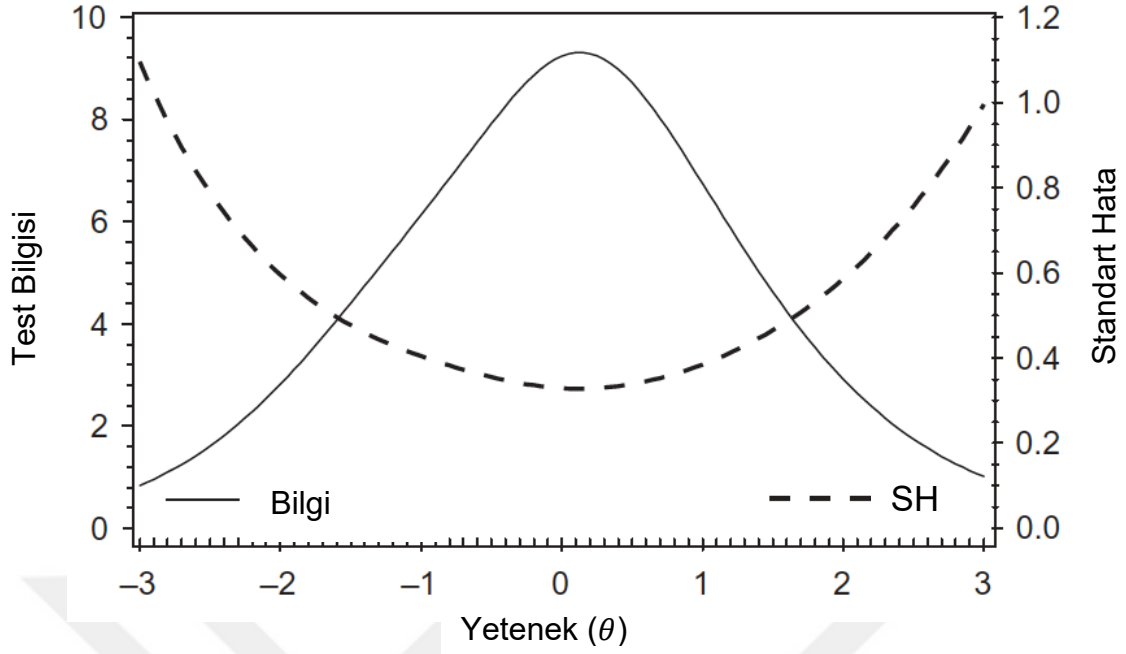
$$P(U_j|\theta_j) = \prod_{i=1}^k P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

Logaritmanın özelliklerinden faydalanarak, olabilirlik fonksiyonunun doğal logaritması alındığında aşağıdaki eşitlikteki log-olabilirlik (log-likelihood) fonksiyona ulaşılır.

$$L = \log P(U_j|\theta_j) = \sum_{i=1}^k u_{ij} \log P_{ij} + (1 - u_{ij}) \log Q_{ij}$$

Bu olabilirlik fonksiyonunun en yüksek değer aldığı θ_j değeri j bireyinin yetenek düzeyi için en çok olabilirlik kestirimi (maksimum likelihood estimate – MLE) olarak tanımlanır. MLE yöntemi yetenek kestiriminde sıklıkla kullanılan bir yöntemdir ancak bireyin test maddelerinin hepsine doğru veya yanlış verdiği durumlarda θ kestirimini $+\infty$ veya $-\infty$ olarak yapar. MLE ile θ kestiriminin yapılmadığı bu durumlarda, θ 'nın sonsal dağılımına dayanan Bayesian kestirim yöntemleri uygulanmaktadır. Bu Bayesian yöntemlerden sıklıkla kullanılanı, beklenen sonsal (Expected a Posteriori – EAP) kestirim yöntemidir (Bock & Mislevy, 1982).

Bir test maddesi veya tüm teste göre yapılan θ kestiriminin varyansı ve standart hatası aynı zamanda, bir maddenin veya testin tüm yetenek kestirimi için ne kadar iyi bilgi verdiği ile ilişkilendirilir (Baker & Kim, 2004). MTK'da bu bilgi miktarı $I(\theta)$ bilgi fonksiyonu ile belirtilir. Standart hata da bilgi fonksiyonu cinsinden $\sqrt{1/I(\theta)}$ olarak ifade edilir. Şekil 2'de yetenek ölçeğine göre test bilgi fonksiyonu ve standart hata grafiklerinin birbirlerine göre konumları verilmiştir.



Şekil 2. Test bilgi fonksiyonu ve standart hata grafiği (DeMars, 2010, ss.92)

Birnbaum (1968) *test bilgi fonksiyonunu* $I(\theta)$ aşağıdaki eşitlikte gösterildiği gibi tanımlamaktadır. $I(\theta)$ bir toplam fonksiyonu olduğu için, her bir maddeye ait *madde bilgi fonksiyonu* $I_i(\theta)$ da aşağıdaki eşitlikte gösterildiği gibi olur.

$$I(\theta) = \sum_{i=1}^k \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

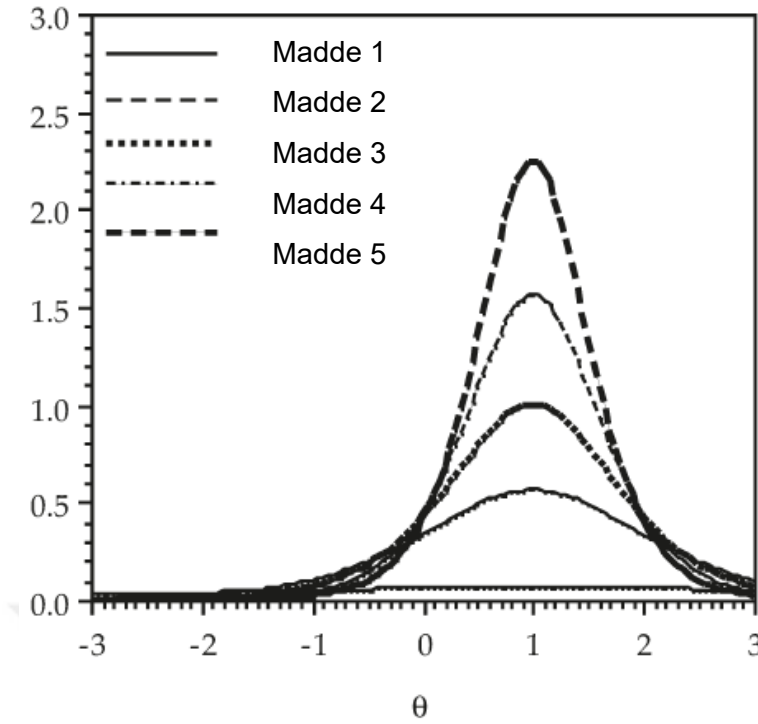
$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} = P'_i(\theta) \left[\frac{P'_i(\theta)}{P_i(\theta)Q_i(\theta)} \right]$$

$$I(\theta) = \sum_{i=1}^k I_i(\theta)$$

$P'_i(\theta)$, madde karakteristik eğrisinin θ noktasındaki eğimi olarak tanımlanabilir. Bu değer de madde ayırt edicilik parametresine karşılık gelmektedir. Bu yüzden 2PLM için madde bilgi fonksiyonu aşağıdaki eşitlikte gösterildiği gibi yazılabilir.

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta)$$

Maddenin ayırt edicilik parametresi arttığında, θ yetenek kestirimi için maddenin verdiği en yüksek bilgi de artar. Şekil 3'te 2PLM örneği olarak; madde güçlük değerleri aynı ($b = 1$), madde ayırt edicilik değerler $a_1 = 0.5$, $a_2 = 1.5$, $a_3 = 2.0$, $a_4 = 2.5$ ve $a_5 = 3.0$ olan beş maddeye ait madde bilgi fonksiyonları verilmiştir.

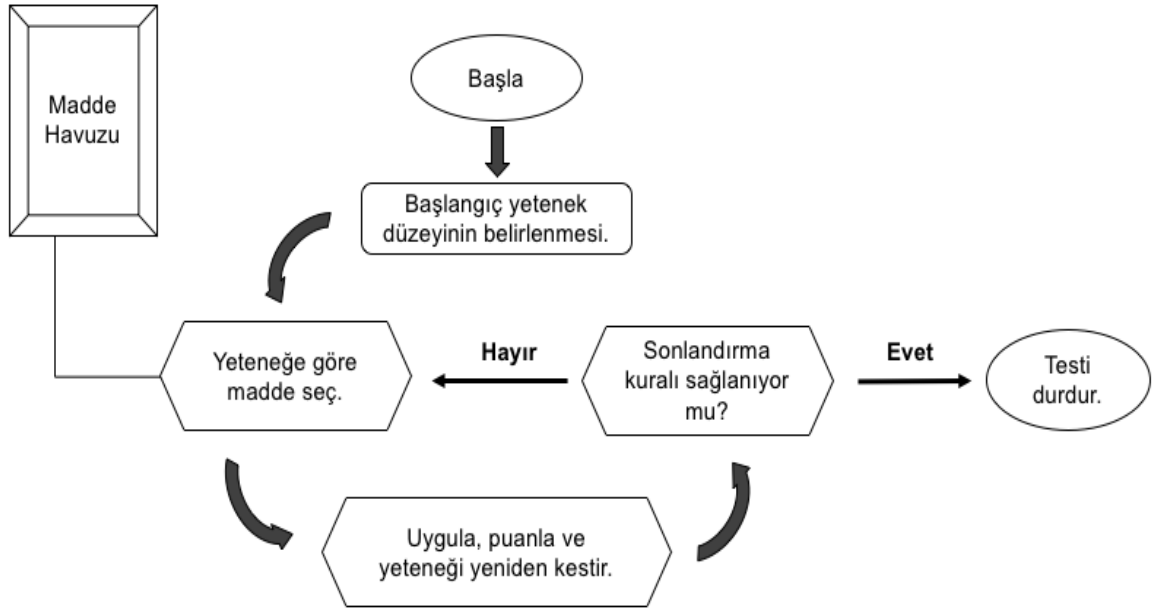


Şekil 3. Madde bilgi fonksiyonları (De Ayala, 2008, ss.102)

Bilgisayar Ortamında Bireye Uyarlanmış Testler (Computer Adaptive Testing)

Günümüzde halen önemli kararların alındığı geniş ölçekli birçok test, kağıt kalem ortamında klasik lineer test olarak uygulanmaktadır. Bu testlerde de amaç bireyin performansını en az hata ile ölçebilmektedir. Bunun için madde analizleri yapılır, test sonuçlarının geçerlik ve güvenirliği için kanıtlar toplanır. Test geliştiriciler, yaptıkları analizler sonucunda testin iç tutarlığını en yüksek düzeyde tutacak maddeleri seçerler ve bu maddeler de genellikle bireylerin ortalama grubuna yönelik orta güçlükte maddeler olur (Hendrickson, 2007; Weiss, 2004). Bu yüzden klasik testler, ortalama yetenek düzeyindeki bireylerin performansını ölçmede daha başarılı iken ortalamanın altında veya üstünde kalan bireyler için bu testler çok zor veya çok kolay olmaktadır (Hambleton & Swaminathan, 1985; Lord, 1980; Wainer, 2000; Weiss, 1983). Her bireyin aynı sıra ve sayıda maddeleri aldığı klasik testlere göre, bireye uyarlanmış testler, daha az soru ile tüm yetenek dağılımı boyunca bireyin yetenek düzeyine ilişkin daha kesin ve etkili ölçme yapabilmektedir (Hendrickson, 2007) .

Bilgisayar ortamında bireye uyarlanmış testlerde, bilgisayarlar maddeleri seçmek ve bireylere atamak için kullanılır. Teste başlamadan önce madde parametreleri belli, geniş bir madde havuzuna ihtiyaç vardır. Bu havuzdan maddeler bireyin yetenek düzeyine göre bilgisayar tarafından seçilir. Bireyin bir önceki maddeye verdiği cevaptaki performansına göre yetenek düzeyi kestirilir ve bireyin karşılaşacağı bir sonraki madde yetenek düzeyine uygun güçlükte, tekrar bilgisayar tarafından seçilir ve uygulanır. Bu süreç belirlenen sonlandırma kuralı sağlanana kadar devam eder (Şekil 4). Böylece her birey performansına göre, yetenek düzeyine uygun farklı bir bireye uyarlanmış bir test almış olur. BOBUT'un klasik testlere göre en büyük avantajlarından biri, bireylerin çok zor ve çok kolay maddeler ile karşılaşmıyor olması, böylece daha kesin bir ölçme yapıyor olmasıdır. BOBUT uygulamasında en az madde ve en az hata ile bireyin yeteneğini kestirmek amaçlanmaktadır.



Şekil 4. Bilgisayar ortamında bireye uyarlanmış test döngüsü

Testlerin bireye uyarlanması fikri madde tepki kuramı öncesinde ortaya çıkmış olsa da bireye uyarlanmış testlerin etkili uygulamaları, madde tepki kuramının (MTK) kullanılması ile gerçekleşmektedir (Weiss, 1983). MTK'nın bireyin yeteneği ile madde güçlüğü aynı ölçekte ele alıyor olması, bireylerin yetenekleri ile karşılaşacakları maddelerin güçlüklerini eşleştirebilmeyi sağlar. Bireye uyarlanmış test uygulamasının ilk örneği Alfred Binet tarafından Stanford Binet Zeka Testi (Binet-Simon ölçeği) uygulamasında görülmektedir (Binet & Simon, 1915). MTK'ya dayanan bireye uyarlanmış testlerle karşılaştırıldığında oldukça basit gibi görünse de Binet'in

uygulanmasında şimdiki bireye uyarlanmış test özelliklerinin tamamı görülmektedir (Weiss, 1983). Burada madde güçlükleri kronolojik olarak yaşlara göre belirlenir. Geniş bir madde havuzundan, bireyin biyolojik yaşına uygun bir madde ile, eğer birey ile ilgili bir ön bilgi varsa (yaşından ileri veya geri) ona uygun bir madde ile teste başlanır. Bireyin maddeye verdiği cevaba göre belirlenen yaş seviyesine uygun yeni maddeler (madde grubu) uygulanır ve aynı anda uygulayıcı tarafından puanlanır. Eğer birey, maddelerin bir kısmını ya da çoğunu doğru cevaplıyorsa daha üst seviyeden bir grup madde ile yanlış cevaplıysa bir alt seviyeden bir grup madde ile devam edilir. Birey verilen madde grubu içerisindeki tüm maddeleri doğru cevapladığında ya da tüm maddeleri yanlış cevapladığında ise test sonlandırılır. Bireyin son puanı da doğru cevaplamış olduğu maddelere göre belirlenir. Şu an bilgisayar ortamında bireye uyarlanmış testlerde de benzer basamaklar izlenmektedir. Şekil 4'te de resmedildiği gibi madde havuzu belirlendikten sonra teste nasıl başlanılacağına, nasıl devam edileceğine (madde seçim ve yetenek kestirim yöntemleri) ve nasıl durdurulacağına karar verilir.

Madde Havuzu. Bireye uyarlanmış testlerde test özelliklerine ve nasıl uygulanacağına karar verildikten sonra madde havuzunun belirlenmesi, geliştirilmesi ya da güncellenmesi gerekmektedir. Bireye uyarlanmış testlerin başarısı, madde havuzunun kalitesi ile doğrudan ilişkilidir. Uygulama ve program ne kadar iyi, ne kadar gelişmiş olsa da sınırlandırılmış bir madde havuzu ya da düşük kalitede maddeler varsa bireye uyarlanmış test doğru işlemez (Flaugher, 2000). Madde havuzuna seçilecek her başarılı madde, bireyin yetenek kestiriminde en yüksek bilgiyi sağlamak ve en az hata ile kestirim yapmak amaçlarına hizmet etmelidir (Veldkamp & van der Linden, 2010).

Thompson ve Weiss (2011), her zaman madde havuzunu yeniden oluşturmaya gerek olmadığını, varolan havuza yeni maddeler eklenerek madde havuzunun genişletilebileceği ve sürekliliğin sağlanabileceğini vurgulamaktadırlar. Bir BOBUT uygulaması için geliştirilecek 500 maddelik madde havuzundaki 400 maddenin uygulama için yeterli bir sayı olduğuna, diğer 100 maddenin de daha sonra dönüştürülebileceğine dikkat çekmektedirler. Madde havuzunu oluşturan maddeler aynı zamanda kullanılacak MTK modelinin varsayımlarını da sağlamalıdır (Falugher, 2000). Maddelerin psikometrik özelliklerinin yanında, ölçülmek istenilen yapının kapsamının belirlenmesi de gerekmektedir. Madde havuzunda her içerik kategorisinden ve farklı

güçlüklerde yeterli sayıda madde bulunmalıdır (Flaugher, 2000; Weiss & Kingsbury, 1984). BOBUT uygulaması öncesinde tüm bu maddelerin test edilmesi ve madde parametrelerinin belirlenmiş olması gerekir. BOBUT uygulamasında bilgisayar bireyin yetenek düzeyine göre en yüksek bilgiyi verecek en iyi maddeyi seçer. Bu yüzden bazı çok iyi maddeler çok fazla bireyin karşısına çıkabilir. Madde ifşası, uygulama sırasında ve sonrasında uygulayıcıların dikkat etmesi gereken önemli bir durumdur.

Başlama Kuralı. Bireye uyarlanmış test uygulamasında, bireyin karşılaşacağı ilk madde bir kestirim yapılmadan uygulanacak maddedir. İlk maddenin bireyin yetenek düzeyinin ne çok üstünde ne de çok altında olmaması gerekir. Bu yüzden bireylerin yetenek düzeylerine ilişkin bir ön bilgiye sahip olunmadığı durumlarda, bireyin yetenek düzeyi orta, $\theta = 0$, kabul edilerek orta güçlükte bir madde ile başlamak en çok tercih edilen yöntemdir. Ancak bu durumda, eğer çoğu birey teste en çok bilgiyi veren aynı madde ile başlarsa bu durum test güvenliğine tehdit oluşturacak madde ifşası açısından ciddi bir sorun teşkil eder (Thissen & Mislevy, 2000). Bu yüzden madde güçlüğü -0.5 ile 0.5 arasından rastgele seçilecek bir madde ile başlamak madde ifşa oranını düşürecek başka bir yöntemdir (Thompson & Weiss, 2011). Bunların dışında bireyin yetenek düzeyine ilişkin bir ön bilgiye sahip olduğu durumda da bireyin yeteneğine uygun güçlükte bir madde ile teste başlanması BOBUT uygulamasının amacı doğrultusunda daha uygun olabilir.

Madde Seçim Yöntemi. BOBUT uygulamasında her aşamada bireyin yeteneğine uygun bir madde, belirlen madde seçim yöntemine göre bilgisayar tarafından seçilir ve bireye uygulanır. Bireye uygulanacak madde, bireyin yetenek düzeyine ne kadar uygunsa, uygulama boyunca yapılan yetenek kestirimleri de o derece iyi olur (van der Linden & Pashley, 2010). Bu yüzden kullanılan madde seçim yöntemi BOBUT uygulamasının etkinliği için önemlidir. Sıklıkla kullanılan, klasik ilk yöntemler, bireyin verdiği cevaba göre kestirilen ara θ değeri için en yüksek bilgiyi ve en düşük beklenen sonsal hatayı sağlayan yöntemlerdir (Thissen & Mislevy, 2000). En yüksek bilgi madde seçim yönteminde, belirlenmiş θ değeri için, $I_i(\theta)$ bilgi fonksiyonunun en yüksek değeri aldığı madde seçilir. Seçilen madde bireye uygulanır ve bu maddeye göre kestirilen θ değeri için yeniden bilgi fonksiyonu hesaplanarak bir sonraki madde (en yüksek bilgiyi veren) seçilir. Test tamamlanana kadar madde seçimi bu şekilde devam eder. En yüksek bilgi yöntemi dışında önsel ve sonsal dağılımları kullanan Bayesian yöntemler de geliştirilmiştir (Owen, 1975; van der Linden 1998). En

yüksek Fisher bilgi (MFI) yöntemi ve önceki sonsal dağılımı her aşamada önsel dağılım olarak kullanan Owen'in (1975) Bayesian yöntemi, klasik yöntemler olarak ele alınmaktadır (van der Linden & Pashley, 2010). Bunun dışında klasik yöntemlere yeni açılımlar getiren ve yeni geliştirilen yöntemler de kullanılmakta ve bu alan, yapılan simülasyon çalışmalarıyla halen gelişmeye devam etmektedir (Thompson & Weiss, 2011; van der Linden & Pashley, 2010).

Yetenek Kestirimi. BOBUT uygulaması boyunca bireyin karşılaştığı her bir maddeye verdiği cevaba göre anında yetenek kestirimi yapılır ve bireyin kestirilen yeteneğine uygun bir madde belirlenen madde seçim yöntemine göre seçilir ve bireye uygulanır. Madde seçim yöntemlerinde olduğu gibi, yetenek kestirimi için de birçok yöntem kullanılmaktadır (van der Linden & Pashley, 2010). Bunların arasında en yaygın olarak kullanılan yöntemler, bir önceki bölümde bahsedilmiş olan en çok olabilirlik kestirimi (maksimum likelihood estimate – MLE) ve Bayesian yöntemlerden biri olan beklenen sonsal (Expected a Posteriori – EAP) kestirim yöntemidir. MLE diğer yöntemler içinde daha az yanlı olanıdır ancak bu yöntemde de cevap örüntüsü içinde an az bir doğru ve bir yanlış cevap olması gerekir (Thompson & Weiss, 2011). EAP yöntemi ise MLE'in bu sınırlılığını ortadan kaldırdığı için sıklıkla tercih edilen Bayesian yöntemidir.

Sonlandırma Kuralı. Madde havuzunda yer alan maddeler, bireyin kestirilen yeteneğine göre sonlandırma kuralı sağlanana kadar bireye uygulanmaya devam eder. Sonlandırma kuralı uygulayıcı tarafından önceden belirlenir. Testin ve uygulamanın amacına göre *test uzunluğu*, *standart hata değeri*, *madde havuzunda kalan maddeler*, gibi ve bunların bazılarının birlikte ele alındığı farklı sonlandırma kuralları kullanılmaktadır.

Test uzunluğuna göre belirlenen sonlandırma kurallarında, sabit bir test uzunluğu belirlenebilir ve her birey için belli bir madde sayısına ulaşıldığında test sonlandırılır. Örneğin, 20 madde tamamlandığında test durdurulur. Bu yöntemi uygulamak daha kolay olsa da son yetenek kestiriminin standart hatası, her birey için farklı olacaktır. Bu da bazı bireyler için aynı kesinlikte kestirimler yapılamadığını gösterir.

Her birey için yetenek kestiriminde standart hatayı belli bir düzeyde tutmak için en düşük standart hataya ulaşıldığında testin durulduğu yöntemler uygulanır. BOBUT uygulamalarında sıklıkla kullanılan bu yöntemde, yetenek kestiriminde standart hata

değeri daha önceden belirlenmiş değerin altına düştüğünde test durdurulur. Örneğin, standart hata 0.30'dan küçük olduğunda test durdurulur. Bu durumda her birey için yetenek kestiriminde belirlenen standart hata değerine ulaşılan kadar birey teste devam eder. Bu durumda, her birey farklı uzunlukta bir test almış olur. Bazı bireyler için bu değere ulaşılan kadar uygulanacak madde sayısı çok fazla olabilir bu durum da testin kullanılabilirliğine zarar verir.

Bir diğer yöntemde ise madde havuzunda birey hakkında yeterli düzeyde bilgi verecek madde kalmadığında test durdurulur. Madde havuzunun sınırlı olduğu uygulamalarda da tüm madde havuzunun uygulanmasıyla test sonlanabilir. Tüm bunların dışında, uygulayıcılar, BOBUT uygulamasında belli bir süre aşıldığında da testi durdurabilirler. Bu uygulamalarda da testin yeterli kesinlikte ölçme yapıp yapmadığı sorgulanmaktadır.

BOBUT uygulamasından en yüksek verimi alabilmek için ne süre sınırı konulmalı ne de belirli madde sayısında dikkate alınması gerekmektedir. Bunlardan biri içerik dengesidir. İçerik dengesinin korunması, testin kapsam geçerliliği için de gereklidir. Birçok testte farklı içeriklerden maddeler yer alır ve bireyin bu farklı içeriklerden maddelerin hepsi ile karşılaşması istenir. Ancak BOBUT uygulamasında maddeler bilgisayar tarafından otomatik olarak seçildiği için, bireyin karşılaşacağı maddeleri önceden tamamen kontrol etmek mümkün değildir. Yazılımsal açıklar ve tasarım kusurlarından da kaynaklanabilecek sınırlılıklar da düşünüldüğünde BOBUT, içerik bütünlüğü anlamında her zaman en iyi sonuçları vermeyebilir (Mead, 2006).

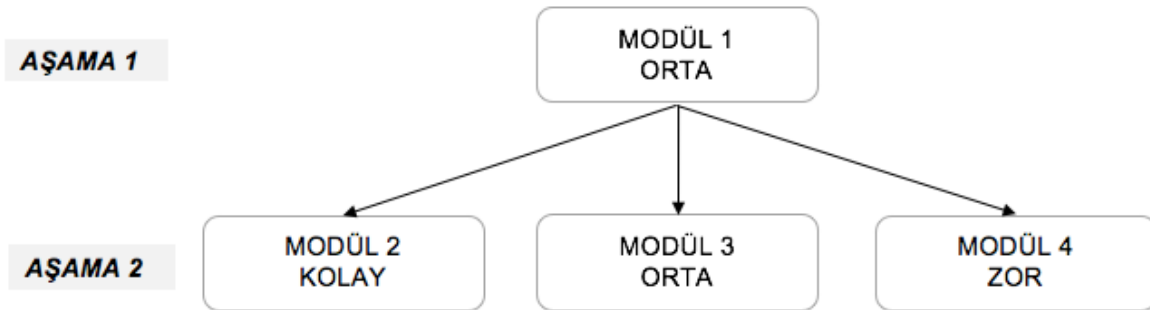
Otomatik madde seçiminin sebep olduğu bir başka sınırlılık da madde ifşasıdır. Bilgisayar otomatik olarak en çok bilgi veren maddeyi seçmeye odaklanmıştır. Hem ilk madde seçiminde hem de diğer maddelerin seçiminde madde ifşası test güvenliğini tehdit eden önemli bir sorundur (Hendrickson, 2007; Thissen & Mislevy, 2000). Maddelerin bireylerin karşısına otomatik olarak tek tek geliyor olması, birbirine ipucu olabilecek maddelerin uygulanmasına ve hatalı ölçümlerin yapılmasına sebep olabilir.

BOBUT'un özellikle uygulamada yaşanan sınırlılıkları, özellikle geniş ölçekli test uygulayıcılarını ve araştırmacıları çok aşamalı bireye uyarlanmış testler (Multistage Testing – MST) üzerinde çalışmaya ve uygulama yapmaya yönlendirmiştir. Bir sonraki bölümde çok aşamalı bireye uyarlanmış testler hakkında detaylı alan yazını sunulmaktadır.

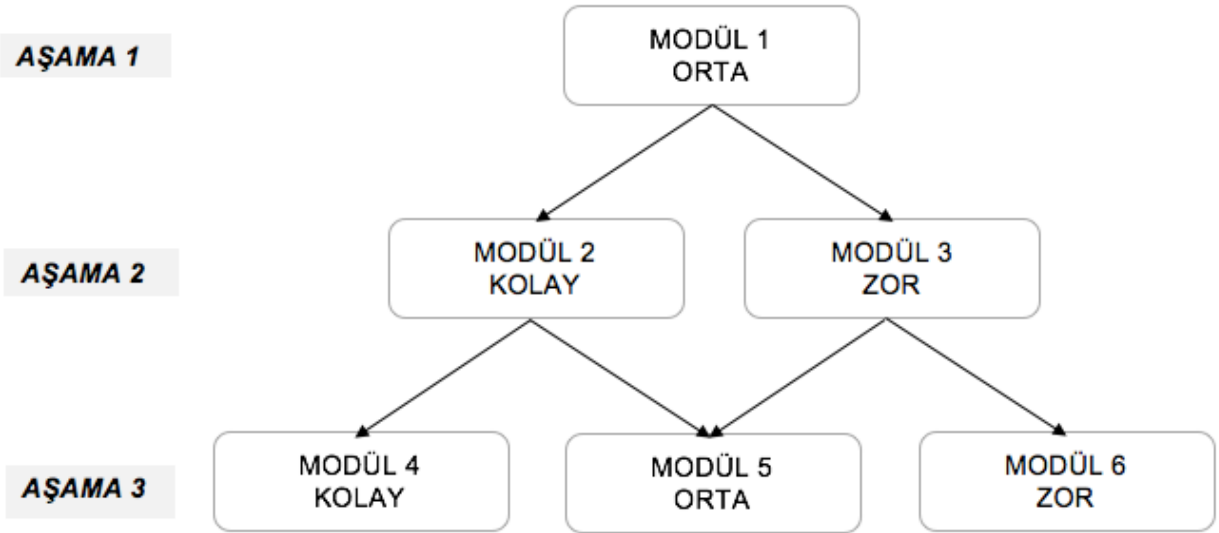
Çok Aşamalı Bireye Uyarlanmış Testler (Multi Stage Testing)

Çok aşamalı bireye uyarlanmış testler, BOBUT uygulamasından farklı olarak bireylerin yeteneklerine göre uygun güçlükte maddelerle tek tek değil belirli ortak özelliklere sahip maddelerden oluşan madde bloklarıyla karşılaştığı uygulamalardır. Çok aşamalı bireye uyarlanmış testlerin günümüzdeki uygulamalarından önceki ilk örnekleri, *iki aşamalı testler (two stages testing)* olarak isimlendirilen uygulamalarda (Lord, 1971; Lord, 1980; Adema, 1990) ve “*testlets*” olarak tanımlanan madde kümlelerinden oluşan bireye uyarlanmış test tasarımlarında görülmektedir. Farklı çalışmalarda farklı terminolojiler kullanılsa da bugün en yaygın olarak Luecht ve Nungester (1998)’in bu madde bloklarını “*modül*” olarak tanımladığı yaklaşım benimsenmiştir (Zenisky, Hambleton & Luecht, 2010).

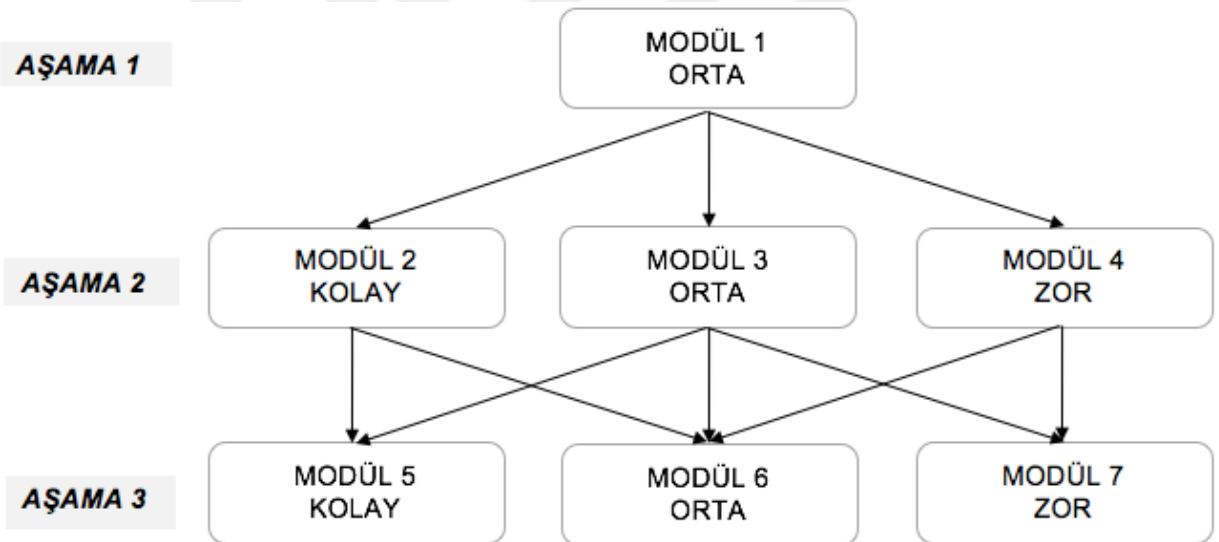
Çok aşamalı bireye uyarlanmış testlerde, bireyin yetenek düzeyine göre belirli sayıda ve özellikteki maddelerden oluşan farklı güçlükteki *modüller*, iki ya da daha fazla *aşamada* uygulanır. İlk aşamada, bireyin modüllerde yer alan tüm maddelere verdiği cevaplara göre birey, ikinci aşamada yer alan uygun modüle yönlendirilir. Benzer şekilde ikinci aşamadan, üçüncü ve varsa sonraki aşamalara bu şekilde bir yol izlenir. Çok aşamalı bireye uyarlanmış bir test oluştururken; aşama sayısı, her bir aşamada yer alacak modül sayısı ve modüllerde yer alacak madde sayıları ile birlikte bireyin performansının nasıl puanlanacağı ve aşmalar arası modüllere nasıl yönlendirileceği önceden tek tek planlanır (Hendrickson, 2007; Yan, Lewis & von Davier, 2014; Zenisky, Hambleton & Luecht, 2010).



Şekil 5. İki aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-3)



Şekil 6. Üç aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-2-3)



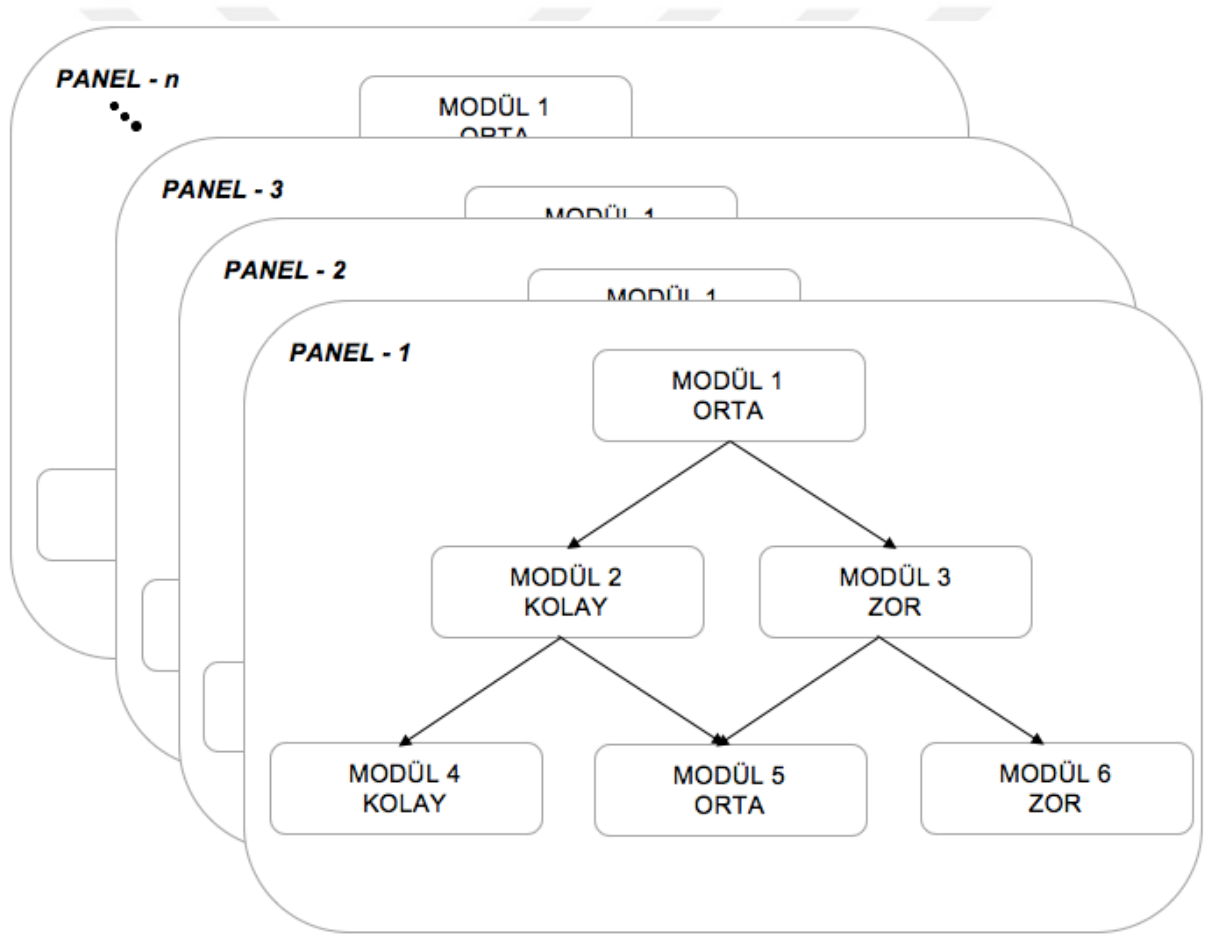
Şekil 7. Üç aşamalı çok aşamalı bireye uyarlanmış test tasarımı (1-3-3)

Çok aşamalı bireye uyarlanmış test tasarımları, aşama sayısı ve her aşamada yer alan modül sayısına göre farklılaşmaktadır. Yapılan çalışmalar ve uygulamalara bakıldığında daha çok iki, üç ve dört aşamalı tasarımların kullanıldığı görülmektedir (Yan vd., 2014). Bireye uygulanan aşamaların ve aşamalardaki modüllerin oluşturduğu bu tasarımlara da *panel* adı verilmektedir. Şekil 5, 6 ve 7'de farklı panel tasarımlarına ait örnekler sunulmuştur.

Test Tasarımı. Şekil 5, 6 ve 7’de de görüldüğü iki ya da daha fazla aşama ve her aşamada farklı sayılardaki modüllerden oluşan farklı test tasarımları olabilmektedir. Örneklerde de görüldüğü gibi test, orta güçlükte bir modül başlar. Tüm bireylerin ortak olarak aldığı bu ilk modüle yönlendirme modülü de denmektedir. Bireyler, birinci aşamadaki modülde yer alan maddelere cevap verdikten sonra bireyin yeteneği kestirilir ve kestirilen yeteneğe göre ikinci aşamada yer alan daha kolay ya da daha zor bir modüle yönlendirilir. İkinci aşamadan üçüncü aşamaya da aynı şekilde birey yönlendirilir ve üç aşamalı bir tasarımda (Şekil 6 ve 7) üçüncü aşamadaki modül de tamamlandıktan sonra bireyin son yeteneği kestirilir. Bireylerin aşama aşamada takip ettikleri modül sırlamasına da yol adı verilir. Birinci aşamada bir, ikinci aşamada iki ve üçüncü aşamada üç modülün yer aldığı bir 1-2-3 panel tasarımında (Şekil 6), bir bireyin izleyebileceği dört yol vardır. Bunlar; Orta-Kolay-Kolay, Orta-Kolay-Orta, Orta-Zor-Orta, Orta-Zor-Zor.

Bireye tanımlanan panelde, bireyin izleyebileceği olası yollar ve alabileceği toplam test bütünü önceden bilinebilmektedir. Modüllerin önceden hazırlandığı bu düzende içerik ile ilgili kontroller yapılabilir ve diğer test istatistikleri önceden hesaplanabilir. Bu sayede test boyunca birbirine ipucu olabilecek maddeler de kontrol edilebilir. Bireyin alacağı olası testlerin önceden hazırlanıyor olması, BOBUT’un içeriği ve madde sırasını kontrol edememe sınırlılığını ortadan kaldıran çok aşamalı bireye uyarlanmış test uygulamalarının en önemli avantajlarından biridir (Hendrickson, 2017; Kim & Plake, 1993). BOBUT uygulamasında, uygulamanın yapısı gereği bireyler karşılaştıkları her maddeyi cevaplamak zorundadır. Bu yüzden bireylerin testi kontrol etme şansı; bir maddeyi boş bırakıp geçip, sonra maddeye geri dönme ya da cevabını değiştirmek için maddeye geri dönme ihtimali yoktur. Bu BOBUT’un testi alan bireyler tarafından en çok eleştirilen önemli bir sınırlılığdır (Visposel, 1998). Bu çok aşamalı bireye uyarlanmış test yapısı ile ortadan kalkmaktadır (Mead, 2006). Çok aşamalı bireye uyarlanmış testlerde, bireylerin modüller ile maddeleri grup halinde alması, tek aşama içindeki modülde yer alan maddeler arasında geçiş yaparak bir maddeye tekrar dönmeyi mümkün kılar. Bu da özellikle testi alan bireyler için esneklik sağlayan bir avantajdır.

Paneller. Çok aşamalı bireye uyarlanmış testlerde birbirine paralel paneller bireylere rastgele olarak atanır. Üç aşamalı, 1-2-3 tasarımı çoklu paneller Şekil 8’de gösterilmektedir. Bu n panelde, bulunduğu aşama ve güçlük düzeyine göre modüller içerik ve istatistik özellikleri açısından birbirine paralel olmalıdır. Çoklu paneller her biri aynı özelliklerde fakat her birini oluşturan modüllerin kombinasyonu da bir diğerinden farklı, eşsiz olur (Jodoin, Zenisky & Hambleton, 2006). Test güvenliğini sağlamak amacıyla, madde ifşası riski panel sayısı artırılarak ve aynı modülü birden fazla panelde kullanmayarak (ya da sınırlandırarak) en aza indirgenebilir (Luecht, 2014). Uygulamadan önce paneller kontrol edilebildiği için de madde ifşası kontrolü, çok aşamalı bireye uyarlanmış testlerde BOBUT’a göre çok daha üst düzeydedir (Hendrickson, 2007; Jodoin, Zenisky & Hambleton, 2006).



Şekil 8. 1-2-3 tasarımı çoklu paneller

Aşamalar. Panel tasarımlarının kaç aşamadan oluşacağı, testin bireye uyarlanma düzeyi ile ilişkilidir. Bireyin yetenek kestirimi sadece her bir aşamadan diğerine yönlendirilirken yapıldığından, aşama sayısı arttıkça yetenek kestirimi sayısı ve dolayısı ile testin bireye uyarlanma düzeyi de artacaktır. Ancak teste çok fazla

aşama eklemek, son test formlarının ölçme kesinliğine ciddi bir katkı sağlamayacağı gibi test tasarımının karmaşıklığını artırır (Luecht & Nungester, 1998). Bu yüzden birçok çalışmada ve çok aşamalı test uygulamasında iki, üç ya da dört aşamanın kullanıldığı görülmektedir. Tek yönlendirme noktası olmasına rağmen iki aşamalı testler de kullanım amacına bağlı olarak yeteri kadar bilgi verici olabilmektedir (Zenisky & Hambleton, 2014). Özellikle geniş ölçekli (large scale) ve yüksek beklentili (high-stakes) test uygulayıcıları için, birden fazla yönlendirme noktası olması sebebiyle en az üç aşamalı testleri uygulaması önerilmektedir (Zenisky, Hambleton & Luecht, 2010; Zenisky & Hambleton, 2014). Çok aşamalı bireye uyarlanmış testler, test geliştiricilere oldukça esneklik sağlar, burada belirleyici olan testin amacı ve test sonuçlarının kullanım amacına göre bir test geliştirmektir. Her bir aşamadaki modül sayısı belirlenirken de aynı durum geçerlidir. Her bir aşamadaki modül sayısı, bireylerin yeteneklerine göre ayrıştırılma derecesini göstermektedir. Bu yüzden bir aşamadaki modül sayısını arttırmak, aşama sayısında olduğu gibi bireye uyarlanma düzeyini arttırıyorsa da son kestirimlere ciddi bir katkı sağlamayacağı gibi test tasarımının karmaşıklığını arttıracaktır (Luecht & Nungester, 1998). Bunun için bir aşamadaki modül sayısı da makul bir düzeyde tutulmalıdır. Yapılan çalışmalar ve gerçek uygulamalarda daha çok bir aşamada en fazla üç modülün olduğu tasarımlar kullanılmaktadır.

Modüller. Modüller, madde güçlük parametresine göre gruplandırılmış maddeleri içeren madde bloklarıdır. Çok aşamalı bireye uyarlanmış test tasarımlarında modül uzunluğuna, yani her bir modülün kaç madde içereceğine de önceden karar verilmektedir. Test tasarımında, tüm modül uzunlukları birbirine eşit olabileceği gibi her bir aşamada farklı uzunlukta modüller de kullanılabilir. Aşamalar boyunca modül uzunluklarının değişmesi, bazı aşamalarda diğerlerine göre daha uzun modüller tercih edilmesi yine testin amacına bağlı olarak değişmektedir (Patsula, 1999). Yönlendirme modülü olarak tanımlanan ilk modülün diğerlerinden daha uzun olması ilk yetenek kestirimlerinin daha iyi yapılmasını sağlayabilmektedir (Kim & Plake, 1993). Bir başka açıdan ise, son aşamalardaki madde sayısının fazlalığı son kestirime yaklaşıldığı için daha önemli olabilmektedir (Zenisky, Hambleton & Luecht, 2010). Çok aşamalı bireye uyarlanmış testlerin esnek yapısı farklı amaçlarla, farklı modül uzunlukları ile test tasarımlarının hazırlanmasına olanak sağlar. Modül uzunluklarını çok arttırmak testin karmaşılaşmasına ve testin kullanılabilirlik amacından

uzaklaşmasına sebep olur. Her modülde en fazla 15 ya da 20 maddenin olması önerilmektedir (Stark & Chernyshenko, 2006).

Modül tasarımı çok aşamalı bireye uyarlanmış test tasarımının en zorlu süreçlerinden biridir (Zenisky, Hambleton & Luecht, 2010). Birinci aşamada bir, ikinci ve üçüncü aşamalarda iki veya üç modülün olduğu tasarımlar yaygın olarak kullanılmaktadır. İlk modül, yönlendirme modülü, orta güçlükte maddelerden oluşur. Birey hakkında ilk bilgi, bireyin yönlendirme modülünde yer alan maddelere verdiği cevaba göre edinilir. İkinci ve üçüncü aşamalarda, bireyin kestirilen yeteneğine göre madde güçlüklerine göre modüller kolay, orta ve zor olarak ayrışmaktadır (Şekil 6 ve 7). Testin amacına bağlı olarak her bir modülün madde istatistikleri, madde güçlükleri ortalaması ve güçlük ranjı belirlenir (Hendrickson, 2007). Bu anlamda test bilgisi, madde istatistikleri ve modüllerin paralelliğini kontrol etmek için kullanılır.

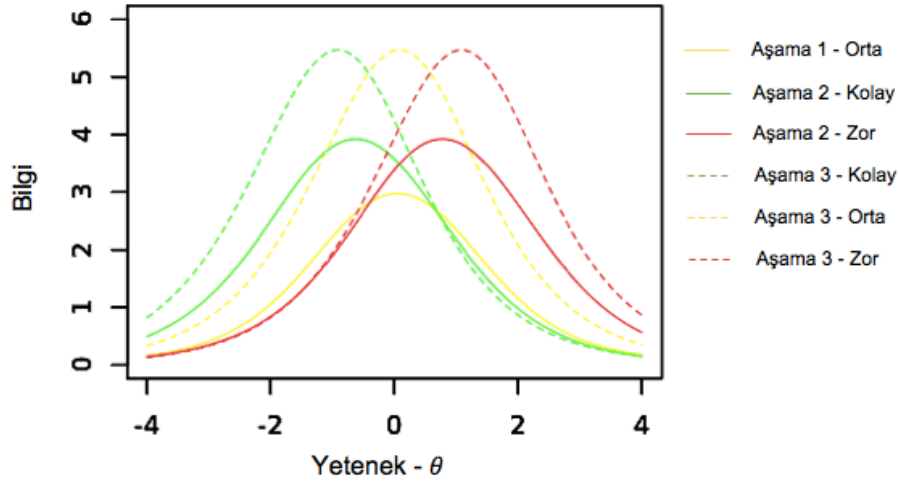
Her bir modülde yer alan maddelerin madde bilgi fonksiyonları toplamı, modül bilgi fonksiyonlarını oluşturur. $I_i(\theta)$, m modülünde yer alan bir i maddesinin madde bilgi fonksiyonu ise, modül bilgi fonksiyonu aşağıdaki gibi tanımlanır.

$$I_m(\theta) = \sum_{i \in m} I_i(\theta)$$

Buna göre, çok aşamalı bireye uyarlanmış test uygulamasında p paneli atanan bir bireyin izlediği r yolu sonundaki, $r(p)$, test bilgi fonksiyonu da bireyin yol boyunca aldığı modüllerin bilgi fonksiyonları toplamına eşit olur.

$$I(\theta) = \sum_{m \in r(p)} \sum_{i \in m} I_i(\theta)$$

Çok aşamalı bireye uyarlanmış test uygulamasındaki tüm panellerde, aynı aşama ve aynı güçlükteki modüllerin (ör.: her paneldeki ikinci aşamada yer alan kolay modüllerin) bilgi fonksiyonlarının birbirine eş olması ve bir panelde bir aşamadaki farklı güçlükteki modüllerin (ör.: üçüncü aşamada yer alan kolay, orta ve zor modüllerin) bilgi fonksiyonlarının birbirlerine paralel olması beklenmektedir (Hendrickson, 2007). 1-2-3 panel tasarımında beklenen modül bilgi fonksiyonlarını gösteren grafikler Şekil 9'da sunulmuştur.



Şekil 9. 1-2-3 tasarımı modül bilgi fonksiyonları

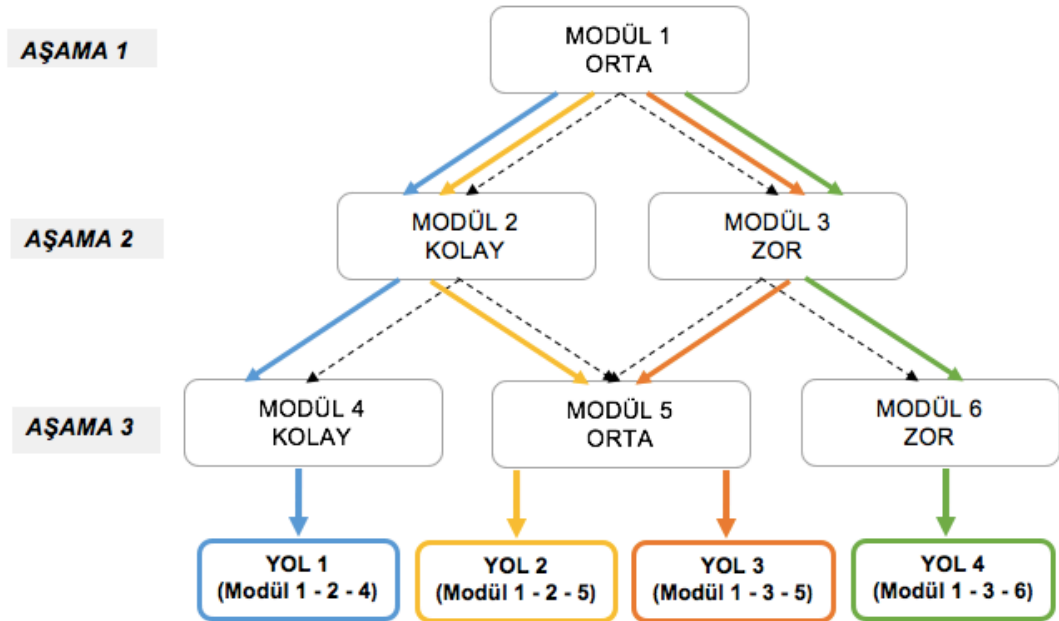
Test bilgi fonksiyonları, geçti kaldı kararlarının verildiği testlerde kesme puanı belirlemek ve test sonunda test bilgisinin yetenek ranjını görmek için de kullanılmaktadır (Luecht, 2014). Modül bilgi fonksiyonları ise test bilgisine olan katkısı dışında bireyin bir aşamadan diğerine nasıl yönlendirileceğini belirlemede kullanılmaktadır.

Bununla beraber bu çalışmanın kapsamına alınmayan, madde ve modül bilgilerinin kullanıldığı bir başka koşul da farklı test oluşturma yöntemleridir. Modül bilgi fonksiyonlarının aşamalar boyunca paralelliğinin yanı sıra daha çok bilgi veren maddelerin hangi aşamalarda olacağına da önceden karar verilebilir. Bu test oluşturma yöntemleri baştan sona ve sondan başa olarak ikiye ayrılmıştır (Luecht & Nungester, 1998). Baştan sonra test oluşturma yönteminde en çok bilgi veren maddeler yönlendirme modülünü oluşturur ve sonraki aşamalarda da daha çok bilgi veren maddelerden başlanarak modüller oluşturulur. Sondan başa test oluşturma yönteminde ise en çok bilgi veren maddeler son aşamadaki modüllerde bulunur ve önceki aşamalarda sırayla daha az bilgi veren maddelerden modüller oluşturulur. Baştan sona test oluşturma yönteminde ilk aşamadaki yönlendirme modülünün verdiği bilgi son aşamaya göre daha fazladır, sondan başa test oluşturma yönteminde ise son aşamadaki modüllerin verdiği bilgi ilk aşamadan daha fazladır.

Yönlendirme. Yönlendirme, çok aşamalı bireye uyarlanmış testlerin bireye uyarlanma basamağıdır. Bireylerin bir aşamadaki modülü tamamladıktan sonra, bir sonraki aşamada hangi modüle yönlendirileceklerine karar veren kurallar *yönlendirme kuralları* olarak tanımlanır. Yönlendirme kuralları, bireylerin test boyunca izleyeceği

yolları belirler ve bu yüzden bu kurallar, çok aşamalı bireye uyarlanmış test uygulamasından çıkacak sonuçların kullanılabilirliği açısından merkezi bir öneme sahiptir (Zenisky & Hambleton, 2014). Üç aşamalı bir 1-2-3 panel tasarımında, bireyin izleyeceği dört yol vardır (Şekil 10). Modül 1 tamamlandıktan sonra, iki seçenek vardır; modül 2 ya da modül 3'e devam edilir. Modül 2'ye devam eden bir birey bu modülü tamamladıktan sonra modül 4 veya modül 5'e giderek testi tamamlayabilir. Modül 3'e devam eden bir birey ise bu modülü tamamladıktan sonra modül 5 ya da modül 6'ya giderek testi tamamlar.

Yönlendirme kuralları, genellikle bilgiyi en yükseğe çıkarmak ve standart hatayı da en aza indirmek yoluyla MTK'ya dayalı olarak yapılsa da testin amacına ve uygulama yöntemine bağlı olarak MTK'ya dayalı ve parametrik olmayan farklı yönlendirme kuralları da kullanılmaktadır. Farklı yönlendirme kurallarını karşılaştıran çalışmalarda, her yöntemin farklı avantaj ve dezavantajları olduğuna vurgu yapılır (Luecht, Brumfield & Breithaupt, 2006; Svetina, Liaw, Rutowski ve Rutowski, 2019; Weissman, Belov ve Armstrong, 2007). MTK'ya dayalı yönlendirme yöntemleri, araştırmacılar tarafından farklı şekillerde isimlendirilmiş ve sınıflandırılmıştır (Luecht vd., 2006; Weissman, 2014; Zenisky & Hambleton, 2014). Bu çalışma için de benimsenen en genel sınıflandırmayı, statik ve dinamik yöntemler olarak ayırarak Weissman (2014) yapmaktadır.

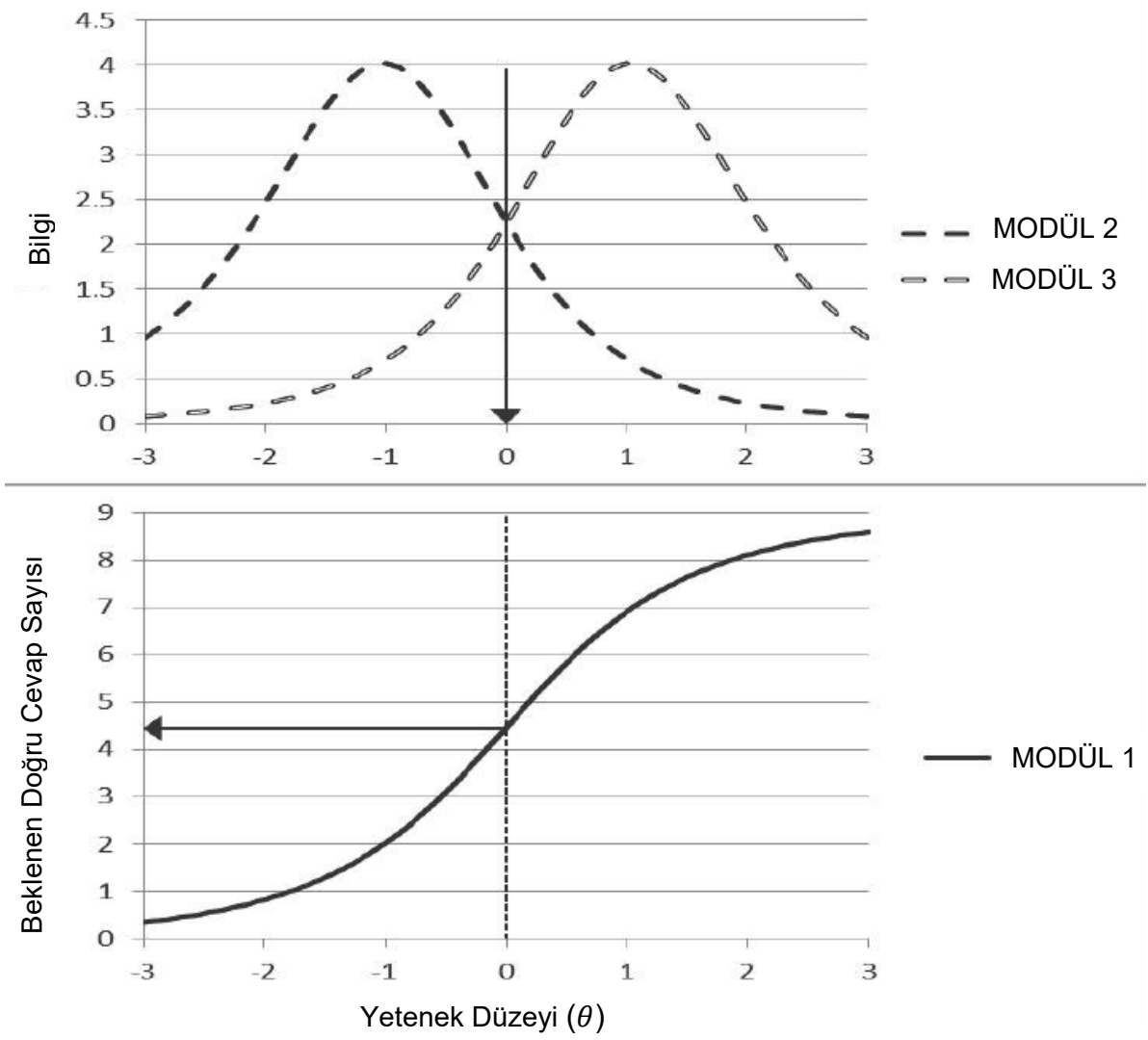


Şekil 10. 1-2-3 tasarımında izlenebilecek yollar

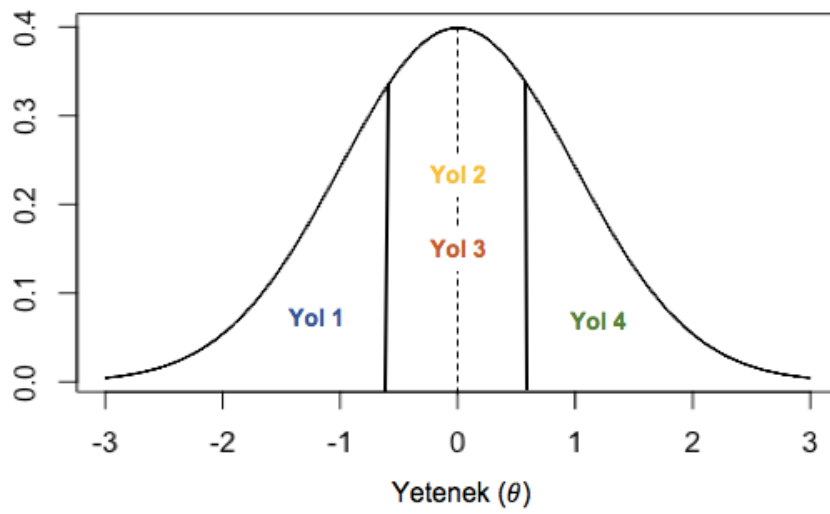
Statik Yönlendirme. Statik yönlendirme kuralında, adından da anlaşılacağı gibi önceden belirlenen bir eşik değere göre yönlendirme yapılır. Bu eşik değer her bir modülde doğru cevaplanan madde sayısıdır. Örneğin birey, modül 1 için belirlenen doğru cevap sayısından az sayıda maddeyi doğru cevaplarsa modül 2'ye, belirlenen doğru cevap sayısından çok sayıda maddeyi doğru cevaplarsa modül 3'e yönlendirilir. Eşik değer belirlenmesinde öncelikle odağın birey mi yoksa grup mu olduğuna karar verilmelidir. Eğer odak bireyse, bilgi fonksiyonları üzerinden eşik değerler hesaplanır. Ancak, odak grup olduğunda popülasyonun yetenek dağılımına göre yapılacak bir hesaplama ile eşik değeri belirlenir.

Birey odaklı statik yöntemde modül bilgi fonksiyonlarından faydalanılarak eşik değeri olacak doğru cevap sayısı belirlenir. Birinci aşamadan ikinci aşamaya geçişte, öncelikle ikinci aşamadaki modüllerin modül bilgi fonksiyonlarının kesiştiği noktaya karşılık gelen θ değeri belirlenir. Bu θ değerinin birinci aşamada yer alan modüldeki maddelere doğru cevap verme olasılıkları üzerinden, θ yetenek düzeyindeki bir bireyin vereceği doğru cevap sayısının beklenen değeri hesaplanır (Şekil 11). Bu yöntem beklenen en yüksek bilgi yöntemi (Approximate Maximum Information - AMI) olarak da isimlendirilmektedir (Luecht, Brumfield & Breithaupt, 2006).

Grup odaklı statik yöntemde ise popülasyonun belli gruplarını belli yollara yönlendirecek eşik değeri önceden belirlenir. Şekil 10'da resmedildiği gibi ikinci aşamada iki modülün olduğu tasarımda, toplamda dört ve birinci aşamadan ikinci aşamaya geçiş için ise iki yol vardır. Bu yöntemde birinci aşamadan ikinci aşamaya geçişte grubun alt %50'sinin birinci yolu (Modül 2) ve üst %50'sinin de ikinci yolu (Modül 3) takip etmesi istenir. Bu yüzden bireylerin normal dağıldığı bir popülasyon için $\theta = 0$ değeri üzerinden ilk modülde beklenen doğru cevap sayısı hesaplanarak eşik değeri belirlenir. Benzer şekilde başlangıçta popülasyon dörde bölünerek diğer yollar için de eşik değerler belirlenir. Bu yöntem, tanımlanmış popülasyon aralıkları yöntemi (Defined Population Intervals - DPI) olarak da isimlendirilmektedir (Luecht, Brumfield & Breithaupt, 2006).

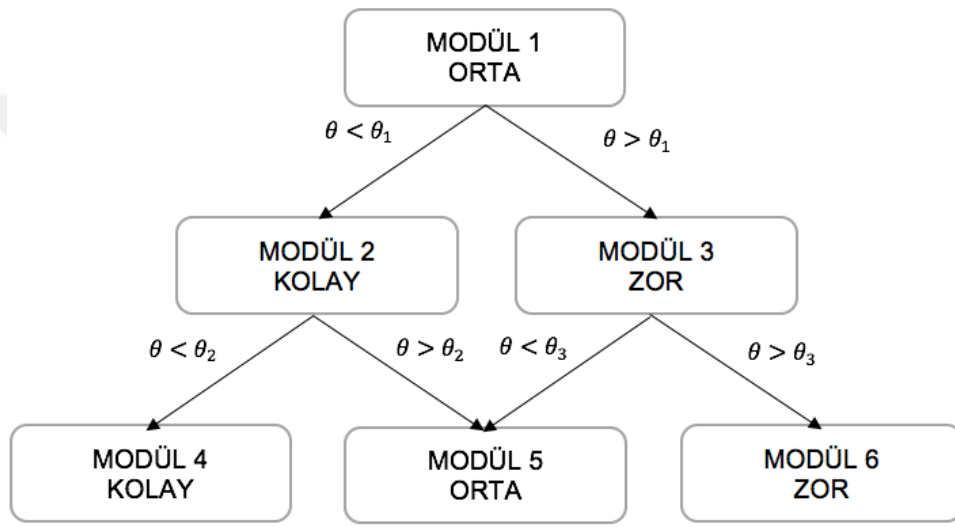


Şekil 11. Birey odaklı statik yönlendirme kuralı (Weissman, 2014, s. 165)



Şekil 12. Grup odaklı statik yönlendirme kuralı

Dinamik Yönlendirme. Dinamik yönlendirme kuralında, bireyin yetenek kestirimi yapılarak, birey bir sonraki aşamada en yüksek bilgiyi veren modüle yönlendirilir. Birinci aşamadan ikinci aşamaya geçişte, ikinci aşamada yer alan modüllerin bilgi fonksiyonlarının kesiştiği noktaya karşılık gelen yetenek düzeyi θ_1 olarak belirlenir. Bireyin birinci aşamadaki modülü tamamladıktan sonra hesaplanan θ yetenek düzeyi θ_1 değerinden küçükse kolay modül (Modül 2)'ye ve büyükse zor modül (Modül 3)'e yönlendirilir (Şekil 13). Benzer şekilde üçüncü aşamadaki modül bilgi fonksiyonlarının kesişim noktalarına karşılık gelen θ_2 ve θ_3 değerleri hesaplanarak Şekil 11'de resmedildiği gibi yönlendirme yapılır.



Şekil 13. Dinamik yönlendirme kuralı

İlgili Araştırmalar

Çok aşamalı bireye uyarlanmış testlerin esnek yapısı, testlerin amaçlarına uygun çok farklı tasarımlar geliştirilmesine olanak sağlar. Bununla beraber, alandaki ve teknolojiye yeni gelişmeler yeni yöntemlerin uygulanmasına fırsat sağladığından çok aşamalı bireye uyarlanmış test alan yazınında sıklıkla, farklı yöntemlerin de test edildiği ve karşılaştırıldığı çalışmalar görülmektedir. Bu bölümde bu çalışmalar ve sonuçları ele alınmıştır. Ayrıca, gerçek uygulama denemeleri ve geniş ölçekli test uygulamalarına da yer verilmiştir.

Öncelikle, çok aşamalı bireye uyarlanmış testlerin, madde düzeyinde BOBUT'un ölçme kesinliği açısından kabul edilebilir bir alternatifi olarak hangi özelliklere sahip olması gerektiğini araştırmak amacıyla BOBUT ile karşılaştırma

alıřmaları yapılmıřtır (Kim & Plake,1993; Kim, Chung, Dodd & Park, 2012; Patsula, 1999; Wang, 2017). Bu alıřmalarda beklendik bir řekilde BOBUT'un daha kesin lmeler yaptıđı bulunmuř ancak ok ařamalı testlerin avantajlarıyla beraber iyi bir alternatif olabilecek dzeyde lmeler yaptıđına ve farklı tasarımların da incelenebileceđine vurgu yapılmıřtır. ok ařmalı testlerin esnek yapısı birok farklı tasarımın performansını da arařtırmayı gerektirmektedir. Patsula (1999), doktora tez alıřmasında diđer kořulları sabit tutarak ařama sayısı, her ařamadaki modl sayısı ve her modldeki madde sayısı faktrlerini BOBUT ile karřılařtırarak farklı ok ařamalı test tasarımlarının performanslarını arařtırmayı amalamıřtır. Arařtırma sonunda yetenek kestiriminin kesinliđi aısından, her bir modldeki madde sayısının nemli bir etkisi olmadıđı ve  ařamalı modellerin ise daha ok tercih edilen modeller olması gerektiđi sonucuna varmıřtır. Yaklařık en yksek bilgi ve tanımlanmıř poplasyon aralıkları ynlendirme yntemleri ile bireye uyarlanacak, bařtan sona ve sondan bařa test oluřturma yntemleriyle oluřturulacak 45 veya 60 madde uzunluđundaki 1-2-3 ve 1-3-3 panel yapısındaki tasarımları BOBUT ile karřılařtırdıđı alıřmasında Wang (2017), tm ok ařamalı test tasarımlarının BOBUT ile benzer yanlılık deđerlerine sahip olduđu, ok ařamalı bireye uyarlanmıř farklı test tasarımlarında ise panel yapısı, test oluřturma ve ynlendirme yntemlerinin bir fark oluřturmadıđı ancak test uzunluđunun fark yarattıđı sonucuna varmıřtır.

ok ařamalı bireye uyarlanmıř testlerin geliřtirilme ve uygulama yntemi, tipik bir BOBUT uygulamasından olduka farklı olduđundan, birok BOBUT simlasyon yazılımı ok ařamalı bireye uyarlanmıř test simlasyonu tasarlamak iin yetersiz ve verimsiz kalmaktadır (Han & Kosinski, 2014). Arařtırmacıların ok ařamalı test simlasyonu iin sıklıkla kullandıđı, bir aık kaynak istatistiksel hesaplama ve grafik yazılımı olan R, aynı zamanda bir programlama dili olduđundan arařtırmacıların kod yazarak simlasyon alıřmaları ve analiz yapmalarına olanak sađlamaktadır. Magis ve von Davier'in (2018) geliřtirdikleri mstR paketi ok ařamalı bireye uyarlanmıř test simlasyonları yapmak iin tasarlanmıřtır. R yazılımı dıřında Han (2013) tarafından geliřtirilen MSTGen yazılımı da ok ařamalı bireye uyarlanmıř test simlasyonları iin kullanılmaktadır.

ok ařamalı bireye uyarlanmıř test simlasyonlarında, ařama sayısı ve her ařamadaki modl sayısına gre belirlenen farklı panel yapıları, her bir ařamadaki modllerin uzunluđu (madde sayısı), bir ařamadan diđerine geiřte diđer modle

yönlendirme yöntemleri, kestirim yöntemleri gibi koşulların yetenek kestirimine etkisinin yanı sıra madde ifşası ve içerik dengesi gibi test yapısı ve güvenliği ile ilgili parametreleri nasıl etkilediği üzerine birçok çalışma yapılmıştır. Çok aşamalı bireye uyarlanmış testlerin bireye uyarlanma basamağı olan yönlendirme için geliştirilen farklı yöntemler sıklıkla araştırılmakta ve farklı yönlendirme yöntemleri birbirleriyle karşılaştırılarak avantajları ve dezavantajları ortaya konmaktadır.

Armstrong (2002), on dört modülden oluşan altı aşamalı bir çok aşamalı test tasarımı için dört farklı yönlendirme yöntemini test etmiştir. Bunlar; yetenek kestirimine dayalı, modül bilgi fonksiyonuna dayalı yönlendirme, sınıflandırmaya dayalı ve kümülatif doğru cevap sayısına dayalı yöntemlerdir. Hepsinin farklı avantajları olduğuna dikkate çekerken, yetenek kestirimine dayalı yöntemin diğerlerinden biraz daha iyi çalıştığını ama bu çalışmadaki tasarım için bu yöntemler arasında anlamlı bir fark bulunmadığını ortaya koymaktadır. Weissman, Belov ve Armstrong (2007)'un yaptığı bir başka çalışmada yine modül bilgisine dayalı (en yüksek Fisher bilgi ve en yüksek ortak bilgi) ve doğru cevap sayısına dayalı üç yönlendirme yöntemi karşılaştırılmıştır. Bunun için, 1-1-2-3 panel tasarımlı, her bir modülün 10 ve 14 madde içerdiği, bireylerin 4 aşama sonunda toplamda 44-50 maddeyi cevapladığı bir model kullanmışlardır. Normal dağılımdan, $N(0,1)$, türetilen θ değerleri için yaptıkları simülasyon sonucunda, bireyleri sınıflandırma amacıyla yapılan test için üç yöntemin de benzer sonuçlar verdiği sonucuna ulaşmışlardır.

Yönlendirme yöntemlerini farklı tasarımlar altında araştıran ve farklı parametrelere etkilerini karşılaştıran birçok farklı çalışma da vardır. Svetina, Liaw, Rutkowski ve Rutkowski (2019) geniş ölçekli test simülasyonu olarak tasarladıkları çalışmalarında modül uzunluğuna göre farklılaşan tasarımları ve yönlendirme yöntemlerini yetenek kestirimleri üzerinden karşılaştırmayı ve bunların madde ifşası üzerine etkisini test etmeyi amaçlamışlardır. Doğru cevap sayısına göre ve MTK'ya dayalı yönlendirme yöntemleri ile uygun ve uygun olmayan yönlendirmeleri ele almışlardır. 1-2-3 panel tasarımında uygun yönlendirme; ikinci aşamada kolay modülle karşılaşan bireyin üçüncü aşamada kolay ya da orta modüle, ikinci aşamada zor modülle karşılaşan bireyin üçüncü aşamada zor ya da orta modüle gitmesidir. Uygun olmayan yönlendirme ise ikinci aşamada hangi modülü alırsa alsın bireyin üçüncü aşamadaki her modüle rastgele gidebiliyor olmasıdır. MTK'ya dayalı yönlendirme yöntemlerinin, doğru cevap sayısına göre daha iyi çalıştığını ancak doğru cevap sayısı

yöntemlerinin de daha basit bir algoritması olduğunu ve uygulamasının kolay olduğunu vurgulamaktadırlar. Çalışma sonunda araştırmacılar bütün amaçlar için en iyi olan tek bir yaklaşımın olmadığını da altını çizmişlerdir (Svetina, Liaw, Rutkowski & Rutkowski, 2019).

En iyi tek bir yaklaşımın olmaması araştırmacıları, çok aşamalı bireye uyarlanmış testlerin farklı tasarımlarını ve farklı koşullar altındaki performanslarını araştırmaya yönlendirmiştir. Farklı panel tasarımlarında, yönlendirme modülünün özellikleri ve uzunluğunun yetenek kestirimlerine etkisini araştırdığı çalışmasında Bozuncü Öztürk (2019), yönlendirme modülünün uzunluğu artıka RMSE değerinin düştüğünü ve gerçek ve kestirilen yetenek kestirimleri arasındaki korelasyonun arttığı sonucuna ulaşmıştır. Buna ek olarak üç aşamalı modellerin de daha iyi sonuçlar verdiğinin altını çizmektedir. Hepsi üç aşamadan oluşan farklı panel tasarımlarını, farklı yönlendirme yöntemleri ve farklı test uzunluğu koşulları altında inceledikleri çalışmalarında Kim, Chung, Park ve Dodd (2013), yönlendirme yönteminden bağımsız olarak aynı panel tasarımı içinde aynı test uzunluğu koşulu altında verilen sınıflandırma kararlarının birbirine bezer sonuçlar verdiğini görmüşlerdir.

Yan, Lewis ve von Davier (2014), çok aşamalı bireye uyarlanmış testlerin farklı modül tasarımlarını, kalibre edilmiş geniş bir veri setinden rastgele seçilen 250 bireyin 100 maddeye verdiği cevaplar üzerinden karşılaştırmışlardır. 100 maddelik madde havuzundan, 1-2-3 panel yapısındaki 45 maddelik farklı modül uzunlukları ve farklı güçlük ranjlarındaki çok aşamalı bireye uyarlanmış testleri, parametrik olmayan yönlendirme yöntemi kullanarak karşılaştırmışlardır. Yaptıkları çalışma ile Yan vd. (2014), daraltılmış güçlük ranji modelinin ve uzundan kısaya azalan modül uzunluğundaki tasarımların daha iyi sonuç verdiğini göstermişlerdir. Parametrik olmayan yönlendirme yöntemleri kullanılarak tasarlanacak çok aşamalı bireye uyarlanmış testlerin, özellikle varsayımların sağlanmadığı ya da örneklem sayısının düşük olduğu durumlarda etkili bir alternatif yaklaşım olabileceğini savunmaktadırlar.

Geniş Ölçekli Test Uygulamaları.

Yapılan çalışmalar, geniş ölçekli ve yüksek beklentili testlerin test gereklilikleri ve uygulama süreci dikkate alındığında çok aşamalı bireye uyarlanmış test olarak uygulanmasının avantajlarını ortaya koymaktadır. Uygulanacak testin test gereklilikleri, amacı ve uygulama esaslarına göre farklı tasarımlar için simülasyon çalışmaları ve gerçek uygulamalar yapılmaktadır.

Birleşik Devletler Ulusal Test İstatistikleri Merkezi bünyesinde, geniş ölçekli testlerin çok aşamalı bireye uyarlanmış test olarak uygulanmasının yararlarını araştırmak amacıyla deneysel bir çalışma yürütülmüştür (Oranje vd., 2014). İkinci aşamada üç modülün yer aldığı iki aşamalı bir test, deney grubuna çok aşamalı bireye uyarlanmış test olarak, birinci aşamadaki performansına göre yönlendirme yapılarak uygulanırken; kontrol grubuna ikinci aşamadaki modüller rastgele atanarak uygulanmıştır. Araştırmanın sonuçları, özellikle grup düzeyinde değerlendirmelerin yapılacağı testlerde çok aşamalı bireye uyarlanmış testlerin en doğru tasarım olacağını göstermektedir (Oranje vd., 2014).

Grup düzeyinde puanların raporlandığı, en büyük uluslararası geniş ölçekli sınav olan Uluslararası Öğrenci Değerlendirme Programı (The Programme for International Student Assessment – PISA) için çok aşamalı bireye uyarlanmış test iyi bir test tasarımı seçeneği olarak görülmüştür (Yamamoto vd., 2019). Uluslararası geniş ölçekli testlerde, özellikle ortalama öğrenci yeterliliğinin uluslararası ortalamanın oldukça altında olduğu düşük performans gösteren ülkeler için bu testlerin bireye uyarlanmış olarak uygulanması daha büyük önem taşımaktadır (Svetina vd., 2019). Bireylerin yeteneğinin çok üstünde, onlar için çok zor maddeler ile karşılaşılıyor olmaları, testlere katılımlarını ve test motivasyonlarını düşürecek bir etken olacaktır. Bireylerin yetenek düzeylerine uygun maddelerle karşılaşılıyor olmalarının, teste katılımlarını ve test motivasyonlarını arttıracak ve bunun bir sonucu olarak, çok aşamalı bireye uyarlanmış test uygulamalarının da cevaplanmayan ya da rastgele cevaplanan madde oranlarını düşürecek olduğu düşünülmektedir (Yamamoto vd., 2018). Bununla beraber birçok başka avantajları da olduğu düşünülerek, PISA 2018 üç aşamadan oluşan çok aşamalı bireye uyarlanmış test olarak tasarlanmış ve uygulanmıştır (Yamamoto vd., 2019). Yamamoto, Shin ve Khorramdel (2019), PISA 2018 verileriyle yaptıkları simülasyon çalışmasında, çok aşamalı bireye uyarlanmış test tasarımının bariz bir şekilde, yetenek kestiriminin doğruluğuna ve çok düşük ve çok yüksek yetenek düzeylerindeki gruplar için ölçme kesinliğine daha fazla katkıda bulunduğunun altını çizmektedirler. OECD tarafından uygulanan bir diğer uluslararası geniş ölçekli sınav olan, Yetişkin Becerileri Araştırması (The Programme for the International Assessment of Adult Competence - PIAAC) da PISA ile benzer yapıda çok aşamalı bireye uyarlanmış test olarak uygulanmaktadır (OECD, 2016).

Çalışmalar, sadece grup düzeyinde değil, birey düzeyinde kararların verildiği geniş ölçekli ve yüksek beklentili testler için de çok aşamalı bireye uyarlanmış testlerin doğru tasarımlar olduğunu göstermektedir. Dünya genelinde lisansüstü programlara başvurular için kullanılan, Lisansüstü Kayıt Sınavı (Graduate Record Examinations - GRE), şu an çok aşamalı bireye uyarlanmış test olarak uygulanmaktadır (Robin vd, 2014). GRE, analitik yazma, sayısal akıl yürütme ve sözel akıl yürütme olmak üzere birbirinden bağımsız üç bölümden oluşmaktadır. Sayısal akıl yürütme ve sözel akıl yürütme testleri, 20 maddelik modüllerden oluşan, iki aşamalı 1-3 panel tasarımına sahip aynı çok aşamalı bireye uyarlanmış test yapısındadır (Educational Testing Service, 2012). GRE için çok aşamalı bireye uyarlanmış test tasarımının, program gerekliliklerini sağlayacak ve testi alanlar için de kolay anlaşılır bir deneyim olacak en basit ve en etkili tasarım olduğu düşünülmektedir (Robin vd, 2014).

Geniş, orta ve küçük ölçekli mesleki lisans, yeterlilik ve sertifika sınavlarının da çok aşamalı bireye uyarlanmış test olarak uygulanmasına yönelik karşılaştırmalı çalışmalar yapılmaktadır (Breithaupt vd., 2014; Brossman & Guille, 2014; Hambleton & Xing, 2006; Jodoin vd., 2006). Bu çalışmalar, daha küçük gruplara uygulanan bu sınavların klasik lineer testler yerine çok aşamalı bireye uyarlanmış test olarak uygulanması sayesinde daha doğru yetenek kestiriminin yapıldığını ve daha doğru ve tutarlı geçti-kaldı kararlarının verildiğini doğrular niteliktedir (Brossman & Guille, 2014).

K-12 değerlendirmelerinin de klasik lineer testler yerine çok aşamalı bireye uyarlanmış testler olarak uygulanabilirliğini tartıştıkları çalışmalarında Wentzel, Mills ve Meara (2014), böyle bir değişimin puanları anında raporlamaya ve öğretmenler ve test uygulayıcılar için test sürecinin basitleştirilmesine olanak sağlayacağını vurgulamaktadır. Biçimlendirici değerlendirmelerin de çok aşamalı bireye uyarlanmış test olarak uygulanmasının hem öğrenciler hem de öğretmenler için öğrencilerin durum teşhisinde faydalı olacağı düşünülmektedir (Luecht, 2014).

Tüm bu çalışmalar ve uygulamalar çok aşamalı bireye uyarlanmış testlerle daha çok karşılaşacağımızı göstermektedir. Kağıt kalem ortamında uygulanan klasik lineer testler ile BOBUT'un ayrı ayrı avantajlarını kullanarak dengeli bir model oluşturan çok aşamalı bireye uyarlanmış testler esnek yapısı sayesinde her alanda uygulanabilmektedir. Farklı amaçlar için farklı tasarım yöntemleri araştırmaya devam

eden alıřmalarda bu alanda daha ok ve farklı yntem ve tasarımların test edildiđi alıřmaların yapılması onerilmektedir.



Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemine ilişkin bilgiler sunulmuştur. Bu çalışma, farklı madde güçlük dağılımlarına göre oluşturulan çok aşamalı bireye uyarlanmış test tasarımlarını; farklı modül uzunlukları, farklı yönlendirme yöntemleri ve farklı kestirim yöntemleri çerçevesinde gerçek ve kestirilen yetenek düzeyleri bazında karşılaştırmayı amaçlamaktadır. Bu amaç doğrultusunda bu bölümde araştırma yöntemine ilişkin bilgiler sunulmuş, çok aşamalı bireye uyarlanmış test simülasyonun adımları (model, madde havuzu, panel tasarımı, modül tasarımı, yönlendirme kuralı ve yetenek kestirimi ile verilerin türetilmesi) anlatılmış ve veri analizinden bahsedilmiştir.

Araştırma Yöntemi

Bu çalışmada çok aşamalı bireye uyarlanmış testlerde farklı simülasyon koşulları altında elde edilen yetenek kestirimleri karşılaştırılacaktır. Bunun için, tüm koşullardan elde edilen kestirimlerin ortalama hata ve yanlılık değerleri hesaplanmıştır. Tüm veriler bu çalışma için bilgisayar programı kullanılarak türetileceğinden bu çalışma bir Monte Carlo simülasyon çalışmasıdır.

Çok Aşamalı Bireye Uyarlanmış Test Simülasyonu

Bu çalışma için veriler, açık kaynak kodlu ve profesyonel bir yazılım olan RStudio'da üretilmiş ve çok aşamalı bireye uyarlanmış test simülasyonu, yazılım içerisinde yer alan mstR paketi (Magis, Yan & von Davier, 2018) yardımıyla yapılmıştır. Madde tepki kuramı bağlamında yapılan Monte Carlo simülasyon çalışmalarında izlenmesi gereken adımlar Harwell vd. (1996) tarafından şöyle sıralanmaktadır: (1) araştırma problemin belirlenmesi, (2) koşulların belirlenmesi, (3) uygun deneysel tasarımın seçilmesi, (4) belirlenen modele göre verilerin türetilmesi, (5) parametrelerin kestirilmesi, (6) gerçek ve kestirilen parametrelerin karşılaştırılması, (7) Bu işlemin her koşul için belirli bir sayıda tekrarlanması (replikasyon), (8) sonuçların analiz edilmesi.

Bu bağlamda bu bölümde belirtilen simülasyon basamaklarına ve koşulların nasıl belirlendiğine dair açıklamalar yer almaktadır. Bu çalışmada, farklı madde güçlük dağılımları (normal ve uniform dağılım) ile oluşturulacak tasarımlar modül uzunluğu, yönlendirme yöntemi ve kestirim yöntemi koşulları altında test edilecektir. Bu altı

simülasyon koşulu tablo 1’de gösterilmektedir. Sabit tutulan koşullar ise MTK modeli, panel yapısı, toplam test uzunluğu, katılımcı sayısı ve madde ayırt edicilik parametresi olarak sıralanabilir. Bu çalışmada, farklı madde güçlük dağılımlarından tasarımlar ayrı ayrı oluşturulmuş ve her bir güçlük dağılımı için 20 koşul, toplamda da 40 koşul test edilmiştir.

Tablo 1

Simülasyon Koşulları

Madde Güçlük Dağılımı	Modül Uzunluğu	Yönlendirme Yöntemi	Kestirim Yöntemi	Koşul
Normal Dağılım	ARTAN (10-15-20)	STATİK	MLE	1
			EAP	2
		DİNAMİK	MLE	3
			EAP	4
	AZALAN (20-15-10)	STATİK	MLE	5
			EAP	6
		DİNAMİK	MLE	7
			EAP	8
	EŞİT (15-15-15)	STATİK	MLE	9
			EAP	10
		DİNAMİK	MLE	11
			EAP	12
	İKİNCİ AŞAMA EN UZUN (10-20-15)	STATİK	MLE	13
			EAP	14
		DİNAMİK	MLE	15
			EAP	16
	İKİNCİ AŞAMA EN UZUN (15-20-10)	STATİK	MLE	17
			EAP	18
		DİNAMİK	MLE	19
			EAP	20
Uniform Dağılım	ARTAN (10-15-20)	STATİK	MLE	21
			EAP	22
		DİNAMİK	MLE	23
			EAP	24
	AZALAN (20-15-10)	STATİK	MLE	25
			EAP	26
		DİNAMİK	MLE	27
			EAP	28
	EŞİT (15-15-15)	STATİK	MLE	29
			EAP	30
		DİNAMİK	MLE	31
			EAP	32
	İKİNCİ AŞAMA EN UZUN (10-20-15)	STATİK	MLE	33
			EAP	34
		DİNAMİK	MLE	35
			EAP	36
	İKİNCİ AŞAMA EN UZUN (15-20-10)	STATİK	MLE	37
			EAP	38
		DİNAMİK	MLE	39
			EAP	40

Model. Bu çalışma Türkiye’de uygulanan yükseköğretime geçiş sınavı olan YKS’nin bir alt testinin (matematik alt testi) simülasyonu olarak kurgulanmıştır. Çalışmada, bu çerçeveye göz önünde bulundurularak madde parametreleri belirlenmiş ve maddeler türetilerek kullanılmıştır. Bu anlamda çok aşamalı bireye uyarlanmış test olarak uygulanan geniş ölçekli testler (GRE, PIACC, PISA) incelenmiş ve MTK modeli olarak 2PL modelin seçilmiş olduğu görülmüştür (Educational Testing Service, 2019; Kirsch & Lennon, 2017; Robin vd., 2014). Geniş ölçekli test simülasyonu olarak yapılan çalışmalarda da 2PL modelin seçilmiş olması (Svetina vd., 2019) da göz önünde bulundurularak bu çalışmada da 2PL model temel alınmıştır.

Madde Havuzu. Çok aşamalı bireye uyarlanmış testler için madde havuzu modüller olarak da düşünülmelidir. Modüllerden oluşan bu havuzda modüllerin paralelliği ya da istatistiksel olarak eşliği (modüllerde bulunan maddelerin MTK özellikleri ve içerik dengesi) önem kazanmaktadır (Ariel vd., 2006). Bu yüzden madde havuzu hazırlanırken her aşamada aynı güçlük düzeyinde birbiri yerine kullanılabilecek olan modüllerin bilgi fonksiyonlarının eş (benzer yetenek düzeyleri için benzer modül bilgisine sahip olması) ve aynı aşamada yer alacak, fakat farklı güçlük düzeyindeki modüllerin de test özelliklerinin (içerik dengesi, madde formatları, cevap anahtarı örüntüsü vb.) eş olması gerekmektedir (Ariel vd., 2006; Armstrong vd., 2004; Hendrickson, 2007; Veldkamp, 2014). Çok aşamalı bireye uyarlanmış bir test hazırlarken, istatistiksel hedeflerin belirlenmesi verilen en önemli kararlardan biridir (Hendrickson, 2007). Her bir modül için madde güçlükleri ortalamasının ve güçlük ranjının nasıl olması istendiğine karar verilmelidir.

Madde ayırt edicilik parametresi, teorik olarak $-\infty < a < +\infty$ aralığında yer alabilir fakat pratikte 2.5’ten küçük ve pozitif değer alır (Baker & Kim, 2004). Kim ve Plake (1993) iki aşamalı tasarımları simülatif olarak karşılaştırdıkları çalışmalarında madde ayırt edicilik değerini tüm maddeler için 0.7 olarak sabit bir değer belirlemişlerdir. Çok aşamalı bireye uyarlanmış test modellerinin simülatif olarak karşılaştırıldığı birçok çalışmada ayırt edicilik parametresi lognormal dağılımdan türetilmektedir (Luo & Kim, 2018; Wang, 2017). Bu bilgiler ışığında ve bu çalışmanın da amacı doğrultusunda, geniş ölçekli bir test uygulamasına uygun olabilecek (Luo & Kim, 2018) madde ayırt ediciliği parametresi tüm madde havuzu için aynı lognormal dağılımından *lognormal* (0.0, 0.3) olarak türetilmiştir.

Her bir modüldeki maddelerin güçlük dağılımlarının normal uniform dağılımdan türetildiği durumlar bu çalışmanın araştırma konusudur. Bu yüzden iki farklı dağılımından türetilen madde güçlüklerine ilişkin bilgiler ayrı ayrı sunulacak ve değerlendirilecektir.

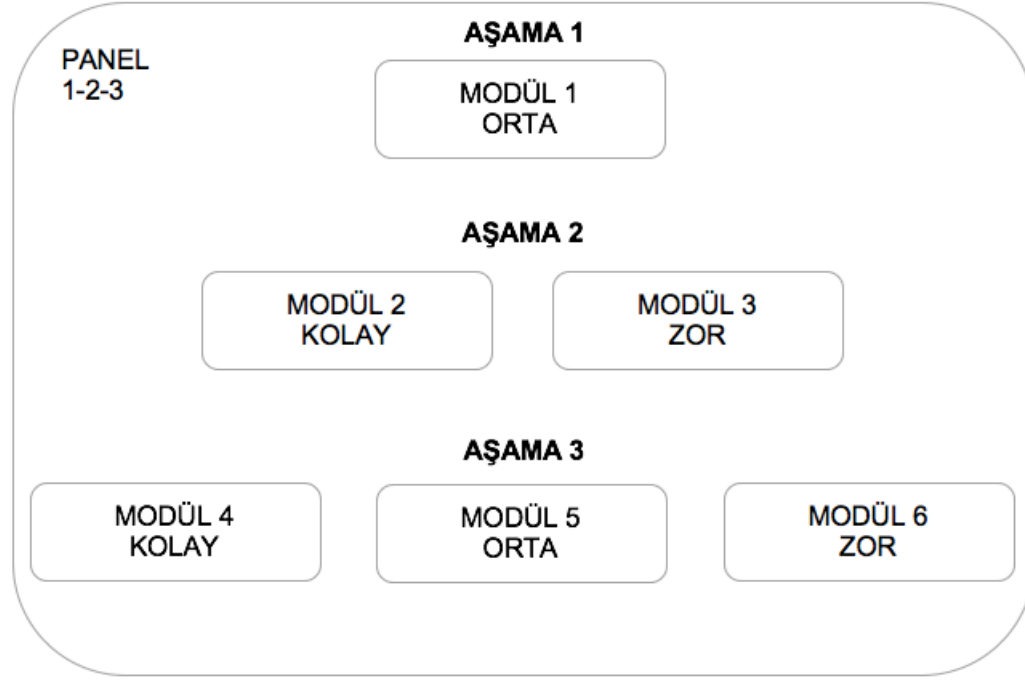
Normal Dağılım. Madde güçlük parametreleri normal dağılımın ortalama ve standart sapma değerleri üzerinden $N(\text{ortalama}, \text{standart sapma})$ olarak ifade edilmiştir. İlk aşamadaki yönlendirme modülünün güçlük dağılımı $N(0,1)$ olarak belirlenmiştir. İkinci aşamadaki modüller için, kolay modül için madde güçlüğü dağılımı $N(-0.7, 0.6)$, zor modül için $N(0.7, 0.6)$ olarak; üçüncü aşama için madde güçlük dağılımları standart sapma değeri ikinci aşamaya göre daha düşük olacak şekilde $N(-1,0.3)$, $N(0,0.3)$ ve $N(1,0.3)$ olarak belirlenmiştir (Wang, 2017).

Uniform Dağılım. Madde güçlük parametreleri uniform dağılımın sınır değerleri üzerinden $U([\text{en düşük}, \text{en yüksek}])$ olarak ifade edilmiştir. İlk aşamadaki yönlendirme modülünün güçlük dağılımı $U([-1,1])$ olarak belirlenmiştir. İkinci aşamadaki nodüllerin madde güçlük dağılımları daha kolay ve daha zor modüller için sırasıyla $U([-1.5, -0.5])$ ve $U([0.5, 1.5])$ olarak, 3. aşamadaki nodüllerin madde güçlük dağılımları kolay, orta ve zor modüller için sırasıyla $U([-2, -1])$, $U([-1,1])$ ve $U([1,2])$ olarak belirlenmiştir (Svetina vd., 2019).

Belirlenen dağılımlara göre madde parametreleri mstR paketinde yer alan ve madde havuzu üretmek için kullanılan `gendichomatrix` fonksiyonu ile türetilmiştir. Madde güçlük dağılımları normal dağılımdan ve uniform dağılımdan olmak üzere iki farklı madde havuzu oluşturulmuştur.

Panel Tasarımı. Çok aşamalı bireye uyarlanmış test tasarımında, bir grup maddeden oluşan test bloğuna modül denmektedir. Modüller, aşamaları ve aşamalar da panelleri oluşturmaktadır. Genellikle ilk aşamada bir modül olmak üzere, her aşamada farklı sayıda modül bulunan iki, üç ya da dört aşamalı panel tasarımları görülmektedir (Yan, Lewis, and von Davier, 2014). Panel tasarımları; aşama sayısı (bir, iki veya daha fazla), her bir aşamadaki modül sayısı, her bir modülün test uzunluğu, her bir modülün aşama içinde ve aşamalar arasındaki istatistiksel özellikleri (güçlük düzeyi, verdiği bilgi vb.) ve kapsam olmak üzere beş farklı özelliğe göre değişiklik gösterir (Luecht & Burgin, 2003). Aşama sayısı ve aşamadaki modül sayısı

arttıkça modüllerin madde güçlük çeşitliliği artacağından ve daha çok kestirim yapılacağından testin bireye uyarlama düzeyi de artar. Ancak bu durum test düzeneğinin karmaşıklığını arttıracak (Hedrickson, 2007; Luecht & Nungester, 1998; Patsula, 1999; Yan vd., 2014) ve madde havuzunda çok daha kolay ve çok daha zor maddelere ihtiyaç duyulacaktır (Zenisky vd., 2010). Bu yüzden kullanılabilirlik da düşünülerek testin ideal aşama ve modül sayısı ile birlikte her bir modülde yer alacak madde sayısı da belirlenmelidir. Özellikle yüksek beklentili (high-stake) testlerde, iki aşamalı bir tasarım olduğunda, birey performansına göre ikinci aşamada test geleceğini bildiği ve eğer düşük performans gösterirse düzeltmeyeceği algısına kapılacağı için iki aşamalı tasarımın çok uygun olmayacağı düşünülmektedir (Zenisky vd., 2010). Buna ek olarak ilk modülün güçlük düzeyi ve yönlendirme yöntemine bağlı olarak bireyin ikinci aşamada uygun olmayan bir modül ile karşılaşma potansiyeli düşünülerek de panel tasarımlarının iki aşamadan fazla olması önerilmektedir (Yan vd., 2014). Farklı aşama sayıları ve her aşamada farklı modül sayılarından oluşan panel tasarımlarını değerlendiren çalışmalarında Armstrong ve arkadaşları (2004), son aşamada en fazla dört modülün uygun olduğunu ancak üç aşamanın da yeterli olabileceğini belirtmektedirler. Buna göre 3 aşamadan oluşan panel tasarımı için, ilk aşamada bir modül (yönlendirme modülü), ikinci aşamada iki modül ve son aşamada da üç modülün yer alacağı modeller oluşturulabilir. Alan yazınındaki bilgiler ışığında, geniş ölçekli testlerin çok aşamalı bireye uyarlanmış test olarak tasarlandığı çalışmalarda da 1-2-3 panel tasarımının kullanıldığı görülmektedir (Svetina vd., 2019; Yan vd., 2014). Buna göre, bu çalışma kapsamında, Şekil 14'te resmedilen üç aşamalı 1-2-3 panel tasarımına göre oluşturulan farklı modeller test edilmiştir.



Şekil 14. 1-2-3 panel tasarımı

Modül Tasarımı. Modül uzunluğu ile ilgili olarak yapılan çalışmalarda farklı sayıda madde içeren modüller kullanılmıştır. Patsula (1999), farklı aşamalarda daha uzun test kullanımı ile ilgili farklı gerekçelerin olduğuna işaret etmektedir. Kim ve Plake (1993)'in yaptığı iki aşamalı bireye uyarlanmış test çalışmasına göre de ilk aşamadaki madde sayısını daha fazla olması daha doğru θ kestirimleri yapılmasını sağlamaktadır. Ancak çok uzun modüllerin de verimliliği düşüreceğine vurgu yaparak ilk aşamada 10 maddelik ya da 15 maddelik modüllerden, 20 maddelik bir modülle başlanması önerilmektedir (Kim & Plake, 1993). Öte yandan, sonraki aşamalarda madde sayısını arttırmanın gerekliliği, testin sonuna doğru bireyin yetenek kestirimine yaklaşıldığı ve bu aşamalarda kestirilen yeteneğe göre daha çok maddenin uygulanmasının daha önemli olduğu düşüncesiyle vurgulanmaktadır (Zenisky vd., 2010). Genel olarak yapılan çalışmalarda ise her modülde de en fazla 15-20 madde olması önerilmektedir (Stark & Chernyshenko, 2006). Modül tasarımı çok aşamalı bireye uyarlanmış test tasarımının en zorlu sürecidir denebilir (Zenisky vd., 2010). Doğru bir yetenek kestirimi için, modüllerin hem içerik ile ilgili gereklilikleri sağlayan, güçlükleri istenilen düzeyde, en çok bilgiyi veren maddelerden oluşması ve bununla birlikte bir aşamada yer alan tüm modüllerin de içeriklerinin aynı, ayırt edicilik ve test bilgilerinin benzer olması gerekecektir. Bu çalışma simülatif bir çalışma olduğundan madde parametreleri ve modül bilgileri dikkate alınarak tasarımlar hazırlanmıştır. Bu çalışmada, ilk aşamadan son aşamaya artan (10-15-20) ve azalan (20-15-10) modül

uzunlukları, en uzun modülün ikinci aşamada olduğu (10-20-15 ve 15-20-10) ve her aşamada eşit uzunlukta (15-15-15) modüllerin bulunduğu tasarımlar farklı madde güçlük dağılımlarından oluşturulan panel tasarımları için ayrı ayrı karşılaştırılacaktır.

Yönlendirme Kuralı ve Yetenek Kestirimi. Madde havuzu, panel tasarımları ve modül özellikleri belirlendikten sonra çok aşamalı testlerin nasıl işleyeceğini belirlemek için yönlendirme yöntemlerinin, testin nasıl puanlanacağını ve yetenek kestiriminin nasıl yapılacağını da belirlenmesi gerekmektedir.

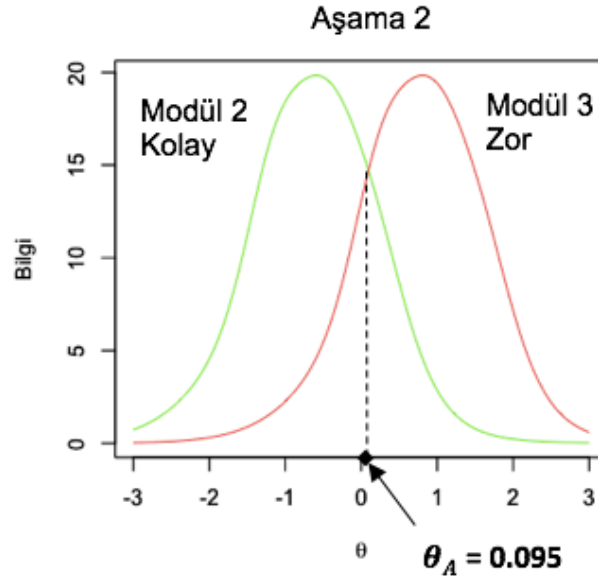
Yönlendirme, hangi bireyin hangi modül ile eşleşeceğini belirlediği ve çok aşamalı testlerin nasıl işlediğini gösteren temel adımlardan biridir (Zenisky & Hambleton, 2014). Bireyin bir önceki aşamadaki performansına göre bir sonraki aşamada hangi modülün bireyin performansı ile eşleştiği genellikle bilgiyi en yükseğe çıkarmak ve standart hatayı da en aza indirmek yoluyla yapılmaktadır (Yan vd., 2014). Çoğunlukla madde tepki kuramına dayanan yöntemler kullanılsa da amaca bağlı olarak parametrik olmayan yönlendirme yöntemlerinin de kullanıldığı görülmektedir. Bu farklı yönlendirme yöntemlerini karşılaştıran çalışmalarda (Luecht vd., 2006; Svetina, Liaw, Rutowski ve Rutowski, 2019; Weissman, Belov ve Armstrong, 2007) her yöntemin farklı avantaj ve dezavantajları olduğuna vurgu yapılmaktadır. Bu yöntemler, birey ve grup odaklı (Zenisky & Hambleton, 2014) olarak sınıflandırıldığı gibi statik ve dinamik yöntemler (Weissman, 2014) olarak da sınıflandırılmaktadır. Grup odaklı yöntemlerde öncelikle bireyin izleyebileceği olası yolların sayısı belirlenir (1-2-3 panel tasarımında örneğin olası yol sayısı 4'tür). Popülasyon dağılımına bağlı olarak bu yollar popülasyon alt gruplarına önceden atanır. Tanımlanmış popülasyon aralıkları olarak da isimlendirilen (Luecht vd., 2006) bu yöntem test uygulayıcısına modül ifşasını kontrol etme imkânı sunmasına rağmen eğer popülasyon dağılımı ile ilgili tahmin doğru değilse bu defa dezavantajlı bir yöntem olacaktır (Zenisky & Hambleton, 2014). Birey odaklı yöntemlerde ise amaç bilgiyi en yükseğe çıkarmaktır. Yaklaşık en yüksek bilgi yöntemi (Luecht, vd., 2006) kullanılarak bireyin performansına göre bir sonraki aşamadaki modüle geçilir. Bireyin performansının bir sonraki modüle geçişte nasıl hesaplandığına ilişkin yöntemler statik ve dinamik yöntemler olarak sınıflandırılmaktadır (Weissman, 2014). Statik yönlendirme yöntemlerinin temel özelliği bir eşik değerin önceden belirleniyor olmasıdır. Bireyin modüldeki maddelere verdiği doğru cevap sayısı temel alınır. Dinamik yönlendirme yöntemlerinde ise bireyin hangi modüle gideceğinin kararı testin

uygulandığı gerçek zamanda bireyin yetenek kestirimi temel alınarak belirlenir. Bu çalışma amacı doğrultusunda birey odaklı statik ve dinamik yöntemler karşılaştırılacaktır.

Birey odaklı statik yöntemde yani bireyin doğru cevap sayısına göre yapılan yönlendirmede, öncelikle ikinci ve daha sonraki aşamalarda modül bilgi fonksiyonlarının kesişim noktası belirlenir. Bu kesişim noktalarına karşılık gelen θ yetenek düzeyleri ile bir önceki aşamada bireyin devam ettiği modülde ulaşılması gereken doğru cevap sayısı eşleştirilir. Böylece bireyin bir sonraki aşamada devam edeceği modüller için eşik değer doğru cevap sayısının beklenen değeri olarak belirlenir. θ yetenek düzeyine sahip bir birey için, modüldeki tüm maddelere (n) vereceği doğru cevap sayısının beklenen değeri, bireyin o θ yetenek düzeyinde her bir madde için doğru cevap verme olasılıklarının toplamı olarak hesaplanır (Colvin vd., 2016). Bulunan değer ya da değerler bir sonraki aşamadaki modüller için kesme puanı olarak tanımlanır.

$$E(\text{doğru cevap}|\theta_j) = \sum_{i=1}^n P_i(\theta_j)$$

Örnek olarak 1-2-3 panel tasarımında, 10 – 15 – 20 modül uzunluğu modeli ele alınabilir. Birinci aşamadan ikinci aşamaya geçişte bireyi, doğru cevap sayısına göre yönlendirmek için öncelikle ikinci aşamada yer alan iki modülün bilgi fonksiyonlarının kesişim noktası belirlenir (Şekil 15). Bu değer bu tasarım için 0.095, θ yetenek düzeyidir. Daha sonra 0.095 yetenek düzeyinde bir bireyin 1. aşamada yer alan 10 maddenin her biri için doğru cevap verme olasılıkları hesaplanır. 10 madde için hesaplanan olasılıkların toplamı 4.89 olmaktadır. Bu durumda yönlendirme kuralı şöyle belirlenebilir: 5 doğrudan az yapan bireyler (1, 2, 3 ya da 4 doğru cevabı olanlar) 2. aşamadaki kolay modüle, 5 doğru ve daha fazla yapan bireyler (5, 6, 7, 8, 9 ya da 10 doğru cevabı olanlar) ise 2. aşamadaki zor modüle yönlendirilecektir. Benzer basamaklar 3. aşamadaki modüllere geçiş için de uygulanır.



Aşama 1 Modül 1

Madde Parametreleri

	<i>a</i>	<i>b</i>
Madde 1	4,28	-0,63
Madde 2	3,06	0,18
Madde 3	2,26	-0,84
Madde 4	1,40	1,60
Madde 5	3,81	0,33
Madde 6	2,68	-0,82
Madde 7	2,71	0,49
Madde 8	3,61	0,74
Madde 9	3,48	0,58
Madde 10	3,25	-0,31

Beklenen Doğru Cevap Sasyısı

$$E(\text{doğru sayısı}|\theta_A) = \sum_{i=1}^{10} P_i(0.095)$$

$$P_1 = 0.96$$

$$P_2 = 0.43$$

$$P_3 = 0.89$$

$$P_4 = 0.11$$

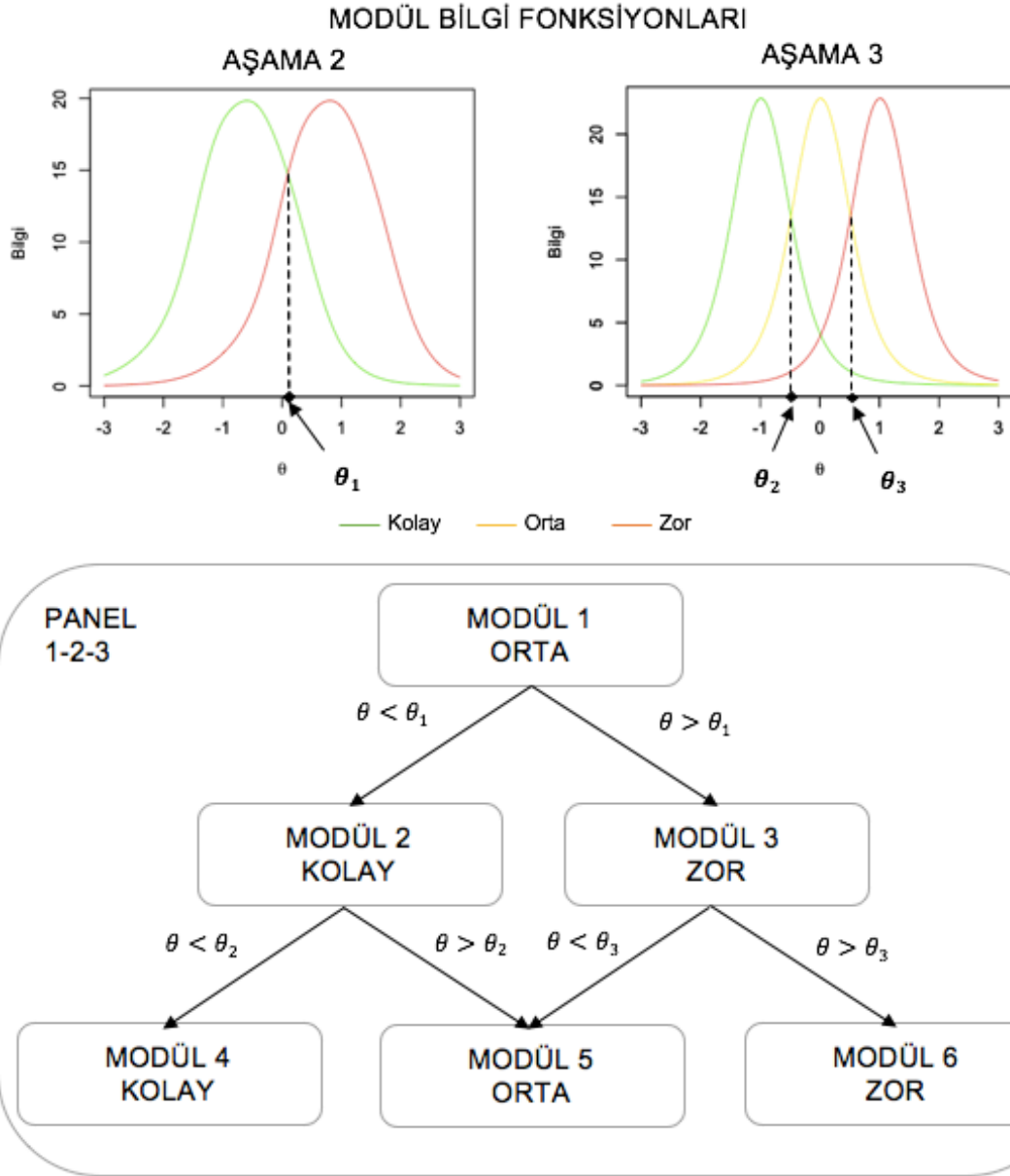
$$\vdots$$

$$\vdots$$

$$\sum_{i=1}^{10} P_i(0.095) = 4.89$$

Şekil 15. Statik yönlendirme hesaplama örneği

Dinamik yönlendirme yönteminde ise bireylerin yetenek kestirimi sonucu bir sonraki aşamadaki modüllerden hangisinin en yüksek bilgiyi verdiği göre yönlendirme yapılır. Bu yöntemde de modül bilgi fonksiyonlarının kesişim noktasındaki θ düzeyi eşik değer olacaktır. Her birey her aşmayı tamladıktan sonra yeteneği EAP yöntemi ile kestirilir ve bir sonraki aşamaya yönlendirme hesaplanan θ düzeyine göre yapılır. Şekil 16'da dinamik yönlendirme yöntemi, 1-2-3 panel tasarımı için resmedilmiştir. Tüm aşamalar tamamlanıp test sonlandığında ise bireylerin yetenek kestirimi için de EAP ve MLE yöntemleri kullanılacak ve son yetenek kestirimleri karşılaştırılacaktır.



Şekil 16. 1-2-3 Panel Tasarımı Dinamik Yönlendirme

Verilerin Türetilmesi. Bu bölümde madde havuzunun oluşturulması, gerçek yetenek düzeyleri ve cevap örüntülerinin türetilmesine ilişkin bilgiler verilmektedir.

Madde Parametreleri ve Modüller. Test uzunluğu, panel ve modül tasarımı, yönlendirme ve yetenek kestirimi yöntemleri ile ilgili yukarıda açıklananlar ışığında bu çalışmaya ait veriler türetilmiştir. Bu çalışma yüksek beklentili testlerin uygulamasına bir örnek teşkil etmek amacıyla üniversite seçme sınavının bir alt testinin simülasyonu olarak kurgulanmıştır. Türkiye’de son 20 yıldır yapılan üniversiteye geçiş sınavlarında alt testler en fazla 45 madde içermektedir. Buradan yola çıkarak bu çalışmada her bir bireyin izleyeceği yolda karşılaşılabilecek toplam madde sayısı 45 olarak belirlenmiştir.

1-2-3 panel tasarımında yer alan altı modül; 1. Aşama – Orta(1O), 2. Aşama – Kolay (2K), 2. Aşama – Zor (2Z), 3. Aşama – Kolay (3K), 3. Aşama – Orta (3O) ve 3.Aşama – Zor (3Z) olarak 3 aşama olacak şekilde belirlenmiştir. Simülasyon koşullarından biri de her aşamadaki modüllerin farklılaşan madde sayılarıdır. En uzun testin hangi aşamada olduğuna göre değişen ve eşit madde sayılarının olduğu modüllere göre, her bir modülde 10, 15 ya da 20 madde olacak şekilde toplam beş farklı panel tasarımı oluşturulmuştur. Bu yüzden her bir modülden (1O, 2K, 2Z, 3K, 3O, 3Z) 10, 15 ve 20’şer olmak üzere 45’er madde türetilmiştir. Her bir güçlük dağılımı için 270 ve toplamda 540 madde türetilmiştir. Normal ve uniform dağılımdan oluşturulan modüllere ait madde güçlük parametresi dağılımları Tablo 2’de sunulmuştur.

Tablo 2
Modül Madde Güçlük Dağılımları

Aşama	Modül	Normal Dağılım $N(Ort.,SD)$	Unifrom Dağılım $U([min,maks])$
1. Aşama	1O	$N(0,1)$	$U([-1,1])$
2. Aşama	2K	$N(-0.7,0.6)$	$U([-1.5,-0.5])$
	2Z	$N(0.7,0.6)$	$U([0.5,1.5])$
	3K	$N(-1,0.3)$	$U([-2,-1])$
3. Aşama	3O	$N(0,0.3)$	$U([-1,1])$
	3Z	$N(1,0.3)$	$U([1,2])$

Gerçek Yetenekler ve Cevap Örüntüleri. Bireylerin gerçek yeteneklerinin uniform dağılım veya normal dağılımdan türetilmesinin yetenek kestirimlerine etkisi göz ardı edilecek düzeyde olması (Chen vd., 1997) ve popülasyondaki bireylerin gerçek yetenek dağılımlarını yansıtması (Jodoin vd., 2006) sebebiyle yetenek dağılımları -3, 3 aralığında standart normal dağılımdan $N(0,1)$ türetilmiştir. Gerçek yetenekleri türetilmiş her bireyin karşılaştığı her bir maddeye cevap verme olasılıkları, madde parametreleri ve θ değerleri bilindiği için bellidir. Cevap örüntülerini oluşturmak için 0 ile 1 arasında uniform dağılımdan rastgele bir sayı seçilir. Eğer bireyin maddeye doğru cevap verme olasılığı bu rastgele seçilen sayıdan büyükse bireyin maddeyi doğru cevapladığı, küçükse yanlış cevapladığı kabul edilir. Tüm veri seti için cevap örüntüleri bu şekilde, *mstR* paketinde simülasyonu gerçekleştiren *randomMst* fonksiyonu içinde yer alan *genPattern* argümanı ile oluşturulmuştur.

Verilerin Analizi

Monte Carlo simülasyon çalışmalarında replikasyon sayısının parametre kestirimlerinin kesinliği ve güvenirliği ile doğrudan bir ilişkisi vardır (Feinberg & Rubright, 2016; Harwell, vd., 1996). Araştırma sorusuna bağlı olarak belirlenecek olan replikasyon sayısının çok az veya çok fazla olmaması önemlidir. Madde tepki kuramı bağlamında yapılan çalışmalarda replikasyon sayısının en az 25 olması önerilirken, 500'den fazla olmasının da gerekli olmayabileceği vurgulanmaktadır (Harwell, vd., 1996). Diğer çok aşamalı bireye uyarlanmış test simülasyon çalışmalarında kullanıldığı gibi (Belov & Armstrong, 2008; Colvin vd., 2016; Reckase vd., 2019), bu çalışmada da 100 replikasyon sayısının yeterli olacağı düşünülmektedir.

Her bir aşamadaki modül uzunluklarına göre birbirinden farklılaşan beş çok aşamalı bireye uyarlanmış test tasarımı, iki farklı yönlendirme yöntemi ve iki farklı son kestirim yöntemine göre ayrı ayrı karşılaştırılmıştır. Tüm koşullar için 100 replikasyon sonucu elde edilmiş yetenek kestirimleri ile gerçek yeteneklere ait yanlışlık (bias) ve hata kareleri ortalamasının karekökü (Root Mean Square Error - RMSE) değerleri hesaplanmıştır. Pozitif ve negatif değerler alabilen yanlışlık değerinin sıfıra yakın olması istenir. RMSE değeri sıfıra eşit veya pozitif değer alabilir ve daha yüksek doğruluk için RMSE değerinin sıfıra yakın olması gerekir.

N , toplam birey sayısını, θ_i , i . bireyin gerçek yetenek düzeyini ve $\hat{\theta}_i$, i . bireyin kestirilen yetenek değerini göstermek üzere yanlışlık (bias) ve RMSE değerlerinin formülleri aşağıda gösterilmektedir.

$$Yanlışlık = \frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)}{N}$$

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

RMSE ve yanlışlık değerleri arasındaki farklılığın istatistiksel olarak anlamlı olup olmadığı varyans analizi (ANOVA) ile test edilmiştir. Bu çalışmada, $p < 0.05$ istatistiksel olarak anlamlı kabul edilmiştir. Örneklemin büyük olduğu durumlarda, sadece

istatistiksel olarak anlamlılığın göstergesi olan p değeri yanıltıcı olabilmekte ve önemsiz farkların bile etkiliymiş gibi görünmesine sebep olabilmektedir (Levine ve Hullet, 2002). Bu yüzden, gruplar arasındaki farklılığın ne kadar büyük olduğu, etki büyüklüğünün göstergesi olan eta kare (η^2) değeri ile hesaplanabilir. Araştırmacılar, örneklem büyüklüğünün 300'ün üzerinde olmasının dahi p değerinin yanı sıra etki büyüklüğü değerinin de raporlanması ve yorumlanması için yeterli bir büyüklükte olduğunu vurgulanmaktadır (Khalilzadeh & Tasci, 2017). Bu çalışmada da ANOVA sonuçlarını yorumlamak için p değerinin yanı sıra etki büyüklüğünün göstergesi olan eta kare (η^2) değeri de hesaplanmıştır. Eta karenin eşik değerleri; 0.1 düşük etki, 0.6 orta etki ve 0.14 yüksek etki üzerinden bulgular yorumlanmıştır (Cohen, 1988).



Bölüm 4

Bulgular ve Yorum

Bu bölümde araştırmanın alt problemine ilişkin bulgular ve yorumlar sunulmaktadır. Bu araştırmada, farklı madde güçlük dağılımları kullanılarak tasarlanmış çok aşamalı bireye uyarlanmış test modellerini; farklı modül uzunlukları ile farklı yönlendirme yöntemleri ve farklı son yetenek kestirim yöntemleri çerçevesinde gerçek ve kestirilen yetenek düzeyleri üzerinden karşılaştırılması amaçlanmıştır. Beş farklı modül uzunluğu, iki farklı yönlendirme yöntemi ve iki farklı son kestirim yöntemine göre, iki farklı güçlük dağılımı ile oluşturulmuş tasarımlara ilişkin veri analizleri iki alt problem için ayrı ayrı sunulmuştur.

Alt problemlere ilişkin bulguların yorumlanabilmesi için, her bir alt problemde belirtilen beş tasarıma ait ayrı ayrı çok aşamalı test modelleri oluşturulmuştur. Her bir alt problem için panelleri oluşturan modüllerin modül bilgi fonksiyonları grafikleri ve modül madde istatistikleri ekte sunulmaktadır. Türetilen gerçek yeteneklere ait bu maddeler için cevap örüntüleri de otomatik olarak oluşturulmuştur. Farklı koşullar altında bu beş tasarımdan elde edilen yetenek kestirimlerinin, yanlılık ve RMSE değerlerine ilişkin bulgular tablo ve grafikler halinde sunulmuş ve yorumlanmıştır. RMSE ve yanlılık değerlerine ait varyans analizi (ANOVA) bulguları da tablo olarak sunulmuş ve yorumlanmıştır.

Birinci Alt Probleme İlişkin Bulgular ve Yorum

Araştırmanın birinci alt problemi şu şekilde belirtilmiştir:

1. Çok aşamalı bireye uyarlanmış test uygulamasında, madde güçlülüklerinin normal dağılımdan türetildiği 1-2-3 panel tasarımından elde edilen yetenek kestirimlerinin yanlılık (bias) ve RMSE değerleri; yönlendirme yöntemine (statik ve dinamik), modül uzunluklarına (artan, azalan, eşit, ikinci aşama en uzun ilk aşama en kısa ve ikinci aşama en uzun son aşama en kısa) ve son kestirim yöntemine göre (MLE ve EAP) nasıl değişmektedir?

Birinci alt probleme ilişkin yapılan veri analizleri sonucunda, belirlenen koşullar altında beş farklı tasarımdan elde edilen yetenek kestirimlerine ilişkin RMSE ve yanlılık değerleri Tablo 3'te verilmiştir. RMSE ve yanlılık değerlerine ilişkin grafikler sırasıyla Şekil 17 ve Şekil 18'de sunulmuştur.

Tablo 3

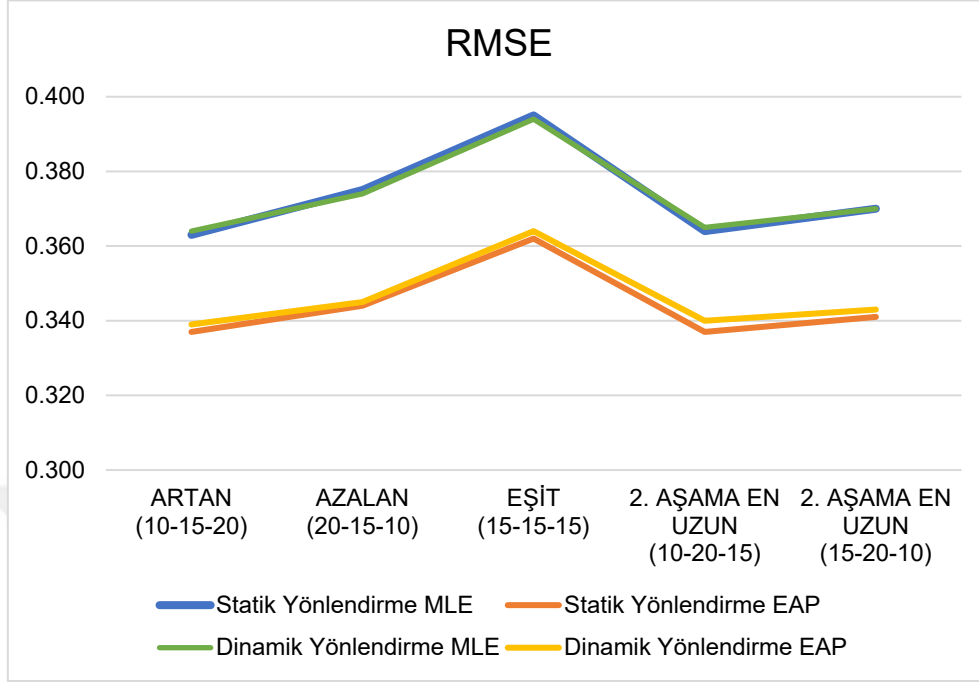
Madde Güçlükleri Normal Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerleri

Modül Uzunluğu	Son Kestirim Yöntemi	Statik Yönlendirme		Dinamik Yönlendirme	
		RMSE	Yanlılık	RMSE	Yanlılık
Artan (10-15-20)	MLE	0,363	-0,006	0,364	-0,002
	EAP	0,337	-0,005	0,339	-0,002
Azalan (20-15-10)	MLE	0,375	-0,001	0,374	-0,005
	EAP	0,344	0,001	0,345	-0,003
Eşit (15-15-15)	MLE	0,395	0,004	0,394	0,006
	EAP	0,362	0,002	0,364	0,007
2. aşama en uzun (10-20-15)	MLE	0,364	-0,006	0,365	-0,005
	EAP	0,337	-0,005	0,340	-0,004
2. aşama en uzun (15-20-10)	MLE	0,370	-0,001	0,370	-0,004
	EAP	0,341	0,001	0,343	-0,003

0,337 ile 0,395 arasında değişen RMSE değerlerinin en düşük olduğu tasarım, EAP son kestirim ve statik yönlendirme yöntemlerinin kullanıldığı artan modül uzunluğu (10-15-20) tasarımı ve en yüksek olduğu tasarım, MLE son kestirim ve statik yönlendirme yöntemlerinin kullanıldığı eşit modül uzunluğu (15-15-15) tasarımı olmuştur. 1-2-3 panel tasarımlı çok aşamalı bireye uyarlanmış testte modüllerin güçlük dağılımlarının normal dağılımdan türetildiği bu durumda, modül uzunlukları farklılaşan beş tasarım için de, tüm koşullar altında ikinci ve üçüncü aşamalarda daha uzun modüller kullanılmasının ortalama hata değerini düşürdüğü görülmektedir. Modül uzunlukları farklılaşan beş tasarım arasında, dört koşul altında da en yüksek RMSE değeri eşit modül uzunluğu (15-15-15) olduğunda hesaplanmıştır. Tüm aşamalarda eşit sayıda madde olduğu durumda ise her koşulda ortalama hata değerinin yükseldiği görülmektedir.

Son kestirim yöntemi EAP olduğunda hem statik hem de dinamik yönlendirme yönteminin kullanıldığı tüm tasarımlarda ortalama hata değeri daha düşük olarak hesaplanmıştır. Dört koşul altında da beş tasarıma ait RMSE değerlerine ilişkin grafikler benzer bir örüntüdedir (Şekil 17). EAP son kestirim ve MLE son kestirim koşullarında statik ve dinamik yöntemlerin grafikleri ise neredeyse çakıştır. Bu durum, madde güçlük dağılımları normal dağılımdan türetilen modüllerden oluşan tasarımlar

için yönlendirme yöntemine göre ortalama hata değerlerinin birbirine çok yakın değerler olduğunu göstermektedir.

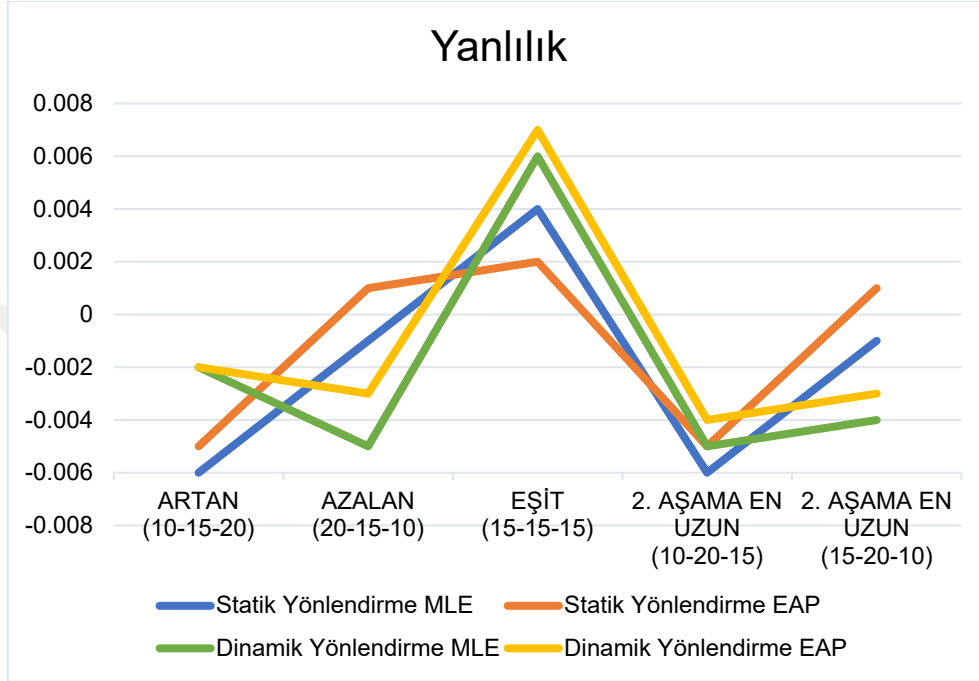


Şekil 17. Normal dağılımdan türetilmiş modül tasarımları için RMSE değerleri

Tablo 3'te yer alan ortalama yanlılık değerleri incelendiğinde genel olarak tüm modeller için yanlılık değerlerinin düşük olduğu ve -0,006 ile 0,007 arasında değiştiği görülmektedir. -0,001 ve 0,001 ile yanlılık değeri sıfıra en yakın, statik yönlendirme kullanılan hem MLE hem de EAP kestirimleri için azalan modül uzunluğu (20-15-10) ve ikinci aşama en uzun (15-20-10) tasarımlarında görülmüştür. Buna göre bu koşullar için daha yansız hesaplamalar yapıldığı söylenebilir. Yanlılık değerinin sıfıra en uzak olduğu tasarım, EAP son kestirim ve dinamik yönlendirme yöntemlerinin kullanıldığı eşit modül uzunluğu (15-15-15) tasarımı olmuştur. Modül uzunlukları farklılaşan beş tasarım arasında, en yüksek yanlılık değerine sahip tasarımlar iki kestirim yöntemi için; dinamik yönlendirme yöntemi kullanıldığı zaman eşit modül uzunluğu (15-15-15) tasarımı, statik yönlendirme yöntemi kullanıldığı zaman artan modül uzunluğu (10-15-20) ve ikinci aşama en uzun (10-20-15) tasarımları olmuştur. Modül uzunlukları farklılaşan beş tasarım için de statik yönlendirme yönteminin kullanıldığı koşullar altında ilk aşamada daha kısa modüllerin kullanıldığı tasarımların ortalama yanlılık değerlerinin daha yüksek olduğu görülmektedir.

Şekil 18'deki grafik incelendiğinde dinamik yönlendirme koşulu altında MLE ve EAP son kestirim yöntemleri için grafiklerin benzer yapıda ve beş farklı modül uzunluğu

tasarımının ortalama yanlılık değeri birbirine çok yakın olduğu görülmektedir. Ancak statik yönlendirme koşullarından elde edilen yetnek kestirimleri için bu beş modül uzunluğu tasarımı için kestirim yöntemlerine bağlı benzer bir örüntü söz konusu değildir.



Şekil 18. Normal dağılımdan türetilmiş modül tasarımları için yanlılık değeri

Araştırmada, madde güçlük parametresi normal dağılımdan türetilmiş modüllerden oluşan çok aşamalı test tasarımlardan elde edilen RMSE ve yanlılık değerlerinin; modül uzunluğuna, son kestirim yöntemine ve yönlendirme yöntemine göre nasıl değiştiğini ve tasarımların birbirlerinden nasıl farklılaştığını belirlemek için varyans analizi (ANOVA) yapılmıştır. Varyans analizinden elde edilen bulgular Tablo 4'te verilmiştir.

Tablo 4 incelendiğinde RMSE değeri için, farklı modül uzunluklarından oluşan tasarımlar, kestirim yönteminin MLE veya EAP olması ve yönlendirme yönteminin statik veya dinamik olması arasındaki farklar ile bu koşulların karşılıklı olarak birbirleriyle etkileşimlerinden doğan farklar istatistiksel olarak anlamlı bulunmuştur. İstatistiksel anlamlılığın yanında gruplar arasındaki farklılığın ne kadar büyük olduğu, etki büyüklüğünün göstergesi olan eta kare (η^2) değeri ile hesaplanmıştır.

Tablo 4

Madde Güçlükleri Normal Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları

Tasarım Özellikleri	Sd	RMSE		Yanlılık	
		F	Eta kare (η^2)	F	Eta kare (η^2)
Modül Uzunluğu (MU)	4	1951,970*	0,313	98,288*	0,158
Kestirim Yöntemi (KY)	1	15057,950*	0,603	8,701*	0,003
Yönlendirme Yöntemi (YY)	1	11,650*	0,000	0,242	0,000
MU*KY	4	22,472*	0,004	1,430	0,002
MU*YY	4	2,795*	0,005	24,052*	0,039
KY*YY	1	16,318*	0,001	0,681	0,000
MU*KY*YY	4	0,816	0,000	0,862	0,001

* $p < 0,05$

Modül uzunluğunun, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,31$) yüksek düzeyde olduğu görülmektedir. Farklı aşamalarda farklı uzunluktaki modüllerden oluşan beş tasarımın birbirinden nasıl farklılaştığını görmek için Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre ilk aşamada 10 modül ile başlanan iki tasarım, artan (10-15-20) modül uzunluğu ve ikinci aşama en uzun (10-20-15) modül uzunluğu tasarımlarından elde edilen RMSE değerleri arasında istatistiksel olarak anlamlı farklılık olmadığı tespit edilmiştir. Diğer üç tasarım; azalan (20-15-10) modül uzunluğu, eşit (15-15-15) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımlarına ait RMSE değerlerinin, kendileri dışındaki dört tasarımın RMSE değerlerinden istatistiksel olarak anlamlı farklılaştığı görülmüştür. Kestirim yöntemlerinin (EAP ve MLE), RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,61$) yüksek düzeyde olduğu görülmektedir. 0,61 değeri ile en büyük etkiye sahip faktör kestirim yöntemi olmuştur. EAP ile yapılan kestirimlerin RMSE değerleri, MLE ile yapılan kestirimlerin RMSE değerlerinden anlamlı olarak daha düşüktür. Yönlendirme yönteminin RMSE değerlerinin farklılaşması üzerindeki etkisi ise 0.001'den küçüktür. Bu yüzden yönlendirme yönteminin bu farklılaşma üzerindeki etkisinden söz etmek pek mümkün değildir. Üç faktörün birbirleriyle olan karşılıklı etkileşimleri incelendiğinde ise tüm etkileşimlerin etki büyüklüğünün eşik değer olan 0,01'den düşük olduğu görülmektedir.

Yanlılık değerleri için ANOVA tablosu incelendiğinde, modül uzunluğu ve kestirim yöntemi faktörlerinin oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunurken, farklı yönlendirme yöntemlerinin yanlılık üzerinde oluşturduğu farkın

istatistiksel olarak anlamlı olmadığı görülmüştür. Bu üç faktörün karşılıklı olarak birbirleriyle olan etkileşimlerinin etkisine bakıldığında ise modül uzunluğu ve yönlendirme yöntemi faktörlerinin etkileşimlerinin yanlılık değeri üzerinde oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunmuştur. Diğer etkileşimlerin ise istatistiksel olarak anlamlı fark oluşturmadığı görülmektedir.

Etki büyüklükleri incelendiğinde modül uzunluğunun, yanlılık değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,16$) yüksek düzeyde olduğu, kestirim yönteminin ise etki büyüklüğünün ($\eta^2 = 0,003$) eşik değer olan 0,01'den de düşük düzeyde olduğu görülmektedir. Modül uzunluğu ve yönlendirme yönteminin etkileşiminden doğan farkın etki büyüklüğünün ($\eta^2 = 0,04$) ise düşük düzeyde olduğu bulunmuştur. Farklı aşamalarda farklı uzunluktaki modüllerden oluşan beş tasarımın birbirinden nasıl farklılaştığını görmek için ise Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre ilk aşamada 10 modül ile başlanan iki tasarım; artan (10-15-20) modül uzunluğu ile ikinci aşama en uzun (10-20-15) modül uzunluğu tasarımları arasında ve son aşamada 10 modül ile biten iki tasarım; azalan (20-15-10) modül uzunluğu ile ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımları arasında yanlılık değerleri üzerinde istatistiksel olarak anlamlı bir farklılaşmanın olmadığı tespit edilmiştir. Eşit (15-15-15) modül uzunluğu tasarımına ait yanlılık değerleri diğer tüm tasarımların yanlılık değerlerinden istatistiksel olarak anlamlı farklılaşmaktadır.

İkinci Alt Probleme İlişkin Bulgular ve Yorum

Araştırmanın ikinci alt problemi şu şekilde belirtilmiştir:

2. Çok aşamalı bireye uyarlanmış test uygulamasında, madde güçlülüklerinin uniform dağılımdan türetildiği 1-2-3 panel tasarımından elde edilen yetenek kestirimlerinin yanlılık (bias) ve RMSE değerleri; yönlendirme yöntemine (statik ve dinamik), modül uzunluklarına (artan, azalan, eşit, ikinci aşama en uzun ilk aşama en kısa ve ikinci aşama en uzun son aşama en kısa) ve son kestirim yöntemine göre (MLE ve EAP) nasıl değişmektedir?

İkinci, alt probleme ilişkin yapılan veri analizleri sonucunda, belirlenen koşullar altında beş farklı tasarımdan elde edilen yetenek kestirimlerine ilişkin RMSE ve yanlılık değerleri Tablo 5'da verilmiştir. RMSE ve yanlılık değerlerine ilişkin grafikler sırasıyla Şekil 19 ve Şekil 20'de sunulmuştur.

Tablo 5

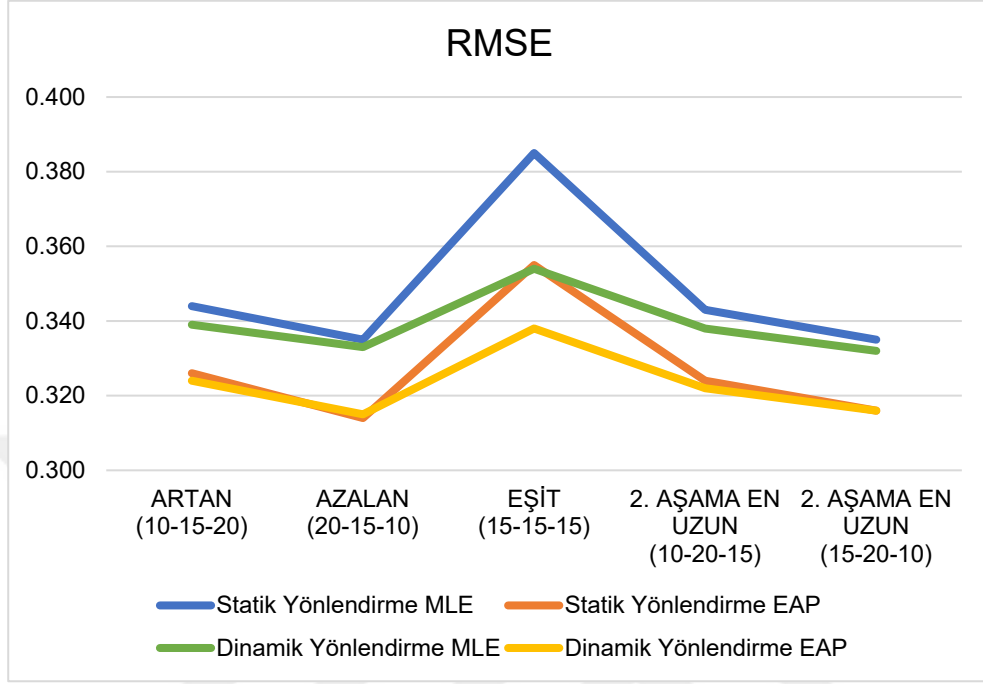
Madde Güçlükleri Uniform Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerleri

Modül Uzunluğu	Son Kestirim Yöntemi	Statik Yönlendirme		Dinamik Yönlendirme	
		RMSE	Yanlılık	RMSE	Yanlılık
Artan (10-15-20)	MLE	0,344	0,010	0,339	0,007
	EAP	0,326	0,010	0,324	0,007
Azalan (20-15-10)	MLE	0,335	0,005	0,333	0,000
	EAP	0,314	0,004	0,315	-0,001
Eşit (15-15-15)	MLE	0,385	0,005	0,354	0,002
	EAP	0,355	0,000	0,338	0,001
2. aşama en uzun (10-20-15)	MLE	0,343	0,007	0,338	0,004
	EAP	0,324	0,006	0,322	0,004
2. aşama en uzun (15-20-10)	MLE	0,335	0,001	0,332	0,002
	EAP	0,316	0,001	0,316	0,001

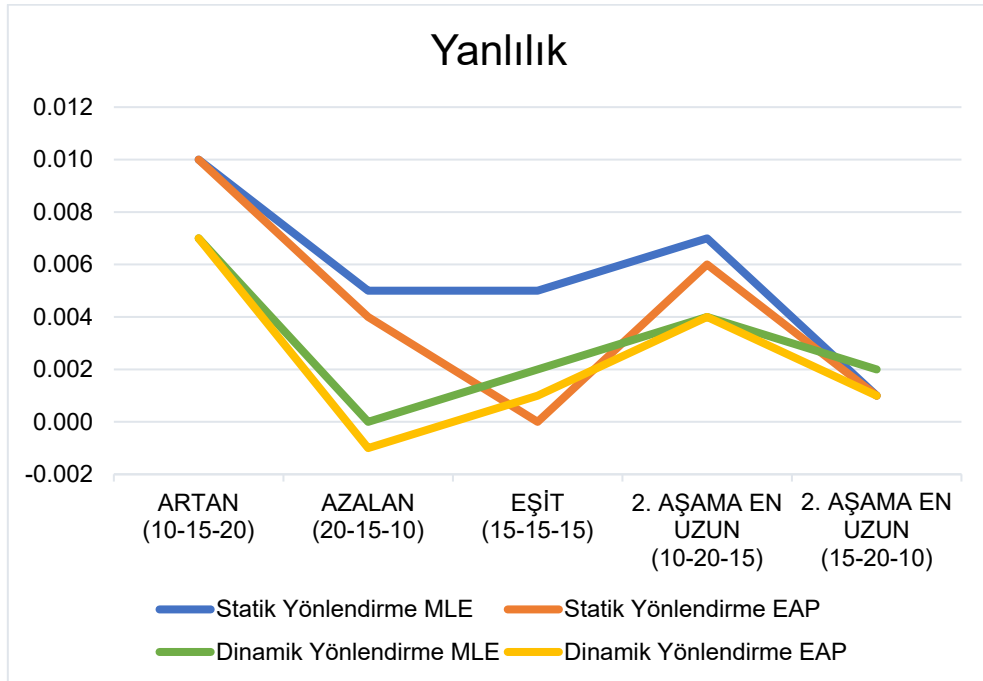
0,314 ile 0,385 arasında değişen RMSE değerlerinin en düşük olduğu tasarım, EAP son kestirim ve statik yönlendirme yöntemlerinin kullanıldığı azalan modül uzunluğu (20-15-10) tasarımı ve en yüksek olduğu tasarım, MLE son kestirim ve statik yönlendirme yöntemlerinin kullanıldığı eşit modül uzunluğu (15-15-15) tasarımı olmuştur. 1-2-3 panel tasarımlı çok aşamalı bireye uyarlanmış testte modüllerin güçlük dağılımlarının uniform dağılımdan türetildiği bu durumda, modül uzunlukları farklılaşan beş tasarım için de, tüm koşullar altında üçüncü aşamada daha kısa modüller kullanılmasının ortalama hatayı düşürdüğü görülmektedir. Modül uzunlukları farklılaşan beş tasarım arasında, dört koşul altında da en yüksek RMSE değeri eşit modül uzunluğu (15-15-15) olduğunda hesaplanmıştır. Tüm aşamalarda eşit sayıda madde olduğu durumda her koşulda ortalama hatanın yükseldiği görülmektedir.

Yönlendirme yöntemlerine göre bakıldığında beş modül uzunluğu için de, statik yöntemlerin (MLE ve EAP) ve dinamik yöntemlerin (MLE ve EAP) RMSE grafikleri benzer bir örüntüdedir (Şekil 19). Hem dinamik hem de statik yönlendirme yöntemleri için EAP ile yapılan son kestirimin RMSE değeri MLE ile yapılan son kestirimin RMSE değerinden düşük olarak hesaplanmıştır. Son kestirim yöntemi ayrı ayrı MLE ve EAP olduğunda statik ve dinamik yönlendirmeden elde edilen RMSE değerleri modül uzunluğu eşit (15-15-15) dışında neredeyse aynıdır. Eşit modül uzunluğu içi ise bu fark 0,03'tür. Modül madde güçlük dağılımları uniform dağılımdan oluşturulan tasarımlarda

yönlendirme yöntemine göre hata değerlerinin birbirine çok yakın değerler olduğu görülmektedir.



Şekil 19. Uniform dağılımdan türetilmiş modül tasarımları için RMSE değerleri



Şekil 20. Uniform dağılımdan türetilmiş modül tasarımları için yanlılık değerleri

Tablo 5'te yer alan ortalama yanlılık deęerleri incelendięinde genel olarak tm modeller iin yanlılık deęerlerinin olduka dřk olduęu ve -0.001 ile 0.010 arasında deęiřtięi grlmektedir. Son kestirim yntemi EAP, ynlendirme yntemi statik olduęunda eřit modl uzunluęu (15-15-15) tasarımı ve son kestirim yntemi MLE, ynlendirme yntemi dinamik olduęunda ise azalan modl uzunluęu (20-15-10) tasarımı iin yanlılık deęerleri sıfır olarak hesaplanmıřtır. Buna gre bu kořullar altında bu iki tasarım iin yansız hesaplamalar yapıldıęı sylenebilir. Modl uzunlukları farklılařan beř tasarım arasında, drt kořul altında da en yksek yanlılık deęeri artan modl uzunluęu (10-15-20) tasarımında grlmektedir. Modl uzunlukları farklılařan beř tasarım iin de tm kořullar altında ilk ařamada daha kısa modller kullanılması nın ortalama yanlılık deęrini arttırdıęı grlmektedir.

řekil 20'deki grafik incelendięinde dinamik ynlendirme kořulu altında MLE ve EAP son kestirim yntemleri iin grafiklerin benzer yapıda bazı tasarımlar iin de akıřık olduęu grlmektedir. Bu grafiklerin rntsne bakarak, dinamik ynlendirme yntemi kořullarında, ilk modldeki madde sayısı arttıka yanlılıęın dřtę gzlemlenebilir. Statik ynlendirme ynteminin kullanıldıęı kořullar iin ise byle bir rnt sz konusu deęildir.

Arařtırmada, madde glk parametresi uniform daęılımdan tretilmiř modllerden oluřan ok ařamalı test tasarımlardan elde edilen RMSE ve yanlılık deęerlerinin; modl uzunluęuna, son kestirim yntemine ve ynlendirme yntemine gre nasıl deęiřtięini tasarımların birbirlerinden nasıl farklılařtıęını belirlemek iin varyans analizi (ANOVA) yapılmıřtır. Varyans analizinden elde edilen bulgular Tablo 6'de verilmiřtir.

Tablo 6 incelendięinde RMSE deęeri iin, farklı modl uzunluklarından oluřan tasarımlar, kestirim ynteminin MLE veya EAP olması ve ynlendirme ynteminin statik veya dinamik olması arasındaki farklar ile bu kořulların karřılıklı olarak birbirleriyle etkileřimlerinden doęan farklar istatistiksel olarak anlamlı bulunmuřtur. İstatistiksel anlamlılıęın yanında gruplar arasındaki farklılıęın ne kadar byk olduęu, etki byklęnn gstergesi olan eta kare (η^2) deęeri ile hesaplanmıřtır.

Tablo 6

Madde Güçlükleri Uniform Dağılımdan Türetilmiş Modüllerden Oluşan Tasarımlar için RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları

Tasarım Özellikleri	Sd	RMSE		Yanlılık	
		F	Eta kare (η^2)	F	Eta kare (η^2)
Modül Uzunluğu (MU)	4	3000,9*	0,502	86,034*	0,140
Kestirim Yöntemi (KY)	1	7280,6*	0,304	12,881*	0,005
Yönlendirme Yöntemi (YY)	1	861,3*	0,036	51,927*	0,021
MU*KY	4	24,7*	0,004	2,620*	0,004
MU*YY	4	374,4*	0,063	12,453*	0,020
KY*YY	1	116,7*	0,005	2,672	0,001
MU*KY*YY	4	23,0*	0,004	1,802	0,003

* $p < 0,05$

Modül uzunluğunun, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,50$) yüksek düzeyde olduğu görülmektedir. 0,50 değeri ile en büyük etkiye sahip faktör modül uzunluğu olmuştur. Farklı aşamalarda farklı uzunluktaki modüllerden oluşan beş tasarımın birbirinden nasıl farklılaştığını görmek için Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre son aşamada 10 modül ile biten iki tasarım, azalan (20-15-10) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımlarından elde edilen RMSE değerlerinin arasında istatistiksel olarak anlamlı farklılık olmadığı tespit edilmiştir. Diğer üç tasarım; artan (10-15-20) modül uzunluğu, eşit (15-15-15) modül uzunluğu ve ikinci aşama en uzun (10-20-15) modül uzunluğu tasarımlarına ait RMSE değerlerinin, kendileri dışındaki dört tasarımın RMSE değerlerinden istatistiksel olarak anlamlı farklılaştığı görülmüştür. Kestirim yöntemlerinin (EAP ve MLE), RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,30$) yüksek düzeyde olduğu görülmektedir. EAP ile yapılan kestirimlerin RMSE değerleri, MLE ile yapılan kestirimlerin RMSE değerlerinden anlamlı olarak daha düşüktür. Yönlendirme yönteminin RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,04$) ise düşük düzeyde olduğu görülmektedir. Üç faktörün birbiri ile karşılıklı etkileşimlerinden doğan farkların etki büyüklükleri incelendiğinde, modül uzunluğu ve yönlendirme yöntemi faktörlerinin etkileşiminin etki büyüklüğünün ($\eta^2 = 0,06$) orta düzeyde olduğu bulunmuştur. Bunun dışında üç faktörün birbirleriyle olan etkileşimlerinden doğan farkların etki büyüklükleri

eta kare değerlerinin eşik değeri olan 0,01'den küçük olduğu görülmektedir. Bu da bu etkileşimlerin RMSE değerleri arasındaki farklara bir etkisi olmadığını göstermektedir.

Yanlılık değerleri için ANOVA tablosu incelendiğinde, modül uzunluğu, kestirim yöntemi ve yönlendirme yöntemi faktörlerinin ayrı ayrı oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunmuştur. Bu üç faktörün birbirleriyle karşılıklı etkileşimlerinin etkisine bakıldığında ise modül uzunluğu ile kestirim yöntemi ve modül uzunluğu ile yönlendirme yöntemi faktörlerinin etkileşimlerinin yanlılık değeri üzerinde oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunmuştur. Diğer etkileşimlerin ise istatistiksel olarak anlamlı fark oluşturmadığı görülmektedir.

Etki büyüklükleri incelendiğinde modül uzunluğunun, yanlılık değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2 = 0,14$) yüksek düzeyde olduğu, kestirim yönteminin etki büyüklüğünün ($\eta^2 = 0,005$) eşik değeri olan 0,01'den de düşük düzeyde olduğu ve yönlendirme yönteminin etki büyüklüğünün ($\eta^2 = 0,02$) ise düşük düzeyde olduğu görülmektedir. Modül uzunluğu ve yönlendirme yönteminin etkileşiminden doğan farkın etki büyüklüğünün ($\eta^2 = 0,02$) düşük düzeyde olduğu ve modül uzunluğu ve kestirim yönteminin etkileşiminden doğan farkın etki büyüklüğünün ise ($\eta^2 = 0,004$) eşik değeri olan 0,01'den de düşük düzeyde olduğu bulunmuştur.

Farklı aşamalarda farklı uzunluktaki modüllerden oluşan beş tasarımın birbirinden nasıl farklılaştığını görmek için ise Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre ilk aşamada 10 modül ile başlanan iki tasarım, artan (10-15-20) modül uzunluğu ve ikinci aşama en uzun (10-20-15) modül uzunluğu tasarımlarından elde edilen yanlılık değerleri tüm diğer tasarımlardan istatistiksel olarak anlamlı farklılaşmaktadır. Azalan (20-15-10) modül uzunluğu, eşit (15-15-15) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımlarından elde edilen yanlılık değerleri, istatistiksel olarak birbirinden farklılaşmamaktadır.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu bölümde araştırmanın bulgularından elde edilen sonuçlara, sonuçlara dayalı tartışma ve önerilere yer verilmiştir.

Sonuçlar

Bu çalışmada, madde güçlük parametreleri normal ve uniform dağılımdan türetilerek oluşturulan 1-2-3 panel tasarımlı çok aşamalı bireye uyarlanmış testlerin farklı modül uzunluğu, farklı kestirim yöntemi ve farklı yönlendirme yöntemi koşulları altındaki performansının, gerçek ve kestirilen yetenek düzeyleri üzerinden RMSE ve yanlılık değerleri hesaplanarak karşılaştırılması amaçlanmıştır. Bu amaç doğrultusunda da bulgulardan elde edilen sonuçlar alt problemler başlıkları altında sunulmaktadır.

Birinci Alt Probleme İlişkin Sonuçlar. Araştırmanın birinci alt probleminde madde güçlükleri normal dağılımdan türetilerek oluşturulan çok aşamalı bireye uyarlanmış test tasarımlarından elde edilen yetenek kestirimlerinin modül uzunluğu, kestirim yöntemi ve yönlendirme yöntemine göre nasıl değiştiğine ilişkin sonuçlar maddeler halinde aşağıda sunulmuştur.

- İkinci ve üçüncü aşamalardaki modül uzunluğu, ilk aşamadaki yönlendirme modülünden daha fazla olduğunda (en uzun modülün ikinci veya üçüncü aşamada olmasından bağımsız olarak) ortalama hata değerinin düştüğü görülmüştür.
- Her tasarımdaki toplam test uzunluğunun 45 madde ile sınırlandırıldığı bu modelde, her aşamada eşit modül uzunluğunda (15 madde) modüllerin kullanıldığı tasarımın ortalama hata değerinin her koşulda diğer tüm tasarımların ortalama hata değerlerinden daha yüksek olduğu saptanmıştır. Farklı aşamalarda daha kısa (10 madde) veya daha uzun (20 madde) modüllerin kullanılmasının ortalama hatayı düşürdüğü gözlenmiştir.
- Son yetenek kestirim yöntemi EAP olarak seçildiğinde tüm tasarımlar için ortalama hata değeri daha düşük olarak hesaplanmıştır.
- Madde güçlükleri normal dağılımdan türetilen modüllerden oluşan bu tasarımlar için bu çalışmanın koşulları altında, yönlendirme yönteminin statik veya dinamik olmasının ortalama hata üzerinde bir fark oluşturmadığı bulunmuştur.

- Statik yönlendirme yöntemi koşulunda ilk iki modülde üçüncü aşamaya göre daha uzun modüllerin olması ortalama yanlılık değerini daha çok düşürdüğü ve ayrıca ikinci ve üçüncü aşamada, ilk aşamaya göre daha uzun modüller kullanılmasının ise ortalama yanlılık değerini daha çok arttırdığı görülmüştür.
- Dinamik yönlendirme yöntemi koşulunda her aşama eşit (15 madde) modül uzunluğunda modüllerin kullanıldığı tasarımın ortalama yanlılık değerinin diğer tasarımların ortalama yanlılık değerlerinden daha yüksek olduğu saptanmıştır. Farklı aşamalarda daha kısa (10 madde) veya daha uzun (20 madde) modüllerin kullanılmasının ortalama yanlılığı düşürdüğü gözlenmiştir.
- Bu çalışma koşulları altında, son kestirim yönteminin EAP ya da MLE olmasının, farklı modül uzunluklarından oluşan tasarımların ortalama yanlılık değerlerindeki farklılaşmaya bir etkisi olmadığı görülmüştür.
- İlk aşamadaki modül uzunluğunun en kısa olduğu tasarımlar; artan (10-15-20) modül uzunluğu ve ikinci aşama en uzun (10-20-15) modül uzunluğu tasarımlarının ortalama hata ve yanlılık değerleri açısından birbirinden farklılaşmadığı bulunmuştur.
- Üçüncü aşamadaki modül uzunluğunun en kısa olduğu tasarımlar; azalan (20-15-10) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımlarının ortalama yanlılık değeri açısından birbirinden farklılaşmadığı bulunmuştur.

İkinci Alt Probleme İlişkin Sonuçlar. Araştırmanın birinci alt probleminde madde güçlükleri normal dağılımdan türetilerek oluşturulan çok aşamalı bireye uyarlanmış test tasarımlarından elde edilen yetenek kestirimlerinin modül uzunluğu, kestirim yöntemi ve yönlendirme yöntemine göre nasıl değiştiğine ilişkin sonuçlar maddeler halinde aşağıda sunulmuştur.

- İlk iki aşamalardaki modül uzunluğu, üçüncü aşamadaki son modülden daha fazla olduğunda (en uzun modülün, 20 maddelik modülün, birinci veya ikinci aşamada olmasından bağımsız olarak) ortalama hata değerinin düştüğü görülmüştür.
- Her tasarımdaki toplam test uzunluğunun 45 madde ile sınırlandırıldığı bu modelde, her aşamada eşit modül uzunluğunda (15 madde) modüllerin kullanıldığı tasarımın ortalama hata değerinin her koşulda diğer tüm tasarımların ortalama hata değerlerinden daha yüksek olduğu saptanmıştır. Farklı aşamalarda daha kısa (10

madde) veya daha uzun (20 madde) modüllerin kullanılmasının ortalama hatayı düşürdüğü gözlenmiştir.

- Son yetenek kestirim yöntemi EAP olarak seçildiğinde tüm tasarımlar için ortalama hata değeri daha düşük olarak hesaplanmıştır.
- Madde güçlükleri uniform dağılımdan türetilen modüllerden oluşan bu tasarımlar için yönlendirme yönteminin statik veya dinamik olmasının ortalama hata üzerinde oluşturduğu farkın çok düşük olduğu bulunmuştur. Bu farkın en fazla olduğu eşit (15-15-15) modül uzunluğu tasarımında ise, statik ve dinamik yönlendirme yöntemleri için RMSE değerleri farkı 0,03 düzeyinde hesaplanmıştır.
- İlk aşamada, ikinci ve üçüncü aşamadan daha kısa modüller kullanılması ortalama yanlılığı daha da arttırdığı görülmüştür.
- Dinamik yönlendirme yöntemi altındaki koşullarda ilk modüldeki madde sayısı arttıkça ortama yanlılık değerinin daha çok düştüğü gözlenmiştir.
- Bu çalışma koşulları altında, son kestirim yönteminin EAP ya da MLE olmasının, dinamik yönlendirme yöntemi koşulunda farklı modül uzunluklarından oluşan tasarımların yanlılık değerlerindeki farklılaşmaya bir etkisi olmadığı bulunmuştur.
- Statik yönlendirme yöntemi koşulunda, son kestirim yönteminin EAP ya da MLE olmasının, farklı modül uzunluklarından oluşan tasarımların ortalama yanlılık değerlerindeki farklılaşmayı az da olsa etkilediği saptanmıştır. Diğer tüm tasarımlardaki son kestirim yöntemine göre yanlılık değerleri eşit veya fark 0,001 düzeyinde hesaplanırken, eşit (15-15-15) modül uzunluğu tasarımında bu fark 0,005 düzeyinde hesaplanmıştır.
- Son aşamadaki modül uzunluğunun en kısa olduğu (10 madde) tasarımlar; azalan (20-15-10) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımları ortalama hata ve yanlılık değerleri açısından birbirinden farklılaşmamaktadır.
- Eşit (15-15-15) modül uzunluğu tasarımı, son aşamadaki modül uzunluğunun en kısa (10 madde) olduğu tasarımlar; azalan (20-15-10) modül uzunluğu ve ikinci aşama en uzun (15-20-10) modül uzunluğu tasarımlarının birbirleri ile ortalama yanlılık değeri açısından farklılaşmadığı bulunmuştur.

Tartışma

İçinde bulunduğumuz dijital çağda, okullar, öğretmenler ve araştırmacılar tarafından eğitim faaliyetlerini desteklemek ve geliştirmek için yenilikçi teknolojiler kullanılmaya başlanmıştır (Yan, 2020). Test uygulayıcıları için de, bu gelişmelere ayak uydurmak ve eğitim alanındaki modernleşmenin bir göstergesi olan bilgisayar ortamında bireye uyarlanmış testleri kullanmak kaçınılmaz olmuştur (Breithaupt, Zang & Hare, 2014). Eğitimde ölçme ve değerlendirmenin etkinliğini arttırmak, daha doğru ölçme ve değerlendirme yapabilmek için öğrenciler hakkında daha çok bilgi edinmeye, bireye uyarlanmış tekniklerin kullanılmasına ve yenilikçi yöntemlerin geliştirilmesine ihtiyaç duyulur (Yan, 2020). Bu ihtiyacı karşılamak için de klasik lineer testlerin yerini bilgisayar ortamında uygulanan madde düzeyinde ve çok aşamalı bireye uyarlanmış testlerin aldığı görülmektedir. Yapılan araştırmalar sonucunda çok aşamalı bireye uyarlanmış testlerin klasik lineer testlerin avantajları ile madde düzeyinde bireye uyarlanmış testlerin (BOBUT) avantajlarını dengeli bir şekilde birleştiren yeni bir model ortaya koyduğu görülmektedir (Hendrickson ,2007). Bir başka deyişle, çok aşamalı bireye uyarlanmış testler hem bireye uyarlanma düzeyi hem pratiklik hem ölçme kesinliği hem de test formları üzerindeki kontrol açısından en dengeli modeli sunmaktadır (Zenisky, Hambleton & Luecht, 2010). Bu durum da araştırmacıları, çok aşamalı bireye uyarlanmış testler üzerinde daha çok çalışmaya ve uygulama yapmaya yönlendirmiştir.

Günümüzde uluslararası düzeyde büyük test merkezlerinin (ör. ETS) yaptığı sınavlar (GRE, TOEFL) ve Ekonomik İş Birliği ve Kalkınma Örgütü OECD'nin yaptığı testler (PISA, PIAAC) ile birlikte daha küçük ölçekli mesleki lisans sınavları (CPA, LSAT, vb.) da çok aşamalı bireye uyarlanmış test formatında hazırlanıp uygulanmaktadır. Önemli kararların verildiği geniş ölçekli ve yüksek beklenti testler özelinde yapılan çalışmalar, hem test gerekliliklerini sağlamak hem de testi alacak bireylerin yaşayacağı deneyimin daha az kaygı ve stresle kolay anlaşılır ve kabul edilebilir olmasını sağlamak açısından en basit ve en etkili tasarımın çok aşamalı bireye uyarlanmış testler olduğunu ortaya koymaktadır (Melican, Breithaupt & Zhang, 2010; Oranje, Mazzeo, Xu & Kulick, 2014; Robin, Steffen & Liang, 2014; Wang, 2017; Yamamoto, Shin & Khorramdel, 2019; Zheng, Nozawa, Gao & Chang, 2012). Bu anlamda hem gelişen dünyaya ayak uydurmak hem de daha kesin ölçme sonuçlarına çok daha hızlı puanlama yaparak ulaşmak amacıyla Türkiye'de uygulanan önemli

kararların verildiği merkezi sınavlar için de çok aşamalı bireye uyarlanmış testlerin uygulanabilirliğini araştırmanın önemli olduğu düşünülmektedir.

Çok aşamalı bireye uyarlanmış test tasarımlarının farklı koşullar altında etkinliğinin araştırıldığı birçok farklı çalışma yapılmıştır. Ancak bu çalışmaların ötesinde araştırmacılar, aşama sayısı, modül sayısı, modül uzunluğu ve kullanılacak yöntemler gibi kararların verildiği çok aşamalı test tasarımı sürecinin hala oldukça karmaşık olduğu ve önemli olduğunu vurgulamaktadırlar (Luo & Kim, 2018; Zenisky, Hambleton & Luecht, 2010). Bu karmaşıklık bir anlamda her testin istenilen özelliklerine en uygun çok aşamalı bireye uyarlanmış test tasarımı oluşturmak için bir avantaj olarak da görülmektedir. Bu anlamda, çok aşamalı bireye uyarlanmış test olarak uygulanması tasarlanan sınavlar için hangi test tasarımının en uygun olacağına ilişkin çalışmalar yapılmalıdır. Bu çalışmada da Türkiye’de uygulanan üniversiteye geçiş sınavlarının bir alt testinin çok aşamalı bireye uyarlanmış test olarak tasarlanması ve modüllerin madde güçlük dağılımları farklılaştırarak, modül uzunluğu, yönlendirme yöntemi ve kestirim yöntemi koşulları altında oluşturulan tasarımların etkinliğinin değerlendirilmesi amaçlanmıştır.

Farklı aşamalarda farklı uzunlukta modüllerin kullanılmasının farklı gerekçeleri vardır (Patsula, 1999). Bireyin yetenek düzeyini başlangıçta daha doğru kestirmek ve olası yanlış yönlendirmelerin önüne geçmek için ilk aşamadaki yönlendirme modülünün diğer aşamadaki modüllerden daha uzun olması gerektiği düşünülmektedir (Kim & Plake, 1993). Bir başka düşünce de son kesitirime yaklaştıkça daha doğru sonuçlar elde edebilmek için son aşamalarda daha uzun modüller kullanılması gerektiğini savunmaktadır (Zenisky, Hambleton & Luecht, 2010; Zheng, Nozawa, Gao & Chang, 2012). Bu çalışmada da madde güçlükleri normal dağılımdan türetilen modüller ve uniform dağılımdan türetilen modüllerden oluşturulan tasarımlar için farklı sonuçlar elde edilmiştir. Normal dağılımdan türetilen, ilk modülün en kısa, ikinci ve üçüncü modüllerin daha uzun olduğu tasarımların ortalama hata değerleri Zenisky, Hambleton ve Luecht (2010) ve Zheng, Nozawa, Gao ve Chang (2012)’in görüşlerine paralel olarak diğer tasarımlardan daha düşük bulunmuştur. Öte yandan uniform dağılımdan türetilen, ilk iki modülün daha uzun ve üçüncü modülün en kısa olduğu tasarımların ortalama hata değerleri diğer tasarımlardan daha düşük çıkmıştır. Bu sonuç da Kim ve Plake (1993) ve Bozuncü Öztürk (2019)’ün çalışmalarının sonuçlarıyla paralellik taşımaktadır. Farklı güçlük aralıklarından farklı dağılımlardan türetilen

modüllerle oluşturulan tasarımlar farklı sonuçlar vermiştir. Madde güçlük dağılımlarının normal dağılımdan türetildiği tasarımda ilk aşamadan son aşamaya madde güçlük ranjları genişten dar doğru gitmektedir. Birinci aşama için standart sapma değeri 1, ikinci aşama için 0.6 ve üçüncü aşama için 0.3 olarak belirlenmiştir. Daha dar madde güçlük ranjlarından türetilen modüllerin daha uzun olduğu tasarımların ortalama hatası daha düşük bulunmuştur. Normal dağılım üzerinden her aşamadaki modüllerin güçlük ranjlarını genişleterek ve daraltarak oluşturduğu tasarımları karşılaştırdığı çalışmada Pohl (2013), modüllerin madde güçlük ranjı dar tasarımların geniş tasarımlara göre daha iyi sonuçlar verdiğini bulmuştur. Bu anlamda bu sonuç araştırmanın bulguları ile de paralellik taşımaktadır. Alan yazınındaki farklı görüşler ve bu çalışmanın sonuçları farklı tasarımlardan farklı sonuçların elde edileceğinin örnekleri olarak sunulabilir.

MTK'da bireylerin yeteneklerini kestirmek için farklı yöntemler kullanılmaktadır. Yetenek kestirimi için seçilen farklı kestirim yöntemlerinin, MTK uygulamalarında hesaplanan puanların psikometrik özelliklerini manidar bir şekilde etkilediği görülmektedir (Kolen & Tong, 2010). Bu yüzden Kim, Moses ve Yoo (2015), bireyin raporlanan son puanını doğrudan etkileyen bu koşulun, pratik uygulamalardaki etkisinin araştırılması gerektiğini savunmakta ve çok aşamalı bireye uyarlanmış testler için farklı kestirim yöntemlerinin etkinliğini araştırmaktadır. Bu çalışmada da hem normal hem de uniform dağılımdan oluşturulan tasarımların hepsinde EAP ile yapılan kestirimlerin ortalama hata değeri MLE ile yapılan kestirimlerin ortalama hata değerinden düşüktür. MLE yönteminde kestirim, olabilirlik fonksiyonunu en yüksek aldığı değere dayanırken, EAP yönteminde ise sonsal dağılımın ortalamasına dayanır. Çalışma için türetilen yetenek parametreleri -3 ile +3 aralığında standart normal dağılımdan türetilmiştir. Normal dağılım altında, EAP yönteminin düşük hata kareleri ortalaması ile kestirim yaptığı bilinmektedir (Bock & Mislevy, 1982). MLE yöntemi, maddelerin hepsine doğru veya hepsine yanlış cevap verildiği durumlarda çalışmamaktadır. Bu çalışmada da uç değerlerdeki yetenek değerleri için bu koşulların oluşmuş olması MLE ile yapılan kestirimlerde hatanın yükselmesine sebep olmuş olabilir. Hem yetenek dağılımının normal dağılımdan türetilmiş olması hem de MLE yönteminin uç değerlerdeki dezavantajları EAP ile yapılan kestirimlerin hata değerlerinin daha düşük çıkmasını sağlamış olabilir. Öte yandan MLE ve EAP yöntemlerinin bu çalışmada yanlılık açısından benzer sonuçlar verdiği gözlenmiştir.

Bock ve Mislevy (1982) gvenirliklerin aynı olduėu, aynı yetnek dzeylerinde EAP ve MLE kestirim yntemlerinin yanlılık deėerlerinin yaklařık olarak aynı olduėunu vurgulamaktadır. Bu alıřmada da nırmal daėılımdan tretilen yetnekler altında MLE ve EAP yntemlerini yanlılık deėerlerinin benzer olması beklenen bir durum olmuřtur.

Ynlendirme, ok ařamalı bireye uyarlanmış testlerde bireye uyarlamanın yapıldıėı, bireyin panel boyunca izleyeceėi yolu belirleyen en nemli kararlarından biridir. Bu yzden, alan yazınında sıklıkla farklı tasarımlar iin ve farklı kořullar altında ynlendirme yntemlerini karřılařtıran alıřmalarla oka karřılařılmaktadır. Bu alıřmada kullanılan iki yntem de MTK'ya baėlı olarak belirlenen yntemlerdir. Statik olarak adlandırılan yntemde, modl bilgi fonksiyonları kullanılarak doėru cevap sayısı cinsinden bir kesme puanı, uygulama ncesinde belirlenir ve bireyin doėru cevap sayısına gre ynlendirme yapılır. Dinamik olarak adlandırılan yntemde ise test uygulaması anında bireyin verdiėi cevaplar zerinden yeteneėi kestirilir ve kestirilen yeteneėe gre bir sonraki ařamada en ok bilgiyi veren modle ynlendirme yapılır. Bu alıřmada, madde glkleri normal daėılımdan tretilen modllerden oluřan tasarımlarda ynlendirme ynteminin statik veya dinamik olması yetenek kestiriminde bir fark yaratmamıřtır. Her bir modlde yer alan maddelerin madde glklerinin normal daėılımdan tretilmesi ideale yakın kořulları saėladıėından statik ve dinamik ynlendirme yntemlerinin hesaplımalarında benzer sonuların ıkmasına sebep olmuřtur. Uniform daėılımdan tretilen modllerden oluřan tasarımlarda ise ynlendirme ynteminin statik veya dinamik olması dřk de olsa bir fark oluřurmaktadır. Dinamik ynlendirme yntemi dřk dzeyde de olsa statik ynlendirme ynteminden daha az hatayla kestirim yapmaktadır. Bu farkın en fazla olduėu eřit (15-15-15) modl uzunluėu tasarımında bu fark 0.003 dzeyindedir. Armstrong (2002), on drt modlden oluřan altı ařamalı bir ok ařamalı test tasarımı iin statik ve dinamik farklı ynlendirme yntemlerini test etmiř ve bu yntemler arasında anlamlı bir fark bulamamıřtır. Benzer řekilde, Weissman, Belov ve Armstrong (2007), bireyleri sınıflandırma amacıyla tasarladıkları ok ařamalı bireye uyarlanmış test iin farklı ynlendirme yntemlerinin benzer sonular verdiėi sonucuna ulařmıřlardır. Svetina, Liav, Rutowski ve Rutkowski (2019) madde glkleri unifrom daėılımdan tretilen modllerden oluřan ok ařamalı test tasarımlarında dinamik ynlendirme yntemlerinin statik ynlendirme yntemlerinden az daha olsa daha iyi

sonular verdiđini vurgulamaktadır. Bu anlamda bu alıřmadan elde edilen bulgular da alan yazınındaki alıřmaların bulguları ile paralellik gstermektedir.

Bu alıřmada her sene sayısı giderek artarak, bir milyondan fazla bireyin katıldıđı niversiteye geiř sınavlarının bir alt testi temel alınarak ok ařamalı bireye uyarlanmış test tasarımların farklı kořullardaki performansı incelenmiř ve bu alıřma ile sınırlı kořullar altında en iyi modelin ne olabileceđi tartıřılmıřtır. Test uzunluđunun 45 madde ile sınırlı tutulduđu bu modellerde modl uzunluđuna bađlı olarak deđiřen tasarımların modllerin madde glk dađılımlarından etkilendiđi sylenebilir. Svetina, Liav, Rutowski ve Rutkowski (2019)'nin geniř lekli bir ok ařamalı bireye uyarlanmış test simlasyonu olarak farklı tasarımları farklı kořullar altında karřılařtırdıkları alıřmalarında, her kořulda en iyi alıřan bir yntemin olmadıđına dikkat ekilmektedir. Bu alıřmada da farklı kořullar altında en iyi sonucu veren tasarımların farklılařtıđı grlmřtr. ok ařamalı bireye uyarlanmış testlerin, test gerekliliklerine ve uygulayıcıların amalarına gre tasarlanabilecek esneklikte olduđu yapılan alıřmalardan da grlebilmektedir. Bu esneklik avantajının, test uygulayıcılarının kendi amalarına gre daha nce kâđıt kalem formatında klasik lineer testler olarak uyguladıkları testleri ok ařamalı bireye uyarlanmış test olarak uygulamalarına olanak sađlayacak nitelikte olduđu sylenebilir. Kâđıt kalem uygulamalarına gre testi hazırlama ařamasında daha ok emek harcanacađı beklendik bir durumdur ancak ok ařamalı bireye uyarlanmış test uygulamasıyla, test btnn tamamen kontrol edilebiliyor olması ve daha dođru lcmeler yapılarak test sonuları anında puanlanabiliyor olmasının nemli avantajlar sađlayacađı da gz ardı edilmemelidir. Bu kadar nemli kararların alındıđı merkezi sınavlar iin ok ařamalı bireye uyarlanmış test uygulamalarına geilmesinin birok aıdan avantaj sađlayacađı dřnlmektedir. Bu anlamda uygulayıcılara ve ileri arařtırmalara ynelik neriler sonraki blmde sunulmuřtur.

neriler

Uygulayıcılara ynelik neriler.

- Toplam test uzunluđunun 45 madde ile sınırlandırıldıđı bu alıřmada, farklı kořullar altında farklı ařamalarda deđiřen uzunluktaki modllerin (artan, azalan, ikinci ařama en uzun) bulunduđu, bir bařka deyiřle herhangi bir ařamada 20 madde uzunluđunda modllerin bulunduđu tasarımların diđer

tasarımlara göre daha az hata ile yetenek kestirimi yaptığı görülmüştür. Bu anlamda eşit modül uzunluğu tasarımları yerine değişen modül uzunluğu tasarımlarının ve herhangi bir aşamada en az 20 maddelik modüllerin olduğu tasarımların kullanılması önerilebilir.

- EAP kestirim yönteminin her koşulda MLE kestirim yönteminden daha az hata ile kestirim yaptığı ancak ortalama yanlışlık değeri açısından iki kestirim yöntemi arasında bir fark olmadığı bulunmuştur. Bu yüzden çalışmanın veri seti de göz önünde bulundurularak EAP kestirim yönteminin son kestirim yöntemi olarak seçilmesi önerilebilir.
- Alan yazınındaki çalışma sonuçlarıyla da desteklenen dinamik ve statik yöntemler arasındaki farkların ihmal edilebileceği sonucu ve statik yöntemin sunduğu uygulama kolaylığı avantajı da düşünüldüğünde, yönlendirme yöntemi olarak önceden MTK' ya dayalı olarak belirlenen doğru cevap sayısının kesme puanı olarak kullanılması önerilebilir.

İleri araştırmalara yönelik öneriler.

- Bu çalışmada simülasyon verisi ile test edilen koşullar, uygulanacak testin amaçlarına ve gerekliliklerine uygun olarak tasarlanarak gerçek veri seti ile de test edilebilir.
- Bu çalışmada, standart normal dağılımdan türetilen gerçek yetenek düzeylerindeki bireylerin farklı koşullar altında farklı tasarımlardan elde edilen yetenek kestirimlerinin kesinliği tüm grubun ortalaması alınarak karşılaştırılmıştır. Bu tasarımların performanslarının, farklı yetenek grupları için nasıl farklılaştığı da araştırılabilir.
- Bu çalışmada, normal dağılımdan ve uniform dağılımdan türetilen iki ayrı madde havuzundan oluşturulan panel tasarımları tek bir panel kullanılarak karşılaştırılmıştır. Madde havuzu genişletilerek, farklı paneller geliştirilebilir ve panel atamaları da çalışmaya dahil edilebilir.
- Farklı güçlük dağılımlarından oluşturulan tasarımların, modül uzunluğu, yönlendirme yöntemi ve kestirim yöntemi koşulları altındaki performansını araştırmayı amaçlayan bu çalışmaya içerik dengesi ve madde ifşası konuları dahil edilmemiştir. İçerik dengesi ile ilgili çalışmalar yapılabilir ve

farklı tasarımların madde ifşasını nasıl etkilediğine yönelik çalışmalar yapılabilir.

- Aynı koşullar altında daha farklı madde güçlük dağılımları kullanılarak farklı sonuçlar elde edilebilir. Bu tasarımlar daraltılmış ve genişletilmiş madde güçlük aralıkları altında tekrar test edilebilir.
- Bu çalışma, gerçek uygulamalar temel alınarak 45 madde ile sınırlandırılmıştır. Farklı test uzunlukları ile oluşturulacak tasarımların sonuçları, avantaj ve dezavantajları da değerlendirilebilir.
- Bu çalışma, madde güçlük dağılımları, modül uzunluğu, yönlendirme yöntemi ve kestirim yöntemi koşulları altında farklı tasarımları ele almıştır. Alan yazınından görüldüğü gibi en iyi performansı gösteren tek bir tasarımdan söz edilememektedir. O yüzden farklı koşullar altında farklı tasarımların performansları çalışmaya devam edilmeli, uygulanacak testlerin amacına en uygun tasarımı belirlemek için çalışmalar yapılmalıdır.

Kaynaklar

- AERA, APA, & NCME. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Akdemir, O., & Oguz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education*, 51(3), 1198-1204.
- Ariel, A., Veldkamp, B., & Breithaupt, K. (2006). Optimal Testlet Pool Assembly for Multistage Testing Designs. *Applied Psychological Measurement*, 30(3), 204-215.
- Armstrong, R. D. (2002). *Routing Rules for Multiple-Form Structures* (LSAC Research Report Series: Computerized Testing Report 02-08). Law School Admission Council.
- Armstrong, R. D., Jones, D. H., Koppel, N. B., & Pashley, P. J. (2004). Computerized Adaptive Testing With Multiple-Form Structures. *Applied Psychological Measurement*, 28(3), 147-164.
- Atkinson, R. (2001). *Standardized Tests and Access to American Universities*. <https://escholarship.org/uc/item/6182126z>
- Aybek, E. C., & Demirtaşlı, R. N. (2014). Bilgisayar Ortamında ve Kağıt-Kalem Formunda Uygulanan Genel Yetenek Testinin Psikometrik Özelliklerinin Karşılaştırılması. *İlköğretim Online*, 13(4).
- Aybek, E. C. (2016). *Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) Olarak Uygulanabilirliğinin Araştırılması*. Ankara Üniversitesi, Ankara.
- Baker, F. B., & Kim, S.H. (2004). *Item Response Theory: Parameter Estimation Techniques, Second Edition* (2nd edition). CRC Press.

- Baykul, Y. (2010). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması* (2. bs). Ankara: Pegem Akademi.
- Belov, D. I., & Armstrong, R. D. (2008). A Monte Carlo Approach to the Design, Assembly, and Evaluation of Multistage Adaptive Tests. *Applied Psychological Measurement*, 32(2), 119-137.
- Binet, A., & Simon, Th. (1915). A Method of Measuring the Development of the Intelligence of Young Children. İçinde *A method of measuring the development of intelligence of young children* (ss. 1-107). Chicago Medical Book Company.
- Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. İçinde F. M. Lord & M. R. Novick (Ed.), *Statistical Theories of Mental Test Scores* (ss. 397-479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444.
- Bozunç Öztürk, N. (2014). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin İncelemesi*. Hacettepe Üniversitesi, Ankara.
- Bozunç Öztürk, N. (2019). *How the Length and Characteristics of Routing Module Affect Ability Estimation in ca-MST?*
- Breithaupt, K., Zang, O. Y., & Hare, D. R. (2014). The multistage testing approach to the AICPA uniform certified public accounting examinations. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 343-354). Chapman and Hall/CRC.
- Breithaupt, K., & Hare, D. R. (2007). Automated Simultaneous Assembly of Multistage Testlets for a High-Stakes Licensing Examination. *Educational and Psychological Measurement*, 67(1), 5-20.

- Brossman, B. G., & Guille, R. A. (2014). A Comparison of Multi-Stage and Linear Test Designs for Medium-Size Licensure and Certification Examinations. *Journal of Computerized Adaptive Testing*, 2(3), 18-36.
- Bulut, O., & Kan, A. (2012). Application of Computerized Adaptive Testing to Entrance Examination for Graduate Studies in Turkey. *Eurasian Journal of Educational Research*.
- Bulut, O., & Sünbül, Ö. (2017). R Programlama Dili ile Madde Tepki Kuramında Monte Carlo Simülasyon Çalışmaları. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 266-287.
- Chen, S.-K., Hou, L., Fitzpatrick, S. J., & Dodd, B. G. (1997). The Effect of Population Distribution and Method of Theta Estimation on Computerized Adaptive Testing (CAT) Using the Rating Scale Model. *Educational and Psychological Measurement*, 57(3), 422-439.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). Routledge.
- Colvin, K. F., Keller, L. A., & Robin, F. (2016). Effect of Imprecise Parameter Estimates on Ability Estimates in a Multistage Test in an Automatic Item Generation Context. *Journal of Computerized Adaptive Testing*, 4(1), Article 1.
- Crotts, K. M., Zenisky, A. L., Sireci, S. G., & Li, X. (2013). Estimating Measurement Precision in Reduced-Length Multistage-Adaptive Testing. *Journal of Computerized Adaptive Testing*, 1(0), 67-87-87.

- De Ayala, R. J. (2008). *The Theory and Practice of Item Response Theory*. The Guilford Press.
- DeMars, C. (2010). *Item Response Theory*. Oxford University Press.
- Educational Testing Service. (2019). *Introduction of multistage adaptive testing design in PISA 2018* (Paper OECD Education Working Paper No. 209).
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory*. Psychology Press.
- Eroğlu, M. G. (2013). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının Ölçme Kesinliği ve Test Uzunluğu Açısından Karşılaştırılması*. Hacettepe Üniversitesi, Ankara.
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49.
- Flaugher, R. (2000). Item Pools. İçinde H. Wainer (Ed.), *Computerized Adaptive Testing: A Primer* (2nd Edition, ss. 41-64). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Springer Science & Business Media.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory* (1st edition). SAGE Publications, Inc.
- Hambleton, R. K., & Xing, D. (2006). Optimal and Nonoptimal Computer-Based Test Designs for Making Pass-Fail Decisions. *Applied Measurement in Education*, 19(3), 221-239.
- Han, K., T., & Kosinski, M. (2014). Software Tools for Multistage Testing Simulations. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 411-420). Chapman and Hall/CRC.

- Harwell, M., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement*, 20(2), 101-125.
- Hendrickson, A. (2007). An NCME Instructional Module on Multistage Testing. *Educational Measurement Issues and Practice*, 26(2), 44-52.
- Jodoin, M. G., Zenisky, A., & Hambleton, R. K. (2006). Comparison of the Psychometric Properties of Several Computer-Based Test Designs for Credentialing Exams With Multiple Purposes. *Applied Measurement in Education*, 19(3), 203-220.
- Kalender, İ. (2011). *Effects of Different Computerized Adaptive Testing Strategies on Recovery of Ability*. Middle East Technical University, Ankara.
- Kezer, F. (2013). *Bilgisayar Ortamında Bireye Uyarlanmış Test Stratejilerinin Karşılaştırılması*. Ankara Üniversitesi, Ankara.
- Khalilzadeh, J., & Tasci, A. D. A. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, 62, 89-96.
- Kim, D. H., & Plake, B. S. (1993). *Monte Carlo Simulation Comparison of Two-Stage Testing and Computerized Adaptive Testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Atlanta, GA.
- Kim, D. H., & Huynh, H. (2007). Comparability of Computer and Paper-and-Pencil Versions of Algebra and Biology Assessments. *The Journal of Technology, Learning and Assessment*, 6(4).
- Kim, J., Chung, H., Dodd, B. G., & Park, R. (2012). Panel Design Variations in the Multistage Test Using the Mixed-Format Tests. *Educational and Psychological Measurement*, 72(4), 574-588.

- Kim, J., Chung, H., Park, R., & Dodd, B. G. (2013). A comparison of panel designs with routing methods in the multistage test with the partial credit model. *Behavior Research Methods*, 45(4), 1087-1098.
- Kim, S., Moses, T., & Yoo, H. (2015). A Comparison of IRT Proficiency Estimation Methods Under Adaptive Multistage Testing. *Journal of Educational Measurement*, 52(1), 70-79.
- Kirsch, I., & Lennon, M. L. (2017). PIAAC: A new design for a new era. *Large-scale Assessments in Education*, 5(1), 11.
- Kolen, M. J., & Tong, Y. (2010). Psychometric Properties of IRT Proficiency Estimates. *Educational Measurement: Issues and Practice*, 29(3), 8-14.
- Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*, 28(4), 612-625.
- Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika*, 36, 227-242.
- Lord, F. M. (1971). The Self-Scoring Flexilevel Test. *Journal of Educational Measurement*, 8(3), 147-151.
- Lord, F. M. (1980). *Applications of Item Response Theory To Practical Testing Problems*. Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Luecht, R. (2014). Design and Implementation of Large-Scale Multistage Testing System. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 69-83). Chapman and Hall/CRC.

- Luecht, R. M., & Nungester, R. J. (1998). Some Practical Examples of Computer-Adaptive Sequential Testing. *Journal of Educational Measurement*, 35(3), 229-249.
- Luecht, R. M., & Burgin, W. (2003). *Test Information Targeting Strategies for Adaptive Multistage Testing Designs* [Paper]. The Annual Meeting of National Council on Measurement in Education (NCME), Chicago, IL.
- Luecht, R., Brumfield, T., & Breithaupt, K. (2006). A Testlet Assembly Design for Adaptive Multistage Tests. *Applied Measurement in Education*, 19(3), 189-202.
- Luo, X., & Kim, D. (2018). A Top-Down Approach to Designing the Computerized Adaptive Multistage Test. *Journal of Educational Measurement*, 55(2), 243-263.
- Magis, D., & Raîche, G. (2012). Random Generation of Response Patterns under Computerized Adaptive Testing with the R Package catR. *Journal of Statistical Software*, 48(1), 1-31.
- Magis, D., & Barrada, J. R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software*, 76(1), 1-19.
- Magis, D., Yan, D., & von Davier, A. (2017). *Computerized Adaptive and Multistage Testing with R: Using Packages catR and mstR*. Springer International Publishing.
- Magis, D., Yan, D., & von Davier, A. (2018). mstR: Procedures to Generate Patterns under Multistage Testing. R package version 1.2.
- McClelland, T., & Cuevas, J. (2020). A comparison of computer based testing and paper and pencil testing in mathematics assessment. *The Online Journal of New Horizons in Education*, 10(2), 12.
- Mead, A. D. (2006). An Introduction to Multistage Testing. *Applied Measurement in Education*, 19(3), 185–187.

- Melican, G. J., Breithaupt, K., & Zhang, Y. (2010). Designing and Implementing a Multistage Adaptive Test: The Uniform CPA Exam. İçinde W. J. van der Linden & C. A. W. Glas (Ed.), *Elements of Adaptive Testing* (ss. 167–189). New York: Springer.
- Molenaar, I. W. (1995). Some Background for Item Response Theory and the Rasch Model. İçinde G. H. Fischer & I. W. Molenaar (Ed.), *Rasch Models: Foundations, Recent Developments, and Applications* (ss. 3-14). Springer.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological Testing: Principles and Applications*. Pearson/Prentice Hall.
- OECD. (2016). *Technical Report of the Survey of Adult Skills (PIAAC)*. OECD Publishing.
- Oranje, A., Mazzeo, J., Xu, X., & Kulick, E. (2014). A Multistage Testing Approach to Group-Score Assessments. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 371-390). Chapman and Hall/CRC.
- Owen, R. J. (1975). A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing. *Journal of the American Statistical Association*, 70(350), 351-356.
- Özdemir, B. (2015). *Madde Düzeyinde Boyutluluk Modellerinin Bilgisayar Ortamında Bireyselleştirilmiş Test Yöntemleri Üzerindeki Etkisinin İncelenmesi*. Hacettepe Üniversitesi, Ankara.
- Özbaşı, D., & Demirtaşlı, N. (2015). Bilgisayar Okuryazarlığı Testinin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Geliştirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2).

- Patsula, L. (1999). *A comparison of computerized adaptive testing and multi-stage testing* [Unpublished doctoral dissertation]. University of Massachusetts Amherst.
- Piaw Chua, Y. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 28(5), 1580-1586.
- Pohl, S. (2013). Longitudinal Multistage Testing. *Journal of Educational Measurement*, 50(4), 447-468.
- Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior*, 77, 1-10.
- Rasch, G. (1960). Probabilistic model for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (2009). Computerized Adaptive Testing Using MIRT. İçinde M. D. Reckase (Ed.), *Multidimensional Item Response Theory* (ss. 311-339). Springer.
- Reckase, M. D., Ju, U., & Kim, S. (2019). How Adaptive Is an Adaptive Test: Are All Adaptive Tests Adaptive? *Journal of Computerized Adaptive Testing*, 7(1), 1-14-14.
- Robin, F., Steffen, M., & Liang, L. (2014). The Multistage Test Implementation of GRE Revised General Test. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 325-342). Chapman and Hall/CRC.
- Rotou, O., Patsula, L., Steffen, M., & Rizavi, S. (2007). Comparison of Multistage Tests with Computerized Adaptive and Paper and Pencil Tests. *ETS Research Report Series*, 2007(1), i-27.

- Sari, H. İ., Yahsi-Sari, H., & Huggins-Manley, A. C. (2016). Computer Adaptive Multistage Testing: Practical Issues, Challenges and Principles. *Journal of Measurement and Evaluation in Education and Psychology*, 7(2), 388-406.
- Schaeffer, G. A., Bridgeman, B., Golub-Smith, M. L., Lewis, C., Potenza, M. T., & Steffen, M. (1998). Comparability of Paper-and-Pencil and Computer Adaptive Test Scores on the Gre® General Test. *ETS Research Report Series*, 1998(2), i-25.
- Sireci, S. G. (2007). Standardized Testing is Useful. İçinde D. A. Henningfeld (Ed.), *Standardized Testing*. Greenhaven Press.
- Stark, S., & Chernyshenko, O. S. (2006). Multistage Testing: Widely or Narrowly Applicable? *Applied Measurement in Education*, 19(3), 257-260.
- Svetina, D., Liaw, Y.-L., Rutkowski, L., & Rutkowski, D. (2019). Routing Strategies and Optimizing Design for Multistage Testing in International Large-Scale Assessments. *Journal of Educational Measurement*, 56(1), 192-213.
- Thissen, D., & Mislevy, R. J. (2000). Testing Algorithms. İçinde H. Wainer (Ed.), *Computerized Adaptive Testing: A Primer* (2nd Edition, ss. 101-133). Lawrence Erlbaum Associates.
- Thompson, N., & Weiss, D. (2011). A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research, and Evaluation*, 16(1).
- Turgut, M. F. (1977). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara.
- van der Linden, W. J., & Glas, C. A. W. (Ed.). (2010). *Elements of Adaptive Testing*. New York: Springer.
- van der Linden, W. J., & Pashley, P. J. (2010). Item Selection and Ability Estimation in Adaptive Testing. İçinde W. J. van der Linden & C. A. W. Glas (Ed.), *Elements of Adaptive Testing* (ss. 3–30). New York: Springer.

- Veldkamp, B. P., & van der Linden, W. J. (2010). Designing Item Pools for Adaptive Testing. İçinde W. J. van der Linden & C. A. W. Glas (Ed.), *Elements of Adaptive Testing* (ss. 231-246). New York: Springer.
- Veldkamp, B. P. (2014). Item Pool Design and Maintenance for Multistage Testing. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 39-54). New York: Chapman and Hall/CRC.
- Vispoel, W. P. (1998). Reviewing and Changing Answers on Computer-Adaptive and Self-Adaptive Vocabulary Tests. *Journal of Educational Measurement*, 35(4), 328-345.
- Wainer, H. (2000). Introduction and History. İçinde H. Wainer (Ed.), *Computerized Adaptive Testing: A Primer* (ss. 1-22). Lawrence Erlbaum Associates.
- Wang, K. (2017). *A Fair Comparison of the Performance of Computerized Adaptive Testing and Multistage Adaptive Testing*. Michigan State University, Michigan.
- Weiss, D. J. (1983). *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. Academic Press.
- 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.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Weissman, A., Belov, D. I., & Armstrong, R. D. (2007). *Information-Based Versus Number-Correct Routing in Multistage Classification Tests* [Research Report]. Law School Admission Council.

- Weissman, A. (2014). IRT-Based Multistage Testing. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 153-168). Chapman and Hall/CRC.
- Wentzel, C., Mills, C. M., & Meara, K. C. (2014). Transitioning a K–12 Assessment from Linear to Multistage Tests. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 355-370). Chapman and Hall/CRC.
- Woolfolk, A. E. (2011). *Educational Psychology: Global Edition*. Pearson/Allyn and Bacon Publishers.
- Yamamoto, K., Shin, H. J., & Khorramdel, L. (2018). Multistage Adaptive Testing Design in International Large-Scale Assessments. *Educational Measurement: Issues and Practice*, 37(4), 16-27.
- Yan, D. (2020). Multistage Testing in Practice. İçinde H. Jiao & R. W. Lissitz (Ed.), *Application of Artificial Intelligence to Assessment* (ss. 141-160). Information Age Publishing Inc.
- Yan, D., Lewis, C., & von Davier, A. A. (2014). Overview of Computerized Multistage Tests. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 3–20). New York: Chapman and Hall/CRC.
- Yan, D., Lewis, C., & von Davier, A. A. (2014). Multistage Test Design and Scoring with Small Samples. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 303-324). Chapman and Hall/CRC.
- Zheng, Y., Nozawa, Y., Gao, X., & Chang, H.-H. (2012). *Multistage Adaptive Testing for a Large-Scale Classification Test: Design, Heuristic Assembly, and*

Comparison with Other Testing Modes (ACT Research Report Series). Iowa City: ACT, Inc.

Zheng, Y., & Chang, H.-H. (2015). On-the-Fly Assembled Multistage Adaptive Testing.

Applied Psychological Measurement, 39(2), 104-118.

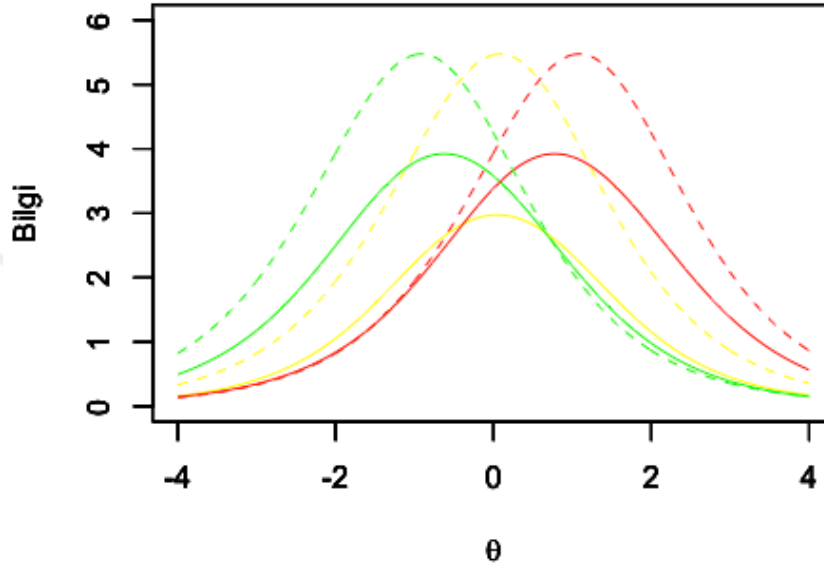
Zenisky, A., Hambleton, R. K., & Luecht, R. M. (2010). Multistage Testing: Issues, Design and Research. İçinde W. J. van der Linden & C. A. W. Glas (Ed.), *Elements of Adaptive Testing* (ss. 355-372). New York: Springer.

Zenisky, A., & Hambleton, R. K. (2014). Multistage Test Designs: Moving Research Results into Practice. İçinde D. Yan, C. Lewis, & A. A. von Davier (Ed.), *Computerized Multistage Testing: Theory and Applications* (ss. 21–37). New York: Chapman and Hall/CRC.

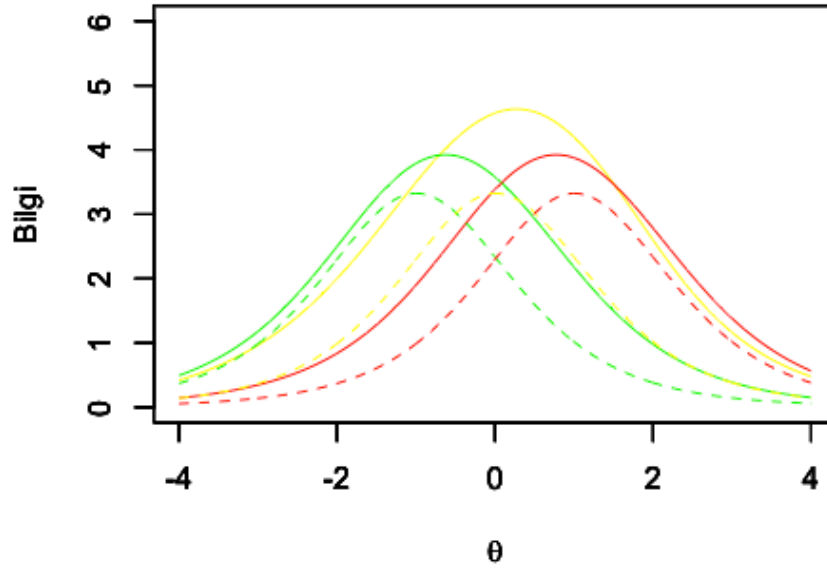
Zwick, R. (2006). Higher Education Admission Testing. İçinde R. L. Brennan, *Educational Measurement* (Fourth Edition, ss. 647-680). American Council on Education/Praeger Publishers.

EK-A1: Modül Bilgi Fonksiyonları (Normal Dağılım)

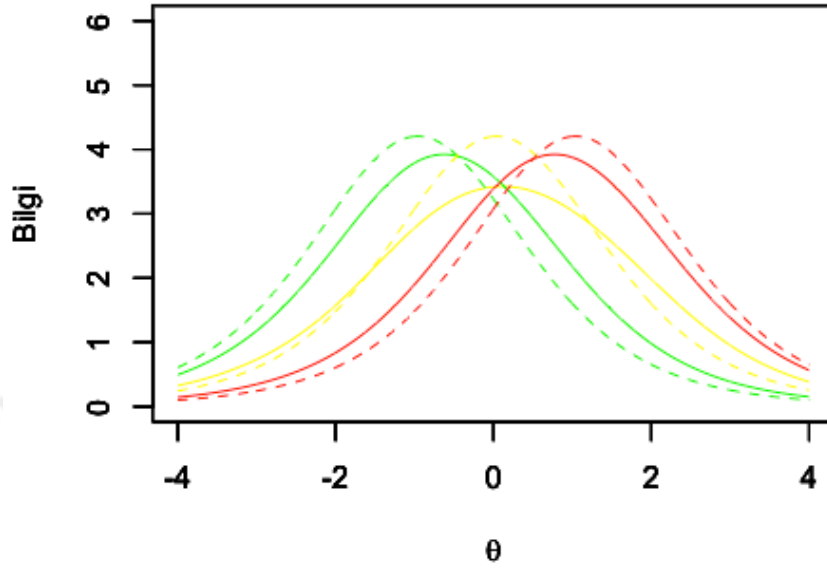
Tasarım 1 (10-15-20)



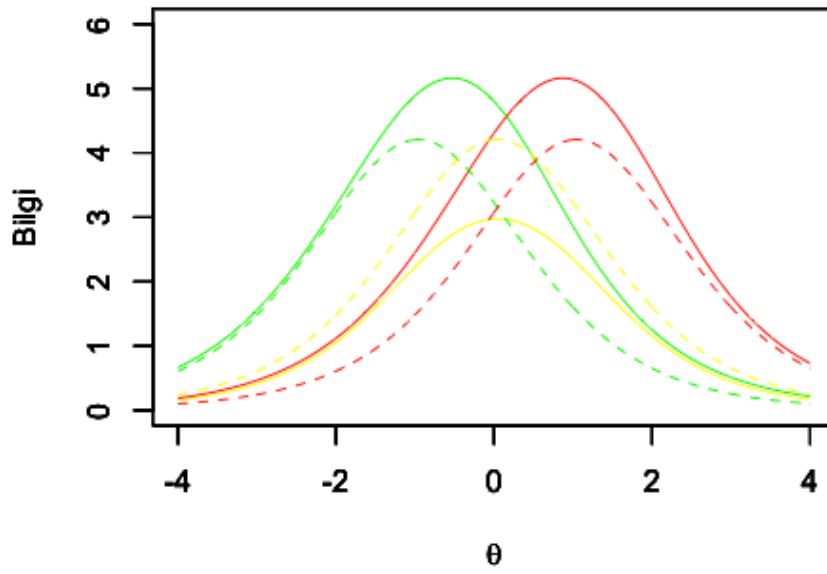
Tasarım 2 (20-15-10)



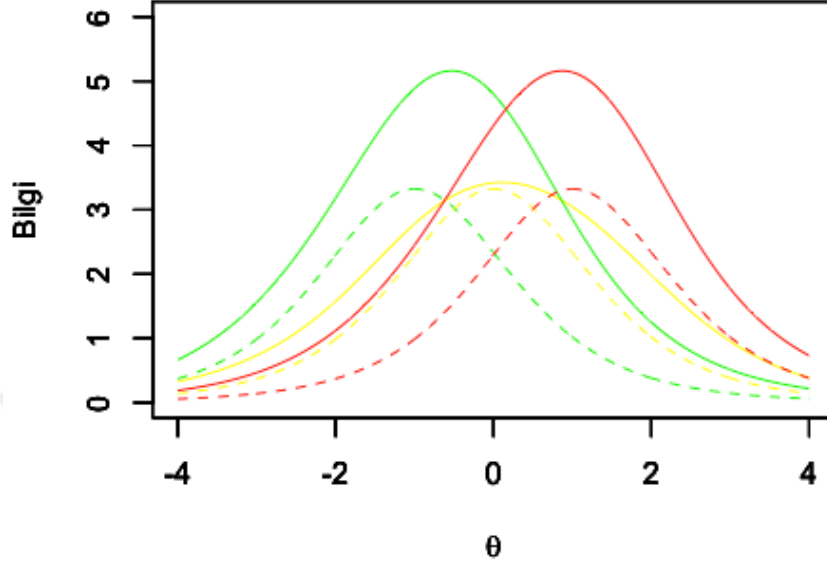
Tasarım 3 (15-15-15)



Tasarım 4 (10-20-15)

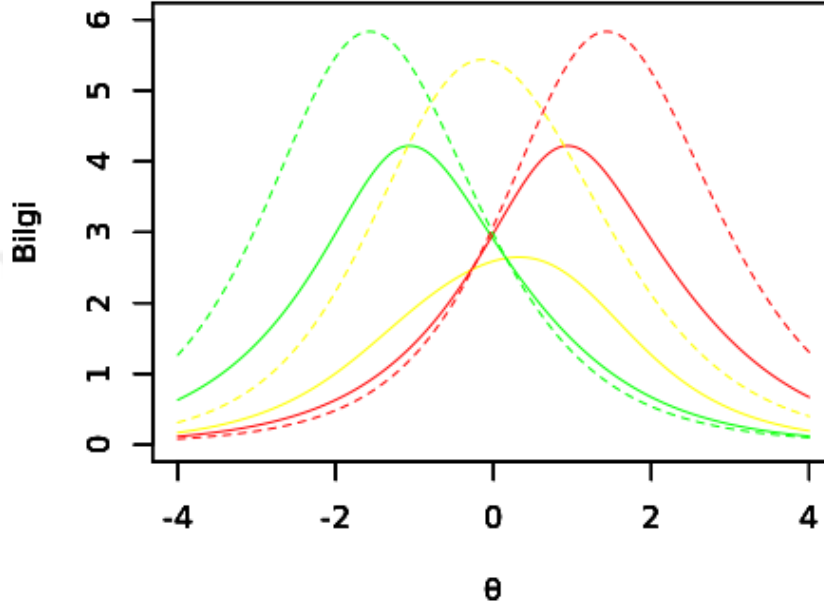


Tasarım 5 (15-20-10)

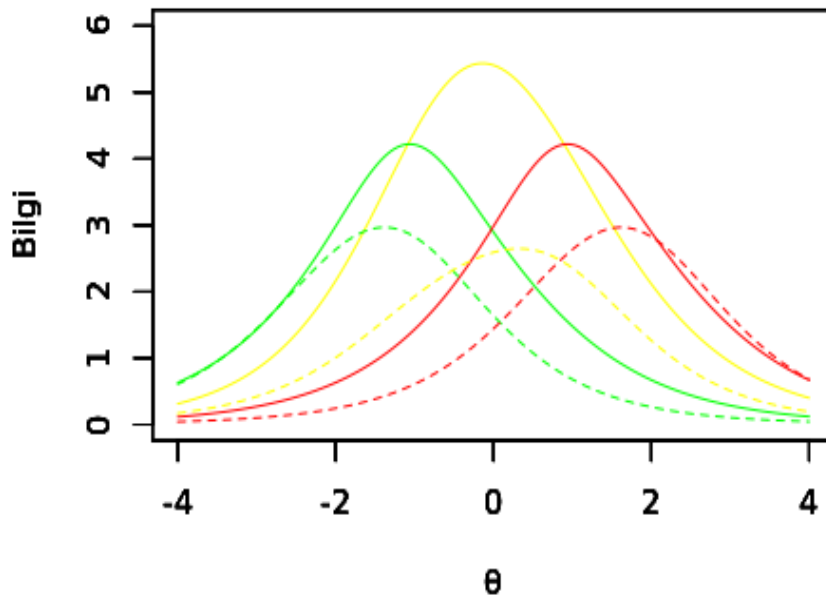


EK-A2: Modül Bilgi Fonksiyonları (Uniform Dağılım)

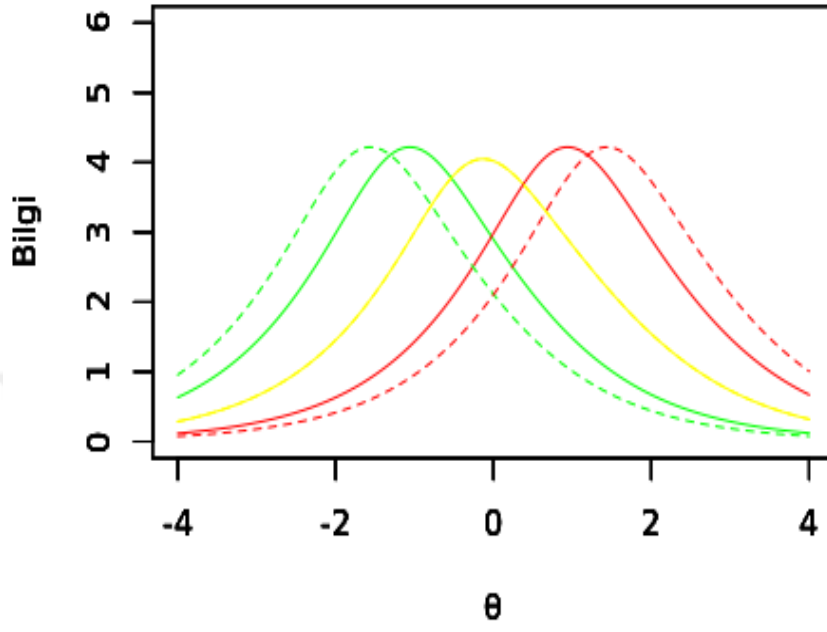
Tasarım 1 (10-15-20)



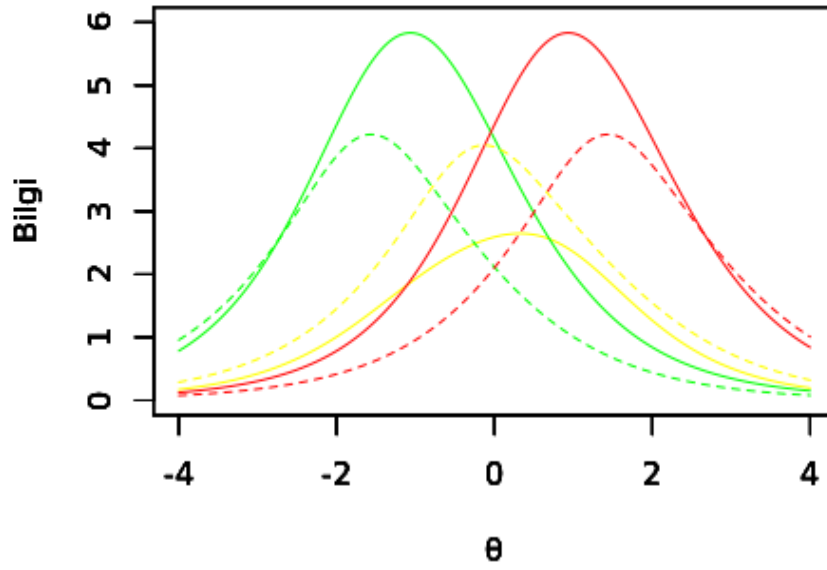
Tasarım 2 (20-15-10)



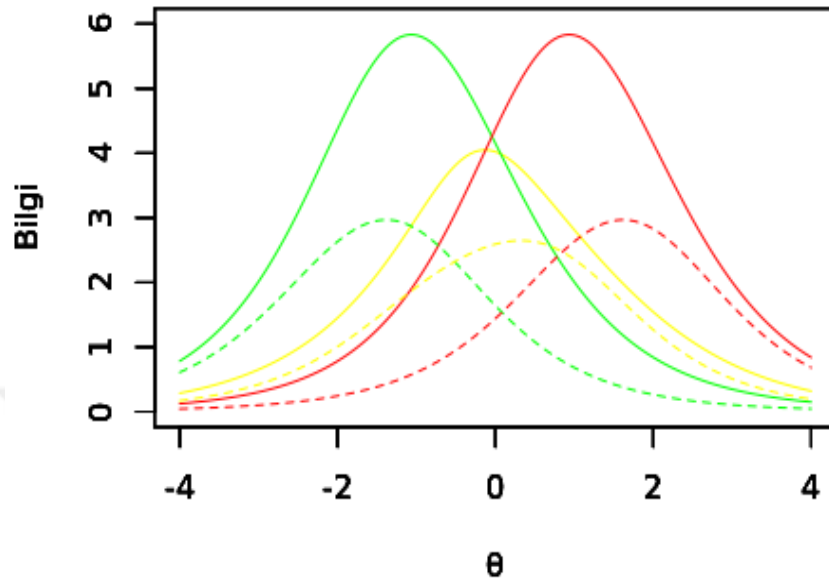
Tasarım 3 (15-15-15)



Tasarım 4 (10-20-15)



Tasarım 5 (15-20-10)



EK-B1: Madde Parametreleri (Normal Dağılım)

Tasarım 1 (10-15-20) için madde güçlük ve ayırt edicilik değerleri

1. Aşama	Orta	
<i>a</i>	<i>b</i>	
1,57	-0,63	
1,12	0,18	
0,83	-0,84	
0,51	1,60	
1,40	0,33	
0,99	-0,82	
1,00	0,49	
1,33	0,74	
1,28	0,58	
1,20	-0,31	

2. Aşama	Kolay	Zor
<i>a</i>	<i>b</i>	<i>b</i>
0,99	-1,08	0,32
1,00	-0,59	0,81
1,33	-1,20	0,20
1,28	0,26	1,66
1,20	-0,50	0,90
1,32	-1,19	0,21
1,26	-0,41	0,99
1,02	-0,26	1,14
0,55	-0,35	1,05
1,20	-0,88	0,52
0,98	0,21	1,61
0,95	-0,47	0,93
0,64	-1,07	0,33
0,87	-2,03	-0,63
1,13	-0,03	1,37

3. Aşama	Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>
1,32	-1,19	-0,19	0,81
1,26	-0,94	0,06	1,06
1,02	-1,25	-0,25	0,75
0,55	-0,52	0,48	1,48
1,20	-0,90	0,10	1,10
0,98	-1,25	-0,25	0,75
0,95	-0,85	0,15	1,15
0,64	-0,78	0,22	1,22
0,87	-0,83	0,17	1,17
1,13	-1,09	-0,09	0,91
1,50	-0,55	0,45	1,45
0,97	-0,88	0,12	1,12
1,12	-1,19	-0,19	0,81
0,98	-1,66	-0,66	0,34
0,66	-0,66	0,34	1,34
0,88	-1,01	-0,01	0,99
0,89	-1,00	0,00	1,00
0,98	-0,72	0,28	1,28
1,39	-0,75	0,25	1,25
1,26	-0,82	0,18	1,18

Tasarım 2 (20-15-10) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
1,32	-0,63			
1,26	0,18			
1,02	-0,84			
0,55	1,60			
1,20	0,33			
0,98	-0,82			
0,95	0,49			
0,64	0,74			
0,87	0,58			
1,13	-0,31			
1,50	1,51			
0,97	0,39			
1,12	-0,62			
0,98	-2,21			
0,66	1,12			
0,88	-0,04			
0,89	-0,02			
0,98	0,94			
1,39	0,82			
1,26	0,59			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
0,99	-1,08	0,32		
1,00	-0,59	0,81		
1,33	-1,20	0,20		
1,28	0,26	1,66		
1,20	-0,50	0,90		
1,32	-1,19	0,21		
1,26	-0,41	0,99		
1,02	-0,26	1,14		
0,55	-0,35	1,05		
1,20	-0,88	0,52		
0,98	0,21	1,61		
0,95	-0,47	0,93		
0,64	-1,07	0,33		
0,87	-2,03	-0,63		
1,13	-0,03	1,37		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
1,57	-1,19	-0,19	0,81	
1,12	-0,94	0,06	1,06	
0,83	-1,25	-0,25	0,75	
0,51	-0,52	0,48	1,48	
1,40	-0,90	0,10	1,10	
0,99	-1,25	-0,25	0,75	
1,00	-0,85	0,15	1,15	
1,33	-0,78	0,22	1,22	
1,28	-0,83	0,17	1,17	
1,20	-1,09	-0,09	0,91	

Tasarım 3 (15-15-15) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
0,99	-0,63			
1,00	0,18			
1,33	-0,84			
1,28	1,60			
1,20	0,33			
1,32	-0,82			
1,26	0,49			
1,02	0,74			
0,55	0,58			
1,20	-0,31			
0,98	1,51			
0,95	0,39			
0,64	-0,62			
0,87	-2,21			
1,13	1,12			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
0,99	-1,08	0,32		
1,00	-0,59	0,81		
1,33	-1,20	0,20		
1,28	0,26	1,66		
1,20	-0,50	0,90		
1,32	-1,19	0,21		
1,26	-0,41	0,99		
1,02	-0,26	1,14		
0,55	-0,35	1,05		
1,20	-0,88	0,52		
0,98	0,21	1,61		
0,95	-0,47	0,93		
0,64	-1,07	0,33		
0,87	-2,03	-0,63		
1,13	-0,03	1,37		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
0,99	-1,19	-0,19	0,81	
1,00	-0,94	0,06	1,06	
1,33	-1,25	-0,25	0,75	
1,28	-0,52	0,48	1,48	
1,20	-0,90	0,10	1,10	
1,32	-1,25	-0,25	0,75	
1,26	-0,85	0,15	1,15	
1,02	-0,78	0,22	1,22	
0,55	-0,83	0,17	1,17	
1,20	-1,09	-0,09	0,91	
0,98	-0,55	0,45	1,45	
0,95	-0,88	0,12	1,12	
0,64	-1,19	-0,19	0,81	
0,87	-1,66	-0,66	0,34	
1,13	-0,66	0,34	1,34	

Tasarım 4 (10-20-15) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
1,57	-0,63			
1,12	0,18			
0,83	-0,84			
0,51	1,60			
1,40	0,33			
0,99	-0,82			
1,00	0,49			
1,33	0,74			
1,28	0,58			
1,20	-0,31			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
1,32	-1,08	0,32		
1,26	-0,59	0,81		
1,02	-1,20	0,20		
0,55	0,26	1,66		
1,20	-0,50	0,90		
0,98	-1,19	0,21		
0,95	-0,41	0,99		
0,64	-0,26	1,14		
0,87	-0,35	1,05		
1,13	-0,88	0,52		
1,50	0,21	1,61		
0,97	-0,47	0,93		
1,12	-1,07	0,33		
0,98	-2,03	-0,63		
0,66	-0,03	1,37		
0,88	-0,73	0,67		
0,89	-0,71	0,69		
0,98	-0,13	1,27		
1,39	-0,21	1,19		
1,26	-0,34	1,06		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
0,99	-1,19	-0,19	0,81	
1,00	-0,94	0,06	1,06	
1,33	-1,25	-0,25	0,75	
1,28	-0,52	0,48	1,48	
1,20	-0,90	0,10	1,10	
1,32	-1,25	-0,25	0,75	
1,26	-0,85	0,15	1,15	
1,02	-0,78	0,22	1,22	
0,55	-0,83	0,17	1,17	
1,20	-1,09	-0,09	0,91	
0,98	-0,55	0,45	1,45	
0,95	-0,88	0,12	1,12	
0,64	-1,19	-0,19	0,81	
0,87	-1,66	-0,66	0,34	
1,13	-0,66	0,34	1,34	

Tasarım 5 (15-20-10) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
0,99	-0,63			
1,00	0,18			
1,33	-0,84			
1,28	1,60			
1,20	0,33			
1,32	-0,82			
1,26	0,49			
1,02	0,74			
0,55	0,58			
1,20	-0,31			
0,98	1,51			
0,95	0,39			
0,64	-0,62			
0,87	-2,21			
1,13	1,12			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
1,32	-1,08	0,32		
1,26	-0,59	0,81		
1,02	-1,20	0,20		
0,55	0,26	1,66		
1,20	-0,50	0,90		
0,98	-1,19	0,21		
0,95	-0,41	0,99		
0,64	-0,26	1,14		
0,87	-0,35	1,05		
1,13	-0,88	0,52		
1,50	0,21	1,61		
0,97	-0,47	0,93		
1,12	-1,07	0,33		
0,98	-2,03	-0,63		
0,66	-0,03	1,37		
0,88	-0,73	0,67		
0,89	-0,71	0,69		
0,98	-0,13	1,27		
1,39	-0,21	1,19		
1,26	-0,34	1,06		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
1,57	-1,19	-0,19	0,81	
1,12	-0,94	0,06	1,06	
0,83	-1,25	-0,25	0,75	
0,51	-0,52	0,48	1,48	
1,40	-0,90	0,10	1,10	
0,99	-1,25	-0,25	0,75	
1,00	-0,85	0,15	1,15	
1,33	-0,78	0,22	1,22	
1,28	-0,83	0,17	1,17	
1,20	-1,09	-0,09	0,91	

EK-B2: Madde Parametreleri (Uniform Dağılım)

Tasarım 1 (10-15-20) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
a	b			
1,57	-0,47			
1,12	-0,26			
0,83	0,15			
0,51	0,82			
1,40	-0,60			
0,99	0,80			
1,00	0,89			
1,33	0,32			
1,28	0,26			
1,20	-0,88			
2. Aşama		Kolay	Zor	
a	b	b		
0,99	-1,23	0,77		
1,00	-1,13	0,87		
1,33	-0,93	1,07		
1,28	-0,59	1,41		
1,20	-1,30	0,70		
1,32	-0,60	1,40		
1,26	-0,56	1,44		
1,02	-0,84	1,16		
0,55	-0,87	1,13		
1,20	-1,44	0,56		
0,98	-1,29	0,71		
0,95	-1,32	0,68		
0,64	-0,81	1,19		
0,87	-1,12	0,88		
1,13	-0,73	1,27		
3. Aşama		Kolay	Orta	Zor
a	b	b	b	
1,32	-1,73	-0,47	1,27	
1,26	-1,63	-0,26	1,37	
1,02	-1,43	0,15	1,57	
0,55	-1,09	0,82	1,91	
1,20	-1,80	-0,60	1,20	
0,98	-1,10	0,80	1,90	
0,95	-1,06	0,89	1,94	
0,64	-1,34	0,32	1,66	
0,87	-1,37	0,26	1,63	
1,13	-1,94	-0,88	1,06	
1,50	-1,79	-0,59	1,21	
0,97	-1,82	-0,65	1,18	
1,12	-1,31	0,37	1,69	
0,98	-1,62	-0,23	1,38	
0,66	-1,23	0,54	1,77	
0,88	-1,50	0,00	1,50	
0,89	-1,28	0,44	1,72	
0,98	-1,01	0,98	1,99	
1,39	-1,62	-0,24	1,38	
1,26	-1,22	0,55	1,78	

Tasarım 2 (20-15-10) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
1,32	-0,47			
1,26	-0,26			
1,02	0,15			
0,55	0,82			
1,20	-0,60			
0,98	0,80			
0,95	0,89			
0,64	0,32			
0,87	0,26			
1,13	-0,88			
1,50	-0,59			
0,97	-0,65			
1,12	0,37			
0,98	-0,23			
0,66	0,54			
0,88	0,00			
0,89	0,44			
0,98	0,98			
1,39	-0,24			
1,26	0,55			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
0,99	-1,23	0,77		
1,00	-1,13	0,87		
1,33	-0,93	1,07		
1,28	-0,59	1,41		
1,20	-1,30	0,70		
1,32	-0,60	1,40		
1,26	-0,56	1,44		
1,02	-0,84	1,16		
0,55	-0,87	1,13		
1,20	-1,44	0,56		
0,98	-1,29	0,71		
0,95	-1,32	0,68		
0,64	-0,81	1,19		
0,87	-1,12	0,88		
1,13	-0,73	1,27		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
1,57	-1,73	-0,47	1,27	
1,12	-1,63	-0,26	1,37	
0,83	-1,43	0,15	1,57	
0,51	-1,09	0,82	1,91	
1,40	-1,80	-0,60	1,20	
0,99	-1,10	0,80	1,90	
1,00	-1,06	0,89	1,94	
1,33	-1,34	0,32	1,66	
1,28	-1,37	0,26	1,63	
1,20	-1,94	-0,88	1,06	

Tasarım 3 (15-15-15) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
0,99	-0,47			
1,00	-0,26			
1,33	0,15			
1,28	0,82			
1,20	-0,60			
1,32	0,80			
1,26	0,89			
1,02	0,32			
0,55	0,26			
1,20	-0,88			
0,98	-0,59			
0,95	-0,65			
0,64	0,37			
0,87	-0,23			
1,13	0,54			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
0,99	-1,23	0,77		
1,00	-1,13	0,87		
1,33	-0,93	1,07		
1,28	-0,59	1,41		
1,20	-1,30	0,70		
1,32	-0,60	1,40		
1,26	-0,56	1,44		
1,02	-0,84	1,16		
0,55	-0,87	1,13		
1,20	-1,44	0,56		
0,98	-1,29	0,71		
0,95	-1,32	0,68		
0,64	-0,81	1,19		
0,87	-1,12	0,88		
1,13	-0,73	1,27		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
0,99	-1,73	-0,47	1,27	
1,00	-1,63	-0,26	1,37	
1,33	-1,43	0,15	1,57	
1,28	-1,09	0,82	1,91	
1,20	-1,80	-0,60	1,20	
1,32	-1,10	0,80	1,90	
1,26	-1,06	0,89	1,94	
1,02	-1,34	0,32	1,66	
0,55	-1,37	0,26	1,63	
1,20	-1,94	-0,88	1,06	
0,98	-1,79	-0,59	1,21	
0,95	-1,82	-0,65	1,18	
0,64	-1,31	0,37	1,69	
0,87	-1,62	-0,23	1,38	
1,13	-1,23	0,54	1,77	

Tasarım 4 (10-20-15) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
1,57	-0,47			
1,12	-0,26			
0,83	0,15			
0,51	0,82			
1,40	-0,60			
0,99	0,80			
1,00	0,89			
1,33	0,32			
1,28	0,26			
1,20	-0,88			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
1,32	-1,23	0,77		
1,26	-1,13	0,87		
1,02	-0,93	1,07		
0,55	-0,59	1,41		
1,20	-1,30	0,70		
0,98	-0,60	1,40		
0,95	-0,56	1,44		
0,64	-0,84	1,16		
0,87	-0,87	1,13		
1,13	-1,44	0,56		
1,50	-1,29	0,71		
0,97	-1,32	0,68		
1,12	-0,81	1,19		
0,98	-1,12	0,88		
0,66	-0,73	1,27		
0,88	-1,00	1,00		
0,89	-0,78	1,22		
0,98	-0,51	1,49		
1,39	-1,12	0,88		
1,26	-0,72	1,28		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
0,99	-1,73	-0,47	1,27	
1,00	-1,63	-0,26	1,37	
1,33	-1,43	0,15	1,57	
1,28	-1,09	0,82	1,91	
1,20	-1,80	-0,60	1,20	
1,32	-1,10	0,80	1,90	
1,26	-1,06	0,89	1,94	
1,02	-1,34	0,32	1,66	
0,55	-1,37	0,26	1,63	
1,20	-1,94	-0,88	1,06	
0,98	-1,79	-0,59	1,21	
0,95	-1,82	-0,65	1,18	
0,64	-1,31	0,37	1,69	
0,87	-1,62	-0,23	1,38	
1,13	-1,23	0,54	1,77	

Tasarım 5 (15-20-10) için madde güçlük ve ayırt edicilik değerleri

1. Aşama		Orta		
<i>a</i>	<i>b</i>			
0,99	-0,47			
1,00	-0,26			
1,33	0,15			
1,28	0,82			
1,20	-0,60			
1,32	0,80			
1,26	0,89			
1,02	0,32			
0,55	0,26			
1,20	-0,88			
0,98	-0,59			
0,95	-0,65			
0,64	0,37			
0,87	-0,23			
1,13	0,54			
2. Aşama		Kolay	Zor	
<i>a</i>	<i>b</i>	<i>b</i>		
1,32	-1,23	0,77		
1,26	-1,13	0,87		
1,02	-0,93	1,07		
0,55	-0,59	1,41		
1,20	-1,30	0,70		
0,98	-0,60	1,40		
0,95	-0,56	1,44		
0,64	-0,84	1,16		
0,87	-0,87	1,13		
1,13	-1,44	0,56		
1,50	-1,29	0,71		
0,97	-1,32	0,68		
1,12	-0,81	1,19		
0,98	-1,12	0,88		
0,66	-0,73	1,27		
0,88	-1,00	1,00		
0,89	-0,78	1,22		
0,98	-0,51	1,49		
1,39	-1,12	0,88		
1,26	-0,72	1,28		
3. Aşama		Kolay	Orta	Zor
<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	
1,57	-1,73	-0,47	1,27	
1,12	-1,63	-0,26	1,37	
0,83	-1,43	0,15	1,57	
0,51	-1,09	0,82	1,91	
1,40	-1,80	-0,60	1,20	
0,99	-1,10	0,80	1,90	
1,00	-1,06	0,89	1,94	
1,33	-1,34	0,32	1,66	
1,28	-1,37	0,26	1,63	
1,20	-1,94	-0,88	1,06	

EK-C: R Kodları

#Madde havuzunun oluşturulması

```
it.2PL <- rbind(genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,1)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,1)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,1)),  
  genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-0.7,0.6)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-0.7,0.6)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-0.7,0.6)),  
  genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0.7,0.6)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0.7,0.6)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0.7,0.6)),  
  genDichoMatrix(items = 10, model = "2PL", aPrior =  
  genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-1,0.3)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-1,0.3)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',-1,0.3)),  
  genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,0.3)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,0.3)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',0,0.3)),  
  genDichoMatrix(items = 10, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',1,0.3)),  
  genDichoMatrix(items = 15, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',1,0.3)),  
  genDichoMatrix(items = 20, model = "2PL", aPrior = c('lnorm',0,0.3), bPrior =  
c('norm',1,0.3)))
```

#1-2-3 Panel Tasarımı için geçiş matrisinin tanımlanması

```
trans <- matrix(0,6,6)  
trans[1,2:3]<-trans[2,4:5]<-trans[3,5:6]<-1
```

#Modüllerin oluşturulması, bilgi fonksiyonlarının ve kesme puanlarının belirlenmesi

#Tasarım1 - (10-15-20)

```
A1 <- c(1:10)
```

```
B1 <- c(56:70)
```

```
C1 <- c(101:115)
```

```
D1 <- c(161:180)
```

```
E1 <- c(206:225)
```

```
F1 <- c(251:270)
```

```
modules1 <- matrix(0,270,6)
```

```
modules1[A1, 1]<- modules1[B1, 2]<- modules1[C1, 3]<- modules1[D1, 4]<-  
modules1[E1, 5]<- modules1[F1, 6]<- 1
```

#Modül Bilgi fonksiyonları

#A1

```
t<-seq(-4,4,0.01)
```

```
infoA1<-function(x) sum(li(x,it.2PL[1:10,])$li)
```

```
TIFA1<-sapply(t,infoA1)
```

```
plot(t,TIFA1,type="l", xlab=expression(theta), ylab="Bilgi", ylim=c(0,6), col='yellow',  
main='Tasarım 1 (10-15-20)')
```

```
par(new=TRUE)
```

#B1

```
infoB1<-function(x) sum(li(x,it.2PL[56:70,])$li)
```

```
TIFB1<-sapply(t,infoB1)
```

```
plot(t,TIFB1,type="l", xlab=expression(theta), ylab="Bilgi", ylim=c(0,6), col='green')
```

```
par(new=TRUE)
```

#C1

```
infoC1<-function(x) sum(li(x,it.2PL[101:115,])$li)
```

```
TIFC1<-sapply(t,infoC1)
```

```
plot(t,TIFC1,type="l", xlab=expression(theta), ylab="Bilgi", ylim=c(0,6), col='red')
```

```
par(new=TRUE)
```

#Kesişim B1 ve C1

```
rtB1C1 <- uniroot(function(x) infoB1(x) - infoC1(x) , c(-3,3), tol=1e-8)
```

```
thB1C1<-rtB1C1[[1]]
```

#Kesme Puanı-Doğru cevap sayısı-A1

```
PA1 <-Pi(th=thB1C1, it=it.2PL[1:10,]) [[1]]
```

```
CRA1<-sum(PA1)
```

#Modül Bilgi fonksiyonları D1 E1 F1

#D1

```
infoD1<-function(x) sum(li(x,it.2PL[296:315,])$li)
```

```
TIFD1<-sapply(t,infoD1)
plot(t,TIFD1,type="l", lty= "dashed", xlab=expression(theta), ylab="Bilgi",
ylim=c(0,6),col='green')
par(new=TRUE)
```

#E1

```
infoE1<-function(x) sum(li(x,it.2PL[341:360,])$li)
TIFE1<-sapply(t,infoE1)
plot(t,TIFE1,type="l", lty= "dashed",xlab=expression(theta), ylab="Bilgi", ylim=c(0,6),
col='yellow')
par(new=TRUE)
```

#F1

```
infoF1<-function(x) sum(li(x,it.2PL[386:405,])$li)
TIFF1<-sapply(t,infoF1)
plot(t,TIFF1,type="l", lty= "dashed",xlab=expression(theta), ylab="Bilgi", ylim=c(0,6),
col='red')
```

#Kesişim D1 ve E1

```
rtD1E1 <- uniroot(function(x) infoD1(x) - infoE1(x) , c(-3,3), tol=1e-8)
thD1E1<-rtD1E1[[1]]
```

#Kesişim E1 ve F1

```
rtE1F1 <- uniroot(function(x) infoE1(x) - infoF1(x) , c(-3,3), tol=1e-8)
thE1F1<-rtE1F1[[1]]
```

#Kesme Puanı-Doğru cevap sayısı-B1

```
PB1i <-Pi(th=thD1E1, it=it.2PL[1:10,]) [[1]]
PB1j <-Pi(th=thD1E1, it=it.2PL[56:70,]) [[1]]
CRB1.last<-sum(PB1j)
```

#Kesme Puanı-Doğru cevap sayısı-C1

```
PC1i <-Pi(th=thE1F1, it=it.2PL[1:10,]) [[1]]
PC1j <-Pi(th=thE1F1, it=it.2PL[101:115,]) [[1]]
CRC1.all<-sum(PC1i,PC1j)
CRC1.last<-sum(PC1j)
```

#Kesme Puanının Tanımlanması-1

```
cut.score1.last <- rbind(c(2, 3, CRA1 ), c(4, 5, CRB1.last), c(5, 6, CRC1.last))
```

#5 tasarım için benzer şekilde kodlar yazılmış ve Tasarım 1'de 1 olarak ifade edilenler diğer tasarımlarda 2,3,4 ve olarak ifade edilmiştir.

#Tasarım Karşılaştırma

#Yönlendirme Kuralı: Dinamik & Yetenek Kestirimi: MLE

```
RES <- matrix(NA, length(TH), 5)
start <- list(theta = 0)
test <- list(method = "EAP", moduleSelect = "MFI", range = c(-3,3))
final <- list(method = "ML", range = c(-3,3))

for (j in 1:5) {
  mod <- get(paste("modules",j,sep = ""))
  for (i in 1:length(TH)){
    prov <- randomMST (trueTheta = TH[i],
                      itemBank = it.2PL, modules = mod,
                      transMatrix = trans, genSeed = i, start = start,
                      test = test, final = final)
    RES [i,j] <- prov$thFinal
  }
}
res.ASB <- res.RMSE <- matrix (NA, length(TH), 5)
ASB <- function (t,th) mean(t-th)
RMSE <- function (t,th) sqrt (mean((t-th)^2))
for (i in 1:length (TH)) {
  for (j in 1:5) {
    ind <- which (TH == TH [i])
    res.ASB [i,j] <- ASB (RES[ind,j], TH[i])
    res.RMSE [i,j] <- RMSE (RES[ind,j], TH[i])
  }
}
ASB<- round(colMeans(res.ASB), 3)
RMSE<- round(colMeans(res.RMSE), 3)

View(ASB)
View(RMSE)
```

#Yönlendirme Kuralı: Statik & Yetenek Kestirimi: MLE

```
RES.cut <- matrix(NA, length(TH), 5)
test.last1 <- list(method = "score", cutoff = cut.score1.last, score.range = 'last', range =
c(-3,3))
test.last2 <- list(method = "score", cutoff = cut.score2.last, score.range = 'last', range =
c(-3,3))
```

```

test.last3 <- list(method = "score", cutoff = cut.score3.last, score.range = 'last', range =
c(-3,3))
test.last4 <- list(method = "score", cutoff = cut.score4.last, score.range = 'last', range =
c(-3,3))
test.last5 <- list(method = "score", cutoff = cut.score5.last, score.range = 'last', range =
c(-3,3))

for (j in 1:5) {
  mod <- get(paste("modules",j,sep = ""))
  test.cut.last <- get(paste("test.last", j, sep = ""))

  for (i in 1:length(TH)){
    prov.cut.last <- randomMST (trueTheta = TH[i],
      itemBank = it.2PL, modules = mod,
      transMatrix = trans, genSeed = i, start = start,
      test = test.cut.last, final = final)
    RES.cut [i,j] <- prov.cut.last$thFinal
  }
}
res.ASB.cut <- res.RMSE.cut <- matrix (NA, length(TH), 5)

for (i in 1:length (TH)) {
  for (j in 1:5) {
    ind <- which (TH == TH [i])
    res.ASB.cut [i,j] <- ASB (RES.cut[ind,j], TH[i])
    res.RMSE.cut [i,j] <- RMSE (RES.cut[ind,j], TH[i])
  }
}
ASB.cut<- round(colMeans(res.ASB.cut), 3)
RMSE.cut<- round(colMeans(res.RMSE.cut), 3)

View(ASB.cut)
View(RMSE.cut)

```

#ANOVA

```

result <- matrix(0, nrow = 100, ncol = 5,
dimnames = list(NULL,c("RMSE1","RMSE2","RMSE2","RMSE4","RMSE5")))

```

```

for(j in 1:5){
  for(i in 0:99){
    temp<-res.RMSE[(1+(i*500)):(500+(i*500)),j]
    result[i+1,j]<-mean(temp)
  }
}
RMSE.V<-c(result)
RMSE.data<-as.data.frame(RMSE.V)
write.csv(RMSE.data, file = "RMSE.data.csv")

RMSE_ANOVA <- aov(RMSE ~ as.factor(Module_Length) +as.factor(Estimation) +
as.factor(Routing)
+as.factor(Module_Length)*as.factor(Estimation)
+as.factor(Module_Length)*as.factor(Routing)
+as.factor(Routing)*as.factor(Estimation)
+as.factor(Module_Length)*as.factor(Estimation)*as.factor(Routing),data = Normal)
summary(RMSE_ANOVA)
etaSquared(RMSE_ANOVA, type = 1)

```

EK-D: Etik Komisyonu Onay Bildirimi



Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

10/ 01/ 2022

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı	Çok Aşamalı Bireye Uyarlanmış Testlerde Farklı Koşullardan Elde Edilen Yetenek Kestirimlerinin Karşılaştırılması
-----------------------	--

Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Çalışmada kullanacağım veriler:

- () Kamusal erişime açık (buraya yazınız):
() Özel izin ve onaya tabi (buraya yazınız):
(X) Üretilmiş veri (buraya yazınız): R yazılımı kullanılarak üretilen yetenek ve madde parametreleri üzerinde çalışılıyor.
() Diğer (buraya yazınız):

Yükseköğretim Kurumları Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyonundan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

F.Güneş Ertaş Polat
(Araştırmacı Adı Soyadı, İmzası)

Araştırmacı Bilgileri

Adı Soyadı	F. Güneş Ertaş Polat
Öğrenci ise No	N14141551
Ana Bilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr. ☐ Diğer

Danışman Görüşü ve Onayı*

Prof. Dr. Duygu Anıl
(İmza)
(Danışmanın ünvanı, Adı ve Soyadı)

*Tez ve tezden üretilen yayınlarda gerekli

EK-E: Etik Beyanı

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

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

beyan ederim.

EK-F: Doktora Tez Çalışması Orijinallik Raporu

.../.../.....

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

Tez Başlığı: Çok Aşamalı Bireye Uyarlanmış Testlerde Farklı Koşullardan Elde Edilen Yetenek Kestirimlerinin Karşılaştırılması

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
28/03 /2022	115	144665	08/04/2022	% 5	1795209819

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden 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: F. Güneş Ertay Polat

Öğrenci No.: N14141551

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Duygu Anıl

EK-G: Dissertation Originality Report

.../.../.....

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

Thesis Title: Comparison of Ability Estimations Obtained From Different Conditions in Multi Stage Adaptive Testing

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
28/03 /2022	115	144665	08/04/2022	% 5	1795209819

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: F. Güneş Ertuş Polat

Student No.: N14141551

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED

Prof. Dr. Duygu Anıl

EK-Ğ: 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 fikrî 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 ilişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

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

..... / /

(imza)

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin 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 tez 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 danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tez 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 tezlere 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.