



**KÜMÜLATİF, İDEAL NOKTA VE KARMA MODELLER ALTINDA
BAYESIAN VE NONBAYESIAN YÖNTEMLER İLE YETENEK
KESTİRİMLERİNİN DOĞRULUĞUNUN VE SINIFLAMA
İNDEKSLERİNİN İNCELENMESİ**

Serpil Çelikten Demirel

DOKTORA TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ, 2021

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 24 (yirmi dört) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Serpil

Soyadı : ÇELİKTEN DEMİREL

Bölümü : Eğitimde Ölçme ve Değerlendirme

İmza :

Teslim tarihi :

TEZİN

Türkçe Adı : Kümülatif, İdeal Nokta ve Karma Modeller Altında Bayesian ve Nonbayesian Yöntemler ile Yetenek Kestirimlerinin Doğruluğunun ve Sınıflama İndekslerinin İncelenmesi

İngilizce Adı : Investigation of Accuracy of Ability Estimations and Classification Indices Under Cumulative, Ideal Point and Mixed Model with Bayesian and Nonbayesian Methods

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduđumu, yararlandıđım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiđimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduđunu beyan ederim.

Yazar Adı Soyadı: Serpil ÇELİKTEN DEMİREL

İmza:.....

JÜRİ ONAY SAYFASI

Serpil ÇELİKTEN DEMİREL tarafından hazırlanan “Kümülatif, İdeal Nokta ve Karma Modeller Altında Bayesian ve Nonbayesian Yöntemler ile Yetenek Kestirimlerinin Doğruluğunun ve Sınıflama İndekslerinin İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı’nda Doktora Tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap ÇAKAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Başkan: Prof. Dr. Nuri Doğan

(Eğitimde Ölçme ve Değerlendirme, Hacettepe Üniversitesi)

Üye: Prof. Dr. Şeref TAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Üye: Doç. Dr. Burcu ATAR

(Eğitimde Ölçme ve Değerlendirme, Hacettepe Üniversitesi)

Üye: Doç. Dr. Dilara BAKAN KALAYCIOĞLU

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Tez Savunma Tarihi: 21/06/2021

Bu tezin Eğitim Bilimleri Anabilim Dalı’nda Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ

Eğitim Bilimleri Enstitüsü Müdürü

TEŞEKKÜR

Doktora eğitimim boyunca yanımda olan, akademik gelişimimi destekleyen, birlikte çalışmaktan onur ve mutluluk duyduğum çok değerli danışman hocam sayın Prof. Dr. Mehtap ÇAKAN'a çok teşekkür ederim.

Tez izleme komitemde ve tez jürimde bulunan, tezimin gelişmesinde değerli katkıları olan kıymetli hocalarım Prof. Dr. Şeref TAN'a ve Doç. Dr. Burcu ATAR'a teşekkürlerimi sunarım. Tez jürimde bulunan, değerli fikir ve önerileriyle tezime katkıda bulunan kıymetli hocalarım Prof. Dr. Nuri DOĞAN'a ve Doç. Dr. Dilara BAKAN KALAYCIOĞLU'na teşekkür ederim.

Doktora tez sürecimizi birlikte geçirdiğimiz, bu sürecimi kolaylaştıran ve güzelleştiren, her konuda fikir ve desteklerini sağlayan Dr. Tuba GÜNDÜZ'e teşekkür ederim. Akademik eğitimimin başından beri hayatımda önemli bir yeri olan, birlikte öğrenmenin keyfine vardığım, sosyal hayatımdaki varlığıyla da mutluluk duyduğum Dr. Ayşenur ERDEMİR'e teşekkür ederim. Gazi Üniversitesinde birlikte keyifle çalıştığım bütün araştırma görevlisi arkadaşlarıma teşekkür ederim.

Bu günlere gelmemi sağlayan, desteklerini her daim hissettiğim ve güç aldığım aileme teşekkür ederim.

Tez yazma sürecimde stresime ortak olan, sevgi ve hoşgörüsüyle sürecimi kolaylaştıran kıymetli eşim Mehmet Remzi DEMİREL'e teşekkür ederim.

**KÜMÜLATİF, İDEAL NOKTA VE KARMA MODELLER ALTINDA
BAYESIAN VE NONBAYESIAN YÖNTEMLER İLE YETENEK
KESTİRİMLERİNİN DOĞRULUĞUNUN VE SINIFLAMA
İNDEKSLERİNİN İNCELENMESİ**

(Doktora Tezi)

Serpil Çelikten Demirel

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Temmuz, 2021

ÖZ

Bu araştırmanın amacı madde yanıt süreçlerinin modellenmesinin bayesian ve nonbayesian yöntemler çerçevesinde yetenek kestirimlerinin doğruluğuna, bu yetenek kestirimlerine dayalı olarak yetenek ölçeği boyunca farklı kesme puan noktalarında sınıflama doğruluğu ve sınıflama tutarlılığına olan etkisini incelemektir. Bu doğrultuda madde yanıt süreçlerinin modellenmesinde kümülatif modellerden ATM (Samejima, 1969), ideal nokta modellerinden GAMOM (Roberts, Donoghue & Laughlin, 2000) ele alınmıştır. Bu modellerin yanı sıra delta parametrelerinin (δ) dağılım ortalamasından sapmasına dayalı olarak kümülatif ve ideal nokta modellerinin birleşimi ile karma model oluşturulmuştur. Bayesian yöntemlerden MAP ve EAP; nonbayesian yöntemlerden WLE ile yetenek kestirimleri elde edilmiştir. Model ve yöntemlere bağlı olarak yetenek kestirimlerinin doğruluğu iki faktörlü tekrarlı ölçümler ANOVA ile incelenmiştir. Yetenek kestirimlerine yönelik incelemelerin ardından elde edilen yetenek kestirimleri üzerinden sınıflama indeksleri yetenek ölçeği üzerinde düşük ve yüksek uçta belirlenmiş kesme puanı noktalarında yöntem ve model türünün üç yönlü etkileşimine dayalı olarak incelenmiştir. Bu doğrultuda $c(0,10, 0,90)$, $c(0,15, 0,85)$ ve $c(0,20, 0,80)$ olmak üzere yetenek ölçeği

üzerinde uç kısımlardan ortalara doğru yaklaşan üç kesme puanı çifti üzerinde ayrı ayrı model ve yöntem açısından sınıflama indekslerinin değişimi ele alınmış olup üç faktörlü tekrarlı ölçümler ANOVA ile farklılıklar test edilmiştir. Araştırmadan elde edilen bulgular incelendiğinde yetenek kestirimlerinin doğruluğunun kullanılan model ve yöntem etkileşimine bağlı olarak değiştiği gözlenmiştir. Yetenek kestirimlerinin doğruluğunun en yüksek GAMOM altında elde edildiği bunu ise karma modelin takip ettiği gözlenmiştir. GAMOM ve karma model altında bayesian ve nonbayesian yöntemler açısından en doğru sonucu sağlayan modelin MAP olduğu bunu WLE'nin takip ettiği gözlenmiştir. Monotonik olmayan maddelerin kestiriminin söz konusu olduğu modellerde EAP'ın diğer yöntemlere göre yüksek hata verdiği sonucuna ulaşılmıştır. Ancak kümülatif model olan ATM altında ise yöntem sıralamasının MAP, EAP ve WLE şeklinde olduğu ve her iki bayesian yöntemin de nonbayesian yöntemle göre daha yüksek doğrulukla sonuç verdiği gözlenmiştir. Yetenek kestirimlerine ilişkin incelemelerin ardından bu kestirimler üzerinden elde edilen sınıflama indekslerine yönelik incelemeler yapılmıştır. Sınıflama indekslerine yönelik incelemelerden elde edilen sonuçlara göre model ve yöntem etkileşiminin yetenek ölçeği üzerinde düşük ve yüksek uçta belirlenen kesme puan noktasına bağlı olarak farklılaştığı ve uç kısımlardaki etkinin orta kısımlara yaklaştıkça azaldığı gözlenmiştir. Yöntem ve modeller arası etkileşim etkilerinin kesme puan lokasyonunun hangi düzeylerinde gözlendiğini ortaya çıkarmak için incelemeler yapılmıştır. Elde edilen sonuçlara göre model ve yöntemlere ilişkin etkileşim etkisinin yetenek ölçeğinin üst kısımlarında ortaya çıkma eğiliminde olduğu, yetenek ölçeğinin düşük ucunda yer alan kesme puan noktalarında ise genel olarak yöntem ve modellerin daha benzer sonuçlar verdiği gözlenmiştir. Modeller açısından inceleme yapıldığında GAMOM'un üst uçlarda karma model ve ATM'den daha iyi sonuç verdiği; GAMOM'u ise karma modelin takip ettiği görülmektedir. Yöntemler açısından inceleme yapıldığında yöntemlerin etkililiğinin seçilen model türüne bağlı değiştiği gözlenmiştir. ATM altında genel olarak WLE, MAP ve EAP benzer sonuç vermesine karşın, GAMOM ve karma model altında EAP'ın daha düşük sınıflama indeksleri; WLE ve MAP' ın ise benzer ve yüksek sınıflama indeksleri ile sonuçlandığı gözlenmiştir. Son olarak yetenek kestirimlerinin doğruluğunda bayesian yöntemlerden MAP öne çıkarken sınıflama indekslerinde nonbayesian yöntemlerden WLE' nin öne çıktığı gözlenmiştir. Araştırmadan elde edilen bulgular gerçek veride karşılaşılabilecek olası durumlar ile ilişkilendirilerek tartışılmış olup pratikte karşılaşılabilecek olası ölçme durumları için model ve yöntem kullanımına ilişkin öneriler sunulmuştur.

Anahtar Kelimeler: Kümülatif Model, İdeal Nokta Modeli, Karma Model, Bayesian ve Nonbayesian Yöntemler, Sınıflama İndeksleri

Sayfa Adedi : xvi + 145

Danışman : Prof. Dr. Mehtap ÇAKAN

**INVESTIGATION OF ACCURACY OF ABILITY ESTIMATIONS
AND CLASSIFICATION INDICES UNDER CUMULATIVE, IDEAL
POINT AND MIXED MODEL WITH BAYESIAN AND
NONBAYESIAN METHODS**

(Ph.D Thesis)

Serpil Çelikten Demirel

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

July, 2021

ABSTRACT

The aim of this study is to examine the effect of modeling item response processes on accuracy of ability estimations in the framework of bayesian and nonbayesian methods, and on classification accuracy and classification consistency at different cut-off points along the ability scale based on these ability estimations. In this direction, GRM (Samejima, 1969), which is one of the cumulative models, and GGUM (Roberts, Donoghue & Laughlin, 2000), one of the ideal point models, were used in modeling the response processes. In addition to these models, a mixed model was created with the combination of cumulative and ideal point models based on the deviation of delta parameters (δ) from the distribution mean. MAP and EAP as bayesian methods with WLE as nonbayesian method were used for the ability estimations. Based on the models and methods, accuracy of ability estimations were investigated through the two-way repeated measures ANOVA. After the investigations of ability estimations, classification indices were investigated at the cut scores located on both lower and higher parts of the theta scales in terms of three way interaction effect of cut score, model and method.

Accordingly, variation of the classification indices in terms of model and method separately on three cut-off point pairs $c(0,10, 0,90)$, $c(0,15, 0,85)$ and $c(0,20, 0,80)$ approaching from the extreme to the middle on the ability scale was examined and statistical differences were tested through three-factor repeated measures ANOVAs. When the findings obtained from the research are examined, it is observed that the accuracy of ability estimates differs depending on the interaction of the model and method used. Highest accuracy of ability estimates is obtained under GGUM and it is followed by mixed model. It is observed that under the GGGUM and mixed model, among the bayesian and nonbayesian methods MAP provides the most accurate results and it is followed by WLE. Then, it is concluded that EAP results in higher errors when the focus is estimation of nonmonotonic items. However, under the GRM, a cumulative model, the order of the methods is as MAP, EAP and WLE which points out that bayesian methods provides higher accuracy than nonbayesian method. After examining the ability estimates, classification indices obtained based on these estimates were examined. According to the results obtained from the examinations for the classification indices, it is observed that interaction of model and method at cut scores determined on lower and upper parts of ability scale, and the effect at the extreme locations is decreasing with moving toward the middle part of the ability scale. Investigations are conducted in order to reveal at which level of the cut score location interaction effect of model and method is observed. Depending on the results, it is observed that interaction effects of model and method are likely to occur at the upper part of ability scale, and models and methods provides similar results at the lower part of ability scale. When examining in terms of models, it is observed that GGUM gives better results than mixed model and GRM at the upper end, and GGUM is followed by mixed model. When examined in terms of methods, it is observed that the effectiveness of the methods depends on the type of model selected. Although WLE, MAP and EAP generally provide similar results under GRM, under GGUM and mixed model EAP results in lower classification indices than WLE and MAP, and WLE and MAP were observed to result in similar and higher classification indices. Finally, it is observed that while MAP stands out from bayesian methods in the accuracy of ability estimates, WLE, one of the nonbayesian methods, stands out in classification indices. The findings obtained from the research were discussed in relation to possible situations that may be encountered in real data, and suggestions for the use of models and methods for possible measurement situations to be encountered in practice are presented.

Key Words: Cumulative Model, Ideal Point Model, Mixed Model, Bayesian ve Nonbayesian Methods, Classification Indices

Page Number : xvi + 145

Supervisor : Prof. Dr. Mehtap ÇAKAN

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
JÜRİ ONAY SAYFASI	iii
TEŞEKKÜR	iv
ÖZ.....	v
ABSTRACT.....	vii
İÇİNDEKİLER	ix
TABLolar LİSTESİ.....	xiii
ŞEKİLLER LİSTESİ	xv
SİMGELER VE KISALTMALAR LİSTESİ.....	xvi
BÖLÜM I.....	1
GİRİŞ	1
Problem Durumu.....	1
Araştırmanın Amacı	7
Araştırmanın Önemi	8
Problem İfadesi.....	10
Araştırma Soruları.....	10
Sınırlılıklar	11
BÖLÜM II	13

KAVRAMSAL ÇERÇEVE.....	13
Madde Yanıt Süreçleri.....	13
Baskınlık Yanıt Süreci.....	14
İdeal Nokta Yanıt Süreci.....	14
Madde Yanıt Süreçlerine Göre Ölçme Modelleri	14
Kümülatif Modeller	15
İdeal Nokta Modelleri	15
<i>Genelleştirilmiş Aşamalı Monotonik Olmayan Model.....</i>	<i>16</i>
ATM ve GAMOM’a dayalı Karma Model Tanımlaması	21
Madde Tepki Kuramına Dayalı Sınıflama Yaklaşımları	21
Rudner’in Yaklaşımı	22
Lee’nin Yaklaşımı.....	25
Rudner ve Lee’nin Yaklaşımlarının Karşılaştırılması.....	25
Bayesian ve Nonbayesian Yetenek Kestirim Yöntemleri	26
İLGİLİ ARAŞTIRMALAR	31
İdeal Nokta Modellerinin Kümülatif Modellerle Karşılaştırmalarına Yönelik	
Araştırmalar	31
İdeal Nokta Modellerinin Kestirim Yöntemlerine Yönelik Uygulama Araştırmaları	
.....	41
İdeal Nokta Modellerine Yönelik Diğer Alanlardaki Uygulama Araştırmaları	44
Sınıflama İndekslerine Yönelik Araştırmalar	48
BÖLÜM III.....	55
YÖNTEM.....	55
Araştırmanın Modeli.....	55
Veri Toplama Aracı	55
Evren ve Örneklem	57
Verilerin Analizi	58

Karma Model Tanımlaması.....	59
Varsayımların Test Edilmesi	60
<i>Tek Boyutluluk</i>	60
<i>Yerel Bağımsızlık</i>	61
Model – Veri Uyumu	62
Model ve Yetenek Kestirimleri.....	64
Kesme Puanlarının Oluşturulması.....	65
Sınıflama İndekslerinin Elde Edilmesi	66
Farklılıkların Test Edilmesi.....	67
BÖLÜM IV	71
BULGULAR VE YORUM.....	71
Birinci Araştırma Sorusuna Yönelik Bulgular.....	71
Genel Yorum.....	76
İkinci Araştırma Sorusuna Yönelik Bulgular	76
Sınıflama Tutarlılıklarına Yönelik Bulgular	77
Sınıflama Doğruluklarına Yönelik Bulgular.....	81
Genel Yorum.....	85
Üçüncü Araştırma Sorusuna Yönelik Bulgular	86
Sınıflama Tutarlılıklarına İlişkin Bulgular.....	87
Sınıflama Doğruluklarına Yönelik Bulgular.....	91
Genel Yorum.....	95
Dördüncü Araştırma Sorusuna Yönelik Bulgular	96
Sınıflama Tutarlılıklarına İlişkin Bulgular.....	96
Sınıflama Doğruluklarına Yönelik Bulgular.....	99
Genel Yorum.....	102
BÖLÜM V.....	103
SONUÇ VE TARTIŞMA	103

Sonuçlar.....	103
Yetenek Kestirimlerinin Doğruluğuna İlişkin Sonuçlar	104
Sınıflama İndekslerine Yönelik Sonuçlar	108
Sınıflama Tutarlılığı ve Sınıflama Doğruluğuna İlişkin Karşılaştırmalı Sonuçlar	112
Yetenek Kestirimlerinin Doğruluğuna ve Sınıflama İndekslerine İlişkin Karşılaştırmalı Sonuçlar	113
Öneriler	115
Uygulayıcılara Yönelik Öneriler	115
Araştırmacılara Yönelik Öneriler.....	116
KAYNAKLAR	119
EKLER.....	135
EK 1. Yükseköğrenim Yaşam Doyumu Ölçeği.....	136
EK 2. Gazi Üniversitesi Etik Komisyonu İzni	138
EK 3. Gazi Üniversitesi Eğitim Fakültesi Uygulama İzni	139
EK 4. GAMOM'a İlişkin Ayırtedicilik, Delta ve Eşik Parametreleri	140
EK 5. ATM'ye İlişkin Ayırtedicilik ve Eşik Parametreleri.....	141
EK 6. ATM Altında Yen' in Q3 İstatistikleri	142
EK 7. Karma Model Altında Yen' in Q3 İstatistikleri.....	143
EK 8. GAMOM Altında Yen' in Q3 İstatistikleri	144
EK 9. Kesme Puanları.....	145

TABLÖLAR LİSTESİ

Tablo 1. Örneklemde Yer Alan Öğretmen Adaylarına İlişkin Bölüm Bilgileri.....	58
Tablo 2. Karma Model Madde Spesifikasyonu.....	60
Tablo 3. Faktör Yükleri ve Ortak Yükler.....	61
Tablo 4. $S-\chi^2$ İstatistikleri, sd , ve $RMSEA$ Değerleri.....	63
Tablo 5. AIC , $AICs$, $SABIC$, BIC , ve loglik Değerleri.....	64
Tablo 6. $GAMOM$, ATM ve Karma Model Altında Bayesian ve Nonbayesian Yetenek Kestirimlerine İlişkin Betimsel İstatistikler.....	72
Tablo 7. $c(0,10)$ ve $c(0,90)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları.....	77
Tablo 8. $c(0,10)$ ve $c(0,90)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları.....	81
Tablo 9. $c(0,15)$ ve $c(0,85)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları.....	87
Tablo 10. $c(0,15)$ ve $c(0,85)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları.....	91

Tablo 11. $c(0,20)$ ve $c(0,80)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları	97
Tablo 12. $c(0,20)$ ve $c(0,80)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları	99
Tablo 13. GAMOM'a İlişkin Ayırtedicilik, Delta ve Eşik Parametreleri	138
Tablo 14. ATM'ye İlişkin Ayırtedicilik ve Eşik Parametreleri	139
Tablo 15. Kesme Puanları	143



ŞEKİLLER LİSTESİ

<i>Şekil 1.</i> Subjektif yanıt kategorileri (Roberts vd., 2000).....	19
<i>Şekil 2.</i> Standart hata ortalamalarına göre modellerin değişimi	73
<i>Şekil 3.</i> Standart hata ortalamalarına göre yöntemlerin değişimi	74
<i>Şekil 4.</i> $c(0,10)$ ve $c(0,90)$ noktalarında yöntemlerin değişimi - γ	78
<i>Şekil 5.</i> $c(0,10)$ ve $c(0,90)$ noktalarında modellerin değişimi - γ	79
<i>Şekil 6.</i> $c(0,10)$ ve $c(0,90)$ noktalarında yöntemlerin değişimi - τ	82
<i>Şekil 7.</i> $c(0,10)$ ve $c(0,90)$ noktalarında modellerin değişimi - τ	83
<i>Şekil 8.</i> $c(0,15)$ ve $c(0,85)$ noktalarında yöntemlerin değişimi - γ	88
<i>Şekil 9.</i> $c(0,15)$ ve $c(0,85)$ noktalarında modellerin değişimi - γ	89
<i>Şekil 10.</i> $c(0,15)$ ve $c(0,85)$ noktalarında yöntemlerin değişimi - τ	92
<i>Şekil 11.</i> $c(0,15)$ ve $c(0,85)$ noktalarında modellerin değişimi - τ	93
<i>Şekil 12.</i> $c(0,20)$ ve $c(0,80)$ noktalarında yöntemlerin değişimi - γ	97
<i>Şekil 13.</i> $c(0,20)$ ve $c(0,80)$ noktalarında modellerin değişimi - γ	98
<i>Şekil 14.</i> $c(0,20)$ ve $c(0,80)$ noktalarında yöntemlerin değişimi - τ	100
<i>Şekil 15.</i> $c(0,20)$ ve $c(0,80)$ noktalarında modellerin değişimi - τ	101

SİMGELER VE KISALTMALAR LİSTESİ

MTK	Madde Tepki Kuramı
GAMOM	Genelleştirilmiş Aşamalı Monotonik Olmayan Model
ATM	Aşamalı Tepki Modeli
MLE	Maksimum Likelihood Estimation
WLE	Weighted Likelihood Estimation
MAP	Maximum A Posteriori
EAP	Expected A Posteriori
$\hat{P}$	Beklenen olasılıklar matrisi
W	Ağırlıklandırma matrisi
$\hat{\gamma}$	Sınıflama Tutarlılığı
$\hat{t}$	Sınıflama Doğruluğu
δ	Delta parametresi

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu belirtilmiş, araştırmanın amacı ve önemi hakkında bilgiler verilmiştir.

Problem Durumu

Test etme (testing) eğitimde ve psikolojide ölçme-değerlendirme süreçlerinde birçok amaç doğrultusunda önemli bir aşama olarak öne çıkmaktadır. Eğitim alanındaki test süreçlerinin temel amaçları arasında “öğretme ve öğrenmeyi haber vermek (inform)”, “öğrenci çıktıları değerlendirme”, “öğrenciler hakkındaki kararlardan haberdar etmek” (AERA, APA, & NCME, 2014, s.184-187) yer alırken; psikolojik ölçme süreçlerinin temel amaçları arasında “teşhis”, “nöropsikolojik değerlendirmeler”, “müdahale ve sonuç değerlendirmesi”, hukuki ve idari kararlar”, ve “kişisel farkındalık, sosyal kimlik ve psikolojik sağlık, gelişme ve eylem” (AERA, APA, & NCME, 2014, s. 159-163) yer almaktadır. Eğitimde ve psikolojide ölçme süreçleri ele alındığında, bireylere yönelik değerlendirme ve karar verme amaçlarının öne çıktığı görülmektedir. Dolayısıyla bireylerin eğitim alanındaki veya psikolojik test puanlarına bakılarak bireyler hakkında çeşitli kararlar verilmektedir (Lee, Hanson & Brennan, 2002; Wyse & Hao, 2012).

Bireyler hakkında doğru kararlar verebilmek için ölçme ve değerlendirme sürecine ilişkin bütün işlem adımlarının dikkatle yerine getirilmesi gerekmektedir. Bu süreç test geliştirmeye ilişkin çalışmalarla başlayıp birey hakkında nihai karar verene kadar devam etmektedir. Ölçme ve değerlendirme sürecinde amacın net olarak belirlenmesi ve bu

doğrultuda uygun bir ölçme aracı ile uygun ortam ve koşullarda bireylerden veri toplanması ölçme sonuçlarının doğruluğu için önem taşımaktadır. Ancak bütün bunların yanı sıra elde edilen ölçme sonuçlarının anlamlandırılması için uygun istatistiksel prosedürlerin yürütülmesi, bireyler hakkında verilecek kararlarda kritik önem taşımaktadır.

Ölçme ve değerlendirme sürecinde uygun istatistiksel prosedürlerin seçilmesinde çeşitli faktörler gözetilmektedir. Bu sürecin yürütülmesindeki uygulamaların temelinde yatan faktörlerden biri de ölçülecek yapıdır. Çünkü yapının operasyonel tanımına göre ilgili yapının özellikleri bağlamında (örneğin, başarı gibi optimal performans veya tutum ve ilgi gibi tipik performans) birey davranışlarının elde edilme süreci için test etme prosedürleri oluşturulmaktadır (Crocker ve Algina, 1986, s.4). Örneğin, uygulamada tutumu ölçen bir madde (bilişsel olmayan) ile başarıyı ölçen bir maddenin (bilişsel) temelde farklılaşması beklenmektedir (Sgammato, 2009, s.23). Bu durum ise bireylerin kendilerini yansıttıkları söz konusu yapılara ilişkin madde tepkilerinin altında yer alan yanıt sürecinin anlaşılmasının uygulayıcılar için önemli bir bakış açısı sunabileceğine işaret etmektedir. Bunun yanı sıra AERA, APA, & NCME’ye (2014, s. 13-21) göre “yanıt süreçlerine dayalı kanıt” geçerlilik kanıtı için kaynaklardan biri olarak ele alınmaktadır. Burada yer alan yanıt süreçlerine dayanan geçerlilik ile ifade edilen şudur: testi alanların yanıt süreçlerinin teorik ve ampirik analizleri, ilgili yapı ve performansın detaylı doğası hakkında kanıt sağlayabilir. Dolayısıyla ölçme değerlendirme sürecinde ele alınan yapının doğasının anlaşılması, uygulayıcının daha doğru ölçümler yapmasına katkı sağlayabilir (Stark, Chernyshenko, Drasgow, & Williams, 2006). Bu doğrultuda test etme sürecinde ölçülmesi amaçlanan yapılar bireylerin yanıt verme sürecindeki tepkilerine dayalı olarak bilişsel ve bilişsel olmayan (noncognitive) yapılar ile ifade edilebilmekte, bunlar ise maksimum ve tipik davranışlar olarak ele alınabilmektedir (Tay, Drasgow, Rounds & Williams, 2009; Tay & Ng, 2018).

Bilişsel yapılar maksimum davranış olarak da ifade edilmekte (Reise, 2010) ve bireylerin belli konulardaki yeterlik ya da performanslarına ilişkin özellikleri içermektedir (Hoftsee, 2001, s. 45-46). Maksimum davranışların ölçülmesinde bireyler yanıt verirken, maddelere “hükmetme (dominate)” girişiminde bulunmakta olup yüksek örtük özellik düzeyindeki bireyler maddelere hükmederek olumlu bir şekilde yanıt verme eğilimindedir (Sun, Fraley & Drasgow, 2020; Zampetakis, Lerakis, Kafetsios & Moustakis, 2015). Dolayısıyla, maksimum davranışların ölçülmesi amacıyla uygulanan bir psikometrik araçta yer alan maddelerin yanıtlanma olasılığı bireyin yeteneğine bağlı olarak artış gösteren bir olasılık

dağılımı ile modellenebilmektedir (Cao, Dragow & Cho, 2014; Scherbaum, Sabet, Kern & Agnello, 2013). Farklı bir ifadeyle maksimum davranışların ölçülmesinde yetenek düzeyi arttıkça maddelerin yanıtlanma olasılığının artma eğiliminin ele alındığı monotonik artış sergileyen bir karakteristik eğrisinin gözlenmesi söz konusudur. Bu tür yapılara ilişkin yanıtlanma süreçleri klasik test kuramı ile ele alınabileceği gibi yaygın bir şekilde kullanılan kümülatif Madde Tepki Kuramı (MTK) modelleri ile de modellenebilmektedir. Örneğin iki kategorili verinin modellenmesinde 1, 2, 3 parametrelili lojistik modeller sıklıkla kullanılırken, çok kategorili verinin ele alınmasında aşamalı tepki modeli (ATM), genelleştirilmiş kısmi puan modeli, dereceli ölçek modeli gibi modeller ele alınmaktadır. Madde tepkilerinin altında yatan sürece göre bu modeller kümülatif (Liu & Chalmers, 2018; Thompson, 2014, s. 10) veya baskınlık (dominance; Liu & Wang, 2019; Zhang, Cao, Tay, Luo & Drasgow, 2019) modelleri olarak ifade edilebilmektedir.

Bilişsel olmayan yapı olarak ele alınan tipik davranışlar ile bireylerin kişiliklerine yönelik ölçüm yapılması amaçlanmaktadır (Hofstee, 2001, s. 48). Doğru/yanlış yanıtların olduğu psikometrik araçlar ile yapılan bilişsel yapılara yönelik ölçümlerden farklı olarak, bilişsel olmayan (tutum, doyum gibi tipik) yapılar için birtakım ifadelerin sağlandığı ve yanıtlayıcıların bu ifadelerin kendilerini tanımlama durumlarına göre işaretleme yaptığı öz raporlama aracılığı ile kişilik ölçümleri gerçekleştirilmektedir (Benson & Campbell, 2007). Bilişsel yapılara yönelik maddeler ile bir bireyin “ne yapabileceğine” ilişkin maksimum performans ölçümü söz konusu iken, bilişsel olmayan yapılara yönelik maddelerde bir bireyin “genellikle ne yapacağına” ilişkin tipik bir davranışın ölçülmesi söz konusudur (Chernyshenko, Stark, Drasgow & Roberts, 2007). Bireylerin kendilerini tanımlamalarına ilişkin göstergeleri elde etmek için dereceli ölçekler ile sağlanan iki kategorili ya da çok kategorili maddelerde bireyin kendini en iyi tanımlayan durumu seçmesi beklenir. Bireylerin bilişsel olmayan maddelere yönelik seçimlerinde ortaya çıkan bu yaklaşım/davranış, iç gözlem (introspection) olarak ele alınır. Öz raporlamanın söz konusu olduğu durumlarda birey iç gözlem yaparak kendisine en yakın derecelendirmeyi işaretler (Stark vd., 2006; Tay, Drasgow, Rounds & Williams, 2009; Tay & Drasgow, 2012). Dolayısıyla bireylerin her bir maddeye ilişkin tepkileri ilgili maddenin kapsamına bağlı olarak farklılaşabilmektedir.

Bilişsel olmayan yapıların ölçülmesine yönelik maddelere ilişkin yanıtlanma süreçlerinin modellenmesinde bilişsel yapılarda gözlenen monotonik olasılık dağılımının yanı sıra monotonik olmayan olasılık dağılımının gözlenmesi de olasıdır (Stark vd., 2006).

Dolayısıyla bilişsel olmayan tipik davranış yapılarına ilişkin madde yanıt süreci maksimum davranışları ölçen bilişsel yapılardan farklılaşmaktadır (LaPalme, Tay & Wang, 2018). Çünkü bilişsel olmayan ölçümlerde ölçülen yapıya ilişkin yetenek düzeyi arttıkça maddeyi tasdikleme olasılığı her zaman artmamaktadır. Örneğin, bireylerin dışadönüklük düzeylerini ölçmeye yönelik bir ölçekte yer alan “arkadaşlarımla kafede oturmaktan hoşlanırım” ifadesinin hem dışadönük bireyler tarafından “katılmıyorum” derecelendirmesi hem de içedönük bireyler tarafından “katılmıyorum” derecelendirmesini işaretlemesi beklenebilir. Çünkü yüksek düzeyde dışadönük bir birey kafede oturma eylemini sıkıcı bulacağı için “katılmıyorum” ifadesini işaretleyerek, düşük düzeydeki bireyler ile benzer bir gözlenen yanıt sergileyebilmektedir. Psikolojik ölçme araçlarında yer alan bu tür maddelerin monotonik dağılım sergileyen kümülatif modellerden farklı olarak tek bir noktada zirve (single-peaked) yaparak düşüş gösterdiği monotonik olmayan bir karakteristik eğri ile sonuçlanması söz konusudur (Roberts, 2008; Speer, Robie & Christiansen, 2016). Monotonik olmayan dağılım sergileyen maddeler genellikle “nötr” maddeler (Sgammato, 2009, s. 161) olarak ele alınmaktadır. Yetenek ölçeğinin orta kısımlarında yayılmış olan bu tür maddeler “monotonik olmayan (unfolding)” (Broadfoat, 2008, s. 34; Dalal & Carter, 2014), “orta düzeyli (intermediate)” (Cao, vd., 2014; Zhang vd., 2019) veya doğrudan “nonmonotonik” (Speer vd., 2016) terimleri ile ifade edilmekte ve bu terimler birbiri yerine kullanılmaktadır. Söz konusu maddelerin ele alındığı modeller ise monotonik olmayan (unfolding) (Roberts, Donoghue & Laughlin, 2000) veya ideal nokta modelleri (Cao vd., 2014) olarak adlandırılmaktadır. Bu araştırmada yetenek ölçeğinin orta kısımlarında yayılan ve monotonik olmayan karakteristiğe sahip olan maddeleri tanımlamak için doğrudan “monotonik olmayan” madde ifadesi kullanılmış olup ilgili ölçme modelleri ideal nokta MTK modelleri olarak ifade edilmiştir. Nonmonotonik davranışa sahip olmayan maddeler ise doğrudan monotonik madde olarak ifade edilmiştir.

Son yıllarda psikometri araştırmalarında ortaya çıkan gelişmeler ideal nokta modellerinin tutum (Roberts, Laughlin & Wedell, 1999; LaPalme, Tay & Wang, 2018), kişilik (Stark vd., 2006; LaPalme vd., 2018), mesleki ilgiler (Tay vd., 2009), mesleki doyum (Carter & Dalal, 2010) ve duygulanım (Tay & Kuykendall, 2016; LaPalme vd., 2018) gibi bilişsel olmayan yapılara iyi uyum sağladığını göstermiştir.

Bilişsel olmayan yapılara yönelik kümülatif ve ideal nokta modelleri ile yapılan karşılaştırmalı çalışmalar (Chernyshenko, Stark, Chan, Drasgow & Williams, 2001; Drasgow, Chernyshenko & Stark, 2010; Roberts vd., 1999; Roberts vd., 2000; Stark vd.,

2006; Roberts, 2008; Tay, Ali, Drasgow & Williams, 2011; Tay & Drasgow, 2012) ideal nokta modellerinin daha uygun olabileceğini öne sürmüştür. Çünkü bilişsel olmayan ölçümlere ilişkin maddelerin yanıtlanma sürecinde yanıtlayıcılar likert-türü maddelere yanıt vermeden önce iç gözlem yaparak bir eşleşme stratejisi ile maddede kendilerini en iyi tanımlayan derecelendirmeyi işaretleme eğilimi gösterir (Drasgow vd., 2010). Bu tür ölçümlere ilişkin yanıt süreçleri çan eğrisi şeklinde olan madde karakteristik eğrilerine yol açabilmekte ve bu tür maddelerin monotonik fonksiyonlara dayalı olarak işlem yapan kümülatif model altında modellenmesi tutarsızlık ile sonuçlanabilmektedir (Cao, vd., 2014). Dolayısıyla maddelerin altında yatan yanıt süreçlerinin modellenmesi ölçümlerin doğruluğu için önemli bir husus olarak baş göstermiş olup buna dayalı olarak ideal nokta ölçme modellerinin çeşitli alanlardaki kullanımı öne çıkmıştır. Bu doğrultuda kullanılan ideal nokta modelleri ele alındığında, genelleştirilmiş aşamalı monotonik olmayan modelin (generalized graded unfolding model – GAMOM) sıklıkla araştırmalarda ele alındığı gözlenmiştir.

İdeal nokta modellerine yönelik çalışmalar incelendiğinde ölçek geliştirme (DeNunzio, 2019; Stark vd., 2006) ve puanlama (Stark vd., 2006), bilgisayara uyarlanmış testler (Wang & Liu, 2011), puanlayıcı değerlendirmeleri (Wang & Engelhard, 2019), yapısal eşitlik modelleme (Lim, 2015; Usami, 2011) gibi çeşitli uygulama alanlarında bu modellerin kullanımının öne çıktığı gözlenmiştir. Genel olarak ideal nokta modellerinin kümülatif modeller ile karşılaştırıldığı çalışmalarda, bilişsel olmayan yapıların ele alınmasında ideal nokta modellerinin etkililiği ele alınmıştır (Tay vd., 2011).

İdeal nokta modellerine ilişkin incelemelerde madde ve yetenek ölçeği arasındaki olasılıksal modele göre madde nitelikleri değerlendirilerek maddeler monotonik-monotonik olmayan olarak değerlendirilmektedir. Monotonik olmayan maddelerin belirlenmesi açısından maddelerin altındaki yanıt süreçlerinin modellenmesi önemli hususlardan biridir. Buna dayalı olarak maddenin baskınlık sürecine göre mi veya ideal nokta yanıt sürecine göre mi ele alınması gerektiğine dair çıkarımlar yapılmaktadır. Genel olarak monotonik olmayan maddeler örtük yetenek dağılımının orta kısımlarında madde lokasyon parametrelerine sahip maddelerdir (Cao vd., 2014). Maddelerin lokasyonu sıfıra yakın oldukça monotonik olmama (nonmonotoniklik) derecesi artarken, sıfırdan uzaklaştıkça monotonik olmama (nonmonotoniklik) derecesi azalmaktadır. Maddelere ilişkin kestirilen yer parametreleri/delta parametrelerinin ortalamaya olan mesafesi değerlendirilerek monotonik olma-monotonik olmama özelliği birtakım aralıklarda

kategorize edilebilmektedir (Roberts, Donoghue & Laughlin, 2002; Stark vd., 2007; Ling, Zhang, Locke, Guangmin & Li, 2016). Ölçme aracında yer alan ideal nokta madde yüzdesine bağlı olarak kümülatif modellerin ve ideal nokta modellerinin farklı uyum ve doğrulukla sonuçlanabildiğini göstermiştir (Cao vd., 2014). Maddelerin sergiledikleri davranışlara göre orta derece (moderate) veya uç (extreme) madde olacak şekilde kategorize edilmesi (Ling vd., 2016) ve orta derecedeki madde yüzdelерinin değişimine bağlı olarak model etkililiklerinin değişmesi (Dalal & Carter, 2014) ölçme aracında yer alan maddelere göre model tanımlamasının önemli olabileceğine işaret etmektedir. Dolayısıyla bu araştırmada monotonik olan ve monotonik olmayan madde türlerine ilişkin modellerin ayrı ayrı kestirimlerinin yanı sıra madde karakteristik fonksiyonlarının altındaki yanıt süreçlerine dayalı olarak sergiledikleri davranışların ele alındığı ve her iki model bileşimine dayalı karma model tanımlamaları ile de kestirimlerin yapılması ele alınmıştır.

Madde Tepki Kuramı çerçevesinde ölçme modelinin madde yanıt sürecine uygun bir şekilde seçilmesi veya oluşturulması birey puanlarının doğru bir şekilde elde edilmesindeki ve dolayısıyla sınıflama kararlarının doğru bir şekilde yapılmasındaki faktörlerden biridir. Sınıflama kararlarının doğru bir şekilde yapılmasındaki önemli faktörlerden bir diğeri ise elde edilecek yetenek kestirimlerinin doğruluğudur (Yang, Poggio & Glasnapp, 2006). Bu doğrultuda ise kestirim için seçilecek yöntem önem kazanmaktadır. MTK altında araştırmanın amacına bağlı olarak çeşitli yetenek kestirim yöntemleri kullanılmakta olup bu yöntemler kullandıkları algoritmalara dayalı olarak farklı şekillerde kategorize edilebilmektedir. Bunlardan biri yöntemlerin önsel dağılım belirleme durumlarına göre bayesian ve nonbayesian yöntemler olarak iki kategoride ele alınmasıdır. MTK'da sıklıkla kullanılan maksimum olabilirlik (Maximum Likelihood estimation-MLE) ve ağırlıklandırılmış olabilirlik kestirim yöntemleri (Weighted Likelihood Estimation-WLE) nonbayesian yöntemler olarak ele alınırken, MAP (Maximum A posteriori) ve EAP (Expected A posteriori) kestirim yöntemleri bayesian yöntemler olarak ele alınmaktadır. MTK'da farklı kestirim yöntemlerinin kullanımı doğruluğu etkileyen önemli bir faktördür (Borgatto, Azevedo, Pinherio & Andrade, 2015; Ching-Fung, 2002; Rose, 2010; Wainer & Thissen, 1987; Wang & Vispoel, 1998; T. Wang & Wang 2001). Bu doğrultuda farklı yöntemlerden elde edilen yetenek kestirimlerinin doğruluğunun incelenmesi ve aynı zamanda sınıflama çalışmalarındaki kullanımı ile karşılaştırılması amaçlanmıştır.

Sınıflama çalışmalarındaki bir diğeri önemli husus seçilecek kesme puan noktasıdır. Birey sınıflamasındaki karar tutarlılığı veya karar doğruluğunun derecesi seçilen (employed)

kesme puanına ve kesme puanının puan dağılımı içerisindeki lokasyonuna özgüdür (AERA, APA, & NCME, 2014, s. 40). Wyse ve Babcock (2016) ve Lathrop ve Cheng (2014) çalışmasındaki koşullarda sınıflama indekslerindeki farklılaşmaya ilişkin sonuçları kesme puanı ve yetenek dağılımının etkileşiminin bir sonucu olarak değerlendirmişlerdir. Bununla birlikte kümülatif ve ideal nokta modelleri arasındaki farklılığın yetenek ölçeğinin üst kısımlarında ortaya çıkabileceği çeşitli çalışmalarda ele alınmaktadır (Roberts vd., 1999; Ling vd., 2016). Ling vd. (2016) çalışmasında birey sıralamasına ilişkin çıkarımlar elde etmek için sırasıyla üst uçta yer alan 100, 200, 300, ve 500 kişi ile bütün örneklemdeki kişileri ele alarak kümülatif ve ideal nokta yanıt süreçlerine dayalı elde edilen yetenek kestirimleri arasındaki ilişkiyi pearson korelasyon ile incelenmiştir. Bu incelemeler ile yüksek puanlı yanıtlayıcıların yer aldığı alt örneklemelerde korelasyonun bütün bireylerin ele alındığı gruptan daha zayıf olduğu dolayısıyla üst uçtaki birey sıralamalarının farklılaştığı gösterilmiştir. Roberts vd. (1999) ideal nokta özelliği gösteren verinin kümülatif modellerle analiz edildiğinde, ilgilenilen yetenek açısından yüksek düzeyde yer bireylere yönelik kestirimlerin olumsuz etkilendiğini simülasyon ve ampirik sonuçlar ile göstermiştir.

Sınıflama çalışmalarının kesme puan noktasına bağlı olması ve çalışmada ele alınan ölçme modelleri arasındaki farklılığın etkisinin yetenek ölçeğinin üst kısımlarında ortaya çıkma potansiyelinden kaynaklı olarak, yetenek ölçeğinin farklı noktalarında yerleştirilmiş kesme puanlarına dayalı incelemeler yapılmıştır. Bu araştırmada farklı kesme puan noktalarında model ve yöntemlerin ele alınması ile üç faktörün eşzamanlı olarak sınıflama indeksleri üzerindeki etkisi yetenek kestirimlerinin doğruluğuna ilişkin incelemeler ile birlikte sunulmuştur. Kesme puanı, yöntem ve model olmak üzere sınıflama kararlarını etkileyecek bu üç faktörün eşzamanlı olarak ele alındığı bu çalışmanın pratikte karşılaşılabilecek durumlar için hangi istatistiksel prosedürlerin takip edilmesi gerektiği hakkında detaylı bir resim sunabileceği düşünülmektedir.

Araştırmanın Amacı

Bu çalışmada bilişsel olmayan yapıların ele alınmasında madde yanıt süreçlerinin değerlendirilmesi ve sınıflama çalışmalarında kümülatif, ideal nokta ve karma ölçme modellerine dayalı incelenmesi amaçlanmıştır. Farklı madde yanıt süreçlerine dayalı kestirimlerde kümülatif modellerden ATM (Samejima, 1969), ideal nokta modellerinden

GAMOM (Roberts vd., 2000) ele alınmıştır. Sınıflama çalışmaları kullanılacak ölçme modelinin yanı sıra yetenek kestirimlerinin doğruluğundan da etkilenmekte olup bu etkileşim de yetenek ölçeğine dayalı olarak farklılaşabilmektedir. Bu doğrultuda araştırmanın amacı ATM, GAMOM ve karma modeller altında WLE, MAP ve EAP kestirim yöntemleri çerçevesinde elde edilen yetenek kestirimlerinin doğruluğu ile yetenek ölçeğinin düşük ve yüksek ucunda yer alan kesme puan noktalarında elde edilen sınıflama indekslerinin incelenmesidir.

Araştırmanın Önemi

Bu araştırmanın amacı MTK'ya dayalı bir sınıflama çalışması yürüterek ölçme modelleri, yetenek kestirim yöntemi ve kesme puanının sınıflama indekslerine olan etkisini incelemektir. Sınıflama çalışmaları genel anlamda ele alınacak olursa, Klasik Test Kuramı'na dayalı sınıflama çalışmalarının 1970'li yıllardan itibaren yapıldığı (Subkoviak, 1976); MTK'ya dayalı sınıflama çalışmalarının ise son yıllarda ön plana çıktığı görülmektedir (Guo, 2006; Lee, 2010; Rudner, 2001, 2010). Her geçen gün karmaşılaşan veri yapıları ve gelişen bilgisayar teknolojileri ile birlikte MTK çeşitli alanlarda sıklıkla tercih edilmektedir. Dolayısıyla MTK çerçevesinde yürütülecek sınıflama çalışmalarına ihtiyaç duyulduğu düşünülmektedir. Bu çalışmada MTK yaklaşımlarından olan Rudner'ın (2001, 2005) yöntemi ile sınıflama indeksleri elde edilerek incelenmiştir. Söz konusu indekslerin elde edilmesinde kullanılan MTK yaklaşımının pratikte çeşitli uygulamada kullanılabilme potansiyeline sahip olması itibariyle sınıflama çalışmalarının yürütülmesindeki kullanımına katkı sağlayacağı düşünülmektedir.

Ölçülen yapılar bilişsel ve bilişsel olmayan yapılar olarak kategorize edilebilmekte ve madde yanıt süreçleri de ilgili yapıya dayalı olarak farklılaşabilmektedir. Bilişsel yapılar alışlageldik modeller ile analiz edilebilirken (ATM, lojistik modeller) bilişsel olmayan yapıların analiz edilmesinde mevcut verinin monotonik olmama özelliklerini de ele alan ölçme modellerinin göz önünde bulundurulması önem kazanmaktadır. Bu çalışma ideal nokta modellerin bilişsel olmayan veriye uygulamasını çeşitli faktörler bağlamında ele alan bir çalışmadır. Dolayısıyla maddelerin altındaki yanıt süreçlerini ele alan monotonik-monotonik olmayan madde davranışlarının incelenmesi uygun model seçimi için önem kazanmaktadır. Çünkü maddeleri puanlamak için yanlış model seçildiğinde, MTK kestirimlerinin daha düşük doğrulukla sonuçlanması beklenir (Ferrando & Chico, 2007). Çalışmada kümülatif model, ideal nokta model ve her iki modele dayalı oluşturulan

kümülatif-ideal nokta karma modele dayalı olarak da kestirimler yapılmıştır. Çalışmada madde yanıt süreçlerini yansıtan farklı türde model kullanımları ile madde davranışlarının kestirimler üzerindeki etkisine katkı sunulduğu için çalışma önem taşımaktadır.

Yetenek kestirimlerinin standart hatası (Cheng, Liu & Behrens, 2015) dolayısıyla ölçme sonuçlarının doğruluğu (Yang, Poggio & Glasnapp, 2006) sınıflama indekslerini etkileyen bir diğer önemli faktördür. Bu çalışmada sınıflama doğruluğu ve tutarlılığını etkileyeceği düşünülen farklı yetenek kestirim yöntemleri dikkate alınmıştır. Kestirim yöntemleri bayesian ve nonbayesian yöntemler olarak ele alınmış olup bu yöntemler arasından uygun olanının seçilmesi sınıflama kararları için önem taşımaktadır. Ayrıca bayesian ve nonbayesian kestirim yöntemlerinin matematiksel alt yapısından dolayı bayesian yöntemlerde karşılaşılan ortalamaya/moda çekilme etkisinin de sınıflama indekslerine olan etkisi ölçme sonuçlarının doğruluğunun ötesinde kestirimin yapıldığı yetenek ölçeğine ilişkin yayılım ölçüsü açısından da ele alınmıştır. Bu doğrultuda elde edilen yetenek kestirimlerinin standart hatası, yetenek dağılımının yayılım ölçüleri bağlamında değerlendirilmiştir. Dolayısıyla farklı yöntemlerden elde edilen farklılıkların altında yatan nedenler yöntemlerin teknik özellikleri de göz önünde bulundurularak incelenmiştir. Yöntemlere ilişkin teknik hususların ele alınmasının pratikteki uygun sınıflama koşullarının belirlenme çalışmalarına katkı sağlayacağı düşünülmektedir.

Sınıflama kararlarını etkileyen faktörler ele alındığında bir diğer önemli faktörün ise kesme puanının yetenek ölçeği üzerindeki yeri olduğu görülmektedir (Lathrop & Cheng, 2014; Wyse & Babcock, 2016; Zhang, Du Chen, Xin & Chen, 2017). Wyse ve Babcock (2016) ve Lathrop ve Cheng (2014) çalışmasındaki koşullarda sınıflama indekslerindeki farklılaşmaya ilişkin sonuçları kesme puanı ve yetenek dağılımının etkileşiminin bir sonucu olarak değerlendirmişlerdir. Bu çalışmada ölçme modelleri ve yetenek kestirimlerinin sınıflama indekslerine olan etkisi farklı kesme puan noktalarına dayalı olarak üç faktörlü olarak ele alınmış olup her bir koşulda kullanılacak yöntem ve model seçimine ilişkin önerilerde bulunulmuştur. Çalışmada model, yöntem ve kesme puanının eşzamanlı olarak değerlendirilerek etkileşimlerine yönelik sonuçların sunulması gerçekte karşılaşılabilecek olası senaryolar için detaylı çıkarımlar sunmasından dolayı araştırma önem taşımaktadır.

Sınıflama çalışmaları incelendiğinde, çok sayıda çalışmanın yapıldığı ancak bu çalışmaların çoğunlukla simülasyon veri üzerinde gerçekleştirildiği görülmektedir (LaFond, 2014; Lathrop & Cheng, 2013; Lathrop & Cheng, 2014). Bu çalışmada psikolojik

bir ölçek kullanılarak veri toplanmış olup söz konusu gerçek veri üzerinde incelemeler yapıldığından alana pratik anlamda katkı sunacağı düşünülmektedir.

Problem İfadesi

Kümülatif (ATM), ideal nokta (GAMOM) ve karma modeller ile bayesian (MAP ve EAP) ve nonbayesian (WLE) yöntemler birlikte ele alındığında yetenek kestirimlerinin doğruluğunu ve farklı kesme puan noktalarında elde edilen sınıflama geçerliğini (doğruluğu) ve sınıflama güvenilirliğini (tutarlılığı) nasıl etkilemektedir?

Araştırma Soruları

1. ATM, GAMOM ve karma modeller altında bayesian ve nonbayesian yöntemler ile elde edilen yetenek kestirimlerinin betimsel istatistikleri nasıldır ve yetenek kestirimlerinin doğruluğu anlamlı olarak farklılaşmakta mıdır?
2. Yetenek ölçeğinin $c(0,10)$ ve $c(0,90)$ noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında
 - a. ATM, GAMOM ve karma modellerin sınıflama tutarlılıklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?
 - b. ATM, GAMOM ve karma modellerin sınıflama doğruluklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?
3. Yetenek ölçeğinin $c(0,15)$ ve $c(0,85)$ noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında
 - a. ATM, GAMOM ve karma modellerin sınıflama tutarlılıklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?
 - b. ATM, GAMOM ve karma modellerin sınıflama doğruluklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?

4. Yetenek ölçeğinin $c(0,20)$ ve $c(0,80)$ noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında
- ATM, GAMOM ve karma modellerin sınıflama tutarlıklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?
 - ATM, GAMOM ve karma modellerin sınıflama doğruluklarına olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?

Sınırlılıklar

Bu çalışmada araştırmacı tarafından lisans öğrencilerine uygulanan psikolojik ölçek ile veri elde edilmiş olup incelemeler elde edilen gerçek veri üzerinde gerçekleştirilmiştir. Araştırmada kullanılan veri delta parametre derecelerine göre 9 tane monotonik ve 20 tane monotonik olmayan maddenin yer aldığı 29 maddelik ölçme aracı ve 1682 kişilik örneklem ile sınırlıdır. Çalışma verisinde tamamı doğru olan veya tamamı yanlış olan yanıt örüntüleri de mevcuttur. Bu doğrultuda araştırma gerçek veri bağlamında gözlenen maddelerin monotonik olma ve monotonik olmama dereceleri, monotonik olan ve olmayan madde sayıları, mevcut birey yanıt örüntüleri ve mevcut örneklem büyüklüğü ile sınırlıdır. Bu nedenle araştırmadan elde edilen sonuçlar araştırma verisinin sahip olduğu bu özellikler çerçevesinde sınırlandırılmıştır.



BÖLÜM II

KAVRAMSAL ÇERÇEVE

Bu bölümde madde yanıt süreçlerine, madde yanıt süreçlerine göre ele alınan ölçme modellerine, MTK çerçevesinde ele alınan sınıflama yaklaşımlarına, ve bayesian ve nonbayesian yetenek kestirim yöntemlerine ilişkin kuramsal bilgilere yer verilmiştir.

Madde Yanıt Süreçleri

Ölçülmesi amaçlanan yapıya bağlı olarak bireylerin maddelere verdikleri tepki davranışları farklı bir ifadeyle madde yanıt süreçleri farklılaşabilmektedir. Madde yanıt sürecinin yanlış tanımlanması yetenek kestirimlerinin doğruluğunu ve yanıtlayıcıların davranışları hakkında yapılan tahminleri olumsuz olarak etkileyebilmektedir (Stark vd., 2006). Bu nedenle madde tepkilerinin altında yatan yanıt süreçlerinin incelenmesi ve anlaşılması kritik öneme sahiptir (Reise, 2010).

MTK, yetenek düzeylerine ilişkin kestirimlerin hem maddelerin özelliklerine hem de bireylerin yanıtlarına dayalı olarak yapıldığı model-temelli ölçme ile ilgili bir kuramdır (Embretson & Reise, 2000). MTK'nın Klasik Test Kuramı'na karşılık birçok avantajının olduğu bilinmektedir. Madde yanıt sürecinin doğru bir şekilde tanımlanmasına izin vermesi de bu avantajlardan biridir (Stark vd, 2006). MTK modellerine ilişkin yanıt süreçleri madde karakteristik eğrilerine ilişkin fonksiyonlara göre “baskınlık (dominance) yanıt süreci” ve “ideal nokta yanıt süreci” olmak üzere iki kategoride ele alınabilmektedir (Wang, 2013).

Baskınlık Yanıt Süreci

MTK modellerinden yaygın bir şekilde kullanılan iki kategorili 1, 2 ve 3 parametrelili lojistik modeller ile ATM ve genelleştirilmiş kısmi puan modeli gibi çok kategorili modellerde bireylerin yetenek düzeyleri arttıkça maddeyi yanıtlama olasılıklarının da artacağı varsayılır. Burada ele alınan yanıtlama sürecinde yetenek düzeyine koşullu olarak kümülatif bir artış söz konusudur ve bu süreç baskınlık süreci olarak ele alınır (Stark vd., 2006).

İdeal Nokta Yanıt Süreci

İdeal nokta modellerinde madde karakteristik eğrisi tek bir noktada zirve yapar ve o noktadan sonra düşüş gerçekleşir. Bu nedenle kümülatif modellerin aksine çan eğrisi şeklinde modellenen monotonik olmayan bir eğri söz konusudur. Dolayısıyla ideal nokta MTK modellerinde maddeyi yanıtlama olasılığı bireyin lokasyonu ile madde lokasyonu arasındaki mesafeye dayalı olarak hesaplanır (Wang, 2013). Bir ideal nokta yanıt sürecinin varsayımı şudur: yanıtlayıcının ilgili maddeyi onaylama eğilimi örtük boyut üzerindeki lokasyonuna ilişkin algılaması (örn; yanıtlayıcının ideal noktası) ile maddenin bu boyuttaki lokasyonunun benzerlik derecesine bağlı olarak artmakta veya azalmaktadır (Stark vd., 2006). İdeal nokta modellerinde ele alınan bu yanıt süreci bireylerin yetenek ölçeği üzerindeki lokasyonlarına dayalı gerçekleştirildiği için ideal nokta yanıt süreci olarak adlandırılmaktadır.

Kişilik ölçümlerine ilişkin ölçeklerdeki maddelerin çoğunun baskınlık yanıt sürecine nispeten ideal nokta sürecine daha iyi uyum sağlayabileceği ele alınan hususlardan biridir (Brown & Maydeu-Olivares, 2010). Buna dayalı olarak yanıt süreci ile kestirilecek özellik arasındaki ilişkinin doğru bir fonksiyonel tanımlamasının sağlanması ile bireylerin örtük özelliklerine dayalı daha doğru kestirimlerin elde edilebileceği gösterilmiştir (Carter & Dalal, 2010; Chernyshenko vd., 2007; Roberts vd., 2000; Stark vd., 2006).

Madde Yanıt Süreçlerine Göre Ölçme Modelleri

MTK modelleri boyutluluk, yanıt örüntülerinin iki veya ikiden fazla kategorili olması, parametrik-nonparametrik gibi çeşitli faktörler ele alınarak kategorize edilebilmektedir. Burada ele alınabilecek faktörlerden biri de ölçülecek yapıdır. Ölçülecek yapının

karakteristiğine (bilişsel ve bilişsel olmayan gibi) göre madde tepki süreçleri dolayısıyla da seçilecek ölçme modelleri de farklılaşabilmektedir. Ölçme modelleri madde yanıt sürecindeki monotonik olma/monotonik olmama varsayımına dayalı olarak kümülatif ve ideal nokta modelleri şeklinde kategorize edilebilmektedir (Fan, 2020).

Kümülatif Modeller

Kümülatif modeller ilk olarak bilişsel özelliklerin ölçülmesi için geliştirilmesine rağmen daha sonra kişilik, tutum ve ilgi gibi bilişsel olmayan örtük özelliklerin ölçülmesi için kullanılmıştır (Tay & Ng, 2018). Bilişsel olmayan veriye uygulanan MTK modelleri yaygın olarak monotonik artış gösteren fonksiyonlar tarafından tanımlanan olasılıksal (probabilistic) modellerdir. Bu MTK modellerinin arkasındaki öncüle göre bir bireyin sahip olduğu yapıya ilişkin düzey arttıkça bir maddeyi seçme (veya doğru yanıtlama) olasılığı artar (Andrich, 1996).

Ancak bilişsel olmayan veri yapısı kümülatif modellerde varsayılan monotonik artış gösteren fonksiyonlardan farklı olarak monotonik olmayan bir fonksiyon ile tanımlanabilmektedir (Cao vd., 2014). Bu tür durumda bilişsel olmayan özelliklerin ölçülmesinden elde edilen ve monotonik olmayan davranış sergileyen veri “ideal nokta” verisi olarak ele alınmakta ve bu doğrultuda geliştirilen ideal nokta modelleri ile analiz edilebilmektedir.

İdeal Nokta Modelleri

Bilişsel davranışların ölçülmesine yönelik maddelerde doğru-yanlış şeklinde değerlendirmeler söz konusudur. Ancak bilişsel olmayan kavramların ölçülmesi için kullanılan ölçme araçlarında yer alan katılıyorum-katılmıyorum şeklindeki derecelendirmeli maddelerde gözlenen yanıtlarda yön daima açıkça belirtilmemektedir. İdeal nokta modellerinde maddeyi tasdikleme (endorsement) bireyin yetenek ölçeği üzerindeki konumu ile maddenin yetenek ölçeği üzerindeki lokasyonunun bir fonksiyonu olarak ele alınır (Roberts vd., 2000). Yanıtlayıcının yetenek ölçeği üzerindeki konumu ile madde lokasyonu arasındaki mesafe azaldıkça bireyin maddeyi tasdikleme ihtimali artarken bu mesafe arttıkça yani bireyin yeteneği madde lokasyonundan uzaklaştıkça bireyin maddeyi tasdikleme olasılığı düşmektedir (LaPalme vd., 2018). Bu durumda

maddelerin işaretlenme olasılığı yetenek düzeyine bağlı olarak monotonik bir şekilde artmamakta hatta bir noktada zirve yapıp düşüş gösterebilmektedir. Dolayısıyla bilişsel olmayan yapılara ilişkin maddelerde madde fonksiyonları monotonik olmayan bir dağılım sergileyebilmektedir.

Bilişsel olmayan yapıların ölçülmesindeki madde yanıtlarına uymak üzere (genelleştirilmiş) hiperbolik cosine model (Andrich, 1988, 1996; Andrich & Luo, 1993), Parella Model (Hoijtink, 1990, 1991), aşamalı monotonik olmayan model (Roberts & Laughlin, 1996; generalized unfolding model) ve genelleştirilmiş aşamalı monotonik olmayan model (Robert vd., 2000; generalized graded unfolding model) olmak üzere birçok ideal nokta modeli geliştirilmiştir. Hem iki kategorili hem çok kategorili yanıtlara genellenebilmesi ve aynı zamanda model kestirimleri için uygun yazılım programlarına (Chalmers, 2012; Roberts, Fang, Cui & Wang, 2006) ulaşılabilirliğinden dolayı GAMOM, ideal nokta modelleri arasında en sık kullanılan popüler bir modeldir (Cao, Son & Tay, 2018; Fan, 2020).

Genelleştirilmiş Aşamalı Monotonik Olmayan Model

GAMOM hem ikili (binary) hem de çok kategorili aşamalı yanıtlara izin veren parametrik bir ideal nokta MTK modelidir. MTK kapsamında ideal nokta modeline göre, bir bireyin negatif ve pozitif etki uçlarını bağlayan bir yetenek ölçeği üzerinde bireyin lokasyonunun ve ilgili ifadenin lokasyonunun birbirine yakın olma derecesine bağlı olarak belli bir ifadeye katılacağı öne sürülmektedir. GAMOM iki parametrelili lojistik modelin çok kategorili durumdaki uzantısı olan Muraki'nin (1992) genelleştirilmiş kısmi puan modelini içinde barındırmaktadır (Roberts vd., 2000).

GAMOM, yanıt sürecine dair dört temel öncüle dayalı olarak geliştirilmiştir (Roberst & Laughlin, 1996; Roberts vd., 1998; Roberts vd., 2000);

- (1) Bir birey yetenek ölçeği üzerindeki tutumuna yakın olma derecesine bağlı olarak bir maddeye katılma eğilimi göstermektedir.
- (2) Her bir gözlenen yanıt için iki tane subjektif yanıt mevcuttur.
- (3) Subjektif yanıtlar gözlenen yanıtların aksine baskınlık süreci ile modellenirler.
- (4) Yanıt kategori eşik parametreleri ideal nokta çevresinde (yetenek kestirimi noktası) maddeye göreli olarak simetriktir. Yani bir bireyin tutumunun lokasyonu ile ilgili maddenin lokasyonu arasındaki mesafe ideal noktanın her iki tarafında da aynıdır.

Birinci öncül kapsamında, bir maddeye ilişkin fikrin bir bireyin görüşünü yansıtmaya derecesi yetenek ölçeği üzerindeki bireyin maddeye olan yakınlığı ile belirlenmekte, bireyin yeteneği ve delta arasındaki mesafenin 0'a yaklaşma derecesine bağlı olarak bireyin maddeye katılma olasılığı artmaktadır.

İkinci öncüle göre bir yanıtlayıcının maddeye verdiği gözlenen yanıt kategorisinin (observed response category) altında iki sebep olabilir. Örneğin bir yanıtlayıcının kürtaja karşı nötr bir tutum sergilediğini düşünelim. Bu birey negatif maddeye de, pozitif maddeye de «tamamen katılmıyorum» yanıtını verebilir. Ancak bireyin maddeye yanıt vermesinde iki durumdan biri söz konusudur: Maddenin kapsamı bireyin tutumundan daha pozitif ise (yanıtlayıcının lokasyonu maddenin yer parametresinin altında ise) «üst düzeyden tamamen katılmıyorum (strongly disagree from above)»; Maddenin kapsamı bireyin tutumundan daha negatif ise (yanıtlayıcının lokasyonu maddenin yer parametresinin üstünde ise) «alt düzeyden tamamen katılmıyorum (strongly disagree from below)» olacak şekilde iki tane olası tepki mevcuttur. Her bir objektif yanıt kategorisi altında yatan bu iki olası tepki subjektif yanıt kategorileri olarak adlandırılır ve objektif yanıt kategorisi bu iki subjektif yanıt kategorisinin toplamı ile ifade edilir. Detaylandırılacak olursa 4'lü dereceye sahip bir maddenin modellenmesinde 8 tane subjektif yanıt kategorisi mevcuttur ve her biri aşağıdaki gibi ifade edilmektedir:

0 = Alt düzeyden tamamen katılmıyorum

1 = Alt düzeyden katılmıyorum

2 = Alt düzeyden katılıyorum

3 = Alt düzeyden tamamen katılıyorum

4 = Üst düzeyden tamamen katılıyorum

5 = Üst düzeyden katılıyorum

6 = Üst düzeyden katılmıyorum

7 = Üst düzeyden tamamen katılmıyorum

Üçüncü öncüle göre gözlenen yanıt olasılıklarının aksine subjektif yanıt kategorisi olasılıklarına ilişkin fonksiyonlar kümülatif bir dağılımı takip eder (Andrich & Luo, 1993). Bu doğrultuda GAMOM'a ilişkin subjektif yanıt kategorilerinin kümülatif modellerden Muraki'nin (1992) genelleştirilmiş kısmi puan modelini takip ettiği varsayılarak olasılıklar

modellenmiştir. Herhangi kümülatif modellerden biri de kullanılabilir olmasına rağmen genelleştirilmiş kısmi puan modelinin genellenebilirliğinden ötürü bu modele dayalı olarak GAMOM geliştirilmiştir. Genelleştirilmiş kısmi puan modeli, GAMOM'daki subjektif yanıtlara uygulandığında aşağıdaki gibi tanımlanır:

$$P(Y_i = y | \theta_j) = \frac{\exp\{\alpha_i[y(\theta_j - \delta_i) - \sum_{k=0}^y \tau_{ik}]\}}{\sum_{w=0}^M \{\exp\{\alpha_i[w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}]\}} \quad (1)$$

Subjektif yanıt kategorilerine ilişkin eşik parametrelerinin toplamı aşağıda gösterildiği gibi 0 olacak şekilde sınırlandırılmıştır.

$$\sum_{k=0}^M \tau_{ik} = 0 \quad (2)$$

Eşitliklerde yer alan semboller aşağıdaki anlamlara gelmektedir:

Y_i – i ifadesine verilen bir subjektif yanıt

$y = 0$ ($y = 0, 1, 2, \dots, M$) – maddenin alt düzeyinden en güçlü katılmama seviyesi

$y = M$ - maddenin üst düzeyinden en güçlü katılmama seviyesi

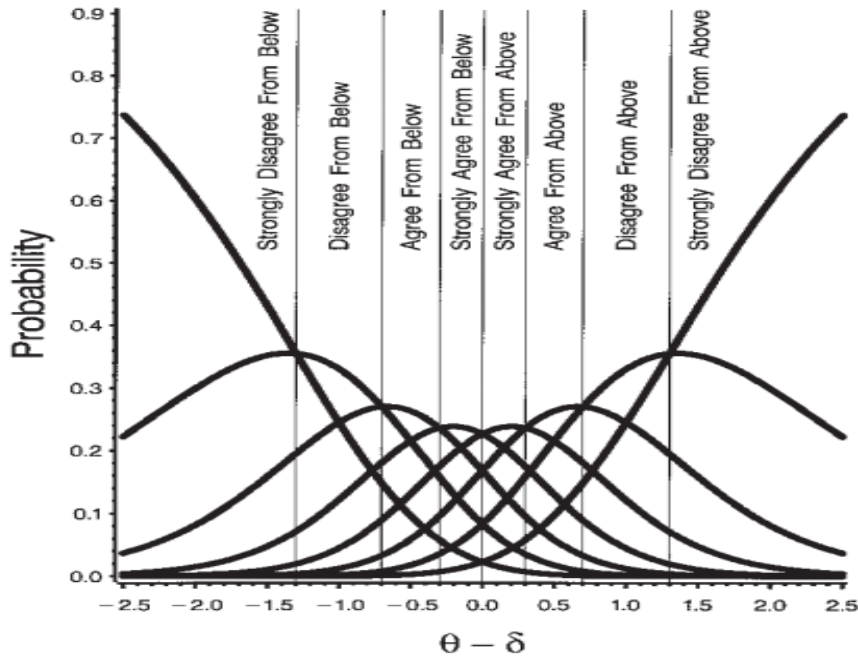
M – subjektif yanıt kategori sayısının bir eksiği (subjektif yanıt kategorisi -1)

α_i – i ifadesinin ayırtediciliği

δ_i – i . maddenin örtük kontinuum üzerindeki yer parametresi

τ_{ik} – i . maddenin lokasyonuna göreli olarak kontinuum üzerinde k . subjektif yanıt kategori eşik parametresinin yeri

GAMOM altında dörtlü likert düzeyinde olan hipotetik bir madde için dört tane gözlenen yanıt kategorisine karşılık gelen sekiz tane subjektif yanıt kategorileri sırasıyla “Alt düzeyden tamamen katılmıyorum”, “Alt düzeyden katılmıyorum”, “Alt düzeyden katılıyorum”, “Alt düzeyden tamamen katılıyorum”, “Üst düzeyden tamamen katılıyorum”, “Üst düzeyden katılmıyorum”, ve “Üst düzeyden tamamen katılmıyorum” şeklinde ifade edilmekte olup ilgili model Şekil 1’de gösterilmiştir:



Şekil 1. Subjektif yanıt kategorileri (Roberts vd., 2000)

Maddede dört tane gözlenen yanıt kategori olduğu için her bir yanıt kategorisine karşılık gelen iki tane subjektif yanıt kategorisi ve dolayısıyla toplamda sekiz tane subjektif yanıt kategorisi modellenmektedir. Yedi tane dikey doğru ise ardışık subjektif yanıt kategorisi olasılıklarının kesişme noktalarını göstermektedir. Bu lokasyonlar subjektif yanıt kategori eşiği ile ifade edilmektedir. Başlangıç τ_{i0} eşiği keyfi olarak 0 ile tanımlanmıştır. Yatay eksen yetenek kontinuumunu temsil etmektedir. Bu eksen birey lokasyonu ve madde lokasyonu arasında işaretlenmiş (signed) farklılığın birimleri ile ölçeklenir. Ordinate ekseni ise bireyin subjektif yanıtının olası sekiz kategoriden birine düşme olasılığını indekslemektedir.

Şekil 1’de yetenek ölçeği farklı bir subjektif yanıtın en olası olduğu 8 ayrı aralığa bölünmüştür. Verilen belli eşik parametre setine göre, her bir aralıktaki en olası subjektif yanıtın baskınlığı ayırt edicilik parametresi tarafından belirlenmektedir. Bu parametre arttıkça, aralıktaki en olası yanıtın olasılığı artan bir şekilde diğer yanıtlarla ilgili olasılıklara baskın gelmektedir.

Eşitlik 1 bir madde tepki modelini bir subjektif yanıt seviyesinde modellemektedir. Ancak model (GAMOM) nihai olarak aşamalı katılma ölçeği ile ilişkili olan gözlenen yanıt kategoriler açısından tanımlanmaktadır. Bir gözlenen yanıt kategorisine karşılık gelen iki subjektif yanıt kategori karşılıklı olarak birbirini dışarda bırakmaktadır. Bir bireyin belli bir

gözlenen yanıt kategorisine yanıt verme olasılığı basit bir şekilde ilgili gözlenen yanıt kategorisine karşılık gelen iki subjektif yanıt kategorisi yanıt olasılığının toplamı ile ifade edilmektedir:

$$P(Z_i = z|\theta_j) = P(Y_i = z|\theta_j) + P[Y_i = (M - z) |\theta_j] \quad (3)$$

Eşitlik 3'te yer alan semboller aşağıdaki anlamlara gelmektedir:

$Z_i - i$ ifadesine verilen bir gözlenen yanıt,

$z = 0$ ($z = 0, 1, 2, \dots, C$) - en güçlü katılmama (disagreement) seviyesi,

$z = C$ - en güçlü katılma (agreement) seviyesi,

C - gözlenen yanıt kategorisinin bir eksiği ($M = 2C + 1$).

Son olarak dördüncü öncül ile τ_{iks} parametrelerinin $(\theta_j - \delta_i) = 0$, noktasına göre simetrik olduğu ifade edilmektedir:

$$\tau_{i(C+1)} = 0, \quad (4)$$

$$\tau_{i(C+1)} = -\tau_{i(M-z+1)}, z \neq 0, \quad (5)$$

Bireyler yetenek ölçeği üzerindeki lokasyonlarının $-h$ ya da $+h$ birim uzağında yerleştirilmiş maddelere benzer şekilde katılma olasılığı sergilemektedir ve durum takip eden aşağıdaki özdeşliğe yol açmaktadır:

$$\sum_{k=0}^z \tau_{ik} = \sum_{k=0}^{M-z} \tau_{ik}, \quad (6)$$

Yukarıdaki Eşitlik 3, Eşitlik 6'ya yerleştirilince GAMOM'un formal tanımı ve formülü ortaya çıkmaktadır:

$$P(Z_i = z|\theta_j) = \frac{\exp\{\alpha_i[z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}]\} + \exp\{\alpha_i[(M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}]\}}{\sum_{w=0}^C \{\exp\{\alpha_i[w(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}]\} + \exp\{\alpha_i[(M-w)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}]\}\}}, \quad (7)$$

ATM ve GAMOM'a dayalı Karma Model Tanımlaması

MTK altında maddeler birtakım özelliklerine göre (kategori sayısı, madde formatı vs.) ele alınarak çeşitli şekillerde karma model tanımlamaları yapılabilmektedir. GAMOM altında elde edilen delta parametrelerine dayalı olarak maddelerin monotonik olma-monotonik olmama dereceleri ele alınarak maddelere ilişkin kestirilen delta parametrelerinin ortalamaya olan mesafesi değerlendirilerek ilgili özellik birtakım aralıklarda kategorize edilebilmektedir (Roberts, Donoghue & Laughlin, 2002; Stark vd., 2006; Ling vd, 2016). Maddelerin altındaki yanıt süreçlerinin modellenerek maddenin baskınlık sürecine göre veya ideal nokta yanıt sürecine göre mi ele alınması gerektiği önemli hususlardan biridir. Bu çalışmanın odak noktalarından biri madde yanıt süreci olduğu için maddeler monotonik olma-monotonik olmama durumlarına göre ele alınarak karma model tanımlaması yapılmıştır. Karma model oluşturulmasının amacı madde karakteristik eğrilerine ilişkin fonksiyon davranışlarının etkisini gözlemleyerek madde spesifikasyonu yapmak ve bu doğrultuda karma modeli tanımlayarak söz konusu monotonik olma ve monotonik olmama derecelerinin etkilerini kümülatif ve ideal nokta modelleri ile karşılaştırmalı olarak incelemektir. Monotonik maddeler kümülatif modellerden ATM altında kestirilirken, monotonik olmayan maddeler ideal nokta modellerinden GAMOM altında kestirilerek karma model kestirimi gerçekleştirilmiştir.

Madde Tepki Kuramına Dayalı Sınıflama Yaklaşımları

Karar verme çalışmalarının önemli aşamalarından biri de bireyleri belli gruplarda sınıflamak için verilen kararların geçerlik ve güvenirliklerinin incelenmesidir. Bu doğrultuda başvurulacak yollardan biri sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin incelenmesidir. Sınıflama doğruluğu tek bir replikasyondan elde edilen sonuçlara dayalı olarak bireyin gözlenen sınıflaması ile gerçek sınıflamasının örtüşme derecesidir. Sınıflama tutarlılığı ise bireylerin gözlenen sınıflamalarının test etme prosedürlerinin çapraz replikasyonlarında benzer olma derecesi olarak ele alınmaktadır (AERA, APA & NCME, 2014, s. 40). Farklı bir ifadeyle, sınıflama tutarlılığı aynı uygulamanın paralel tekrarları üzerinde bireyin aynı kategoride sınıflandırılma derecesini ifade etmektedir (Lee, 2010).

Sınıflama doğruluğuna ilişkin çalışmalar ilk olarak klasik test kuramı (Swaminathan, Hambleton & Algina, 1974) ile başlamış ve 40 yılı aşkın bir süre bu çalışmalar devam etmiştir (Lee, Brennan & Wan, 2009; Lee, Hanson & Brennan, 2002; Livingston & Lewis, 1995; Subkoviak, 1976).

Hambleton ve Novick (1973) ve Swaminathan, Hambleton ve Algina (1974) karar doğruluğu ve karar tutarlılığı ifadelerini kavramsallaştırarak bu indeksleri kestirmeye yönelik prosedürler sağlamışlardır. Bu doğrultuda ilk olarak P ve Kappa uyuma indeksleri sağlanmıştır. Ancak bu indeksler pratikte her zaman mümkün olmayan, iki uygulama verisini gerektirmektedir. Dolayısıyla tek uygulamaya dayalı sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri geliştirilmiştir. Bu doğrultuda Subkoviak (1976) ve Livingston ve Lewis (1995) ve diğer araştırmacılar klasik test kuramı altında tek uygulamaya dayalı indeksler geliştirmişlerdir (Deng, 2011).

Klasik test kuramına dayalı geliştirilen indeksleri MTK'ya dayalı indeksler takip etmektedir. MTK'nın günümüzde yoğun ilgi görmesi ve birçok alanda sıklıkla kullanılması, MTK'ya dayalı sınıflama indekslerinin geliştirilmesinin önünü açmıştır. Bu doğrultuda Hambleton ve Han (2005)'in yaklaşımı, Guo (2006)'nın yaklaşımı, Rudner'in (2001, 2005) yaklaşımı ve Lee'nin (2010) yaklaşımı olmak üzere dört farklı yaklaşım altında sınıflama doğruluğu ve tutarlılığı indeksleri ele alınabilmektedir. Ancak MTK sınıflama yaklaşımlarından olan Rudner'in yaklaşımı ve Lee'nin yaklaşımına yönelik yazılım ve programlara ulaşılabilirliğin daha kolay olduğu ve diğer yaklaşımlara göre daha fazla sıklıkta kullanıldığı görülmektedir.

Rudner'in Yaklaşımı

Rudner'in (2001, 2005) beklenen sınıflama doğruluğu yanlış sınıflamayı, normal dağılan ölçme hatasına dayalı olarak nicelleştirmektedir. Rudner'in sınıflama yaklaşımı altında sınıflama doğruluğunu hesaplamak için üç tane veri vektörü kullanılmaktadır. Birinci vektör C+1 tane kesme puanının olduğu bir vektördür (C tane sınıfın söz konusu olduğu durumda):

$$K = [K_1, K_2, \dots, K_{C+1}], \quad (8)$$

$$K_1 < K_2 < \dots < K_{C+1} \text{ ve } K_1 = -\infty, K_{C+1} = \infty$$

Bu kesme puanı vektörü ölçme üzerindeki (on the assessment) operasyonel kesme puanlarını ve bütün düzeyler için en düşük ve en yüksek sınırları içermektedir. Örneğin eğer üç tane operasyonel kesme puanı varsa Eşitlik 8'deki vektör üç tane kesme puanıyla birlikte artı sonsuz ve eksi sonsuzu da kapsayacak şekilde oluşturulur. Rudner'in yaklaşımı altında indeksin hesaplanması için ele alınan ikinci vektör ise birey yetenek kestirimlerinin vektörüdür:

$$\widehat{\theta} = [\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_{Ne}]^t \quad (9)$$

Buradaki Ne kişi sayısını; $\widehat{\theta}_i$ ise i kişisine ilişkin MTK altında elde edilen yetenek kestirimini göstermektedir.

Sınıflama indeksleri dağılımsal yaklaşımlar ve bireysel yaklaşımlar olarak iki farklı şekilde ele alınmaktadır (Brennan & Wan, 2004; Lee, 2005). Dağılımsal yaklaşım gerçek yetenekler için dağılımsal bir varsayımı dikkate alırken, bireysel yaklaşım tek bir defada her bir birey için karar tutarlılık indeksini hesaplar ve gerçek yetenek hakkında herhangi bir varsayım yapmaksızın bütün bireylerin ortalamasını alır. Eşitlik 9'daki vektörde her bir bireye ilişkin yetenek kestirimi içerildiği için buradaki sınıflama yaklaşımının bir "bireysel yaklaşım" olduğu ifade edilmektedir (Lathrop & Cheng, 2013; Lee, 2010). Rudner'in yaklaşımında ele alınan üçüncü vektör ise standart hata kestirimlerinin bir vektörüdür:

$$\widehat{\sigma}_{\theta} = [\widehat{\sigma}_{\theta_1}, \widehat{\sigma}_{\theta_2}, \dots, \widehat{\sigma}_{\theta_{Ne}}], \quad (10)$$

Yine eşitlikte yer alan Ne kişi sayısı, $\widehat{\sigma}_{\theta_i}$ ise i bireyinin MTK altında elde edilen yetenek kestirimine ilişkin kestirilen standart hata değeridir. Eşitlik 10'da yer alan standart hata değerleri her bir bireyin MTK test bilgi fonksiyonuna dayalı olarak hesaplanabilmektedir. Bu durumda her bir birey için standart hata kestirimi aşağıdaki gibi hesaplanmaktadır:

$$\widehat{\sigma}_{\theta_i} = \frac{1}{\sqrt{I(\widehat{\theta}_i)}} \quad (11)$$

$I(\widehat{\theta}_i)$, i kişisi için test bilgi fonksiyonunu göstermektedir. Daha sonra her bir bireyin yetenek kestiriminin etrafında standart hata kestiriminin koşullu normalliği varsayılarak her bir ardışık kesme puan çifti arasındaki alan bulunur. Normallik varsayımı asimptotik teoriden ve maksimum olabilirlik kestirimi kullanıldığında MTK varsayımlarından gelmektedir. Buna göre birey ve madde sayısı arttıkça bir bireyin maksimum olabilirlik kestiriminin bireyin test bilgi fonksiyonunun karekökünün tersi üzerinde θ 'nın ortalaması ile standart sapması 1 olan normal dağılıma yakınsaması beklenir. Her bir C performans kategorisindeki puanlamanın beklenen olasılığı Eşitlik 12'deki gibi ifade edilmektedir:

$$\hat{P}_{ic} = \phi (K_{ci}, K_{ci+1}, \hat{\theta}_i, \hat{\sigma}_{\theta i}) \quad (12)$$

$\Phi (a, b, \mu, \sigma)$, σ standart sapma μ ortalamaya sahip olan eğri altındaki alanı ifade etmektedir ve Eşitlik 12'deki terimler de bu şekilde anlamlandırılmaktadır. Eşitlik 12'nin altında yatan varsayımlar asimptotik teori ve maksimum olabilirlik kestiriminden gelse de, herhangi bir yetenek kestiricisi ile de kullanılabilir.

Daha sonra her bir C performans kategorisinde yer alan her bir bireyin beklenen olasılıklarının olduğu bir beklenen olasılıklar matrisi, $\hat{P}$, $N_e \times C$ boyutlarında tanımlanır.

Bireyin sınıflandırıldığı performans düzey kategorisine karşılık gelen beklenen olasılığın, doğru sınıflamanın beklenen olasılığı olduğu kabul edilir ve diğer olasılıkların ise beklenen yanlış sınıflama olasılıkları olduğu kabul edilir. Ağırlıklandırma matrisi (W) olarak ele alınan, w ağırlıklarından oluşan bir $N_e \times C$ matrisi tanımlanacak olursa; buradaki w 'ler bireylerin elde ettikleri performans düzey kategorilerine işaret etmek amacıyla kullanılan bir matris olarak ifade edilebilir:

$$W = \begin{bmatrix} w_{11} & w_{12} \dots & w_{1C} \\ w_{Ne1} & w_{Ne2} \dots & w_{NeC} \end{bmatrix} \quad (13)$$

Eğer bireyin puanı C performans düzeyinde sınıflandırılırsa, W_{ci} ağırlığı bire, aksi takdirde sıfıra eşittir.

Rudner'in sınıflama doğruluğu indeksinin hesaplanabilmesi için W ve $\hat{P}$ matrisi çarpılır; elde edilen nihai matristeki bütün elemanlar toplanır ve N_e birey sayısına bölünür:

$$\hat{\tau} = \frac{\sum(\hat{P} * W)}{N_e} \quad (14)$$

Bu yaklaşım altında sınıflama doğruluğu indeksi tek bir uygulamaya dayalı olarak elde edildiği için Eşitlik 14 sadece $\hat{P}$ ve W matrislerini içermektedir, $\hat{P}$ 'nin kendisiyle çarpımını içermemektedir.

Sınıflama tutarlılığı, aynı uygulamanın paralel replikasyonları sonucunda aynı kategoride yer alma oranına ilişkin bir ölçüm sağlamaktadır. Bu yüzden bu indeks $\hat{P}$ 'nin kendisiyle çarpımını içerirken bireylerin gözlenen performans düzeyine işaret eden bir matris içermez. Bu doğrultuda sınıflama tutarlılık indeksi matematiksel olarak Eşitlik 15'deki gibi ifade edilmektedir:

$$\hat{Y} = \frac{\sum(\hat{P}*\hat{P})}{Ne} \quad (15)$$

Eşitlik 15’de yer alan $\hat{P}$ matrisideki elemanların karesinin alınması gerektiğinden sınıflama tutarlılığı her zaman sınıflama doğruluğuna ya eşit ya da sınıflama doğruluğundan daha düşük değerler alır (Wyse & Hao, 2012).

Lee’nin Yaklaşımı

Lee’nin (2010) yaklaşımını Rudner’in yaklaşımından ayıran temel farklılık Lee’nin yaklaşımında sınıflamaların toplam puan ölçeğinde meydana gelmesidir. Bireyin toplam puanı bireyin yetenek kestirimi olarak ele alınır ve kesme puanı da toplam puan ölçeği üzerinde verilir. Ama toplam puan ölçeği üzerinde vurgu olmasına rağmen Lee’nin yaklaşımı, bireyin toplam puanındaki belirsizliği yansıtan dağılımları oluşturmak için MTK modellerini kullanır. Bu süreç iyi bilinen bir tekrarlamalı (recursive) algoritma ile gerçekleştirilir (Lord & Wingersky, 1984). Elde edilen dağılım bireylerden gelen her bir toplam puana ilişkin olasılıkları verir. Bu doğrultuda sınıflama doğruluğu ve tutarlılığı indeksleri hesaplanır (Wyse & Hao, 2012).

Rudner ve Lee’nin Yaklaşımlarının Karşılaştırılması

Hem Rudner hem de Lee’nin yaklaşımı aynı MTK modellerini ve madde parametrelerini kullanmaktadır. Her iki yaklaşım da bütün bireyler için tek tek olasılık dağılımları oluşturmaktadır ve her ikisi de bu dağılımları aynı yolla manipüle etmekte ve toplamaktadır. Bu yaklaşımlar arasındaki temel farklılık sınıflamanın meydana geldiği ölçek (Rudner theta ölçeği; Lee toplam puan ölçeği) ve bireylere ilişkin dağılımın (belirsizlik dağılımı) oluşturulma şeklidir. Karmaşık süreçler gerektiren modellerin olduğu durumlarda ya da farklı kestirim yöntemlerinin söz konusu olduğu durumlarda Rudner’in yaklaşımının kullanımının Lee’nin yaklaşımının kullanımından daha uygun olacağı belirtilmektedir (W. C. Lee, kişisel iletişim, Kasım 7, 2017). Çünkü sınıflama kararlarını gerçekleştirmek için Lee toplam puan ölçeğini kullanırken, Rudner yaklaşımında doğrudan $\hat{\theta}$ değerleri kullanılmaktadır. Rudner’in yaklaşımı doğrudan $\hat{\theta}$ değerlerini kullandığı için her türlü veri yapısında (iki kategorili-çok kategorili, çok boyutlu, karma test vs.) rahatlıkla kullanılabilir (Q. Lathrop, kişisel iletişim, Kasım 25, 2017). Dolayısıyla karmaşık matematiksel süreçler gerektiren veri yapılarında ya da modellerde Rudner’in sınıflama

yaklaşımına göre incelemeler yapmak daha uygun olacaktır. Bu çalışmada bayesian ve nonbayesian yetenek kestirim yöntemleri ve aynı zamanda karmaşık kestirim süreçleri gerektiren ideal nokta model kullanımları bir arada ele alınacağı için Rudner'in yaklaşımına dayalı olarak sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin hesaplanması amaçlanmıştır. Çünkü MTK'ya dayalı bu sınıflama yaklaşımları farklı teknik niteliklere sahip olduğu için seçilecek yaklaşım sınıflama doğruluğu ve tutarlılığını etkilemektedir.

Ölçme sonuçlarının doğruluğunu etkileyen faktörler de sınıflama doğruluğunu ve tutarlılığını önemli ölçüde etkilemektedir (Yang, Poggio & Glasnapp, 2006). Cheng, Liu ve Behrens (2015) çalışmalarında ikili kararın (master/nonmaster) söz konusu olduğu durum için, yetenek kestirimlerine ilişkin standart hatanın beklenen sınıflama doğruluğuna ve tutarlılığına etkisini analitik olarak teoremlerle göstermişlerdir. Analitik bulgulara göre standart hata azaldıkça koşullu sınıflama doğruluğu ve tutarlılığı artmaktadır. Dolayısıyla ele alınacak yetenek kestirim yöntemi de sınıflama doğruluğu ve tutarlılığını etkilemektedir.

Yaygın bir şekilde kullanılan yetenek kestirim yöntemleri ele alındığında MTK altında maksimum olabilirlik kestirim (Maximum Likelihood Estimation-MLE), ağırlıklandırılmış olabilirlik kestirim yöntemi (Weighted Likelihood Estimation-WLE), Maximum a Posteriori (MAP) ve Expected a Posteriori (EAP) kestirim yöntemleri öne çıkmaktadır.

Bayesian ve Nonbayesian Yetenek Kestirim Yöntemleri

MTK'da farklı modellerin kullanımının yanı sıra farklı yetenek kestirim yöntemlerinin kullanımı da sınıflama doğruluğunu etkileyen önemli bir faktördür (Borgatto, Azevedo, Pinherio & Andrade, 2015; Ching-Fung, 2002; Rose, 2010; Wang & Vispoel, 1998; Wainer & Thissen, 1987; Wang & Wang 2001). MTK'da yaygın olarak olabilirlik temelli yaklaşım olan MLE yöntemi kullanılmakla beraber, WLE yöntemi de bu doğrultuda kullanılan nonbayesian yetenek kestirim yöntemleri arasında yer almaktadır. Ancak gelişen dünyayla birlikte veri yapıları her geçen gün daha da karmaşılaşmakta ve bunların üstesinden gelmek için yeni yaklaşımların önünü açmaktadır. MTK da bayesian çerçevesinde ele alınmakta olup bu doğrultuda bayesian temelli yaklaşımlardan MAP ve EAP kestirim yöntemlerine dayalı yetenek kestirimleri elde edilmektedir.

MLE yöntemi MTK'ya dayalı yetenek kestirimlerinde sıklıkla kullanılan kestirim yöntemidir. Bu kestirim yöntemi olabilirlik fonksiyonuna dayanan iteratif süreçlerin kullanıldığı bir kestirim yöntemidir. Olabilirlik fonksiyonunun maksimum olduğu nokta ilgili yanıt örüntüsü için kestirilen maksimum olabilirlik yetenek kestirimini, yani MLE'yi (θ) vermektedir. MLE'ye ilişkin olabilirlik fonksiyonu Eşitlik 16'da yer almaktadır:

$$L(u|\theta) = \prod_{i=1}^n P_i(\theta)^{u_i} \cdot (1 - P_i(\theta))^{1-u_i} \quad (16)$$

$L(u|\theta)$ olabilirlik fonksiyonu, u yanıt örüntüsü için i maddesine doğru yanıt verme (P_i) ve doğru yanıt vermeme (Q_i) durumlarına göre modellenmektedir. Olabilirlik fonksiyonu olasılıkların çarpımına dayalı olarak elde edildiği için küçük bir değer ile sonuçlanır. Pratik kullanım için daha büyük bir değer elde etmek amacıyla bu fonksiyonun doğal logaritması alınır. Çünkü bu fonksiyonun doğal logaritması alındığında ilgili fonksiyon, olasılıkların çarpımı yerine olasılıkların logaritmasının toplamı olarak daha büyük bir değer ile sonuçlanır. Olabilirlik fonksiyonu ile bu fonksiyonun doğal logaritması monotonik olarak ilişkili olduğundan her iki fonksiyonu maksimize eden değer denktir. Doğal logaritması alınmış olabilirlik fonksiyonuna ilişkin eşitlik aşağıda yer almaktadır:

$$\ln L(u|\theta) = \ln P_1 + \ln P_2 + \dots + \ln P_n \quad (17)$$

Eşitlik 17'de yer alan u ($u_1, u_2, \dots, u_n$) yanıt vektörünü; P ise ilgili maddeye ilişkin yanıt olasılığını göstermektedir. Bu eşitliği maksimum yapan yetenek değerini bulmak için eşitliğin birinci dereceden türevinin çözümü yapılır. Çözümü yapılacak eşitlik aşağıdaki gibidir:

$$l_1 = \frac{\partial \ln L(u|\theta)}{\partial \theta} = \frac{\sum_{i=1}^n (u_i - P_i) P_i'}{P_i Q_i} = 0, \quad (18)$$

Eşitlik 18'i sağlayan değer, yeteneğin maksimum olabilirlik kestirimi olarak belirlenmektedir. Eşitlik 18 doğrudan çözülemeyeceğinden bir takım numerik prosedürler kullanılmaktadır. Newton-raphson bu çözüm için kullanılan iterasyona dayalı olan yöntemlerden biridir (Hambleton & Swaminathan, 1985).

Çalışmada ele alınacak bir diğer kestirim yöntemi ise bayesian temelli olan MAP kestirim yöntemidir. Bu kestirim yönteminin matematiksel alt yapısı MLE yöntemine

benzemektedir. Bu yöntemde MLE yönteminden farklılaşan nokta maksimum olabilirlik yetenek değerinin önsel bir dağılım fonksiyonuna dayalı olarak elde edilmesidir. MAP kestirim yöntemine ilişkin eşitlik aşağıda yer almaktadır:

$$f(\theta) \cdot L(u|\theta) = f(\theta) \prod_{i=1}^n P(\theta)^{u_i} \cdot Q(\theta)^{1-u_i} \quad (19)$$

Eşitlik 19’da yer alan $f(x)$ fonksiyonu önsel dağılımı gösteren bir yoğunluk fonksiyonudur. Bu eşitlikte de bu fonksiyonun birinci türevinin çözümüne dayalı olarak maksimum olabilirlik noktası belirlenmektedir:

$$\frac{\sum_{i=1}^n (u-P)P'}{PQ} + \frac{\partial \ln f(\theta)}{\partial \theta} = 0 \quad (20)$$

Eşitlik 20’ yi sıfır yapan yetenek değeri MAP kestirim yöntemi için maksimum olabilirlik değeri olarak ele alınmaktadır. Eşitlikten de görüldüğü gibi önsel dağılım fonksiyonunun sabit bir değer olarak ele alınması durumunda Eşitlik 20, MLE yönteminde kullanılan eşitliğe indirgenmiş olacaktır (Rose, 2010).

Ağırlıklandırılmış olabilirlik kestirim yöntemine ilişkin eşitliği elde etmek amacıyla, Lord (1983), $MLE(\theta)$ ’nin yanlılığına ilişkin asimptotik ifadeyi Eşitlik 21’de vermiştir:

$$BIAS(MLE(\theta)) = \frac{-J}{2I^2} \quad (21)$$

Eşitlik (21)’de yer alan I test bilgi fonksiyonu olup eşitlikte yer alan I , J ve P terimlerine ilişkin eşitlikler $I = \sum P'^2 / PQ$, $J = \sum P'P'' / PQ$, ve $P'' = \partial^2 P / \partial \theta^2$ şeklindedir. Eşitlik 21 çok terimli (multinomially) dağılan bir değişken için $BIAS (MLE(\theta))$ ’ nin genel ifadesine eşittir. Standart normal önsel ile $MAP(\theta)$ ’ nin yanlılık kestirimi için Lord (1984) Eşitlik 22’ yi vermiştir:

$$BIAS(BME(\theta)) = BIAS(MLE(\theta)) - \frac{\theta}{I} \quad (22)$$

Eşitliğin sağında sonda yer alan terim (θ/I) , standart normal yoğunluğun logunun θ ’ya göre türevinin test bilgi fonksiyonuna bölümüdür. Buradan yola çıkarak, Eşitlik 20’deki kestiricinin (estimator) yanlılığı Eşitlik 23’deki gibi ele alınmaktadır:

$$BIAS(\theta^*) = BIAS(MLE(\theta)) + \frac{\frac{\partial \ln f(\theta)}{\partial \theta}}{I} \quad (23)$$

Dolayısıyla, yansız kestiriciyi bulmak için Eşitlik 23’ün sıfıra eşitlenmesi ve eşitliğin $f(\theta)$ için çözülmesine ihtiyaç vardır. Eşitlik 23’ün sağ tarafı sıfıra eşitlenir ve buraya Eşitlik 21 yerleştirilirse aşağıdaki yeni eşitlik elde edilir:

$$\frac{\partial \ln f(\theta)}{\partial \theta} = \frac{J}{2I} \quad (24)$$

Eşitlik 20'deki $f(\theta)$ yerine $w(\theta)$ yerleştirilerek Eşitlik 25 elde edilir:

$$\frac{\sum_{i=1}^n (u-P)P'}{PQ} + \frac{\partial \ln w(\theta)}{\partial \theta} = 0 \quad (25)$$

Eşitlik 24, Eşitlik 20' ye yerleştirilirse Eşitlik 26 elde edilir:

$$\frac{\sum_{i=1}^n (u-P)P'}{PQ} + \frac{J}{2I} = 0 \quad (26)$$

Eşitlik (26)' da yer alan I ve J terimleri Eşitlik 21'de tanımlandığı gibi ele alınır. Eşitlik 26'nın bir kestirimi ağırlıklandırılmış olabilirlik kestirimi (WLE) olarak adlandırılır. WLE bayesian değildir. Çünkü θ 'nın dağılımına ilişkin herhangi bir varsayım söz konusu değildir; $w(\theta)$ testteki madde parametrelerinin bir fonksiyonudur, bayesian kestirimlerde yer alan $f(\theta)$ önsel dağılım fonksiyonu ile karıştırılmamalıdır (Warm, 1989).

Çalışmada ele alınan bir diğer bayesian temelli yaklaşım ise EAP kestirim yöntemidir. Bu kestirim yönteminde de önsel bir dağılım söz konusudur ve buna dayalı olarak yetenek kestirilmektedir. EAP kestirim yöntemine ilişkin eşitlik aşağıda yer almaktadır:

$$E(\theta|u) = \frac{\int_R \theta L(u|\theta).f(\theta)d\theta}{\int_R L(u|\theta).f(\theta)d\theta} \quad (27)$$

Eşitlik 27'de yer alan $f(\theta)$ önsel dağılım fonksiyonunu, L ise olabilirlik fonksiyonunu göstermektedir. Bu eşitliğin de analitik çözümünü elde etmek mümkün değildir. Dolayısıyla burada da numerik yöntemlerden Gaus'un Quadratt İntegrasyon (Gaussian Quadrature Integration) yöntemi kullanılarak EAP kestirimleri elde edilmelidir. Gaus'un Quadratt İntegrasyon yöntemine göre eşitlik aşağıdaki gibi dönüştürülerek yetenek kestirimi yapılmaktadır:

$$\hat{\theta} = \frac{\sum_{k=1}^q X_k L(X_k)F(X_k)}{\sum_{k=1}^q L(X_k)F(X_k)} \quad (28)$$

Eşitlik 28'de yer alan X_k , q quadratt noktasından birini; $f(X_k)$ önsel dağılım fonksiyonundaki her bir quadratt (X_k)' a karşılık gelen değeri; $L(X_k)$ ise her bir quadratt noktasındaki olabilirlik fonksiyonunu göstermektedir (Borgatto, Azevedo, Pinherio & Andrade, 2015).

Bu alıřmada nonbayesian yetenek kestirim yntemlerinden WLE; bayesian yetenek kestirim yntemlerinden ise MAP ve EAP kestirim yntemlerinin kullanılması amalanmaktadır.



İLGİLİ ARAŞTIRMALAR

Bu bölümde ideal nokta modellerinin kümülatif modellerle karşılaştırmalarına, ideal nokta modellerinin kestirim yöntemleri ile diğer alanlardaki uygulamalarına ve sınıflama indekslerine yönelik çalışmalara yer verilmiştir.

İdeal Nokta Modellerinin Kümülatif Modellerle Karşılaştırmalarına Yönelik Araştırmalar

Stark vd. (2006) çalışmalarında ideal nokta yanıt sürecine ilişkin yöntemlerin ölçek geliştirme ve puanlamada geleneksel olarak kullanılan baskınlık yanıt sürecine karşılık geçerli bir alternatif olup olmayacağı hususunu ele almışlardır. Çalışmada MTK yöntemleri altında parametrik ve parametrik olmayan ideal nokta modelleri ile kümülatif modeller kullanılmıştır. Parametrik ideal nokta modellerinden GAMOM ile parametrik olmayan modellerden ideal nokta sınırlandırmaları ile Levine'in maksimum olabilirlik puanlama modeli (Levine, 1984) kullanılmıştır. Kümülatif modellerden ise 2PLM ve kümülatif sınırlandırmaları ile Levine'in maksimum olabilirlik puanlama modeli nonparametrik model olarak kullanılmıştır. Her bir modelin uygunluğunu karşılaştırmalı ele almak için çapraz-geçerleme örneklemesinde hem grafiksel hem de istatistiksel model uyumu incelemeleri yapılmıştır. Araştırmadaki incelemeler Kişilik Faktör Ölçekleri – 16 ölçek (Conn & Rieke, 1994) ile elde edilmiş 13059 yanıtlayıcı grubundan elde edilen örneklem üzerinde gerçekleştirilmiştir. Ölçekler sırasıyla İçtenlik (11 madde), Duygusal kararlılık (10 madde), Hâkimiyet (10 madde), Canlılık (10 madde), Kural-Bilinci (11 madde), Sosyal-Cesaret (10 madde), Duyarlılık (11 madde), İhtiyat (10 madde), Soyutluk (11 madde), Özellik (10 madde), Endişe (10 madde) Değişime açıklık (14 madde), Özgüven (10 madde), Mükemmelliyetçilik (10 madde), Gerginlik (10 madde), ve İzlenim Yönetimi'dir (12 madde). Araştırmadan elde edilen bulgular monotonik olmayan maddenin bulunmadığı Duygusal Kararlılık ve Değişime Açıklık ölçeklerinde ideal nokta ve kümülatif modellere ilişkin uyumun neredeyse benzer olduğunu ortaya çıkarmıştır.

Monotonik olmayan maddelerin gözleendiği endişe ve Mükemmelliyeçilik alt boyutunda ise ideal nokta modellerinin daha iyi uyum sergilediği gözlenmiştir. Model uyumunun yanı sıra bütün maddelerin monotonik olduğu Duygusal Kararlılık ölçeği ile iki tane monotonik olmayan maddenin mevcut olduğu endişe ölçekleri üzerinde 2PLM ve GAMOM ile elde edilen yetenek kestirimleri arasındaki ilişki incelenmiştir. İncelemeler hem 2000 kişilik örneklem üzerinde hem de sırasıyla üst uçtaki 100, 200, 300 ve 500 kişiler üzerinde yapılmıştır. Duygusal kararlılık ölçeğinde bütün örneklemelerde ilişkinin 0,99 olduğu, ideal nokta ve dominant modellerden elde edilen yetenek kestirimlerinin neredeyse mükemmel eşleşmesinin söz konusu olduğu gözlenmiştir. Ancak kaygı ölçeğinde 100 kişilik üst uça ilişkinin negatif (-0,46) olduğu ancak üst uçtaki bireylerin daha genişletilerek oluşturulan 200, 300, 500 kişilik örneklem artışına bağlı olarak ilişkinin de arttığı gözlenmiştir. 2000 kişilik örneklem üzerinde ise ilişkinin 0,99 olduğu gözlenmiştir.

Chernyshenko vd. (2007) kişilik ölçümleri alanını geliştirmek için neden ideal nokta yöntemlerine ilişkin ölçek geliştirmeye geçiş yapılması hususunu açıklamışlardır. Bu çalışmada ideal nokta metodolojisinin asli (substantive) faydaları ampirik olarak KTK ve kümülatif MTK yöntemleri ile karşılaştırılarak gösterilmiştir. Bu doğrultuda 50 maddelik büyük ve heterojen bir madde havuzuna dayalı olarak geleneksel KTK, kümülatif MTK ve ideal nokta MTK yöntemleri ile 20'şer maddelik üç ayrı ölçek geliştirilmiştir. KTK yöntemi ile ölçek geliştirilirken temelde madde istatistikleri; kümülatif MTK yöntemi ile yerel bağımsızlık, maddelere ilişkin uyum istatistikleri ve madde ayırt edicilik değerleri; ideal nokta MTK yöntemi ile ise maddelere ilişkin uyum istatistikleri, madde ayırt edicilik değerleri ve delta parametrelerine dayalı değerlendirmeler ile ölçek geliştirme çalışmaları yürütülmüştür. Madde havuzunda yer alan maddelerin her bir yaklaşım altında kullanılabilirliği ele alındığında ideal nokta MTK altında 48 maddenin de kullanımının uygun olduğu gözlenirken; kümülatif MTK altında daha fazla uyumsuz ve yerel bağımsızlığı ihlal eden maddelerin mevcut olduğu gözlenmiştir. Her bir yaklaşım altında geliştirilen ölçeklere dayalı yetenek kestirimleri arasında hesaplanan korelasyonların yüksek olduğu gözlenmiştir. Ayrıca ideal nokta sürecine dayalı geliştirilen ölçeğin kümülatif model ile kestirildiği ve baskınlık sürecine dayalı geliştirilen ölçekten elde edilen verinin ideal nokta modeli ile kestirildiğinde elde edilen sonuçlar incelenmiştir. Bulgular baskınlık sürecine dayalı geliştirilen ölçek verisi için 2PLM yerine GAMOM kullanıldığında da sonuçların aynı kaldığını ancak ideal nokta sürecine dayalı geliştirilen ölçek verisine dayalı kestirimlerde GAMOM yerine 2PLM'nin kullanılmasının uygun

olmayacağını göstermiştir. Bu doğrultuda araştırmacılar GAMOM' un herhangi kişilik maddeleri için doğru bir şekilde puan elde edilmesinde yeterince esnek olduğu sonucuna ulaşmışlardır.

Tay vd. (2009) tipik davranış ölçümlerine verilen yanıtların baskınlık sürecini takip eden maksimal performans davranışlarının aksine bir ideal nokta yanıt sürecini takip ettiğini ileri sürerek Mesleki Tercih Envanteri, İlgi Profil Oluşturucu (the Interest Profiler) ve İlgi Bulucusu (the Interest Finder) envanterlerine dayalı olarak ideal nokta ve kümülatif modelleri karşılaştırmışlardır. Envanterlere verilen yanıtlara ilişkin derecelendirmeler iki kategoriye düşürülerek (dichotomization) kestirimler GAMOM, 2PLM ve 3PLM ile yapılmıştır. GAMOM analizleri için GGUM2004 yazılımından 2PLM ve 3PLM analizleri için ise BILOG-MG programından faydalanılmıştır. Model veri uyumu için elde edilen grafikler için ise MODFIT bilgisayar yazılımı kullanılmıştır. Uyum iyiliğine ilişkin istatistiksel testlerden ordinary χ^2 istatistiğinin hesaplanması için FORSCORE bilgisayar yazılımından faydalanılmıştır. Düzeltilmiş χ^2/sd değerleri incelendiğinde her bir envanter için elde edilen sonuçlara göre bütün modeller altındaki uyumun birbirine yakın olduğu ancak GAMOM altındaki model-veri uyumunun 2PLM ve 3PLM altındaki uyumdan nispeten daha iyi olduğu gözlenmiştir. Yine birinci envanter için maddelerin nonomonotiklik durumu grafiksel incelemelerle yapıldığında çok sayıda monotonik maddelerin olduğu belirtilmiştir. İkinci envanter için uyum incelendiğinde GAMOM ile 2PLM ve 3PLM arasındaki farklılığın yüksek olduğu; GAMOM altındaki χ^2/sd değerlerinin daha düşük olduğu gözlenmiştir. Grafiksel incelemeler yapıldığında ikinci envanterde yer alan maddelerden biri hariç hemen hepsinin -3 ve 3 aralığında yayılan monotonik olmayan maddeler olduğu ele alınmıştır. Benzer sonuçların üçüncü envanter için de elde edildiği gözlenmiştir. GAMOM ile 2PLM arasındaki ilişki saçılma grafikleri ile incelenerek ideal nokta ve kümülatif modellerin puanlamaya olan etkisi ele alınmıştır. Çok sayıda monotonik maddenin yer aldığı birinci envanter için grafik incelendiğinde GAMOM-2PLM arasında yüksek ve monotonik bir ilişki olduğu sadece yetenek ölçeğinin uç kısımlarında (yaklaşık yeteneğin 2 civarından itibaren) hafif saçılmaların meydana geldiği gözlenmiştir. Ancak monotonik olmayan madde yoğunluğunun yüksek olduğu ikinci ve üçüncü envanterlerde yaklaşık yeteneğin sıfır olduğu noktaya kadar GAMOM-2PLM arasındaki pozitif yöndeki ilişkinin devam ettiği ancak hafif saçılmaların olduğu görülmektedir. Ancak yeteneğin yaklaşık sıfır noktası

itibariyle yetenek kestirimleri arasında keskin bir farklılığın meydana çıktığı ters yönde bir ilişkinin olduğu gözlenmiştir.

Cao vd. (2014) çalışmalarında monotonik olmayan maddelerin psikometrik özelliklerini ele alarak bu maddelerin ideal nokta modelleri ile kalibre edilmesinin kişilik ölçeklerindeki ölçme doğruluğuna olan etkisini açıklamayı amaçlamışlardır. Bu amaç doğrultusunda iki farklı örneklem üzerinde iki ayrı çalışma yürütülmüştür. İlk olarak monotonik olmayan maddelerin kalibrasyonu yapılarak bu maddeler KTK ve MTK ile değerlendirilmiştir. Çalışmaya 20 madde dâhil edilmiş olup bu maddelerin bir kısmını Kapsamlı (Comprehensive) Kişilik Ölçeği'nden (Wang, 2013) uyarlanmış maddeler ve bir kısmını ise araştırmacı tarafından geliştirilmiş monotonik olmayan maddeler oluşturmaktadır. Nihai olarak 22 yüzeyden oluşan Comprehensive Kişilik ölçeğinin üç yüzeyi ele alınmıştır. Her bir yüzeye dayalı ölçek için toplamda 20 madde mevcut olup maddelerin 4 tanesi yetenek ölçeğinin üst ucunda; 4 tanesi yetenek ölçeğinin alt ucunda ve geriye kalan 12 madde ise yetenek ölçeğinin orta kısımlarında olacak şekilde düzenleme yapılmıştır. MTK analizleri için kümülatif modellerden 2PLM; ideal nokta modellerinden ise GAMOM ele alınmış olup her bir model sırasıyla BILOG ve GAMOM2004 programları ile MML kullanılarak madde parametreleri kestirilmiştir. Model uyumları için χ^2/sd kullanılmıştır. Monotonik olmayan maddelerin performansını değerlendirmek için KTK yaklaşımından düzeltilmiş madde toplam korelasyonları; MTK yaklaşımından ise düzeltilmiş χ^2/sd kullanılmıştır. Düzeltilmiş madde toplam korelasyonlarına göre uç maddeler yüksek değerlerle sonuçlanırken monotonik olmayan maddeler düşük ve hatta negatif değerler ile sonuçlanmıştır. Model uyumları incelendiğinde ortalama düzeltilmiş χ^2/sd değerlerinin GAMOM altında bütün ölçekler için 3'ün altında olduğu ve iyi uyum sergilediği; 2PLM için ise 3'ün üzerinde yüksek χ^2/sd değerleri ile sonuçlanarak kötü uyum sergilediği sonucuna ulaşılmıştır. İkinci çalışmada ideal nokta modellerinden GAMOM'un çok kategorili versiyonu ile kümülatif modellerden ATM'ye dayalı olarak analizler yapılmıştır. Order ölçme aracı göz önünde bulundurularak ilk olarak tamamı uç maddelerden oluşan 8 maddelik form oluşturulmuştur. Daha sonra sırasıyla 2 ve 4 uç madde çıkarılarak yerine monotonik olmayan maddelerin eklendiği iki ayrı form daha oluşturulmuştur. Böylelikle %0, %25 ve %50'si monotonik olmayan maddelerden oluşan üç form üzerinde model performanslarına yönelik karşılaştırma yapılmıştır. Model uyumları incelendiğinde monotonik olmayan madde yüzdesinin %0 olduğu durumda GAMOM'un 8,18 ve ATM'nin 7,09 χ^2/sd ; %25 olduğu durumda ise farklılığın ideal nokta modelinin lehine

olacak şekilde GAMOM ve ATM için sırasıyla 6,35 ve 7,28 değerleri ile sonuçlandırıldığı gözlenmiştir. Monotonik olmayan madde yüzdesinin %50 olduğu durumda modeller arası farklılık yükselerek GAMOM ve ATM'nin sırasıyla 7,30 ve 19,65 değerleri ile sonuçlandırıldığı dolayısıyla monotonik olmayan madde yüzdesine bağlı olarak ATM'nin uyumunun kötüye gittiği gözlenmiştir. Benzer durum yakınsak ve yordayıcı geçerlik incelemelerinde de gözlenmiştir. Monotonik olmayan madde yüzdesinin %50 olduğu durumda her iki geçerlik incelemesinde ATM'nin düşüş gösterdiği ancak GAMOM' un madde yüzde değişimine rağmen kararlı olduğu gözlenmiştir. Bunun yanı sıra yetenek ölçeği boyunca sağlanan bilgi ele alındığında monotonik olmayan madde yüzdesinin %50 olduğu durumda monotonik olmayan madde yüzdeslerinin %0 ve %25 olduğu durumlara göre GAMOM altında $\theta = 1,5$ ve üzerindeki noktalarda belirgin iyileşmelerin olduğu yani üst yetenek düzeyinde yüksek ölçme doğruluğu ile sonuçlandırıldığı belirtilmiştir.

Dalal ve Carter (2014) toplama dayalı puanlama ve ideal nokta ölçeğinin bireylerin sıra düzenlemelerine, kritere dayalı geçerlik kestirimlerindeki yanlılığa ve seçim sonuçlarına olan etkisini belirlemek için bir Monte Carlo simülasyon çalışması yürütmüşlerdir. Bu doğrultuda GAMOM altında 20 tane madde için 2000 birey yanıt vektörü üretilmiştir. Delta parametreleri (-1,28; 1,28) aralığında olacak şekilde maddelerin yarısı pozitif ve diğer yarısı negatif değerlere sahip olacak şekilde delta parametreleri üretilmiştir. Monotonik olmayan madde sayısı manipüle edilerek sırasıyla 0, 2, 4, 6, 8, 10, 12 ve 14 monotonik olmayan madde olacak şekilde veri setleri oluşturulmuştur. Bunun yanı sıra her bir yanıtlayıcı için kriter değişkenindeki puanların gerçek theta değerleri ile 0,15, 0,25, 0,35 ve 0,45 düzeyinde ilişkili olduğu durumlar simüle edilmiştir. Her bir gerçek yetenek değeri ortalaması 0 standart sapması 1 olan normal dağılımdan seçkisiz olarak üretilmiştir. Madde ve birey parametreleri GGUM2004 programı ile kestirilmiştir. Madde parametreleri için marjinal maksimum olabilirlik kestirimi, birey parametreleri için ise EAP kestirimi kullanılmıştır. Kritere dayalı geçerlikteki yanlılık evren geçerliği ile örnekleme dayalı kestirilen geçerlik arasındaki fark (evren-gözlenen) alınarak hesaplanmıştır. Çalışmadan elde edilen sonuçlara göre monotonik olmayan madde sayısı %20 veya üzerinde olduğunda toplam puanın kritere dayalı geçerliği gereğinden daha düşük kestirdiği bunun yanı sıra gerçek korelasyon değeri 0,45'e yaklaştıkça yanlılığın arttığı belirtilmiştir. Son olarak birey sıralamaları ele alındığında normal dağılan büyük bir örnekleme monotonik olmayan maddelerin ihmal edilmesinin genel yansımasının çok kötü olmadığı ancak üst

kısımlara göre seçim kriterleri ele alındığında farklılığın daha gözlenebilir olduğu belirtilmiştir.

Zampetakis vd. (2015) çalışmalarında beklenen etki (anticipated affect) olarak adlandırılan bireylerin gelecekteki olaylara karşı duygusal tepkilerine ilişkin kestirimlerin yapılmasında MTK'dan faydalanmışlardır. Beklenen etki kestiriminin tipik davranış sürecine mi yoksa maksimal davranış sürecine mi uyduğunu (conform) incelemişlerdir. Araştırma büyük bir projenin parçası olan ve öğrencilerin iş kurlmalarına yönelik inceleme yapmayı amaçlayan 10 tane pozitif ve 10 tane negatif maddenin yer aldığı ölçme aracından elde edilen veri ile gerçekleştirilmiştir. Uygulama verisi 1624 üniversite öğrencisinden elde edilmiştir. İlgili maddelere öğrencilerin beş farklı derecelendirme ile yanıt vermesi sağlanmıştır. Tipik davranış süreci için ideal nokta MTK modellerinden GAMOM kullanılırken maksimal davranış süreci için monotonik kümülatif MTK modellerinden ATM kullanılmıştır. GAMOM kestirimleri için GAMOM2004 (versiyon 1.1) programından ATM kestirimleri için ise Multilog (versiyon 7.0.3) programından faydalanılmıştır. Modellere ilişkin kestirimlerde marjinal maksimum olabilirlik kestirim yöntemi; birey parametrelerine ilişkin kestirimler için ise MLE yöntemi kullanılmıştır. Modellerin uyumu için MODFIT (versiyon 2.0) ile elde edilen χ^2/sd değerleri kullanılmıştır. Çalışmadan elde edilen sonuçlar ATM'nin daha iyi uyum gösterdiğini ancak hem ATM hem GAMOM'un da ilgili uyum kriterinin üzerinde olduğunu göstermiştir. Bunun yanı sıra ATM ve GAMOM ile elde edilen yetenek kestirimleri arasındaki ilişkinin 0,47 olduğu belirtilerek iki model arasındaki uyumsuzluğun daha çok üst yetenek düzeylerinde meydana geldiği saçılma grafiği ile gösterilmiştir.

Ling vd. (2016) 64 maddelik bireyler arası değerlerin dairesel (circumplex) ölçeğindeki maddelere bireylerin verdikleri yanıtların tanımlanmasında ideal nokta ve kümülatif modellerin hangisinin en iyi seçim olabileceğine dair incelemeler yapmışlardır. İncelemeler söz konusu ölçeğin uygulandığı 1894 üniversite öğrencisinden elde edilmiş veri üzerinde gerçekleştirilmiştir. Bu doğrultuda MTK altında ideal nokta modellerinden GAMOM; kümülatif modellerden ise genelleştirilmiş kısmi puan modeli ile grafiksel (madde karakteristik eğrileri) ve istatistiksel karşılaştırmalara (χ^2/sd , AIC, BIC) dayalı olarak en iyi uyum sergileyen modele ilişkin çıkarımlar yapılmıştır. Bunun yanı sıra ölçme modellerinin bireylerin yerleştirilmesindeki ve yakınsak geçerlik üzerindeki etkisi test edilmiş olup ayrıca ideal nokta modeli ve klasik test kuramına (beklenen toplam puan – KTK “gerçek puan”) dayalı birey sıralamalarının da birbirinden ayrışma derecesi

incelenmiştir. Elde edilen model uyumu sonuçları ile hem madde karakteristik eğrisine ilişkin grafiksel hem de model uyumuna ilişkin istatistiksel karşılaştırmalar ile ölçekteki maddelere verilen yanıtları en iyi tanımlayan modelin ideal nokta modeli olduğu gösterilmiştir. Birey sıralamasına ilişkin çıkarımları netleştirmek için sırasıyla üst uçta yer alan 100, 200, 300, ve 500 kişi ile bütün örneklemdaki kişiler ele alınarak KTK ve GAMOM ile elde edilen yetenek kestirimleri arasındaki ilişki pearson korelasyon ile incelenmiştir. Bu incelemeler yüksek puanlı yanıtlayıcıların yer aldığı alt örneklemelerde korelasyonun bütün bireylerin ele alındığı gruptan daha zayıf olduğunu dolayısıyla üst uçtaki birey sıralamalarının farklılaştığını göstermiştir. Son olarak yakınsak geçerlik incelemeleri için Locke'un (2000) Bem Cinsiyet Rolü envanteri ile tematik kavrama testi verilerinden faydalanılmıştır. Ancak yakınsan geçerlik incelemelerine göre genel olarak kriter değişkenlerin kestiriminde KTK ve GAMOM arasında neredeyse farklılaşma olmadığı sonucuna ulaşılmıştır.

Cao vd. (2018) kümülatif ve ideal nokta puanlama yaklaşımlarının eğrisel ilişkilerin belirlenmesine olan etkisini Monte Carlo simülasyon çalışmalarıyla incelemiştir. Kümülatif puanlama yaklaşımları olarak bütün maddelerin ortalama puanı, regresyon temelli faktör puanları, ATM; ideal nokta puanlama yaklaşımı olarak da GAMOM ele alınmıştır. Puanlama yaklaşımlarının performansları birinci tür hata oranı, güç, kestirim doğrulukları (RMSEA), yanlılık ve R^2 'deki değişim olmak üzere beş indeks üzerinden karşılaştırılmıştır. Çalışmada örneklem büyüklüğü (250, 500, 1000, 2000), yordayıcıda (-0,5) ve dışsal değişkende çarpıklık (-0,5), madde sayısı (15, 30) ve quadratik etki (var, yok) faktörlerine göre incelemeler yapılmıştır. Sonuçlar ele alındığında gücün örneklem büyüklüğüne bağlı olarak bütün puanlama yaklaşımları açısından arttığı gözlenmiştir. Ancak örneklem büyüklüğü küçük olduğunda (250, 500) ideal nokta puanlama yaklaşımının kümülatif puanlama yaklaşımlarına göre daha yüksek güç ile sonuçlandığı belirtilmiştir. Ancak örneklem büyüklüğü arttıkça puanlama yaklaşımları arasındaki güç farklılığının azaldığı gözlenmiştir. Test uzunluğu açısından, GAMOM yaklaşımında tutarlı bir etkinin söz konusu olduğu, madde sayısı arttıkça gücün arttığı gözlenmiştir. Ancak kümülatif yaklaşımda test uzunluğuna dayalı güç değişiminin örneklem büyüklüğüne dayalı olarak değiştiği belirtilmiştir. Bütün puanlama yaklaşımlarına ilişkin gücün yordayıcı ve sonuçtaki (dışsal) çarpıklık durumundan aşırı derecede etkilendiği gözlenmiştir. GAMOM için çarpıklığın hangi değişkende olduğu gözetilmeksizin küçük örneklem koşulunda (250, 500) daima düştüğü gözlenmiştir. Ancak örneklem orta

düzyeydeyken gücün yordayıcı değışkendeki çarpıklıktan daha az etkilendiğı büyük olduğunda ise güç üzerinde çarpıklığın etkisinin olmadığı sonucuna ulaşılmıştır. Kümülatif yaklaşımların ise eğrisel etkilerin belirlenmesinde çarpıklığın negatif etkisinin olmadığı belirtilmiştir. RMSEA açısından inceleme yapıldığında Kümülatif yaklaşımlar arasında tutarlı bir şekilde göze çarpan farklılığın olmadığı gözlenmiştir. İdeal nokta modeline göre veri üretildiğinde GAMOM yaklaşımının daha doğru quadratik katsayı kestirimleri sağladığı görölmektedir. Kümülatif yaklaşıma göre veri üretildiğinde ise küçük örnekleme GAMOM'un kümülatif yaklaşıma göre daha düşük doğrulukla sonuçlandığı ancak örneklem büyüklüğü orta veya yüksek olduğunda kümülatif yaklaşımlar kadar doğru sonuçlar sağladığı gözlenmiştir. Quadratik terimlerin kestiriminde hem ideal nokta hem de kümülatif altındaki puanlama yaklaşımlarının önemsiz negatif yanlılık ürettiğı gözlenmiştir. Genel olarak örneklem büyüklüğü, test uzunluğu, çarpıklık veya puanlama yönteminin yanlılık kestirimi üzerinde sistematik bir etkisi gözlenmemiştir. Benzer şekilde quadratik terimlerden kaynaklı olarak R^2 değışimlerinin de önemsiz olduğu belirtilmiştir.

Kula-Kartal (2018) Akademik Güdülenme Ölçeğı maddelerinin boyutluluğunu ATM, GAMOM, iki faktörlü model ve DIMTEST'e dayalı olarak incelemiştir. İncelemeler Ankara Üniversitesi'nde öğrenimlerini sürdüren 1858 öğrenciden elde edilen veri üzerinde gerçekleştirilmiştir. Veri matrisinin boyutluluğı modeller tarafından sağlanan madde parametre kestirimlerine, genel model veri uyumuna, madde ve birey uyum istatistiklerine dayalı olarak incelenmiştir. Bu doğrultuda madde parametre kestirimleri için ayırtedicilikler; genel model veri uyumu için log-olabilirlik, AIC, BIC ve A-BIC istatistikleri; madde uyumu için χ^2/sd istatistikleri ve birey bazında uyum için I_z istatistikleri ile incelemeler yapılmıştır. Madde ayırt ediciliklerine ilişkin karşılaştırma sonuçlarına göre, maddeler için kuramsal temelle en tutarlı ve en yüksek parametre kestirimlerinin iki faktör modeliyle elde edildiğı görölmüştür. Genel model veri uyumu açısından GAMOM ve iki faktör modelinin ATM'den daha iyi sonuçlar sağladığı ortaya koyulmuştur. Madde uyum istatistiklerine göre en iyi uyumun sırasıyla iki faktör modeli, GAMOM ve ATM olduğu gözlenmiştir. Birey bazındaki model veri uyumuna ilişkin en iyi uyumu sergileyen modellerin sırasıyla ATM, iki faktör model ve GAMOM olduğu gösterilmiştir. Birey bazındaki I_z uyum istatistiğinin yer ve ayırtedicilik parametrelerinden etkilenmesinden dolayı GAMOM, INFIT ve OUTFIT birey uyum istatistikleri ile de incelenmiştir. Buradan elde edilen sonuçlar GAMOM altında bireylerin büyük ölçüde iyi uyum ile sonuçlandığını göstermiştir. Sonuçlar genel olarak değerlendirildiğinde ölçek

maddelerine kuramsal ve istatistiksel açıdan en iyi uyum sağlayan modelin iki faktör model olduğu; GAMOM'un ise ATM'den daha iyi sonuç verdiği gözlenmiştir.

Zhang vd. (2019) kişilik ölçeklerinin uygulamalarına ilişkin yanıt süreçlerinin söz konusu ölçümlerin küçük ölçekli ve büyük ölçekli olma durumlarından etkilenebileceğini öne sürmüş ve bu durumlarda yanıtlayıcıların tepki sürecinin değişebileceğini ele almışlardır. Araştırmacılar bu hususu gözetererek düşük ve yüksek bahisli ölçme durumlarında uygulanan kişilik ölçümlerine yönelik madde yanıt süreçlerini incelemek amacıyla GAMOM ve ATM modellerini ele almışlardır. Bu doğrultuda ideal nokta modeline uygun ölçekler Kapsamlı Kişilik Ölçeğinden (Wang, 2013); kümülatif modele uygun ölçekler ise Uluslararası Kişilik Madde Havuzu'ndan (Goldberg, 1999) elde edilmiştir. Uygulanan ölçme aracındaki yönerge bilgilendirmelerinde farklılıklar yapılarak hileli ve hilesiz olmak üzere iki koşul oluşturulmuştur. Hileli koşulda “çok istediğiniz bir iş için başvurduğunuzu hayal ederek yanıtlayın” bilgilendirmesi verilerek yüksek-riskli; hilesiz koşulda ise “yanıtlar sadece araştırma amacı için kullanılacaktır” bilgilendirmesi verilerek düşük riskli durumlar oluşturulmuştur. Dolayısıyla araştırmada hileli (N=458) ve hilesiz (N=489) olmak üzere iki farklı katılımcı grubu söz konusudur. Araştırmada genel ve madde bazında model-veri uyumu, değişen madde fonksiyonu ve yüksek bahisli durumlarda tepki modellerinin kişilik puanlarının geçerliği üzerindeki etkileri incelenmiştir. Veri analizleri mirt R paketi ile gerçekleştirilmiştir. Genel model-veri uyumuna ilişkin incelemelere göre hem hileli hem hilesiz koşulda GAMOM' un ATM'ye göre veriye daha iyi uyum sağladığı görülmüştür. Ancak madde bazındaki uyum açısından modellerin birbirine karşı belirgin bir üstünlüğü tespit edilmemiştir. Değişen madde yanlılığına ilişkin sonuçlar incelendiğinde GAMOM (% 45,57) altında yanlılık gösteren madde sayısının ATM (%56,76) altında gözlenen yanlı madde sayısından daha düşük olduğu gözlenmiştir. Son olarak hilesiz durumdaki monotonik olmayan maddelerin çoğu (ve hatta bazı uç maddelerin) hileli durumda açık bir şekilde monotonik olmayan bir davranış sergilediği gözlenmiştir. Bu sonuçlar ışığında yüksek riskli koşullarda madde yanıt süreçlerinin tanımlanmasında ideal nokta modelinin iyi bir seçim olabileceği belirtilmiştir.

Sun vd. (2020) kümülatif modellerden ATM ve ideal nokta modellerinden GAMOM' u “Yetişkin Bağlılığı” yapısı üzerinde karşılaştırmışlardır. Çalışmada “Yakın İlişkilerde Deneyimler-Revize edilmiş” ölçme aracı kullanılmış olup ölçme aracının “bağlılıkla ilişkili kaygı” ve “bağlılıkla ilişkili kaçınma” boyutları mevcuttur. İncelemeler için hem ampirik (çalışma-1, çalışma-2) hem de simülasyon (çalışma-3) çalışmaları yürütülmüştür. Her iki

model altında elde edilen yetenek kestirimleri için EAP kestirim yöntemi dikkate alınmıştır. Çalışma 1’de 1293 kişilik örneklem üzerinde boyutlar bazında model-veri uyumu incelemeleri uyum iyiliği istatistiksel testleri ve grafiksel incelemeler aracılığıyla yapılmıştır. Elde edilen sonuçlara göre χ^2/sd açısından her iki modelin de yeterli uyumu göstermediği ancak mutlak uyum değerlerinin makul olduğu belirtilmiştir. Grafiksel incelemelere göre her iki modelin uyum grafiklerinin kestirilen eğrilerin %95 güven aralıkları içerisinde yer aldığı ve her iki modelin de uyum sergilediği ele alınmıştır. Bunun yanı sıra ölçümlerin ideal nokta yanıt özelliklerine sahip olup olmasına ilişkin incelemeler için grafiksel yol izlenilmiştir. GAMOM modeli altında açık bir şekilde monotonik olmayan davranış örüntülerinin gözlemlendiği ele alınmıştır. Çalışma-2’de yeni bir örneklem (N=2000) üzerinde Subjektif İyi Olma’ nın alt boyutu olan “algılanan genel yaşam doyumu” dışsal ölçüt olarak ele alındığı ve yetişkin bağlılığının kaygı alt boyutuyla ilişkilendirildiği kriter geçerliği çalışmasına yönelik incelemeler yapılmıştır. Söz konusu yapıların ters yönde ilişki göstermesi beklendiğinden uzaksak geçerlik kanıtları açısından incelemeler gerçekleştirilmiştir. “Algılanan genel yaşam doyumu”na ilişkin puanlar ise KTK ile elde edilmiştir. Modellere ilişkin yetenek kestirimleri hem bütün örneklem üzerinde (N=2000) hem de örneklemin %10’luk üst kısmı (N =200) üzerinde yürütülmüştür. Elde edilen sonuçlar genel örneklem bazında kümülatif ve ideal nokta modellerinin karşılaştırılabilir olduğunu ancak üst uçta yer alan örneklem bazında GAMOM ile elde edilen kriter geçerlik değerlerinin daha yüksek olduğunu göstermiştir. Bunun yanı sıra kümülatif ve ideal nokta modellere ilişkin yetenek kestirimleri ilişkilendirildiğinde genel olarak yüksek olduğu (0,94-0,98 arasında değişim) ancak üst örneklem bazında ilişkinin düştüğü (0,49-0,54 arasında değişen) gözlenmiştir. Saçılma grafiği ile incelemeler yapıldığında yetenek ölçeğinin orta kısımlarında her iki modelin monotonik olarak ilişkili olduğu ancak üst kısımlarda sert bir farklılığın olduğu gösterilmiştir. Son olarak çalışma-3’te her bir yanıt sürecine göre veri üretilerek kümülatif süreç ile üretilen veri ideal nokta modeli ile ve tam tersi ideal nokta yanıt süreci ile üretilen veri kümülatif model ile analiz edilerek yanıt süreçleri ile doğrudan eşleşmeyen modellerin uygulama sonuçları gösterilmiştir. Gerçek yanıt süreci kümülatif olarak ele alınıp veri üretildiğinde gerçek yetenek puanları (kümülatif) ile ideal nokta modeli ile elde edilen yetenek kestirimleri arasındaki ilişkinin bütün yetenek ölçeği için yüksek ($r=0,99$) ve uç bölge için ise orta düzeyde yüksek ($r=0,65$) olduğu gözlenmiştir. Ancak gerçek yanıt süreci ideal nokta olarak ele alınıp veri üretildiğinde gerçek yetenek puanları (ideal nokta) ile

kümülatif model ile elde edilen yetenek kestirimleri arasındaki ilişkinin genel ölçek bazında -0,75 olduğu ve bireylere ilişkin düzeylerin büyük ölçüde değişme eğilimi gösterdiği gözlenmiştir.

İdeal Nokta Modellerinin Kestirim Yöntemlerine Yönelik Uygulama Araştırmaları

Roberts, Donoghue ve Laughlin (2002) GAMOM altında marjinal maksimum olabilirlik/EAP yöntemi ile birey parametrelerini elde ederek EAP'ın etkililiğini test etmişlerdir. İncelemeler iki çalışma olarak yürütülmüştür. Çalışma 1'de altı farklı test uzunluğu (5, 10, 15, 20, 25, 30) ve altı farklı örneklem büyüklüğü (200, 300, 500, 750, 1000) koşulları faktöryel olarak çaprazlanarak 36 koşulda inceleme yapılmıştır. Madde lokasyonları eşit bir şekilde mesafe pozisyonlarında olacak şekilde -2 ve 2 aralığında yerleştirilmiştir. Gerçek birey parametreleri $N(0,1)$ dağılımından seçkisiz olarak örneklenmiştir. Her bir madde-birey kombinasyonu için gözlenen yanıt 6-noktalı (tamamen katılmıyorum, ... , tamamen katılıyorum) olacak şekilde üretilmiş olup her bir koşul için 30 replikasyon yapılmıştır. Örneklem büyüklüğü ve test uzunluğu replikasyonlar arası iki faktör olarak ele alınmış olup her bir indeks için ANOVA'lar yürütülmüştür. Elde edilen sonuçlara göre hem ana etkilerin hem de etkileşim etkilerinin istatistiksel olarak anlamlı olduğu gözlenmiştir. Genel olarak değerlendirildiğinde doğruluk ölçümlerine ilişkin analizler ile örneklem büyüklüğünün 750; test uzunluğunun 15-20 olduğu ve önsel dağılımın gerçek yetenek dağılımını doğru bir şekilde tanımladığı durumda makul parametre kestirimlerinin elde edilebildiği gösterilmiştir. Çalışma 2'de gerçek yetenek dağılımı ile önsel dağılım arasındaki tutarsızlıkların sonuçlar üzerindeki etkisini gözlemlemek için önsel dağılımın normal olduğu; gerçek yetenek dağılımının ise normal olduğu ve normal olmadığı koşullarda incelemeler yapılmıştır. Bu doğrultuda gerçek yetenek dağılımının türü (normal, bimodal-simetrik, pozitif çarpık ve negatif çarpık), örneklem büyüklüğü (500, 750, 1000, ve 2000) ve quadrat nokta sayısı (15, 20, 30) olmak üzere üç farklı faktör çaprazlanarak 48 koşulda incelemeler yapılmıştır. ANOVA sonuçlarından elde edilen etki büyüklüğü incelemelerine göre yetenek kestirimleri için quadrat sayısı ve gerçek yetenek dağılımının öne çıktığı gözlenmiştir. Yetenek kestirimlerinde hata miktarının gerçek yetenek dağılımının çarpık olduğu koşullarda, özellikle dağılımın pozitif çarpık olduğu koşulda en yüksek olduğu gözlenmiştir. Şekilsel incelemeler üzerinden de en yüksek hata miktarının uç yanıt örüntüleriyle ilişkili olduğu gösterilmiştir. Genel değerlendirme yapıldığında uç yanıt örüntüleriyle ilişkili yetenek

değerleri hariç GAMOM altında 20 maddelik test koşulunda EAP ile elde edilen kestirimlerin makul olduğu ifade edilmiştir.

Luo ve Andrich (2004) çok kategorili yanıtlar için ideal nokta modelinde madde parametrelerinin mevcut olduğu durumda WLE eşitliklerini türeterek WLE' nin kestirimlerdeki etkililiğini MLE ile karşılaştırmalı olarak simülasyon çalışması ile incelemişlerdir. Çalışmada eşik parametreleri arasındaki mesafeler eşit alınmış olup hyperbolic cosine fonksiyonu için sonuçlar raporlanmıştır. Her bir veri simülasyonu için madde lokasyonları $[-2, 2]$ aralığında olacak şekilde 10 madde için yanıtlar üretilmiştir. Madde birimleri seçkisiz olarak $[0,6, 1,0]$ aralığında üretilmiştir. $N(0, 2)$ dağılımına sahip 500 birey lokasyonu (yetenek parametresi) üretilmiş olup 10 replikasyon yapılmıştır. Bütün iterasyonlarda yakınsama kriteri olartak 0,001 seçilmiştir. Kestirim süreci iki adımda ardışık tekrarlar ile gerçekleştirilmiştir: (i) birey parametrelerinin bilindiği veya sabit olduğu durumda madde parametreleri kestirilmiştir; (ii) madde parametrelerinin sabit olduğu durumda ise birey parametreleri kestirilmiştir. Yakınsama gerçekleşene kadar verilen iki adım tekrar edilmiştir. WLE ve MLE birey kestirimleri arasındaki korelasyon ve RMSE değerleri raporlanmıştır. Elde edilen sonuçlara göre MLE'nin madde parametreleri için şişirilmiş kestirimlere yol açtığı; WLE'nin ise maddelerin birçoğu için üretilen değerlere daha yakın sonuçlar verdiği gösterilmiştir. Bunun yanı sıra üretilen ve kestirilen birey parametreleri arasındaki korelasyonun WLE için daha yüksek olduğu ve WLE'nin daha düşük RMSE ile sonuçlandığı gözlenmiştir. Bu doğrultuda ideal nokta model kullanımında WLE'nin etkililiği gösterilmiştir.

Wang ve Liu (2011) GAMOM altında bilgisayarlı (computerized) sınıflama test etmeyi uygulayarak bir takım simülasyon serisi ile modelin uygulanabilirliğini göstermektir. Bu doğrultuda çalışmalarında GAMOM altında bilgisayarlı sınıflama test etme kapsamında yetenek güven aralık yöntemi ve ardışık olasılık oran testi yöntemi olmak üzere iki tane madde seçim yöntemi geliştirerek bu yöntemlerin sınıflamaya yönelik doğruluk ve etkililikleri değerlendirilmiştir. Değerlendirme yapılırken üç farklı bağımsız değişken ele alınmıştır: (a) her bir madde için nokta sayısı: 2, 4 ve 6 (derecelendirme kategori sayısı); (b) sınıflama kategori sayısı: 3 ve 4; ve (c) yetenek kestirim yöntemleri: MLE, MAP ve EAP. Kesme puan noktaları üç düzey için -1 ve 1; dört düzey için ise -1.5, 0 ve 1.5 olarak ele alınmıştır. GAMOM altında veri üretmek üzere kullanılan madde havuzu Roberts et al. (2001) un çalışmasındaki kürtaj hakkındaki maddelerin 750 kişiye uygulanmış ve kalibre edilmiş 6'lı derecelendirilmiş 47 maddeyi içermektedir. Mevcut olan 6'lı derecelendirme

ardışık derecelendirmelerin birleştirilmesi ile 4'lü ve 2'li olarak da düzenlenmiştir. GAMOM2004 bilgisayar programı ile her bir derecelendirme için ayrı ayrı madde kalibrasyonları yapılmıştır. $N(0,1)$ dağılımından toplamda 1000 kişi üretilmiştir. Her bir birey için sınıflama gerçekleştiğinde veya maksimum test uzunluğu 20 maddeye ulaştığında test durmaktadır. Veriyi üretmek ve analiz etmek için Windows-temelli MATLAB programı yazılmıştır. Çalışmadan elde edilen bulgular her iki yöntemin de 10 maddeden daha az sayıdaki madde ile çok yüksek bir doğruluk ile sonuçlandığını göstermiştir. Özellikle maddelerin nokta sayısı arttıkça sınıflamanın daha doğru ve etkili olduğu gözlenmiştir. Yetenek güven aralık yöntemi için MAP ve EAP kestirim yöntemlerinin orta kategoride MLE yönteminden daha iyi olduğu ancak uç kategorilerde ise MLE yönteminin MAP ve EAP yöntemlerinden daha doğru olduğu gözlenmiştir. Bunun yanı sıra özellikle maddelerin iki kategorili olarak ele alındığı durumlarda MAP ve EAP kestirim yöntemlerinin daha iyi sonuç verdiği ve kategori sayısı arttıkça MLE, MAP ve EAP yöntemleri arasındaki farklılığın azaldığı görülmüştür.

Roberts ve Thompson (2011) GAMOM madde parametrelerinin kestirimlerinde marjinal MAP prosedürünü uygulayarak buradan elde edilen sonuçları marjinal maksimum olabilirlik ve MCMC prosedürlerine dayalı elde edilen kestirimler ile karşılaştırmışlardır. Araştırma simülasyon üzerinden yürütülmüş olup değişen örneklem büyüklükleri ($N=500$, $N=2000$), ölçme aracının uzunluğu ($I=10$, $I=30$) ve madde yanıt kategori sayılarına (2,3,4,5,veya 6) göre çeşitli faktörler açısından $2 \times 2 \times 5 \times 3$ karışık ölçüm deseni oluşturularak madde parametre kestirimleri üzerinde inceleme yapılmıştır. Simülasyon için gerçek veriden elde edilmiş parametreler kullanılmıştır. Gerçek veri parametreleri GAMOM altında marjinal MAP tekniği kullanılarak elde edilmiştir. Madde lokasyonları yetenek ölçeği üzerinde düşükten yükseğe doğru oluşturulmuş beş sıralı aralığa atanmıştır. Replikasyonda her bir aralıktan eşit sayıda seçkisiz olarak geriye koyma yöntemi dikkate alınmış ve gerçek parametreler bu şekilde kullanılmıştır. Her replikasyon üzerinde $N(0,1)$ dağılımından gerçek birey parametreleri bağımsız bir şekilde elde edilmiştir. Gerçek madde ve birey parametreleri kullanılarak her bir kategori için bir yanıt olasılığı türetilmiştir. Ardışık kategorilerle ilişkili olasılıklar $(0,1)$ olasılık aralığını dilimlemek için kullanılmıştır. Bu doğrultuda yanıtlar üretilmiştir. Her bir replikasyon üzerinde madde parametreleri marjinal MAP, marjinal maksimum olabilirlik ve MCMC yaklaşımları ile kestirilmiş olup bayesian yaklaşımlar olan MCMC ve marjinal MAP parametre kestirimleri için aynı önsel kullanılmıştır. Marjinal maksimum olabilirlik ve marjinal MAP

GAMOM2004 ile; MCMC ise WinBugs bilgisayar programı ile kestirilmiştir. Elde edilen parametre kestirimleri gerçek parametre metriğinde yeniden ölçeklenmek üzere madde karakteristik eğri yöntemi uygulanmıştır. GAMOM-LINK bilgisayar programı ile yeniden ölçeklenen kestirimler gerçek değerleri ile karşılaştırılmış ve her bir parametrenin doğruluğu ele alınmıştır. Sonuçlar incelendiğinde marjinal MAP parametre kestirimlerinin genel olarak en düşük standart hata ortalamasına sahip olduğu dolayısıyla en doğru parametre kestirimlerinin marjinal MAP ile elde edildiği gözlenmiştir. Madde kategorisi düşük olduğunda ve gerçek madde lokasyonları uç olduğunda marjinal maksimum olabilirlik kestirimleri bozulurken, marjinal MAP'ın bu durumda da iyi sonuç verdiği ve MCMC' ye göre de sayısal (computationally) daha etkin olduğu gösterilmiş olup marjinal MAP' ın GAMOM parametre kestiriminde kullanımı önerilmiştir. Bu doğrultuda madde parametre kestiriminde bayesian yaklaşımlardan marjinal MAP'ın etkililiği gösterilmiştir.

İdeal Nokta Modellerine Yönelik Diğer Alanlardaki Uygulama Araştırmaları

Anderson Koenig ve Roberts (2007) çalışmalarında test karakteristik eğrisi, madde karakteristik eğrisi ve operating karakteristik eğri bağlama (linking) yöntemlerinin simetrik ve simetrik olmayan versiyonlarını geliştirerek bunları GAMOM'a genişletmişlerdir. Bu yöntemlerin GAMOM ile kullanımındaki yeterliği değişen çeşitli faktörlere göre simülasyon çalışması ile değerlendirilmiştir. Altı kategorili yanıt seçenekli 20 tane tutum ifadesi simüle edilmiştir. Parametreler Roberts'in (2001) çalışmasındaki prosedürleri takip eden gerçek bir veri setine benzeyen dağılımlardan elde edilmiştir. GAMOM altında veri üretmek için delta parametreleri -2 ve 2 aralığında tekdüze bir dağılımdan; madde ayırtedicilik parametreleri 0,5 ile 2,0 aralığında değişen bir dağılımdan örneklenmiştir. Her bir bağlama yönteminin performansı dört koşulda incelenmiştir: (a) simetrik ve simetrik olmayan formülasyonlar; (b) yatay ve dikey bağlama senaryoları; (c) küçük ($N = 300$) ve büyük ($N = 1000$) örneklem; ve (d) 5, 10, 15 ve 20 olmak üzere farklı ankor madde sayısı. Bu değişkenler 2 (bağlama senaryoları) x 2 (örneklem büyüklüğü) x 3 (karakteristik eğri yöntemleri) x 2 (simetrik ve simetrik olmayan çözüm) x 4 (ankor madde sayısı) olan iki replikasyonlar arası ve üç replikasyonlar içi olan bir faktöryel desen üretmek üzere birleştirilmiştir. Faktöryel desenin her bir hücresi için 30 replikasyon yapılmıştır. Simülasyon sonuçları beş yönlü karışık faktöryel ANOVA ile analiz edilmiştir. Bağlama katsayılarının her biri olan A ve B için iki ayrı şekilde tek değişkenli ANOVA'lar yürütülmüştür. Gözlenen ve gerçek değer arasındaki farkın karesine ilişkin değerler

bağımlı değişken olarak ele alınmıştır. Analiz sonuçları ele alındığında A katsayısına göre bağlama senaryosu, örneklem büyüklüğü, çözüm türü ve ankor madde seti; B katsayısına göre de bağlama senaryosu, örneklem büyüklüğü, karakteristik eğri yöntemi ve ankor madde seti ana etkiler olarak yorumlanmıştır. B katsayısı için karakteristik eğri yöntemleri ele alındığında TCC yönteminin OCC ve ICC'den daha düşük hataya sahip olduğu ve diğer iki yöntemden anlamlı olarak değiştiği gözlenmiştir. A katsayısı için çözüm türü ele alındığında asimetrik çözüm ile elde edilen kareler ortalamasının biraz daha yüksek olduğu ancak etkisinin düşük olduğu belirtilmiştir. Örneklem büyüklüğü açısından ana etkiler değerlendirildiğinde küçük örneklem koşulunda hata karelerinin ortalamasında biraz daha artış gözlenmiştir. Bununla birlikte saçılma grafikleri incelendiğinde örneklem büyüdüğünde uç birey sayılarında azalma olduğu gözlenmektedir. Son olarak ankor madde sayıları açısından değerlendirildiğinde 15 ve 20 maddelik ankor sayısının olduğu durumlarda manidar farklılık olmadığı ve 15 maddeye kadar ise farklılığın gözlemlendiği belirtilmiştir. Araştırma sonuçları genel olarak değerlendirildiğinde karşılaştırmalar arasında çok küçük manidar farklılıklar olsa bile hataya ilişkin kareler ortalamasının çok düşük olduğu belirtilmiştir. Bu doğrultuda genel değerlendirme yapıldığında GAMOM altında bağlama yöntemlerine ilişkin çeşitli senaryolar dikkate alındığında ilgili katsayılar açısından performansın iyi olduğu ve bağlama çalışmalarında kullanımının uygun olabileceği önerilmiştir.

Wang ve Wu (2016) aşamalı-yanıtlı maddelerden oluşan çoklu testlere yönelik kestirim yapılabilmesi için doğrulayıcı çok boyutlu GAMOM'u geliştirmişlerdir. Söz konusu çoklu testler için maddeler-arası çok boyutluluk yani basit faktör yapısı ele alınmış olup bir takım simülasyon çalışmaları ve gerçek veri uygulamaları ile modelin etkililiği değerlendirilmiştir. Simülasyon desenleri oluşturulurken örneklem büyüklüğü (500 ve 1500) ve örtük özellikler arasındaki korelasyon (düşük: $r = 0,30$, orta: $r = 0,60$ - $0,80$, ve yüksek: $r = 0,90$) olmak üzere iki faktör manipüle edilerek toplamda altı durumda incelemeler yapılmıştır. Veriyi üretmek için Matlab bilgisayar programından faydalanılmış olup 30 replikasyon gerçekleştirilmiştir. Üretilen veriyi analiz etmek için WinBUGS kullanılmıştır. Parametreler için önsel dağılımlar sırasıyla $\zeta_i \sim N(0, 4)$, $a_i \sim \log N(0, 0,25)$, $t_{i1} \sim N(-2, 4)$, $t_{i2} \sim N(-1, 4)$, $t_{i3} \sim N(-0,5, 4)$, $\theta_n \sim MVN(0, \Sigma)$ olarak ele alınmış olup Σ büyük bir varyans ile Wishart dağılımını takip etmektedir. Birey parametreleri için EAP yöntemi kullanılmış olup değerlendirmeler için yanlışlık ve RMSE değerleri hesaplanmıştır. Simülasyon çalışmalarından elde edilen bulgulara göre büyük örneklem daha düşük

ortalama mutlak yanlılık ile sonuçlandı; örtük özellikler arasındaki korelasyonun eğim, yer ve eşik parametreleri üzerinde düşük etkiye sahip olduğu ancak düşük korelasyonun varyans-kovaryans parametreleri üzerinde daha düşük mutlak yanlılık ürettiği gözlenmiştir. RMSE değerleri incelendiğinde ise büyük örneklemin daha düşük RMSE ile sonuçlandı gözlenmiştir. Yanlılığa ilişkin sonuçlara benzer şekilde örtük özellikler arasındaki korelasyonun eğim, yer ve eşik parametreleri üzerinde küçük etkiye sahip olduğu ve düşük korelasyonun varyans-kovaryans parametreleri üzerinde daha düşük ortalama RMSE ürettiği gözlenmiştir. Ancak varyans-kovaryans parametrelerine ilişkin RMSE değerlerinin eğim, yer ve eşik parametrelerinden biraz daha yüksek olduğu gözlenmiştir. Gerçek veriye dayalı inceleme için dövme karşı tutum ölçeği verisi (Lin ve Huang, 2005) kullanılmıştır. Üç alt boyutlu olan ölçeğin çok boyutlu analizinin yanı sıra her bir boyutu ardışık olarak tek boyutlu yaklaşım ile analiz edilmiş ve buna dayalı model karşılaştırmaları yapılmıştır. Gerçek veriye dayalı sonuçlar ele alındığında logaritmik PsBF değerleri ve PPC anlamlılık değerlerine göre çok boyutlu modelin tek boyutlu modele kıyasla veriye daha iyi uyum sergilediği gözlenmiştir. Bunun yanı sıra test güvenilirliklerinin çok boyutlu model için sırasıyla 0,915, 0,899 ve 0,816 olduğu; tek boyutlu model için ise 0,856, 0,832 ve 0,816 olduğu dolayısıyla çok boyutlu analizin güvenilirliği de iyileştirdiği sonucuna ulaşılmıştır. Son olarak boyutlar arası korelasyon değerleri incelenmiştir. Çok boyutlu yaklaşımda doğrudan örtük özellikler arasında hesaplanan ilişkiler tek boyutlu her bir ölçek arasındaki pearson korelasyon değerleri ile karşılaştırılmıştır. Elde edilen ilişkiler çok boyutlu model analizi ve tek boyutlu model analizi (parantez içinde) sırasıyla birinci-ikinci boyut arasında 0,80 (0,56), birinci-üçüncü boyut arasında 0,39 (0,19), ve ikinci-üçüncü boyut arasında 0,30 (0,05) olarak elde edilmiştir. Bu bulgulara göre tek boyutlu yaklaşım ile boyutlar arası ilişkilerin olduğundan düşük hesaplandığı çok boyutlu modelin sağladığı yüksek ilişkilerle de test güvenirliliğinin arttığı belirtilmiştir. Dolayısıyla basit faktör yapısına sahip maddeler-arası çok boyutlu ideal nokta modelin tek boyutlu ideal nokta modele göre model-veri uyumu, güvenilirlik ve boyutlar arası yüksek korelasyon açısından daha etkili olduğu öne sürülmüştür.

Wang ve Engelhard Jr. (2019) değerlendirici aracılı ölçmelerde bireysel değerlendirici tercihlerinin kompozisyonlar için amaçlanan puanlama rubriğine dayalı olarak hakettiğinden daha yüksek veya daha düşük derecelendirmelere yol açabileceğini öne sürmüşlerdir. Bu durumda kompozisyonların differansiyel sıralamasının ve yanlılığın tespit edilmesi için baskınlık yanıt sürecinin yetersiz kalabileceği gerekçesi ile ideal nokta

modeline dayalı olarak değerlendirici aracılı ölçme sisteminin psikometrik kalitesinin değerlendirilmesini amaçlanmıştır. Bu doğrultuda araştırma yedinci sınıflardan elde edilmiş geniş ölçekli bir yazma ölçmesi ile elde edilen derecelendirmelere dayalı veri üzerinde yürütülmüştür. Ölçme aracında yer alan maddeler arasından öğrencilerin hikâyelendirme becerilerini değerlendirmek için iki okuma pasajına dayalı yazmaya ilişkin dizayn edilen maddeler incelenmiştir. Yirmi tane eğitilmiş operasyonel puanlayıcı 100 tane öğrenci kompozisyonunu iki domain üzerinde puanlamışlardır. Detaylı bilgi sağlayabilmek için analizler her iki domain ayrı ayrı yürütülmüştür. Verilerin analizi için ideal nokta modellerden çok kategorili veri için olan hyperbolic cosine model ele alınmıştır. Her bir puanlayıcı, kompozisyonlara yönelik seçim sıralamalarını temsil eden bir Bireysel Puanlayıcı (Individual Rater) ölçeğine sahiptir ve IR olarak ifade edilmiştir. Model altında değerlendirici aracılı ölçmeler kapsamında puanlayıcılar ve kompozisyonların bileşik kalibrasyonunu yansıtmak üzere ele alınan ölçek düzeyi J (Joint) ölçek olarak ifade edilmiştir. Bir IR ölçeği bir puanlayıcının lokasyonunda J ölçeğini katlayarak; bir J ölçeği ise bütün puanlayıcıların IR ölçeğini yayılmasına dayalı (ideal nokta) olarak oluşturulabilir. Analizler için RateFoLD yazılımı (Luo & Andrich, 2003) kullanılmıştır. Genel uyumu test etmek için Pearson'ın χ^2 istatistiğine dayalı değerlendirme yapılmıştır. Çalışma sonuçlarına göre birinci domain ($\chi^2(76) = 58,86$; $p = 0,96$) ve ikinci domain ($\chi^2(76) = 81,90$; $p = 0,30$) olmak üzere kompozisyonların gözlenen sırasının HCM-P'ye dayalı beklenen sıralamalara uyduğu gözlenmiştir. Birinci domaine ilişkin kategori yanıt eğrileri incelendiğinde her bir eğrinin H ölçeği üzerinde bir essayı tercih etme olasılığının puanlayıcının lokasyonundan itibaren zirve yapıp düşüşe geçtiği belirtilmiştir. Puanlayıcı uyum ölçüleri için, beklenen fonksiyon eğrileri ve χ^2 testleri kullanılmıştır. Domain bir altında bütün puanlayıcılara ilişkin χ^2 istatistiklerinin anlamsız olduğu ve dolayısıyla bütün puanlayıcıların iyi model uyumu indekslerine sahip olduğu gözlenmiştir. Domain iki altında ise bir puanlayıcı hariç bütün puanlayıcılara ilişkin χ^2 istatistiklerinin anlamsız olduğu ve dolayısıyla geriye kalan puanlayıcıların iyi model uyumu indekslerine sahip olduğu gözlenmiştir. Domainlere ilişkin polinomial eğriler incelendiğinde birinci domainin % 96,94; ikinci domainin ise % 97,14 varyans açıkladığı ve iyi uyum sergilediği gözlenmiştir. Son olarak ideal nokta J ölçeğinin asli yorumlamasını ortaya çıkarmak için HCP-M'ye dayalı olarak Coh-Metrix kompozisyon nitelik indisleri ve kompozisyon lokasyon ölçümleri arasındaki ilişkiyi incelemişlerdir. İncelemelere dayalı olarak her iki alanda hem doğrusal hem de kuadratik regresyon fonksiyonlarının HCM-P kompozisyon

lokasyon ölçümleri ile her bir essay özelliği için iyi uyum sergilediği grafiksel incelemeler ile gösterilmiştir. Bu sonuçlara dayalı olarak değerlendirici aracılı ölçme durumlarında puanlayıcı kararlarının kalitesinin değerlendirilmesinde diğer yaklaşımlara ek olarak ideal nokta modelin iyi bir tercih olabileceği yönünde öneriler sunulmuştur.

Sınıflama İndekslerine Yönelik Araştırmalar

Ercikan ve Julian (2002) çalışmalarında sınıflama doğruluğunun ölçme kesinliği/doğruluğu, yeterlik düzeylerinin sayısı ve farklı yetenek ranjlarına göre nasıl farklılaştığını simülasyon temelli olarak çeşitli koşullarda incelemişlerdir. Bu doğrultuda 2-5 arasında değişen yeterlik düzeyleri 10 farklı test koşulunda ele alınmıştır. Her bir yeterlik düzeyi için ise iki farklı kesme puan seti toplamda ise 8 farklı kesme puan seti tanımlanmıştır. Ayrıca yetenek ölçeği 8 farklı bölgeye ayrılarak incelemeler farklı yetenek ranjlarında da ele alınmıştır. Güvenirlik katsayıları (cronbach alpha ve KR-20) 0,79 ile 0,92 aralığında değişen 10 farklı testten (iki kategorili, çok kategorili ve karma) elde edilen parametrelere dayalı olarak gerçek yetenek puanları ve simüle edilmiş madde yanıtları iki kategorili ve çok kategorili olarak BSSIM simülasyon programı (Burket, 1995) ile üretilmiştir. İlk olarak normal dağılım gösteren 3000 kişilik gerçek puanlar üretilmiştir. Madde parametreleri, gerçek puanlar kullanılarak simüle edilmiş yanıtlara dayalı olarak üretilmiştir. İki kategorili maddeler için 3 parametrelili lojistik model (3PLM) kullanılırken, çok kategorili maddeler için 2 parametrelili kısmi puan modeli kullanılmıştır. Madde parametrelerinin kestirilmesinde iki kategorili ve çok kategorili madde parametrelerinin kestirilmesini sağlayan PARDUX kullanılmıştır. Bu programda EM algoritması ile uygulanan marjinal maksimum olabilirlik kestirim prosedürleri kullanılmaktadır. Kestirilen madde parametreleri ve simüle edilmiş yanıtlara dayalı olarak MLE'ye dayalı yetenek değerleri elde edilmiştir. Orijinal madde parametreleri ankor olarak ele alınıp Stcoking ve Lord'un eşitleme yöntemi ile kestirilen yetenek değerlerinden gerçek değerlere dönüşüm uygulanmış ve kesme puanları gerçek puan ölçeği üzerinde olacak şekilde oluşturulmuştur. Her bir sınıf düzey sayısı koşulu için 2 farklı kesme puan seti toplamda ise 8 kesme puan seti gerçek yetenek ölçeği üzerinde tanımlanmıştır. Nihai olarak elde edilen gözlenen ve gerçek puanlara dayalı olarak basit bir uyum ölçüsü olan Po ve Cohen'in K katsayısı kullanılarak sınıflama doğrulukları hesaplanmıştır. Çeşitli koşullarda elde edilen güvenirliliğin yüksek olduğu testlerde sınıflama doğruluğunun daha yüksek olma eğiliminde olduğu gözlenmiştir. Yeterlik düzey sayısı arttıkça da sınıflama doğruluğunun ölçmenin

doğruluğuna karşı duyarlılığının arttığı ve sınıflama doğruluklarının düştüğü görülmüştür. Ayrıca farklı puan ranjlarına göre incelemeler yapıldığında kesme puanının 0,5 standart sapma altındaki ve üstündeki bölgelerin kritik bölgeler olduğu ve bu bölgelerin yer aldığı ranjlarda sınıflama doğruluklarının diğer ranjlardakine göre daha düşük olduğu sonucuna ulaşılmıştır.

Wyse ve Hao (2012) çalışmalarında MTK'ye dayalı sınıflama indekslerinden Rudner-temelli, Guo-temelli ve MTK-recursive prosedürüne dayalı olmak üzere üç farklı sınıflama indeksini tanıtmış ve çeşitli koşullarda incelemişlerdir. Bu incelemeler simülasyon veri ve matematik ve okuma olmak üzere iki farklı geniş ölçekli sınav veri setleri üzerinde yapılmıştır. Simülasyon verisi için yetenek dağılımı, test uzunluğu ve farklı kesme puan setleri dikkate alınarak bu faktörlerin manipüle edildiği çeşitli koşullar oluşturulmuştur. Yetenek dağılımı için $N(\bar{X}=0, ss=1)$ ve $N(\bar{X}=0,5, ss=1,25)$ olmak üzere 2 farklı normal dağılım; test uzunluğu için 60 ve 30 olmak üzere iki farklı test uzunluğu ve 3 farklı kesme puan seti dikkate alınmıştır. Kesme puanı için ele alınan birinci kesme puan seti 0'a göre simetrik dağılan $\theta = -0,75, 0,00, 0,75$; ikinci kesme puan seti 0'a göre simetrik dağılmayan $\theta = -0,75, -0,35, 0,75$ puan setleridir. Üçüncü kesme puan seti ise her bir test uzunluğu için ayrı ayrı ele alınmış olup 60 maddelik test için $\theta = -0,827, -0,034, 0,694$ olarak ve 30 maddelik test için ise $\theta = -0,745, -0,042, 0,706$ olarak ele alınmıştır. Yetenek kestirimi için MTK gerçek puan (true score), MLE, MAP ve EAP kestirim yöntemleri kullanılmıştır. MTK gerçek puan kestirimi için madde parametreleri BILOG-MG programıyla kestirilmiş ve her bir doğru sayı puanına (number-correct) Newton-Raphson prosedürü uygulanarak bireylerin yetenekleri kestirilmiştir. Diğer üç kestirim yöntemi için de BILOG-MG programından faydalanılmıştır. Her bir indeksin performansını değerlendirmek için sınıflama doğruluğu ve tutarlıkları R yazılımı kullanılarak hem simüle edilmiş hem de kestirilmiştir. Sınıflama doğruluklarını simüle etmek için üretilmiş yetenek dağılımları gerçek dağılım olarak ele alınmış ve gerçek dağılımların varsaydığı yanıt örüntülerine dayalı olarak yetenek kestirimleri elde edilip buradan elde edilen dağılım ise gözlenen dağılım olarak ele alınmıştır. Bu iki dağılıma kesme puanları uygulanarak simüle edilmiş sınıflama doğrulukları hesaplanmıştır. Simüle edilmiş sınıflama tutarlılıklarının hesaplanması için ise gerçek yetenek kestirimlerine dayalı olarak iki farklı grup için madde örüntüleri oluşturulmuş ve bu iki farklı veri setinden kestirilen yetenek dağılımları gözlenen dağılımlar olarak ele alınmış ve sınıflama tutarlılıkları hesaplanmıştır. Kestirilen sınıflama doğruluğu ve tutarlılığı değerlerini hesaplamak için R yazılım ortamında Rudner-

temelli indeksler, Guo-temelli indeksler ve MTK-recursive-temelli indeksler kestirilen madde ve yetenek parametrelerine uygulanarak elde edilmiştir. Bu doğrultuda her bir koşul için hem simüle edilen hem kestirilen sınıflama doğruluğu ve tutarlılığı olmak üzere 4 tane indeks elde edilmiştir. Çalışmadan elde edilen sonuçlar incelendiğinde Guo-temelli indekslerin en yüksek olma eğiliminde olduğu ancak Rudner-temelli indeksler ile yakın olduğu birçok koşulda ise aynı indeks değerleriyle sonuçlandığı gözlenmiştir. Ayrıca Rudner-temelli indeks koşulları detaylı incelendiğinde dağılımın $N(0,1)$ ve test uzunluğunun 60 olduğu bütün kesme puan setleri koşulunda EAP ve MAP değerlerine dayalı elde edilen indekslerin daha yüksek olduğu gözlenmiştir. Dağılımın $N(0,5, 1,25)$ ve test uzunluğunun 60 olduğu ve dağılımın $N(0,1)$ ve test uzunluğunun 30 olduğu koşullarda ise iki koşul dışındaki bütün koşullarda bayesian yetenek kestiricilerine dayalı elde edilen sınıflama indekslerinin daha yüksek olduğu sonucuna ulaşılmıştır. Ancak dağılımın $N(0,5, 1,25)$ olduğu 30 maddelik test koşullarında kesme puan setlerine bağlı olarak MLE yetenek kestiricisi ile bayesian yetenek kestiricilerine dayalı elde edilen sınıflama indeksleri arasındaki farklılığın MLE'nin lehine artış gösterdiği gözlenmiştir. Bu durum ise özellikle sınıflama tutarlık indekslerinde meydana gelmiştir. Simülasyona dayalı incelemelerin yanı sıra matematik ve okuma olmak üzere iki farklı geniş ölçekli sınav verisi üzerinde incelemeler yapılmıştır. Buradan elde edilen bulgulara göre Guo-temelli ve Rudner-temelli indekslerin en yüksek olduğu ancak simülasyon çalışmasından farklı olarak bazı koşullarda Rudner-temelli indekslerin daha iyi sonuç verdiği gözlenmiştir. MTK-recursive ise simülasyon çalışmasında olduğu gibi en düşük sınıflama doğruluğu ve tutarlıkları ile sonuçlanmıştır. Simülasyon çalışmasından elde edilen bulgulara göre farklılaşan önemli bir nokta ise yetenek kestiricilerine dayalı elde edilen farklılaşmanın yönüdür. Rudner-temelli indeks için simülasyon çalışmasında EAP ve MAP kestiricilerine dayalı hesaplanan indeksler daha iyi sonuç verirken gerçek veriye dayalı çalışmada EAP ve MAP kestiricilerine dayalı elde edilen sınıflama doğruluğu ve tutarlıklarının MTK-gerçek puan kestiriminden ve MLE'den daha düşük olduğu gözlenmiştir.

Lathrop ve Cheng (2014) çalışmalarında parametrik yaklaşımlardan Lee (2010) ve Livingston & Lewis (1995)'in yaklaşımlarını kullanılarak elde edilen sınıflama indeksleri ile nonparametrik yaklaşıma dayalı sınıflama doğruluğu ve tutarlılığı indekslerini çeşitli simülasyon koşullarında test etmişlerdir. Simülasyon koşullarının oluşturulmasında ele alınacak madde tepki kuramı modeli, test uzunluğu ve yetenek dağılımlarına ilişkin faktörler manipüle edilmiştir. Madde tepki kuram modellerinden 3PLM ve ATM

kullanılarak 10, 25 ve 40 maddeli testler oluşturulmuştur. Yeteneklerin elde edilmesinde normal, çarpık ve karma (mixture) dağılım olmak üzere üç farklı dağılım dikkate alınmış ve 1000 kişilik örneklem büyüklüğü sabit olarak göz önünde bulundurulmuştur. Normal dağılım koşulunda yetenek dağılımı için standart normal dağılım; çarpık dağılım koşulunda yetenek dağılımı için ortalaması 0, standart sapması 1, çarpıklığı 0,75 ve basıklığı 0 olan dağılım ele alınmış. Karma koşulda ise yetenekler $N(-0,25, 0,61)$ ve $N(2,19, 1,50)$ olmak üzere iki normal dağılımın karmasından sırasıyla %90 ve %10 oranlarında çekilerek karıştırılmıştır. Kesme puanları örneklemin gözlenen puanlarının %25, %50 ve %75'lik noktalarında yerleştirilmiştir. Bu doğrultuda hem iki kategorili hem çok kategorili (4 düzey) düzeyler göz önünde bulundurularak koşullar test edilmiştir. Bütün simülasyonlar R yazılımı ile gerçekleştirilmiştir. Livingston & Lewis (1995) indekslerinin hesaplanması testin güvenilirliğinin cronbach alfa ile kestirildiği S-plus uygulamasına dayalı olarak yapılmıştır. Parametrik olmayan yaklaşım için ise yine R yazılımından faydalanılmıştır. Her bir koşul için 1000 replikasyon yapılmıştır. Çalışmadan elde edilen bulgulara göre 3PLM altında elde edilen sınıflama doğruluğu indeksleri için en iyi performansı gösteren yaklaşımın parametrik olmayan yaklaşım olduğu ancak Lee'nin yaklaşımıyla çok benzer olduğu sonucuna ulaşılmıştır. Test uzunluğu arttıkça da bu farklılığın gözden kaybolduğu belirtilmiştir. Ancak Livingston & Lewis yaklaşımının diğer yaklaşımlar ile benzer örüntüyü takip etmediği özellikle dağılımın normal olmadığı koşullarda daha yanlış sonuçlar verdiği gözlenmiştir. 3PLM altında elde edilen sınıflama tutarlıkları için de benzer sonuçlar elde edilirken burada Lee'nin yaklaşımıyla elde edilen yanlışlığın daha düşük olduğu belirtilmiştir. ATM altında hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin bulgular, normal dağılım koşulunda üç yaklaşımın benzer sonuç verdiğini ancak çarpık ve karma dağılım koşullarında Lee ve parametrik olmayan yaklaşımların çok iyi sonuç verdiğini göstermiştir. Ayrıca her iki model altında da Livingston & Lewis yaklaşımının karma dağılım koşulunda birçok kesme puan noktasında olduğundan fazla kestirdiği; çarpık dağılım koşulunda ise %25'lik noktada olduğundan fazla kestirirken %75'lik noktada olduğundan az kestirdiği sonucuna ulaşılmıştır. Dolayısıyla Livingston & Lewis yaklaşımının dağılımın normal olmadığı koşullardan etkilendiği ve kesme puanı ile yetenek dağılımı arasındaki etkileşime dayalı olarak yanlış sonuçlar verdiği gözlenmiştir.

Wyse ve Babcock (2016) çalışmalarında sınıflama doğruluğu ve sınıflama tutarlılığının yüksek sonuçlanabileceği test dizaynlarını incelemişlerdir. Çalışmada çeşitli test merkezlerinde bilgisayar aracılığıyla uygulanan Tıbbi Görüntüleme Akreditasyonu programının verisi ile kağıt-kalem ile uygulanan bir K-12 test programının verisi kullanılmıştır. Bu veri setleri için iki düzey söz konusudur. İlk programdan elde edilen verinin kesme puanı için daha önce angoff ile belirlenmiş kesme puanı; kağıt-kalem ile uygulanan programdan elde edilen veri için ise işaretleme yöntemi ile belirlenmiş kesme puanı göz önünde bulundurularak incelemeler yapılmıştır. Her bir veri seti için operasyonel olarak ele alınan kesme puanları ve madde parametreleri sabit tutularak bu veri setlerine dayalı çeşitli simülasyon koşulları ele alınmıştır. Bu doğrultuda kesme puanının yerine göre yetenek dağılımının ortalaması, madde sayısı, çeşitli yetenek (θ) noktalarında sağlanan test bilgi fonksiyonuna ilişkin koşullar manipüle edilerek çeşitli senaryolar oluşturulmuş ve sınıflama doğruluğu ve tutarlılıkları bu senaryolarda incelenmiştir. Çalışmadan elde edilen bulgulara göre kesme puanı yerine başka bir θ noktasında testin bilgisini arttırmanın sınıflama doğruluğu ve sınıflama tutarlılığını arttırma konusunda daha iyi sonuç verdiği sonucuna ulaşılmıştır. Bu durum ise kesme puanının yeri ile yetenek dağılımının etkileşiminin bir sonucu olarak ele alınmıştır. Kısa testler için (< 25) yetenek dağılımının ortalamasının kesme puanının üzerinde olduğu durumda, kesme puanının üzerindeki noktada (location) bilgiyi arttırmak sınıflama performansını (sınıflama doğruluğu ve sınıflama tutarlılığı) arttırmaktadır. Benzer şekilde yetenek dağılımının ortalamasının kesme puanının altında olduğu durumda ise kesme puanının altındaki noktalarda bilgiyi arttırmak sınıflama performansını arttırmaktadır. Ancak test uzunluğu arttıkça bu noktalar kesme puanına yaklaşmakta ve dolayısıyla kesme puanına yakın noktalardaki bilgiyi arttırmak daha iyi sonuç vermektedir. Yetenek dağılımı ortalamasının kesme puanına eşit olduğu durumda ise kesme puanının çok az altındaki noktada bilgiyi arttırmak sınıflama performansını iyileştirmektedir. Genel olarak test uzunluğu arttıkça kesme puanına yakın yerlerdeki bilgiyi arttırmanın sınıflama performansını iyileştirdiği sonucuna ulaşılmıştır. Ayrıca sınıflama doğruluğu ve sınıflama tutarlılıklarına ilişkin maksimum bilginin sağlandığı noktaların farklılaştığı görülmüştür. Özellikle 50 maddenin altındaki testler için sınıflama doğruluğu ve sınıflama tutarlılıkları arasındaki farklılık yüksek olurken; 50 maddenin üzerindeki testlerde bu farklılığın giderek azaldığı gözlenmiştir.

Zhang, Du, Chen, Xin ve Chen (2017) çalışmalarında yetenek dağılımı, kesme puanının sayısı ve yeri ve puan ölçeğinin sınıflama tutarlılığına etkisini simülasyon ve gerçek veri koşullarında incelemişlerdir. Simülasyon çalışması için veri sağa çarpık ($\bar{X}=0$; $SS=1$; basıklık = 0; çarpıklık = 0,8), sola çarpık ($\bar{X}=0$; $SS=1$; basıklık = 0; çarpıklık = -0,8) ve normal dağılım ($\bar{X}=0$; $SS=1$) olmak üzere üç farklı yetenek dağılımı; doğrusal ve doğrusal olmayan olmak üzere iki puan ölçeğine dayalı olarak üretilmiştir. Bu doğrultuda MATLAB 2012 programı kullanılarak 20 tanesi iki kategorili, 4 tanesi ise çok kategorili olmak üzere 24 maddeye dayalı 2000 kişilik bir veri matrisi üretilmiştir. Ayırt edicilik ve güçlük parametrelerinin elde edilmesi için PARSCALE programından faydalanılmıştır. Ayırt edicilik parametresi [0, 2,5] arasında değişen bir normal dağılımdan çekilirken, güçlük parametresi [-3, 3] aralığında değişen normal dağılımdan çekilmiştir. Her bir dağılım için 100 replikasyon yapılmıştır. İncelemelerin yapılması için yetenek ölçeğinin iki farklı ucunda yer alan iki tane kesme puanı ele alınmıştır. Bu kesme puanlarının her birinin ayrı ayrı ve ikisinin birlikte ele alındığı koşullar göz önünde bulundurularak sınıflama tutarlıkları hesaplanmıştır. Çalışmadan elde edilen bulgulara göre sağa çarpık dağılım koşulunda kesme puanının yüksek olması sınıflama tutarlılığını artırırken sola çarpık dağılımda da tam tersi söz konusudur. Ayrıca dağılımın iki ucunda yer alan iki kesme puanının birlikte ele alındığı çoklu düzey koşullarının ele alındığı durumda ise sağa ve sola çarpık dağılımdan elde edilen sınıflama tutarlılıklarının benzer sonuç verdiği ancak normal dağılıma dayalı elde edilen sınıflama tutarlılığının daha düşük olduğu gözlenmiştir. Simülasyon çalışmasının yanı sıra geniş ölçekli 4. ve 5. sınıf düzeyindeki bir sınav verisi üzerinde de incelemeler yapılmıştır. Her iki sınıf düzeyine ilişkin dağılımın da sola çarpık olduğu ve bu dağılıma dayalı elde edilen sınıflama tutarlıklarının simülasyon koşullarındaki bulguları desteklediği görülmüştür. Bu incelemeler aynı zamanda doğrusal olan ve doğrusal olmayan puan ölçekleri de göz önünde bulundurularak yapılmıştır. Elde edilen sonuçlara göre her iki puan ölçeğinde de tutarlı indeks değerlerinin elde edildiği ancak doğrusal olmayan puan ölçeği için yetenek dağılımlarında koşullu sınıflama tutarlık eğrilerinin sola kayma eğiliminde olduğu gözlenmiştir. Buna bağlı olarak da koşullu sınıflama tutarlıklarının düşük yetenek düzeyindeki bireyler için olduğundan az, yüksek yetenek düzeyindeki bireyler için ise olduğundan fazla kestirildiği sonucuna ulaşılmış ancak her iki dönüşümden elde edilen sonuçların da kabul edilebilir düzeyde olduğu belirtilmiştir.

Çalışmalar genel olarak incelendiğinde ideal nokta modellerinin çeşitli alanlarda uygulandığı ve kümülatif modellerle karşılaştırıldığı gözlenmiştir. Literatürde yer alan bu çalışmalardan elde edilen sonuçlar değerlendirildiğinde ideal nokta modellerinin birçok alanda etkili bir şekilde uygulanabileceği gösterilmiştir.

Bunun yanı sıra model-veri uyumu, yetenek kestirimlerinin doğruluğu ve yetenek kestirimleri arasındaki ilişki gibi birçok açıdan yapılan karşılaştırmalı inceleme sonuçları, ideal nokta modellerinin kümülatif modellere göre daha esnek kestirimler sağlayabileceğini ve özellikle yetenek ölçeğinin üst uçlarında modeller arası farklılaşmaların öne çıkabileceğine işaret etmektedir.

Sınıflama indekslerine yönelik literatürde yer alan çalışma sonuçları bireyler hakkında verilecek kararların birçok faktör faktörden etkilenebileceğini göstermiştir. Bunun yanı sıra sınıflama çalışmalarının genel olarak kümülatif modeller üzerinde yürütüldüğü gözlenmiştir. Bu bölümde ele alınan çalışmalara ilişkin sonuçlar genel olarak değerlendirildiğinde ideal nokta modelleri ile kümülatif modellerin karşılaştırıldığı sınıflama çalışmasına ihtiyaç olabileceği düşünülmüştür. Modeller arası farklılaşmaların üst uçlarda olması, kestirim yöntemlerinin etkiliklerinin modele göre değişebilme eğiliminde olması ve sınıflama indekslerinin birden fazla faktörün bileşimine dayanması itibarıyla bu çalışmanın yürütülmesinin alanyazına önemli katkı sunacağı sonucuna ulaşılmıştır.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, veri toplama aracı, araştırmanın evren ve örnekleme ve elde edilen verilerin analizi hakkında bilgiler verilmiştir.

Araştırmanın Modeli

Bu çalışmada ATM, GAMOM ve karma modeller altında WLE, MAP ve EAP kestirim yöntemleri çerçevesinde elde edilen yetenek kestirimlerinin doğruluğu ile yetenek ölçeğinin düşük ve yüksek ucunda yer alan kesme puan noktalarında elde edilen sınıflama indekslerinin incelenmesi amaçlanmıştır. Mevcut faktörler arası etkileşimi ve değişimi gözlemlemek için yetenek kestirimlerinin doğruluğu açısından iki faktörlü model ve yöntem etkileşimleri; sınıflama indeksleri açısından ise üç faktörlü kesme puanı, model ve yöntem etkileşimleri incelenmiştir. Çalışma kapsamında gerçekleştirilen inceleme sonuçlarına dayalı olarak en uygun model ve yöntem seçimlerinin yapılabilmesine yönelik bilgilerin üretilmesinin amaçlanması yönüyle bu araştırma temel araştırma niteliği taşımaktadır. Temel araştırma genellikle kuram olarak ifade edilen hipotezle birlikte alta yatan süreçlerin açıklığa kavuşturulması ile ilgilidir (Fraenkel, Wallen & Hyun, 2012, s. 7).

Veri Toplama Aracı

Bu araştırmada yükseköğrenim kurumlarına devam eden öğrencilerin üniversite yaşam doyum düzeylerini ölçmek amacıyla geliştirilmiş olan Yükseköğrenim Yaşam Doyum Ölçeği (EK-1) kullanılmıştır (Erol & Yıldırım, 2016). Ölçek 29 maddeden oluşmaktadır. Bu maddelerin 10 tanesi olumsuz ifadelerden oluşan maddelerdir. Ölçek likert tipi, beşli

derecelendirmeli (hiç katılmıyorum=1, tamamen katılıyorum=5) olarak geliştirilmiştir. Ölçek 5 alt boyuttan (“genel üniversite doyum”, “öğretim üyelerinden sağlanan doyum”, “üniversite sosyal hizmetlerinden sağlanan doyum”, “üniversite imkânlarından sağlanan doyum”, “üniversite yönetiminden sağlanan doyum”) oluşan ve toplam puanı alınabilen bir ölçektir. Ölçeğin birinci faktörü on üç maddeden (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13), ikinci faktörü beş maddeden (14, 15, 16, 17, 18), üçüncü faktörü dört maddeden (19, 20, 21, 22), dördüncü faktörü dört maddeden (23, 24, 25, 26) ve beşinci faktörü üç maddeden (27, 28, 29) oluşmaktadır. Ölçme aracının yapı geçerliğinin incelenmesinde açımlayıcı faktör analizi ve doğrulayıcı faktör analizi kullanılmıştır. Açımlayıcı faktör analizinde faktör döndürme öncesine ve faktör döndürme sonrasında ilişkin sonuçlar sunulmuştur. Faktör döndürme öncesinde bütün maddelerin tek bir faktöre yüklendiği ve maddelerin faktör yüklerinin 0,31-0,78 aralığında değiştiği gösterilmiştir. Varimaks eksen döndürme tekniği sonrasında birinci faktörün on üç maddeden, ikinci faktörün beş maddeden, üçüncü faktörün dört maddeden, dördüncü faktörün dört maddeden ve beşinci faktörün üç maddeden (26, 27 ve 28) oluştuğu görülmüştür. Ölçekte birinci faktörün tek başına varyansın % 35,71’ini açıkladığı ve açıklanan toplam varyansın ise %58,35 olduğu gösterilmiştir. Çizgi grafiğindeki birinci çizgiden sonraki hızlı düşüş, faktör varyansına ilişkin değerler ve birinci faktördeki yük değerleri birlikte değerlendirilerek ölçeğin beş faktörlü olmasının yanı sıra aynı zamanda genel bir faktöre de sahip olduğu belirtilmiştir. Yapı geçerliği incelemesinin yanı sıra güvenirlik incelemesi için cronbach alfa ve test-tekrar test güvenirliği ile incelemeler yapılmıştır. Ölçeğin tamamı için Cronbach alfa güvenirlik katsayısı 0,93, birinci faktör için 0,92, ikinci faktör için 0,81, üçüncü faktör için 0,73, dördüncü faktör 0,69, beşinci faktör için 0,81 olarak hesaplanmıştır. Test-tekrar test güvenirliği için 118 kişilik bir gruba iki hafta ara ile uygulama yapılmıştır. İki uygulamadan elde edilen puanlar arasındaki pearson momentler korelasyon katsayısının ölçeğin tamamı için $r_{xx}=0,996$ ($p < 0,01$), birinci faktör için $r_{xx}=0,91$ ($p < 0,01$), ikinci faktör için $r_{xx}=0,74$ ($p < 0,01$), üçüncü faktör için $r_{xx}=0,74$ ($p < 0,01$), dördüncü faktör için $r_{xx}=0,84$ ($p < 0,01$), beşinci faktör için $r_{xx}=0,75$ ($p < 0,01$) olduğu gösterilmiştir.

Toplam puanı alınabilen bu ölçek çalışmada tek boyutlu olarak ele alınmıştır. Bu araştırmada tek boyutlu yapıya ilişkin inceleme için MTK çerçevesinde polikorik korelasyon matrisine dayalı faktör analizi yapılmış olup sonuçlar tek boyutluluk varsayımı altında sunulmuştur. Bunun yanı sıra her bir boyut için Cronbach alfa değerleri elde

edilmiş olup ölçeğin geneline ilişkin tabakalı Cronbach alfa hesaplanmıştır. Bu çalışma verisine ilişkin cronbach alfa değerleri sırasıyla birinci faktör için 0,932, ikinci faktör için 0,805, üçüncü faktör için 0,669, dördüncü faktör için 0,527 ve beşinci faktör için 0,785 olarak elde edilmiştir. Ölçeğin geneline ilişkin hesaplanan tabakalı cronbach alfa değerinin 0,950 olduğu gözlenmiştir. Bu doğrultuda ölçeğin güvenirliğine ilişkin ölçümlerin hem alt boyutlar açısından hem de tekboyutlu olarak kabul edilebilir sınırlarda yer aldığı sonucuna ulaşılmıştır (Salvucci, Walter, Conley, Fink, & Saba, 1997, s. 115)

Evren ve Örneklem

Çalışma verisi Gazi Üniversitesinde öğrenim gören öğretmen adaylarından elde edilmiştir. Veri toplama süreci için Gazi Üniversitesi Etik Komisyonundan etik izin (EK-2) ve Gazi Üniversitesi Eğitim Fakültesinden uygulama izni (EK-3) alınmıştır.

Çalışmanın evrenini 2019-2020 Bahar döneminde Gazi Üniversitesi Eğitim Fakültesinde Öğrenim gören öğretmen adayları oluşturmaktadır. Veri toplama sürecinde uygun örnekleme yöntemi izlenilmiştir. Bu doğrultuda çalışmanın örneklemini 2019-2020 Bahar döneminde Gazi Üniversitesi Eğitim Fakültesinde çeşitli bölümlerde öğrenim gören 1682 öğretmen adayı oluşturmakta olup örnekleme ilişkin bölüm, frekans ve yüzde bilgileri Tablo 1’de sunulmuştur:

Tablo 1

Örnekleme Yeri Alan Öğretmen Adaylarına İlişkin Bölüm Bilgileri

Bölüm	Frekans	Yüzde (%)
Psikolojik Danışmanlık ve Rehberlik	188	11,18
Okul Öncesi Öğretmenliği	149	8,86
Sınıf Öğretmenliği	94	5,59
İlköğretim Matematik Öğretmenliği	147	8,74
Türkçe Öğretmenliği	69	4,10
Sosyal Bilgiler Öğretmenliği	88	5,23
Fen Bilgisi Öğretmenliği	35	2,08
İngilizce Öğretmenliği	175	10,40
Özel Eğitim	137	8,15
Ortaöğretim Matematik Öğretmenliği	11	,65
Fizik Öğretmenliği	21	1,25
Kimya Öğretmenliği	25	1,49
Müzik Öğretmenliği	58	3,45
Resim Öğretmenliği	40	2,38
Bilgisayar ve Öğretim Teknolojileri Eğitimi	50	2,97
Felsefe Grubu Öğretmenliği	29	1,72
Tarih Öğretmenliği	10	,59
Türk Dili ve Edebiyatı	11	,65
Coğrafya Öğretmenliği	15	,89
Arapça Öğretmenliği	37	2,20
Fransızca Öğretmenliği	22	1,31
Almanca Öğretmenliği	30	1,78

Verilerin Analizi

Bu çalışmada sınıflama performansı kümülatif, ideal nokta ve karma modeller altında bayesian ve nonbayesian yöntemler çerçevesinde kesme puan noktasının yetenek ölçeği üzerindeki lokasyonuna bağlı düşük ve yüksek uçta yer alan üç farklı kesme puan çifti üzerinde incelenmiştir. Çalışmada kümülatif modellerden ATM; ideal nokta modellerden ise GAMOM kullanılmıştır. İncelemeler tek boyutlu model üzerinde gerçekleştirilmiştir. Model kestirimleri için mirt R paketinden faydalanılmış olup EM algoritması kullanılarak model kestirimleri gerçekleştirilmiştir. ATM (EK-4) ve GAMOM (EK-5) model kestirimleri doğrudan gerçekleştirilirken, karma model için GAMOM altında elde edilen delta parametrelerine dayalı olarak maddelerin monotonik - monotonik olmayan spesifikasyonu ile karma model tanımlaması yapılmıştır.

Karma Model Tanımlaması

Karma model tanımlaması için monotonik olma-monotonik olmama durumları açısından maddelerin davranışlarını ortaya çıkarmak gerekmektedir. Maddelerin monotonik olmayan madde olarak değerlendirildiği çalışmalar incelendiğinde, delta parametrelerine göre çeşitli sınırlar gözetilerek madde tanımlamalarının yapıldığı gözlenmiştir. Roberts vd. (2002), Cao, Song ve Tay (2018) ve Fan (2020) çalışmalarında GAMOM'a dayalı gerçekleştirdikleri veri simülasyonunda delta parametreleri için $[-2, 2]$ aralığını ele almış ve parametreler bu aralık içerisinde yayılacak şekilde veri üretilmiştir. Broadfoat (2008) gerçek veri üzerinde çeşitli psikolojik ölçekler üzerinde çalışmış ve ölçeklerde yer alan maddeleri kalibre ederek 2'den küçük mutlak delta parametresine sahip maddeleri monotonik olmayan madde olarak tanımlamıştır. Stark vd. (2006) ve Tay vd. (2009) çalışmalarında kişilik ölçümlerine yönelik ölçme araçlarındaki maddelerin yanıt süreçlerini değerlendirirken, delta parametreleri $[-3, 3]$ aralığında yer alan maddeleri monotonik olmayan maddeler olarak tanımlamışlardır. Ling vd. (2016) ise monotonik olma-monotonik olmama açısından maddeleri sırasıyla nispeten (relatively) nötr, orta düzeyde (moderately) uç ve çok uç olacak şekilde adlandırmış, buna dayalı olarak da $|\delta| < 2$ olan maddeler oldukça nötr; $|\delta| < 4$ olan maddeler orta düzeyde uç ve $|\delta| > 4$ olan maddeler ise çok uç maddeler olarak değerlendirilebileceğini belirtmişlerdir. Maddelerin yanıt süreçlerinde sergiledikleri davranışın doğru bir şekilde tespit edilip madde spesifikasyonunun yapılması doğru karma modelin oluşturulabilmesine katkı sağlayabilmektedir. Maddelerin yanıt süreçlerinde sergiledikleri davranışın doğru bir şekilde tespit edilmesi ile madde spesifikasyonu yapılarak uygun bir karma model tanımlanabilmektedir. GAMOM'a dayalı yapılan kestirimlerin yer aldığı çalışmalar değerlendirildiğinde, yaygın olarak $[-2, 2]$ aralığında delta parametrelerine sahip maddelerin monotonik olmayan madde olarak değerlendirildiği gözlenmiş olup bu araştırmada da $[-2, 2]$ aralık referans alınmıştır. Bu doğrultuda GAMOM altında maddelerin lokasyonunu gösteren delta parametreleri kestirilmiş ve delta parametreleri $[-2, 2]$ aralığında yer alan maddeler monotonik olmayan; bu aralığın dışındaki maddeler ise monotonik olarak ele alınmıştır. Bu doğrultuda kümülatif ve ideal nokta modeller bağlamında monotonik olan- monotonik olmayan madde spesifikasyonları yapılarak Tablo 2'de sunulmuştur.

Tablo 2

Karma Model Madde Spesifikasyonu

Maddeler	Delta	Karma	Maddeler	Delta	Karma
M1	1,234	GAMOM	M16	1,362	GAMOM
M2	1,289	GAMOM	M17	-5,454	ATM
M3	1,119	GAMOM	M18	1,299	GAMOM
M4	1,121	GAMOM	M19	-3,844	ATM
M5	1,117	GAMOM	M20	1,263	GAMOM
M6	-3,652	ATM	M21	1,247	GAMOM
M7	1,306	GAMOM	M22	-7,001	ATM
M8	1,372	GAMOM	M23	-4,887	ATM
M9	-4,467	ATM	M24	-5,700	ATM
M10	-2,503	ATM	M25	1,065	GAMOM
M11	-3,215	ATM	M26	-4,801	ATM
M12	1,352	GAMOM	M27	1,069	GAMOM
M13	1,330	GAMOM	M28	1,129	GAMOM
M14	1,297	GAMOM	M29	1,115	GAMOM
M15	1,292	GAMOM			

Tablo 2 incelendiğinde, delta parametresine göre 10 maddenin monotonik, 19 maddenin ise monotonik olmayan madde olarak ele alındığı görülmektedir. Bu doğrultuda delta parametrelerine dayalı olarak karma model tanımlaması yapılmış olup model kestirimleri mirt R paketi kullanılarak gerçekleştirilmiştir.

Varsayımların Test Edilmesi***Tek Boyutluluk***

Tek boyutlu madde tepki kuramı varsayımlarından olan tek boyutluluk varsayımının incelenmesi için R psych paketinden faydalanılmıştır. En küçük hatalar (minimum residuals) yöntemi kullanılarak açımlayıcı faktör analizi ile tek boyutlu yapı test edilmiştir. Yapının çıkarılmasında Pearson, tetrakorik ve polikorik olmak üzere çeşitli korelasyon matrislerine dayalı olarak faktör analizi gerçekleştirilebilmektedir. MTK analizlerinde çok kategorili veri yapılarının incelenmesinde polikorik yapıya dayalı faktör analizinin

yapılması önerilmektedir (Revelle, 2021). Polikorik korelasyon matrisine dayalı açımlayıcı faktör analizi ile madde tepki analizi yapılmış olup standartlaştırılmış faktör yükleri ve ortak yükler Tablo 3'te sunulmuştur:

Tablo 3

Faktör Yükleri ve Ortak Yükler

Maddeler	Faktör yükü	Ortak yükler	Maddeler	Faktör yükü	Ortak yükler
M1	0,78	0,62	M16	0,71	0,51
M2	0,73	0,54	M17	0,45	0,20
M3	0,80	0,65	M18	0,72	0,52
M4	0,78	0,60	M19	0,52	0,27
M5	0,79	0,63	M20	0,70	0,49
M6	0,59	0,34	M21	0,73	0,53
M7	0,77	0,59	M22	0,30	0,09
M8	0,86	0,75	M23	0,40	0,16
M9	0,49	0,24	M24	0,39	0,15
M10	0,61	0,38	M25	0,46	0,21
M11	0,63	0,40	M26	0,37	0,14
M12	0,82	0,68	M27	0,50	0,25
M13	0,85	0,71	M28	0,57	0,32
M14	0,71	0,51	M29	0,65	0,42
M15	0,71	0,50			
Açıklanan varyans	%43				

Tablo 3 incelendiğinde, faktör yüklerinin 0,30-0,86 aralığında değiştiği ve açıklanan varyans oranının %43 olduğu gözlenmiştir. Bu doğrultuda veri yapısının tek boyutlu olarak incelenmesinin uygun olabileceği sonucuna ulaşılmıştır.

Yerel Bağımsızlık

Yerel bağımsızlık varsayımı için Yen'in Q3 istatistiği ele alınmış olup incelemeler her bir model altında ayrı ayrı yürütülmüştür. Çalışmada kullanılan ölçme aracında 29 madde yer

alması itibariyle her bir model altında olası 406 madde çifti için Yen'in Q3 istatistikleri hesaplanmıştır. Christensen, Makransky ve Horton (2016) çalışmasında Yen'in Q3 istatistiklerinin değerlendirilmesinde sınır puanın örneklem, yanıt kategori sayısı ve madde sayısı gibi çeşitli faktörlerden etkilenebileceğini öne sürerek çalışma amaçları doğrultusunda çeşitli sınır değerlerin mevcudiyetini ancak bütün çalışmalarda ortak kullanılabilecek kesin bir sınır değer olamayacağı yönünde fikir belirtmişlerdir. Psikolojik ölçümlerin söz konusu olduğu çalışmalar incelendiğinde, çok kategorili veri yapılarında madde kalıntıları arasındaki ilişki için 0,5 sınır değerinin dikkate alındığı gözlenmiştir (Davidson, Keating & Eyres, 2004; Ten Klooster, Taal & van de Laar, 2008). Bu araştırmada da 0,5 sınırı gözetilerek Q3 istatistiklerine yönelik incelemeler yapılmıştır. Her bir model altındaki madde çiftlerine ilişkin Q3 istatistikleri değerlendirildiğinde, ATM (EK-6) ve karma model (EK-7) altındaki bütün değerlerin 0,5 sınırının altında olduğu; GAMOM (EK-8) altında ise M5-M10 madde çifti (0,54) hariç bütün değerlerin ilgili sınırın altında kaldığı gözlenmiştir. Maraun ve Rossi (2001) ideal nokta verisinde “ekstra-faktör olayı (extra-factor phenomenon)” durumunu ele alarak bu durumdan kaynaklı olarak gerçekte tek boyutlu ideal nokta verisinin ekstra bir “ideal nokta faktör” daha sağlayabileceğini ele almaktadır. Bu nedenle ideal nokta verisi gerçekte tek boyutlu bir yapıyı ölçmesine rağmen iki boyut varmış gibi bir sonuç ortaya çıkabilmektedir. Bunun nedeni ideal nokta modelinin sahip olduğu kovaryans yapısının iki boyutlu doğrusal model yapısından ayırt edilememesi ile açıklanmıştır. Dolayısıyla M5-M10 arasında yer alan Q3 istatistiğinin “ekstra-faktör olayı” ile alakalı olabileceği ele alınmış olup ilgili maddelerin diğer maddeler ile olan Q3 istatistiklerinin ilgili sınırın altında olduğu ve yerel bağımsızlığın genel olarak sağlandığı görülmüştür. Bu doğrultuda analizlerin yapılmasına karar verilmiştir.

Model – Veri Uyumu

Model-veri uyumuna yönelik inceleme yapmak için ilk olarak $S-\chi^2$ istatistikleri kullanılmıştır. Bu doğrultuda elde edilen $S-\chi^2$ istatistiği, $S-\chi^2$ istatistiğinin serbestlik derecesi, $S-\chi^2/sd$ ve RMSEA değerleri her bir model için ayrı ayrı elde edilmiş olup Tablo 4'te sunulmuştur:

Tablo 4

S- χ^2 İstatistikleri, sd ve RMSEA Değerleri

	ATM				GAMOM				KARMA			
	S- χ^2	sd	S- χ^2 / sd	RMSEA	S- χ^2	sd	S- χ^2 / sd	RMSEA	S- χ^2	sd	S- χ^2 / sd	RMSEA A
M1	197,46	186	1,06	0,006	115,88	105	1,10	0,008	185,82	184	1,01	0,002
M2	209,35	193	1,08	0,007	142,25	115	1,24	0,012	213,16	195	1,09	0,007
M3	232,96	172	1,35	0,015	110,57	98	1,13	0,009	196,82	167	1,18	0,010
M4	253,74	171	1,48	0,017	161,33	99	1,63	0,019	232,73	168	1,39	0,015
M5	225,24	155	1,45	0,016	146,03	101	1,45	0,016	214,27	159	1,35	0,014
M6	334,29	238	1,40	0,016	198,90	138	1,44	0,016	299,80	238	1,26	0,012
M7	220,63	183	1,21	0,011	132,57	116	1,14	0,009	213,74	185	1,16	0,010
M8	190,99	150	1,27	0,013	121,42	96	1,26	0,013	175,52	152	1,15	0,010
M9	533,31	268	1,99	0,024	284,84	154	1,85	0,022	427,82	248	1,73	0,021
M10	332,01	239	1,39	0,015	158,95	132	1,20	0,011	298,28	228	1,31	0,014
M11	254,15	155	1,64	0,020	219,52	139	1,58	0,019	303,69	226	1,34	0,014
M12	176,55	160	1,10	0,008	130,78	104	1,26	0,012	172,93	164	1,05	0,006
M13	194,30	154	1,26	0,012	110,53	96	1,15	0,009	192,59	155	1,24	0,012
M14	223,76	208	1,08	0,007	129,99	124	1,05	0,005	229,25	213	1,08	0,007
M15	241,06	212	1,14	0,009	130,10	127	1,02	0,004	206,41	201	1,03	0,004
M16	251,73	207	1,22	0,011	137,59	126	1,09	0,007	245,68	205	1,20	0,011
M17	419,38	275	1,53	0,018	211,34	144	1,47	0,017	364,13	258	1,41	0,016
M18	246,68	200	1,23	0,012	136,24	124	1,10	0,008	284,75	204	1,40	0,015
M19	426,73	257	1,66	0,020	245,18	163	1,50	0,017	222,08	139	1,60	0,019
M20	260,08	208	1,25	0,012	137,28	122	1,13	0,009	285,00	211	1,35	0,014
M21	238,47	200	1,19	0,011	170,73	115	1,48	0,017	253,35	203	1,25	0,012
M22	388,01	269	1,44	0,016	230,26	173	1,33	0,014	358,21	260	1,38	0,015
M23	337,44	279	1,21	0,011	248,27	174	1,43	0,016	297,67	268	1,11	0,008
M24	373,06	270	1,38	0,015	285,66	188	1,52	0,018	381,49	274	1,39	0,015
M25	261,65	255	1,03	0,004	205,70	153	1,34	0,014	210,61	187	1,13	0,009
M26	351,05	296	1,19	0,011	298,13	188	1,59	0,019	234,80	189	1,24	0,012
M27	346,12	267	1,30	0,013	190,91	153	1,25	0,012	323,78	246	1,32	0,014
M28	264,80	241	1,10	0,008	176,75	141	1,25	0,012	283,27	247	1,15	0,009
M29	286,53	221	1,30	0,013	157,17	124	1,27	0,013	297,83	224	1,33	0,014

Tablo 4 incelendiğinde, maddelere yönelik elde edilen RMSEA değerlerinin ATM altında 0,006-0,024 aralığında; GAMOM altında 0,004-0,022 aralığında ve karma model altında 0,002-0,021 aralığında değiştiği ve genel olarak düşük olduğu gözlenmiştir. Roberts (2016, s. 382) uyumun değerlendirilmesinde $S-\chi^2$ istatistiğini serbestlik derecesine bölerek literatürde yer alan (Drasgow vd., 1995; Chernyshenko vd., 2007) sınır değerine göre üçün altında yer alan maddelerin iyi uyum sergilediklerine yönelik değerlendirmeyi ele almıştır. Bu doğrultuda her bir model altında elde edilen $S-\chi^2/sd$ değerleri hesaplanarak maddelerin uyumu değerlendirilmiştir. Bütün modeller altında ve bütün maddelerin ilgili sınır değerinin oldukça altında değerler ile sonuçlandığı ve model-veri uyumunun bütün modeller altında bütün maddeler için sağlandığı sonucuna ulaşılmıştır.

Madde bazındaki incelemelerin ardından genel model – veri uyumu ele alınmıştır. Genel model-veri uyumunun değerlendirilmesindeki yaklaşımlardan biri model karşılaştırmasıdır (Ranger & Much, 2020). Bu doğrultuda her bir model çiftinin uyum düzeyleri genel uyum açısından karşılaştırmalı olarak ele alınmış olup sonuçlar Tablo 5’te sunulmuştur.

Tablo 5

AIC, AICs, SABIC, BIC, ve loglik Değerleri

	AIC	AICc	SABIC	BIC	loglik
ATM	126226,7	126254,2	126553	127013,7	-62968,33
Karma	126380,4	126416,1	126749,6	127270,6	-63026,21
GAMOM	127239	127279,4	127630,7	128183,4	-63445,51

Tablo 5’te her bir model altında elde edilen AIC, AICs, SABIC, BIC, ve loglik değerleri yer almaktadır ve model çiftlerinin uyum düzeyleri bu istatistikler doğrultusunda karşılaştırmalı olarak ele alınmıştır. Model çiftleri karşılaştırıldığında genel olarak değerlerin birbirine yakın olduğu gözlenmiştir. ATM’nın karma modelden daha iyi uyum sergilediği ancak ATM ve karma modelin birbirine çok yakın uyum değerleri sergilediği, GAMOM’un ise bunları takip ettiği gözlenmiştir.

Model ve Yetenek Kestirimleri

Araştırmada ATM, GAMOM ve karma model olmak üzere üç farklı model kestirimi yapılmıştır. Model kestirimlerinin ardından her bir modele dayalı olarak yetenek kestirimleri elde edilmiştir. Yetenek kestirimlerinin elde edilmesinde nonbayesian

yöntemlerden WLE; bayesian yöntemlerden ise MAP ve EAP kullanılmıştır. Yetenek kestirimleri ve yetenek kestirimlerine ilişkin standart hata değerleri R mirt paketi ile elde edilmiştir.

Sınıflama çalışmasının yürütülmesi için bir diğer önemli nokta kesme puanlarının oluşturulmasıdır. Bu çalışmada kesme puanlarının yetenek ölçeği üzerindeki lokasyonu da incelenecek faktörlerden bir olması itibarıyla birden fazla kesme puan noktaları ele alınmış olup kesme puan noktalarının oluşturulmasına ilişkin işlem detaylı olarak açıklanmıştır.

Kesme Puanlarının Oluşturulması

Sınıflama indekslerinin incelenmesine yönelik çalışmalarda kesme puanlarının oluşturulmasında çeşitli yaklaşımlar ele alınabilmektedir. Çalışmalarda keyfi kesme puan setleri (Wyse & Hao 2012) ele alındığı gibi yetenek puanlarına dayalı yüzdelikler üzerinden kesme puan noktalarının oluşturulması sınıflama çalışmalarında sıklıkla başvurulan yollardan biridir (Lathrop & Cheng, 2014; Wan, Brennan & Lee, 2007; Deng, 2011; Gundula, 2012; Kim & Lee, 2019; Knupp, 2009; LaFond, 2014). Knupp (2009) çalışmasında 25., 50., 80. ve 90. yüzdelikleri kesme puan noktaları olarak ele almış ve her bir kesme puan noktasında ayrı ayrı inceleme yaptığı gibi çok düzeyli incelemeler için kesme puanlarının eşzamanlı olarak dikkate alındığı durumda da karar doğruluklarını incelemiştir. Wan, Brennan ve Lee (2007) çalışmalarında olası maksimum puanın %50, %65 ve %80 noktalarına karşılık gelen değerlerini kesme puan noktaları olarak ele almışlardır. Bu çalışmada kesme puan lokasyonunun etkisini gözlemlemek için her bir kesme puan noktasında ayrı ayrı incelemeler (iki düzeyli) yapılmış ve bunun yanı sıra çok düzeyli inceleme (dört düzey) için üç kesme puan setinin eşzamanlı olarak ele alındığı durumda sınıflama indeksleri hesaplanmıştır. Lathrop ve Cheng (2014) kesme puanlarını eşit mesafeli olacak şekilde örneklemin gözlenen puanlarının %25, %50 ve %75'lik noktalarına yerleştirerek hem iki düzeyli hem de çok düzeyli incelemeler yapmışlardır. Araştırmalar incelendiğinde yüzdeliklere göre çeşitli kesme puan noktalarında incelemelerin gerek iki düzeyli gerek çok düzeyli olacak şekilde gerçekleştirildiği gözlenmiştir.

Araştırmalarda oluşturulan kesme puan noktaları araştırmacıların amacı doğrultusunda söz konusu yöntemlere ilişkin sonuçları gözlemleyebilecek şekilde oluşturulmaktadır. Sınıflama çalışmalarında kesme puanlarının yüzdeliklere göre oluşturulmasında

benimsenen yaklaşımlardan biri sonuçlar arasında sistematik bir ilişki kurulabilmesi adına benzer aralıklarla kesme puan noktalarının oluşturulması veya dağılıma göre kesme puanlarının simetrik bir şekilde yerleştirilmesidir. McMillan (2019) birey seçme çalışmalarında karşıt yaklaşımlar arasında ortak bağlantı kurulmasını, puanların aşırılıklarının (extremeness) önemli ölçme doğurgularına (implications) sahip olması ile açıklamaktadır. Bu doğrultuda kesme puanlarının oluşturulmasında simetrik bir dağılım varsayımı ile 10., 20., ve 30. yüzdelerdeki ölçme sonuçlarının 90., 80., ve 70. yüzdelerdeki koşulları ile ayrı ayrı yansıtılması gerektiğini ele almıştır.

Kesme puanlarının oluşturulmasına yönelik bu tür durumlar araştırmacının incelemeye konu olan yöntem ve tekniklerinin literatüre göre ortaya çıkarılması gereken özelliklerine dayanmaktadır. İdeal nokta modelleri ve kümülatif modellere ilişkin çalışmalar (Cao vd., 2014; Ling vd., 2016; Roberts vd., 1999; Stark vd., 2006) incelendiğinde, modeller arasında yetenek ölçeğinin düzeyine (düşük/yüksek) bağlı olarak farklılaşmaların olduğu görülmektedir. Dolayısıyla model ve yöntemlerin kesme puan noktalarına göre farklılaşma düzeylerini ortaya çıkarmak amacıyla yetenek ölçeğinin üst ve alt kısmındaki yüzdeler dikkate alınarak kesme puanları oluşturulmuştur. İndekslerdeki değişimin seyrini gözlemlemek için düşük ve yüksek olmak üzere her bir düzeyde ardışık olarak üçer tane kesme puanı oluşturularak iki düzeyli incelemeler yapılmıştır. Bu doğrultuda yetenek ölçeğinin düşük ucunda $c(0,10)$, $c(0,15)$ ve $c(0,20)$; yüksek ucunda ise düşük uçtaki kesme puan noktalarına simetrik olacak şekilde $c(0,90)$, $c(0,85)$ ve $c(0,80)$ yüzdelerine karşılık gelen noktalar dikkate alınmıştır. Araştırmada ele alınan her bir kesme puan noktası ATM, GAMOM ve karma modelleri altında ve WLE, MAP ve EAP olmak üzere her bir yöntemden elde edilen yetenek kestirimleri üzerinden elde edilmiştir (EK-9). Araştırmanın amacına dayalı olarak McMillan'ın (2019) kesme puanlarının oluşturulmasında simetrik dağılım varsayımı ve karşıt yaklaşımlar arasında bağlantı kurulmasına ilişkin husus ele alınmıştır. Bu doğrultuda yetenek ölçeğinin düşük ve yüksek ucunda yer alan kesme puanlarını karşılaştırmalı olarak ele almak için sırasıyla $c(0,10)$ ve $c(0,90)$; $c(0,15)$ ve $c(0,85)$; $c(0,20)$ ve $c(0,80)$ kesme puanı çiftleri üzerinde incelemeler yapılmıştır.

Sınıflama İndekslerinin Elde Edilmesi

Sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin elde edilmesi için R cacIRT paketi kullanılmış olup kestirimlerde Rudner'in (2001, 2005) sınıflama yaklaşımından

faydalanılmıştır. Söz konusu indekslerin elde edilmesi için yetenek kestirimleri vektörü, yetenek kestirimlerine ilişkin standart hata vektörü ve kesme puan vektörü olmak üzere üç vektör oluşturulmuştur. Oluşturulan bu üç vektöre dayalı olarak her bir koşul için marjinal ve koşullu sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri elde edilmiştir. Koşullu sınıflama indeksleri ilgili koşul altında her bir birey için ayrı ayrı elde edilen sınıflama indekslerini ifade ederken marjinal sınıflama indeksi yine ilgili koşul altındaki koşullu sınıflama indekslerinin (sınıflama doğruluğu veya sınıflama tutarlılığı) ortalamasını ifade etmektedir. Farklı bir ifadeyle koşullu indeks birey bazında, marjinal indeks ise genel bazdaki sınıflama sonuçlarını göstermektedir. Marjinal indekse göre genel değerlendirme yapılmış; koşullu indeksler kullanılarak da yöntem, model ve kesme puan noktaları arasındaki istatistiksel farklılıklar test edilmiştir.

Farklılıkların Test Edilmesi

Çalışmanın amacı ATM, GAMOM ve karma modellerin sınıflama indekslerine olan etkisini bayesian ve nonbayesian yöntemler çerçevesinde, farklı kesme puan noktalarında incelemektir. İlk olarak sınıflama sonuçlarını ve dolayısıyla sınıflama indekslerinin değerlendirilmesine ilişkin yorumları etkileyen yetenek kestirimlerinin doğruluğuna ilişkin manidarlık test edilmiştir. Bu aşamada model türü ve kestirim yöntemi olmak üzere iki faktör söz konusudur. Bireylerin ise her bir faktör altında tekrarlı olarak elde edilmiş ölçümleri mevcuttur. Birden fazla faktörün ele alındığı ve her bir faktör düzeyinde birey ölçümlerinin söz konusu olduğu ve faktöryel olarak birleştirildiği araştırma desenine yönelik incelemeler için iki veya üç yönlü bireyler-içi (within-subjects) desen kullanılabilmektedir (Gamst, Meyer & Guarino, 2008, s. 287 & s. 325; Maxwell & Delaney, 2004, s. 531). Dolayısıyla yetenek kestirimlerinin doğruluğu ATM, GAMOM ve karma modeller altında bayesian ve nonbayesian yöntemler çerçevesinde iki faktörlü tekrarlı ölçümler ANOVA ile incelenmiştir. Bunun ardından birey bazında elde edilen koşullu sınıflama indeksleri üzerinden sınıflama indekslerine etki eden faktörler ele alınarak manidarlık test edilmiştir. Sınıflama indekslerinin oluşturulmasındaki sürece yöntem ve model faktörlerinin yanı sıra kesme puan faktörü dâhil edilmiştir. Dolayısıyla sınıflama indekslerinin ele alındığı aşamada model, yöntem ve kesme puanı olmak üzere üç faktör söz konusudur. Her bir bireyin yine her bir koşul altındaki sınıflama indeksi tekrarlı olarak elde edildiği için üç faktörlü tekrarlı ölçümler ANOVA ile istatistiksel manidarlık test edilmiştir. Faktöryel tekrarlı ölçümler ANOVA desende faktörler

araştırmanın amacı doğrultusunda odak (focal) faktör ve moderatör faktör olarak analize dâhil edilerek incelemeler yapılabilmektedir (Laerd, 2015). Bu çalışmada araştırmanın amacı doğrultusunda üç faktörlü ANOVA’da model ve yöntem değişkenleri odak (focal) faktör olarak ayrı ayrı ele alınmış olup kesme puanı faktörü ise moderatör değişken olarak analize dâhil edilerek incelemeler yapılmıştır. Üç faktörlü etkileşim etkilerinin anlamlı olduğu durumda kesme puanının her bir düzeyinde yöntem ve modelin etkileşim etkileri ve ikili karşılaştırma testleri ile yöntem ve model açısından incelemeler yapılmıştır. Üç yönlü etkileşim etkilerinin anlamlı olmadığı durumda ise elde edilen iki yönlü etkileşim etkileri ve buna yönelik ikili karşılaştırma testlerinden elde edilen sonuçlar incelenmiştir. İki faktörlü etkileşim etkisinin anlamlı olmadığı durumda ise sadece ana etkiler yorumlanmıştır. Analizlerin yürütülmesinde SPSS paket programından faydalanılmış olup sekmelerin yanı sıra syntaxlar yazılarak sonuçlar elde edilmiştir.

Tekrarlı ölçümler ANOVA analizlerinin yürütülmesinde söz konusu olan küresellik varsayımı ve normallik varsayımına yönelik incelemeler yapılmıştır. Küresellik varsayımı için Mauchly’nin küresellik testi ele alınmış olup sonuçlar incelendiğinde bütün koşullarda testin istatistiksel olarak anlamlı çıktığı dolayısıyla bu teste göre küresellik varsayımının sağlanmadığı sonucuna ulaşılmaktadır. Ancak Keselman, Rogan, Mendoza ve Breen (1980) Mauchly’nin küresellik testinin çok küçük sapmalara bile duyarlı olduğunu dolayısıyla küresellik ihlalinin belirlenmesinde zayıf bir yöntem olduğunu çalışmalarındaki ampirik incelemeler ile göstermişlerdir. Ampirik incelemeler doğrultusunda testin küçük örneklemelerde küresellikten sapmayı belirlemede başarısız olduğu gibi büyük örneklemelerde ise küresellikten sapma konusunda aşırı tespit (over-detecting) yaptığı ele alınmıştır. Dolayısıyla bu çalışmadaki örneklemin büyük olmasından kaynaklı olarak küresellik varsayımına ilişkin anlamlılık testinin manidar çıkması beklenen bir durumdur. Maxvell ve Delaney (2004, s. 609) küresellik varsayımının ihlal edilmesi durumunda küresellikten olası sapma ihtimali ve tekrarlı ölçümler ANOVA’nın küresellikten sapmaya karşı olası hassasiyetini dikkate almak için Geisser-Greenhouse düzeltmesi, Box’s ϵ düzeltmesi veya Huynh-Felt düzeltmelerinden herhangi birinin kullanılabileceğini belirtilmiştir. Bu doğrultuda Geisser-Green house düzeltmesi kullanılarak anlamlılık sonuçları yorumlanmıştır.

Çalışmadaki örneklemin büyük olması itibarıyla normallikten sapmaların sonuçlar üzerinde hissedilir derecede etkili olmadığı ele alınmaktadır (Waternaux, 1976). Kirk (2008, s. 411) normallikten sapma durumlarında bile ANOVA’nın oldukça güçlü bir teknik

olduğunu ele almıştır. Kirk (2013, s. 111) bir genelleştirilmiş doğrusal modeli kullanmak için araştırmacının yanıt değişkeni ile bir bağlantı fonksiyonu arasında dağılımın belirlendiğini ve bu dağılımın üstel fonksiyon ailesinden herhangi bir üye olabileceğini belirtmiştir. Bu dağılım ailesinin normal, poisson, gamma ve üstel dağılımları içerdiği ele alınmış ve genelleştirilmiş doğrusal modelin normallik varsayımının üstesinden gelebileceğini çeşitli dağılımlara uygulanabilirliği bağlamında öne sürmüştür. Laerd (2015) merkezi limit teoreminden dolayı çok non-normallik durumlarında bile tekrarlı ölçümler ANOVA'nın geçerli sonuçlar sağlayacağını öne sürmüştür. Bunun yanı sıra tekrarlı ölçümler ANOVA desenlerinde dağılımın normalliğinden ziyade her bir hücredeki dağılımların birbirleriyle benzer olmasının önemli olduğunu ele almışlardır. Bu doğrultuda bütün hücrelerdeki dağılım benzer şekilde çarpık ise (hepsi yaklaşık negatif çarpık veya pozitif çarpık gibi) bu durumun farklı dağılım kombinasyonlarına sahip durum kadar zor olmadığını öne sürmüşlerdir. Bu çalışmada sınıflama indekslerinin kesme puan noktasının seçilmesine bağlı olarak bütün hücrelerinin dağılımın benzer olması tekrarlı ölçümler ANOVA'nın uygulanabilirliğinin bir göstergesidir. Bunu yanı sıra örneklemin de büyük olması geçerli sonuçlar elde edilebileceğine işaret etmektedir. Bütün bu bilgiler ışığında tekrarlı ölçümler ANOVA deseninin kullanılabileceği sonucuna ulaşılmıştır.

ANOVA sonuçları 0,01 alfa anlamlılık düzeyinde değerlendirilmiştir. F testi manidar olduğunda etki büyüklüğü olarak da ele alınan bağımsız değişkenin bağımlı değişkeni açıklama oranı kestirilerek yorumlanır (Tan, 2016, s. 348). Bu doğrultuda istatistiksel anlamlılık sonuçları pratikteki etki büyüklükleri dikkate alınarak değerlendirilmiştir. Anlamlılık değerlendirmelerinde etki büyüklükleri için tekrarlı ölçümler ANOVA sonuçlarının ele alınmasında kısmi eta-kare değerleri hesaplanmıştır. Kısmi eta-kare sonuçlarının değerlendirilmesinde 0,02 küçük etki, 0,13 orta etki ve 0,26 büyük etki olarak yorumlanmıştır (Cohen, Cohen, West & Aiken, 2003). Etki büyüklüğünün ilgili analiz bağlamında küçük etki sınırının altında olduğu durumda ise pratikteki etkinin yorumlanabilir olmadığı hususu gözetilerek, değerlendirmeler bu doğrultuda yapılmıştır.



BÖLÜM IV

BULGULAR VE YORUM

Bu bölümde, araştırma sürecinde toplanan veriler ile elde edilen bulgu ve yorumlara yer verilmiştir. Her bir alt problem sırasıyla ele alınmıştır.

Birinci Araştırma Sorusuna Yönelik Bulgular

Bu araştırma sorusu kapsamında “ATM, GAMOM ve karma model altında bayesian ve nonbayesian yöntemler ile elde edilen yetenek kestirimlerinin betimsel istatistikleri nasıldır ve yetenek kestirimlerinin doğruluğu anlamlı olarak farklılaşmakta mıdır?” sorusuna yanıt aranmıştır.

Tablo 6’da ATM, GAMOM ve karma model altında nonbayesian yöntemlerden WLE, bayesian yöntemlerden ise MAP ve EAP kestirim yöntemleri ile elde edilen yetenek kestirimlerine ilişkin standart hata ortalamaları, minimum, maksimum ve ranj değerleri yer almaktadır:

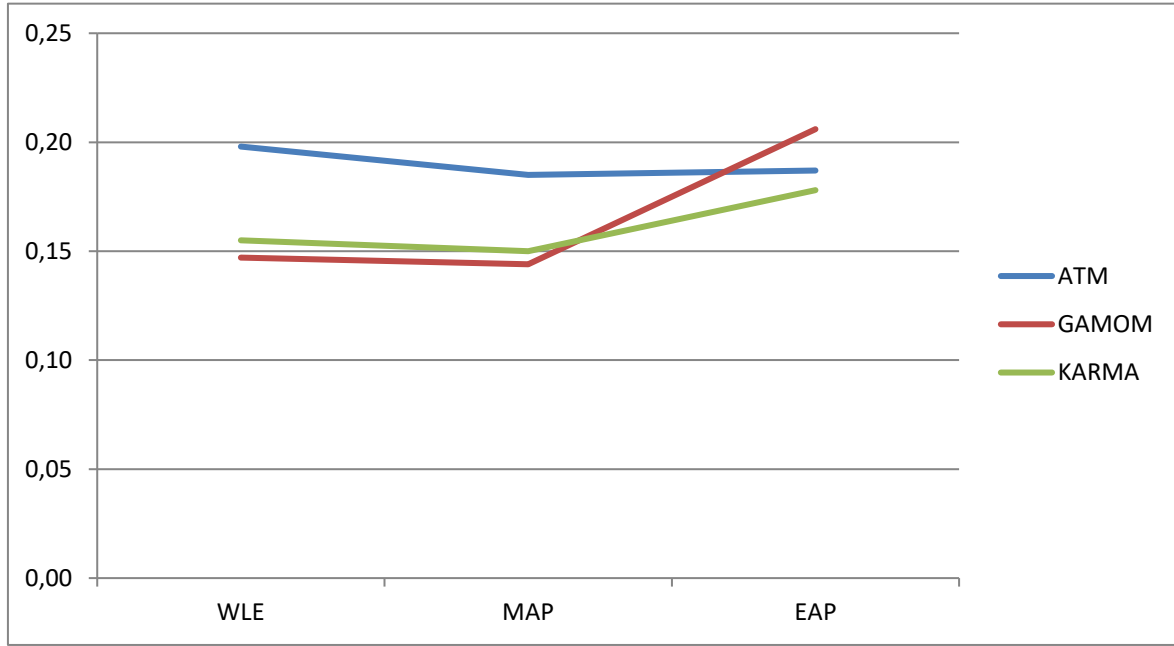
Tablo 6

GAMOM, ATM ve Karma Model Altında Bayesian ve Nonbayesian Yetenek Kestirimlerine İlişkin Betimsel İstatistikler

		WLE	MAP	EAP
ATM	SH _{ort}	0,198	0,185	0,187
	Minimum	-5,571	-3,209	-3,309
	Maksimum	5,773	3,32	3,43
	R	11,345	6,529	6,740
GAMOM	SH _{ort}	0,147	0,144	0,206
	Minimum	-5,564	-3,52	-3,42
	Maksimum	4,917	4,411	4,486
	R	10,481	7,931	7,906
KARMA	SH _{ort}	0,155	0,150	0,178
	Minimum	-4,763	-2,991	-3,099
	Maksimum	5,274	4,74	4,809
	R	10,038	7,731	7,908

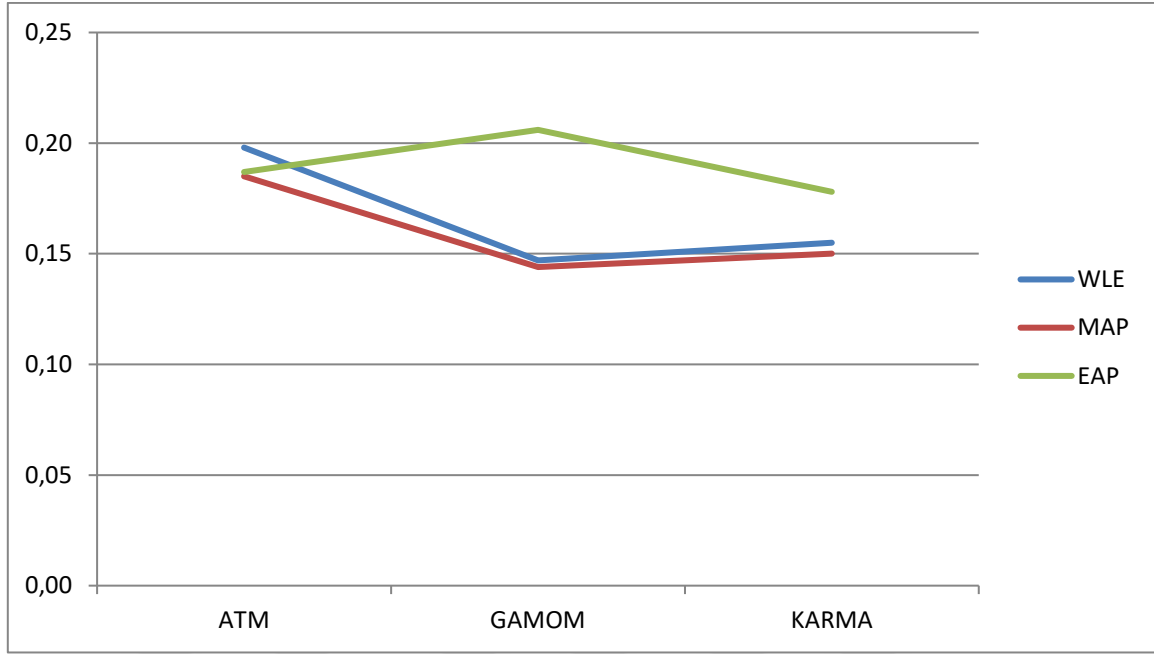
Tablo 6 incelendiğinde, en yüksek standart hata ortalamalarının ATM altında elde edildiği gözlenmiş olup WLE, MAP ve EAP için bu değerler sırasıyla 0,198, 0,185 ve 0,187 şeklindedir. En düşük standart hata ortalama değerlerinin ise GAMOM altında WLE ve MAP kestirim yöntemleri ile edildiği gözlenmiş olup bu değerlerin sırasıyla 0,147 ve 0,144 olduğu görülmektedir. Ancak GAMOM altında EAP'a dayalı elde edilen yetenek kestirimlerinin standart hata ortalamasının 0,206 değeri ile bütün modeller altında ve bütün yöntemler arasında en yüksek standart hata ortalaması olduğu gözlenmiştir. Karma model altında elde edilen standart hata ortalamalarının ise WLE, MAP ve EAP kestirim yöntemleri için sırasıyla 0,155, 0,150 ve 0,178 olduğu gözlenmiş olup her iki modelin standart hata ortalaması arasında değerler alan bu modelin ATM'ye nispeten GAMOM'a daha yakın değerler aldığı da görülmüştür. Ancak GAMOM'da olduğu gibi karma model altında da EAP ile elde edilen yetenek kestirimlerinin standart hata ortalamalarının diğer yöntemlere göre yüksek olduğu gözlenmiştir. Modellere ilişkin standart hata ortalamaları genel olarak değerlendirildiğinde, bayesian yöntemlerden MAP ile elde edilen yetenek kestirimlerinin standart hata ortalamasının bütün modeller altında en düşük olduğu gözlenmiştir. ATM altında bunu EAP takip ederken monotonik olmayan madde

kestirimlerinin söz konusu olduğu GAMOM ve karma model altında MAP'ı WLE yönteminin takip ettiği; EAP ile elde edilen yetenek kestirimlerinin standart hata değerlerinin ise yükseldiği gözlenmiştir. Dolayısıyla modellerin yetenek kestirimlerine ilişkin standart hata ortalamaları açısından kestirim doğruluklarının kullanılan kestirim yöntemine bağlı olarak betimsel düzeyde farklılaştığı gözlenmiştir. Söz konusu farklılığın model ve yöntem açısından farklılaşma etkisini grafiksel olarak gözlemlemek için her bir model altında yöntemlerin değişimi Şekil 2'de ve her bir yöntem altında modellerin değişimi Şekil 3'te sunulmuştur.



Şekil 2. Standart hata ortalamalarına göre modellerin değişimi

Şekil 2 incelendiğinde kestirim yöntemine bağlı olarak farklı modellerden elde edilen standart hata ortalamalarının farklılaştığı gözlenmektedir. WLE ve MAP yöntemleri altında en düşük standart hata ortalamalarının GAMOM ile elde edildiği, bunu karma modelin takip ettiği ve karma model altında elde edilen değerlerin GAMOM'a yakın olduğu görülmektedir. WLE ve MAP altında ATM'nin ise en yüksek standart hata ortalamalarıyla sonuçlandığı görülmektedir. Bununla birlikte ATM'nin EAP ile elde edilen standart hata ortalamalarının WLE ve MAP'a benzer olduğu gözlenmiştir. Ancak GAMOM ve karma modellere yönelik EAP ile yapılan kestirimlerde standart hata ortalamasının yükseldiği görülmektedir.



Şekil 3. Standart hata ortalamalarına göre yöntemlerin değişimi

Şekil 3 incelendiğinde ATM altında neredeyse yöntemlerin benzer sonuç verdiği ancak GAMOM altında yöntemler arası farklılığın en yüksek olduğu gözlenmiştir. Bunun yanı sıra MAP'ın bütün modeller altında en düşük standart hataya sahip olduğu ve yine bütün modeller altında MAP ve WLE'nin değişiminin benzer olduğu gözlenmiştir.

Yetenek kestirimlerinin doğruluğunun değişimine ilişkin farklılığı model ve yöntem açısından birlikte ele almak için iki faktörlü tekrarlı ölçümler ANOVA yürütülmüştür.

Yürütülen ANOVA sonucuna göre kestirim yöntemleri ve model türleri arasında yetenek kestirimlerinin doğruluğu açısından etkileşim etkisinin anlamlı olduğu görülmektedir ($F_{(2,085, 3505,569)} = 144,069$, $p < 0,01$, $\eta_p^2 = 0,079$). Faktörler arasındaki etkileşim etkisine yönelik ana etkiler için model ve yöntemin düzeylerinde fark incelemeleri yapılmıştır. İlk olarak kestirim yönteminin her bir düzeyinde model türlerinin farklılaşma düzeylerini ortaya çıkarmak için WLE, MAP ve EAP olmak üzere her bir kestirim yöntemi düzeyinde modeller arası farklılıklara yönelik incelemeler yapılmıştır.

WLE altında modellerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir ($F_{(1,218, 2046,731)} = 321,483$, $p < 0,01$, $\eta_p^2 = 0,161$). ATM'nin GAMOM'dan 0,051 farklılık %95 GA [0,045, 0,058] ile ($p < 0,01$); karma modelden 0,043 farklılık %95 GA [0,037, 0,048] ile ($p < 0,01$) daha yüksek standart hata ortalamasıyla sonuçlandığı gözlenmiştir. Karma modelin ise GAMOM'dan 0,008 farklılık %95 GA [0,006, 0,011] ile ($p < 0,01$) daha yüksek standart hata ortalamasıyla sonuçlandığı görülmüştür.

MAP altında modellerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir ($F_{(1,107, 1860,369)} = 3010,216$, $p < 0,01$, $\eta_p^2 = 0,642$). ATM'nin GAMOM'dan 0,040 farklılık %95 GA [0,039, 0,042] ile ($p < 0,01$); karma modelden 0,035 farklılık %95 GA [0,033, 0,036] ile ($p < 0,01$) daha yüksek standart hata ortalaması ile sonuçlandığı gözlenmiştir. Karma modelin ise GAMOM'dan 0,006 farklılık %95 GA [0,005, 0,006] ile ($p < 0,01$) daha yüksek standart hata ortalaması ile sonuçlandığı görülmüştür.

EAP altında modellerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir ($F_{(1,559, 2620,972)} = 33,832$, $p < 0,01$, $\eta_p^2 = 0,02$). GAMOM'un ATM'den 0,019 farklılık %95 GA [0,009, 0,029] ile ($p < 0,01$); karma modelden 0,028 farklılık %95 GA [0,020, 0,037] ile ($p < 0,01$) daha yüksek standart hata ortalaması ile sonuçlandığı gözlenmiştir. ATM'nin ise karma modelden 0,009 farklılık %95 GA [0,003, 0,016] ile ($p < 0,01$) daha yüksek standart hata ortalaması ile sonuçlandığı görülmüştür.

Yöntemlerin düzeylerinde yapılan model farklılaşmalarına yönelik incelemelerinin ardından her bir model düzeyinde yöntemler arasında istatistiksel olarak anlamlı farklılaşmanın olup olmadığı değerlendirilmiş olup varsa ikili karşılaştırma testleri ile farklılaşmanın yönü incelenmiştir.

ATM altında yöntemlerin birbirinden istatistiksel olarak farklılaştığı ancak elde edilen etki büyüklüğünün yorumlanabilir olmadığı gözlenmiştir ($F_{(1,001, 1682,133)} = 27,557$, $p < 0,01$, $\eta_p^2 = 0,016$).

GAMOM altında yöntemlerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir ($F_{(1,018, 1711,236)} = 205,371$, $p < 0,01$, $\eta_p^2 = 0,109$). EAP'ın WLE'den 0,059 farklılık %95 GA [0,049, 0,069] ile ($p < 0,01$); MAP'tan 0,062 farklılık %95 GA [0,052, 0,072] ile ($p < 0,01$) ve WLE'nin ise MAP'tan 0,003 farklılık %95 GA [0,002, 0,004] ile ($p < 0,01$) daha yüksek standart hata ortalamasıyla sonuçlandığı gözlenmiştir.

Karma model altında yöntemlerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir ($F_{(1,290, 2169,107)} = 81,614$, $p < 0,01$, $\eta_p^2 = 0,046$). EAP'ın WLE'den 0,022 farklılık %95 GA [0,015, 0,029] ile ($p < 0,01$); MAP'tan 0,027 farklılık %95 GA [0,021, 0,033] ile ($p < 0,01$) ve WLE'nin ise MAP'tan 0,005 farklılık %95 GA [0,002, 0,008] ile ($p < 0,01$) daha yüksek standart hata ortalamasıyla sonuçlandığı gözlenmiştir.

Genel Yorum

Yöntem ve model açısından yetenek kestirimlerinin standart hatalarına dayalı olarak kestirim doğruluğu iki faktörlü tekrarlı ölçümler ANOVA ile incelenmiş olup etkileşim etkisi pratikte düşük düzeyde etki büyüklüğü ile istatistiksel olarak manidar çıkmıştır. Yöntem düzeyinde modeller ve model düzeyinde yöntemler karşılaştırılarak genel bir değerlendirme yapılmıştır. Elde edilen sonuçlara göre WLE ve MAP ile yapılan kestirimlerde modeller arasında GAMOM daha düşük standart hata ile en yüksek doğrulukla sonuçlar vermesine rağmen, EAP yönteminin kullanıldığı durumda GAMOM altında elde edilen yetenek kestirimlerinin standart hata değerlerinin arttığı ve doğruluğun düştüğü gözlenmiştir. Farklı bir ifade ile GAMOM altındaki kestirimlerde EAP kestirim yönteminin diğer yöntemler kadar etkili olmadığı görülmüştür. Genel olarak modeller altında bayesian yöntemlerden MAP daha yüksek doğrulukla sonuçlanmasına rağmen EAP'ın etkililiğinin modele bağlı olarak değiştiği gözlenmiştir. ATM altında MAP'ı EAP takip ederken; GAMOM ve karma model altında MAP'ı WLE'nin takip ettiği; EAP kestirim yönteminin en düşük doğrulukla sonuçlandığı gözlenmiştir. Modeller arasında en yüksek değişim miktarının GAMOM altında; en düşük değişim miktarının ise ATM altında olduğu ve ATM altındaki istatistiksel farklılaşmaya ilişkin etki büyüklüğünün yorumlanabilir olmadığı görülmüştür.

İkinci Araştırma Sorusuna Yönelik Bulgular

Bu araştırma sorusu kapsamında “Yetenek ölçeğinin $c(0,10)$ ve $c(0,90)$ noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında ATM, GAMOM ve karma modellerin sınıflama indekslerine olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?” sorusuna yanıt aranmıştır.

Sınıflama indeksleri sınıflama tutarlılıkları ve sınıflama doğrulukları olmak üzere iki indeks olarak ele alınmış olup her biri için sonuçlar ayrı ayrı sunulmuştur.

Sınıflama Tutarlılıklarına Yönelik Bulgular

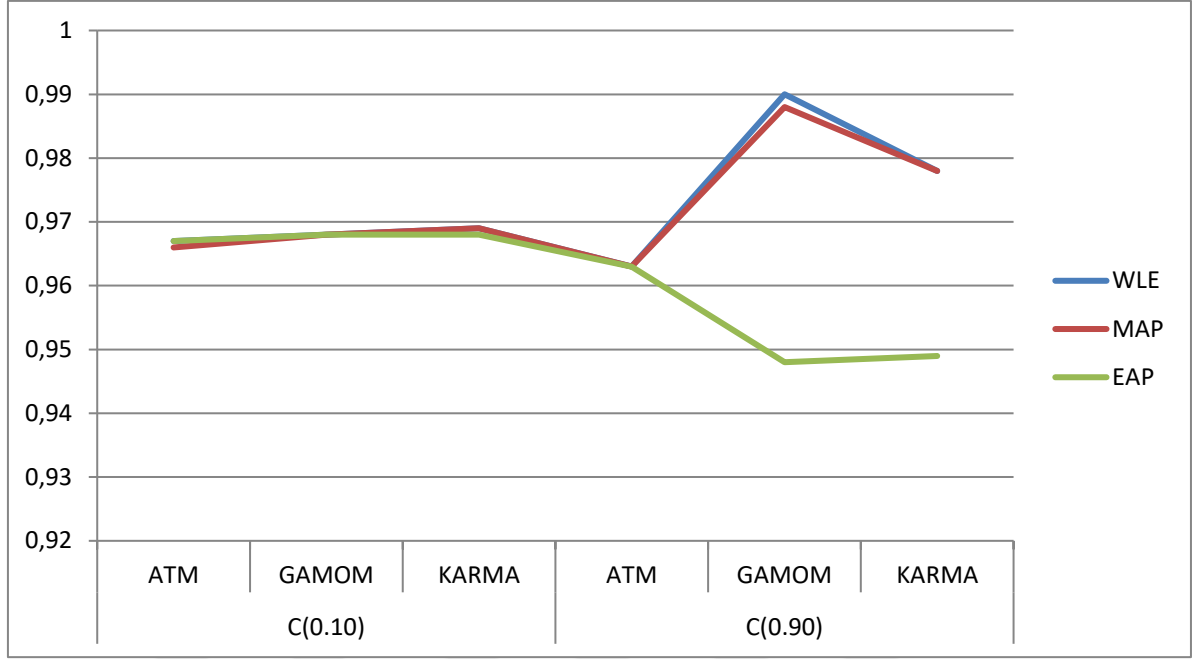
Tablo 7’de yetenek ölçeğinin $c(0,10)$ ve $c(0,90)$ kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilmiş sınıflama tutarlılığı indeksleri yer almaktadır:

Tablo 7

$c(0,10)$ ve $c(0,90)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları

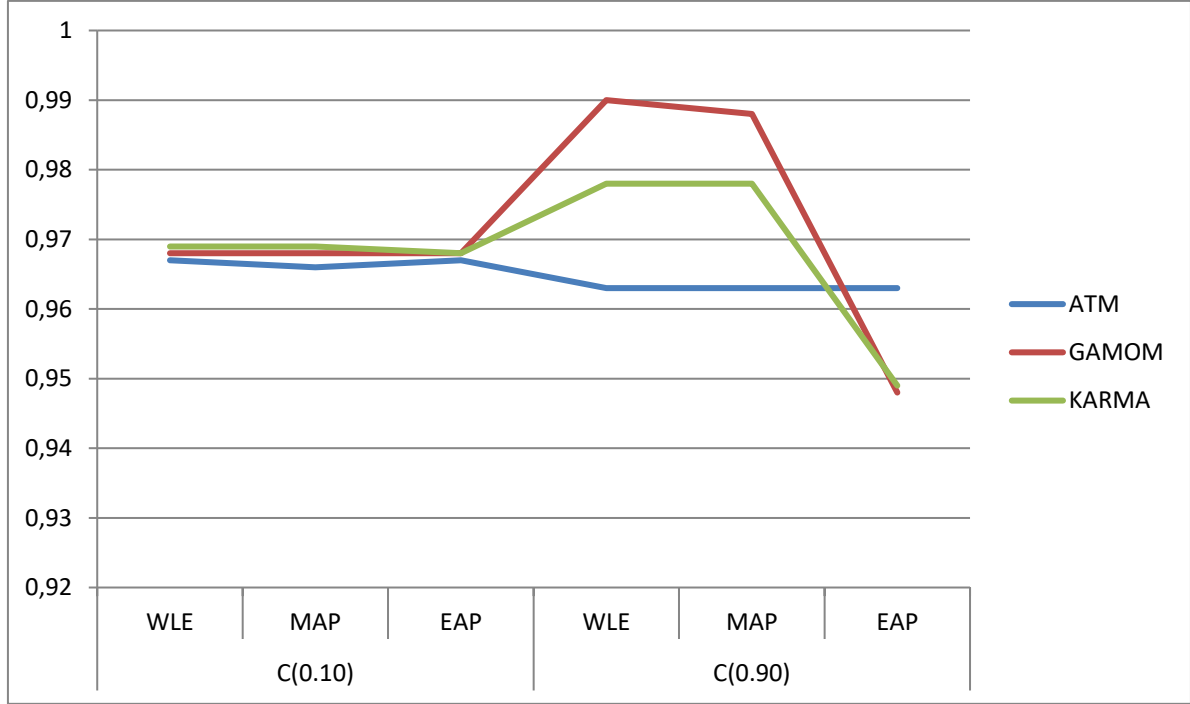
		WLE	MAP	EAP
C(0,10)	ATM	0,967	0,966	0,967
	GAMOM	0,968	0,968	0,968
	KARMA	0,969	0,969	0,968
C(0,90)	ATM	0,963	0,963	0,963
	GAMOM	0,990	0,988	0,948
	KARMA	0,978	0,978	0,949

Tablo 7’deki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama tutarlılığına ilişkin değişimi gözlemlemek için grafikler elde edilmiş olup Şekil 4 ve Şekil 5’te sunulmuştur. Şekil 4’te her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 5’te her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişim grafiği yer almaktadır:



Şekil 4. $c(0,10)$ ve $c(0,90)$ noktalarında yöntemlerin değişimi - $\hat{\gamma}$

Şekil 4 incelendiğinde $c(0,10)$ kesme puan noktasında modeller altında üç yöntemin değişiminin benzer olduğu gözlenmektedir. Yüksek uçta yer alan $c(0,90)$ noktasında yine bütün modeller altında WLE ve MAP değişiminin benzer olduğu; WLE'nin MAP'a nispeten GAMOM altında daha yüksek değerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler birbirine yakın değerlerle sonuçlanmasına rağmen, GAMOM ve karma model altında EAP'ın WLE ve MAP a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 5. c(0,10) ve c(0,90) noktalarında modellerin değişimi - $\hat{y}$

Şekil 5 incelendiğinde c(0,10) kesme puan noktasında yöntemler altında üç modelin de değişiminin benzer olduğu ATM'nin diğer modellere nispeten daha düşük sınıflama tutarlılığı ile sonuçlandığı görülmektedir. Yetenek ölçeğinin yüksek ucunda yer alan c(0,90) kesme puan noktasında ise modellere ilişkin farklılığın ortaya çıktığı; WLE ve MAP altında GAMOM'un en yüksek sınıflama tutarlılığı ile sonuçlandığı gözlenmiştir. GAMOM'u ise karma model takip etmektedir. EAP altındaki kestirimlerde GAMOM ve karma modelde düşüş yaşandığı, ATM altında elde edilen sonuçların ise yöntemlere göre değişiminin belirgin olmadığı gözlenmiştir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama tutarlılığı üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Elde edilen sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu görülmüştür ($F_{(1,885, 3169,488)} = 125,427$, $p < 0,01$, $\eta_p^2 = 0,069$). Bu sonuç yöntem ve model arasındaki iki yönlü etkileşim etkilerinin kesme puan noktasının düzeylerine bağlı olarak farklılaştığına işaret etmektedir. Söz konusu iki yönlü etkileşim etkisinin kesme puan noktasının hangi düzeyinde farklılaştığını ortaya çıkarmak için iki düzeyli olan kesme puan noktasının her bir düzeyinde yöntem ve model arasında iki yönlü tekrarlı ölçümler ANOVA analizleri yürütülmüştür.

Kesme puan değişkeninin c(0,10) noktasında yöntem ve model değişkenleri için yürütülen iki faktörlü tekrarlı ölçümler ANOVA sonuçlarına ilişkin etkileşim etkisinin anlamlı olduğu ($F_{(2,288,3845,912)} = 5,143, p < 0,01, \eta_p^2 = 0,003$) ancak etki büyüklüğünün pratikte yorumlanabilir etkiye sahip olmadığı görülmüş olup söz konusu etkileşim etkisine yönelik anlamlılık yorumlanmamıştır. Bununla birlikte c(0,10) kesme puan noktasında hem yöntem değişkeni açısından ($F_{(1,315, 2210,217)} = 4,251, p > 0,01$) hem de model türü değişkeni açısından ($F_{(1,029, 1730,347)} = 1,617, p > 0,01$), istatistiksel olarak farklılaşmanın olmadığı gözlenmiştir.

Kesme puanı değişkeninin c(0,90) düzeyinde yöntem ve model arasında yürütülen iki yönlü tekrarlı ölçümler ANOVA sonucunun ise anlamlı çıktığı gözlenmiştir ($F_{(1,823, 3063,683)} = 132,325, p < 0,01, \eta_p^2 = 0,073$). Yöntem ve model arasındaki etkileşim etkisine ilişkin ana etkiler hem model hem de yöntem düzeyinde incelenmiştir.

İlk olarak model düzeyinde incelemeler yapılmış olup ATM ($F_{(1,099, 1847,900)} = 41,782, p < 0,01, \eta_p^2 = 0,024$) , GAMOM ($F_{(1,028, 1727,304)} = 61,269, p < 0,01, \eta_p^2 = 0,111$) ve karma model ($F_{(1,033, 1710,305)} = 28,082, p < 0,01, \eta_p^2 = 0,060$) olmak üzere model türünün bütün düzeylerinde yöntemlerin istatistiksel olarak farklılaştığı sonucuna ulaşılmıştır.

ATM düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin MAP'tan 0,001 farklılık %95 GA [0,001, 0,001] ile ($p < 0,01$); WLE'nin EAP'tan 0,001 farklılık %95 GA[0,000, 0,001] ile ($p < 0,01$); ve EAP'ın MAP'tan 0,0003 farklılık GA [0,0002, 0,0003] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

GAMOM düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin MAP'tan 0,002 farklılık %95 GA [0,001, 0,003] ile ($p < 0,01$); WLE'nin EAP'tan 0,042 farklılık %95 GA [0,035, 0,049] ile ($p < 0,01$); ve MAP'ın EAP'tan 0,040 farklılık GA [0,034, 0,047] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Karma model düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin EAP'tan 0,029 farklılık %95 GA [0,022, 0,036] ile ($p < 0,01$); ve MAP'ın EAP'tan 0,029 farklılık GA [0,022, 0,036] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği WLE ile MAP arasında ise istatistiksel olarak anlamlı farklılık olmadığı gözlenmiştir.

Model düzeyinde yöntemler karşılaştırıldıktan sonra yöntem düzeyinde modeller karşılaştırılmıştır. Yöntemin WLE düzeyinde, $(F_{(1,831, 3078,072)} = 52,620, p < 0,01, \eta_p^2 = 0,030)$, MAP düzeyinde $(F_{(1,794, 3015,782)} = 48,891, p < 0,01, \eta_p^2 = 0,028)$, EAP düzeyinde $(F_{(1,200, 2017,118)} = 20,701, p < 0,01, \eta_p^2 = 0,012)$ modellerin birbirinden istatistiksel olarak farklılaştığı ancak EAP düzeyindeki farklılaşmanın yorumlanabilir etki büyüklüğüne sahip olmadığı gözlenmiştir.

WLE düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,027 farklılık %95 GA [0,019, 0,034] ile ($p < 0,01$); GAMOM'un karma modelden 0,012 farklılık %95 GA [0,006, 0,018] ile ($p < 0,01$); karma modelin ATM'den 0,015 farklılık %95 GA [0,009, 0,021] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

MAP düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,026 farklılık %95 GA [0,018, 0,03] ile ($p < 0,01$); GAMOM'un karma modelden 0,010 farklılık %95 GA [0,004, 0,016] ile ($p < 0,01$); karma modelin ATM'den 0,015 farklılık %95 GA [0,010, 0,021] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Sınıflama Doğruluklarına Yönelik Bulgular

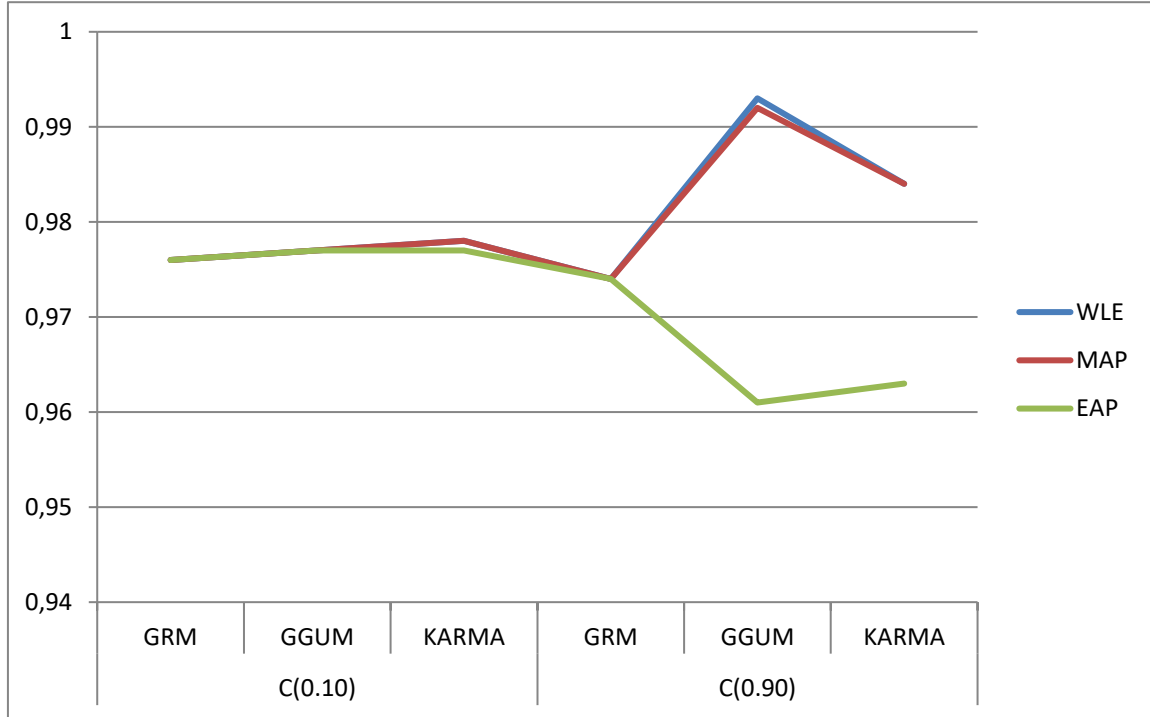
Tablo 8'de yetenek ölçeğinin $c(0,10)$ ve $c(0,90)$ kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilmiş sınıflama doğruluğu indeksleri yer almaktadır.

Tablo 8

c(0,10) ve c(0,90) Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları

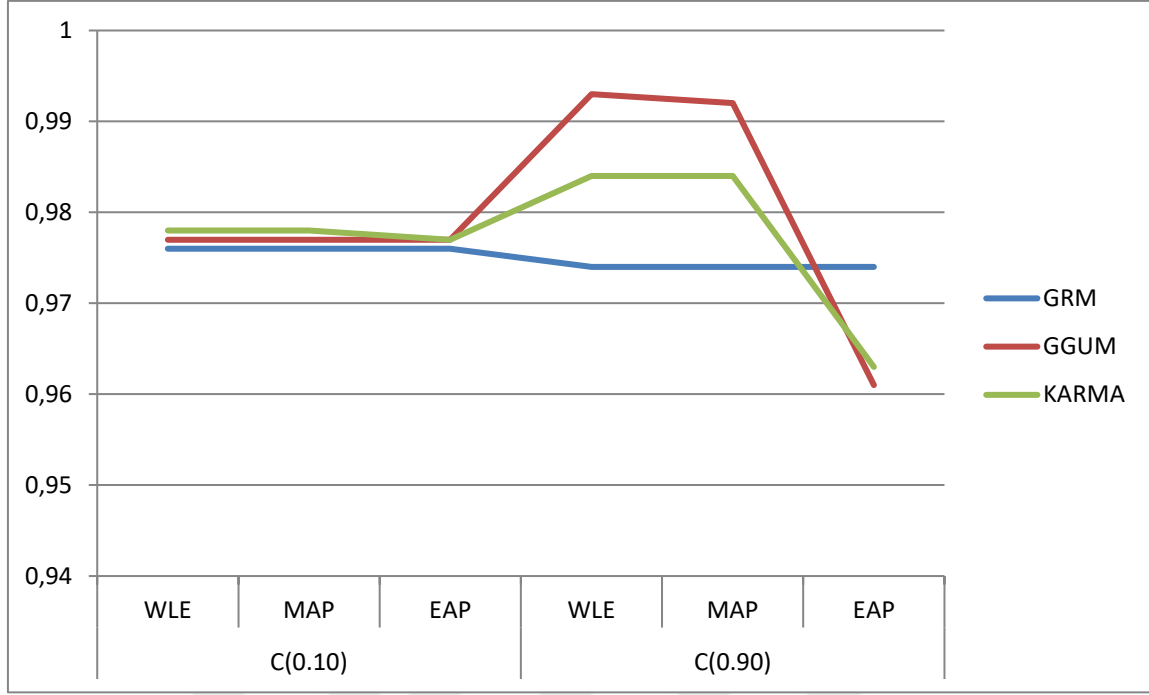
		WLE	MAP	EAP
C(0,10)	ATM	0,976	0,976	0,976
	GAMOM	0,977	0,977	0,977
	KARMA	0,978	0,978	0,977
C(0,90)	ATM	0,974	0,974	0,974
	GAMOM	0,993	0,992	0,961
	KARMA	0,984	0,984	0,963

Tablo 8’deki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama doğruluğuna ilişkin değişimi gözlemlemek için grafikler elde edilmiş olup Şekil 6 ve Şekil 7’de sunulmuştur. Şekil 6’da her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 7’de her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişimine ilişkin grafik yer almaktadır:



Şekil 6. $c(0,10)$ ve $c(0,90)$ noktalarında yöntemlerin değişimi - $\hat{t}$

Şekil 6 incelendiğinde $c(0,10)$ kesme puan noktasında modeller altında üç yöntemin de değişiminin yaklaşık olarak benzer olduğu gözlenmektedir. Yetenek ölçeğinin yüksek ucunda yer alan $c(0,90)$ noktasında yine bütün modeller altında WLE ve MAP’ın değişiminin benzer olduğu; WLE’nin MAP’a nispeten daha GAMOM altında yüksek değerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler benzer sonuçlanmasına rağmen, GAMOM ve karma model altında EAP’ın WLE ve MAP’a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 7. $c(0,10)$ ve $c(0,90)$ noktalarında modellerin değişimi - $\hat{\tau}$

Şekil 7 incelendiğinde $c(0,10)$ kesme puan noktasında yöntemler altında üç modelin değişiminin benzer olduğu, karma modelin diğer modellere nispeten daha yüksek sınıflama doğruluğu ile sonuçlandığı gözlenmiştir. Yetenek ölçeğinin yüksek ucunda yer alan $c(0,90)$ kesme puan noktasında ise modeller arası farklılığın ortaya çıktığı; WLE ve MAP altında GAMOM'un en yüksek sınıflama doğruluğu ile sonuçlandığı gözlenmiştir. GAMOM'u ise benzer değişim ile karma model takip etmektedir. EAP altındaki kestirimlerde ise GAMOM ve karma modelde düşüş yaşandığı; ATM altında elde edilen sonuçların ise yöntemlere göre değişiminin belirgin olmadığı gözlenmiştir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama doğruluğu üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Elde edilen sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu görülmüştür ($F_{(1,912, 3213,328)} = 108,120$, $p < 0,01$, $\eta_p^2 = 0,060$). Bu sonuç yöntem ve model arasındaki iki yönlü etkileşim etkilerinin kesme puan noktasının düzeylerine bağlı olarak farklılaştığına işaret etmektedir. Söz konusu iki yönlü etkileşim etkisinin kesme puan noktasının hangi düzeyinde farklılaştığını ortaya çıkarmak için iki düzeyli olan kesme puan noktasının her bir düzeyinde yöntem ve model arasında iki yönlü tekrarlı ölçümler ANOVA analizleri yürütülmüştür.

Kesme puan değişkeninin c(0,10) noktasında yöntem ve model değişkenleri için yürütülen iki faktörlü tekrarlı ölçümler ANOVA sonuçlarına ilişkin etkileşim etkisinin anlamsız olduğu görülmüştür ($F_{(2,098, 3527,456)} = 3,107, p > 0,01$). Bunun yanı sıra c(0,10) kesme puan noktasında hem yöntem değişkeni açısından ($F_{(1,325, 2227,867)} = 1,871, p > 0,01$) hem de model türü değişkeni açısından ($F_{(1,036, 1742,309)} = 1,346, p > 0,01$) istatistiksel olarak farklılaşmanın olmadığı gözlenmiştir.

Kesme puanı değişkeninin c(0,90) düzeyinde ise yöntem ve model arasında yürütülen iki yönlü tekrarlı ölçümler ANOVA sonucunun anlamlı çıktığı gözlenmiştir ($F_{(1,855, 3117,999)} = 114,067, p < 0,01, \eta_p^2 = 0,064$). Yöntem ve model arasındaki etkileşim etkisine ilişkin ana etkiler hem model hem de yöntem düzeyinde incelenmiştir.

İlk olarak model düzeyinde incelemeler yapılmış olup ATM ($F_{(1,079, 1813,901)} = 23,209, p < 0,01, \eta_p^2 = 0,014$), GAMOM ($F_{(1,031, 1733,829)} = 187,595, p < 0,01, \eta_p^2 = 0,100$) ve karma model ($F_{(1,026, 1724,665)} = 87,327, p < 0,01, \eta_p^2 = 0,049$) olmak üzere model türünün bütün düzeylerinde yöntemlerin istatistiksel olarak farklılaştığı sonucuna ulaşılmıştır. Ancak ATM altında gözlenen farklılığın pratikte yorumlanabilir etki büyüklüğüne sahip olmadığı gözlenmiştir.

GAMOM düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin MAP'tan 0,001 farklılık %95 GA [0,001, 0,002] ile ($p < 0,01$); WLE'nin EAP'tan 0,032 farklılık %95 GA [0,026, 0,037] ile ($p < 0,01$); ve MAP'ın EAP'tan 0,031 farklılık GA [0,025, 0,031] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Karma model düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin ve MAP'ın EAP'tan 0,021 farklılık %95 GA [0,016, 0,027] ile ($p < 0,01$), daha iyi sonuç verdiği WLE ile MAP arasında ise istatistiksel olarak anlamlı farklılık olmadığı gözlenmiştir.

Model düzeyinde yöntemler karşılaştırıldıktan sonra yöntem düzeyinde modeller karşılaştırılmıştır. Yöntemin WLE düzeyinde, ($F_{(1,909, 3209,392)} = 39,690, p < 0,01, \eta_p^2 = 0,023$), MAP düzeyinde ($F_{(1,887, 3172,061)} = 34,928, p < 0,01, \eta_p^2 = 0,020$), EAP düzeyinde ($F_{(1,308, 2199,495)} = 21,054, p < 0,01, \eta_p^2 = 0,012$) modellerin birbirinden istatistiksel olarak farklılaştığı gözlenmiştir. Ancak EAP düzeyindeki farklılaşmanın pratikte yorumlanabilir etki büyüklüğüne sahip olmadığı gözlenmiştir.

WLE düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,019 farklılık %95 GA [0,013, 0,024] ile ($p < 0,01$); GAMOM'un karma modelden 0,009 farklılık %95 GA [0,004, 0,013] ile ($p < 0,01$); karma modelin ATM'den 0,010 farklılık %95 GA [0,005, 0,015] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

MAP düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,018 farklılık %95 GA [0,012, 0,023] ile ($p < 0,01$); karma modelin ATM'den 0,010 farklılık %95 GA [0,006, 0,015] ile ($p < 0,01$) ve GAMOM'un karma modelden 0,007 farklılık %95 GA [0,003, 0,012] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği ancak gözlenmiştir.

Genel Yorum

Sınıflama doğrulukları ve sınıflama tutarlılıkları incelendiğinde, sınıflama doğruluklarının sınıflama tutarlılıklarından daha yüksek değerler ile sonuçlandığı gözlenmiştir. Kesme puan noktası, yöntem ve model değişkenleri açısından karşılaştırmalar hem sınıflama doğruluğu hem de sınıflama tutarlılığı için yapılmış olup koşullardaki değişiminin her iki indeks için de benzer olduğu gözlenmiştir. Bu doğrultuda sınıflama doğruluğu ve sınıflama tutarlılığı “sınıflama indeksi” ifadesi ile genel olarak değerlendirilmiştir. Kesme puan değişkeni düşük uçta $c(0,10)$ ve yüksek uçta $c(0,90)$ olmak üzere dikkate alınıp yetenek kestirim yöntemleri ve model türleri çerçevesinde sınıflama indeksleri üç faktörlü olarak incelenmiş olup üç yönlü etkileşim etkisi her iki indeks için de anlamlı çıkmıştır. Bu durum kullanılan kestirim yöntemi ve model türünün etkileşimine dayalı olarak hesaplanan sınıflama indekslerinin kesme puan noktasına göre değiştiğini ortaya çıkarmıştır. Model ve yöntemin etkileşimine dayalı bu farklılığın kesme puan noktalarının hangi düzeyinde mevcut olduğuna yönelik incelemeler için kesme puan noktasının her bir düzeyinde iki faktörlü tekrarlı ölçümler ANOVA ile incelemeler yapılmıştır. Kesme puan düzeyinin düşük ucunda etkileşim etkisine yönelik pratikte yorumlanabilir istatistiksel bir farklılık elde edilmemiş olup bununla birlikte model ve yönteme ilişkin ana etkilerin de anlamsız olduğu görülmüştür. Bu durum yetenek ölçeğinin düşük ucundaki kesme puan noktasında bütün yöntemlerin ve modellerin benzer sonuç verdiğine işaret etmektedir. Ancak yetenek ölçeğinin yüksek ucunda yöntem ve model türünün etkileşiminin istatistiksel olarak anlamlı olduğu gözlenmiştir. Modellerin her bir düzeyinde yöntemler karşılaştırıldığında

bütün modeller altında yöntemlerin birbirinden istatistiksel olarak farklılaştığı gözlenmiş olup en belirgin farklılığın GAMOM altında ortaya çıktığı, en düşük farklılığın ise ATM altında ortaya çıktığı gözlenmiştir. Model ve yöntemler birlikte ele alınacak olursa yöntemlerin etkililiğinin kullanılan modele göre farklılaştığı gözlenmiştir. GAMOM altında WLE ve MAP yüksek indekslerle sonuçlanırken GAMOM altında EAP'a dayalı yapılan kestirimlerin düştüğü gözlenmiştir. Karma model altında da yöntemlerin değişimine ilişkin durum GAMOM'a benzemektedir. ATM altında yöntemler arasındaki miktarsal farklılığın diğer modellere göre daha düşük olduğu ve yöntemlerin yaklaşık olarak benzer sonuçlar verdiği gözlenmiştir. Model düzeyinde yapılan incelemelerin ardından yöntem düzeyindeki inceleme sonuçlarında WLE ve MAP altında GAMOM öne çıkarken, EAP altında modeller arası farklılaşmanın istatistiksel olarak anlamsız olduğu gözlenmiştir. Dolayısıyla sınıflama indekslerinin elde edilmesinde kesme puan noktası, model ve yöntem etkilerinin eşzamanlı değerlendirildiği durumda monotonik olmayan madde analizlerinin söz konusu olduğu modellerde yöntem seçiminin önem kazandığı gözlenmiştir. Sınıflama tutarlılığı ve sınıflama doğruluğu indeksleri karşılaştırıldığında sınıflama tutarlılığı indekslerine dayalı yapılan farklılık incelemelerine ilişkin etki büyüklüklerinin sınıflama doğruluklarındakinden daha yüksek olduğu ancak değişiminin benzer olduğu gözlenmiştir.

Üçüncü Araştırma Sorusuna Yönelik Bulgular

Bu araştırma sorusu kapsamında “Yetenek ölçeğinin $c(0,15)$ ve $c(0,85)$ noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında ATM, GAMOM ve karma modellerin sınıflama indekslerine olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?” sorusuna yanıt aranmıştır.

Sınıflama indeksleri sınıflama tutarlılıkları ve sınıflama doğrulukları olmak üzere iki indeks olarak ele alınmış olup her biri için sonuçlar sunulmuştur.

Sınıflama Tutarlılıklarına İlişkin Bulgular

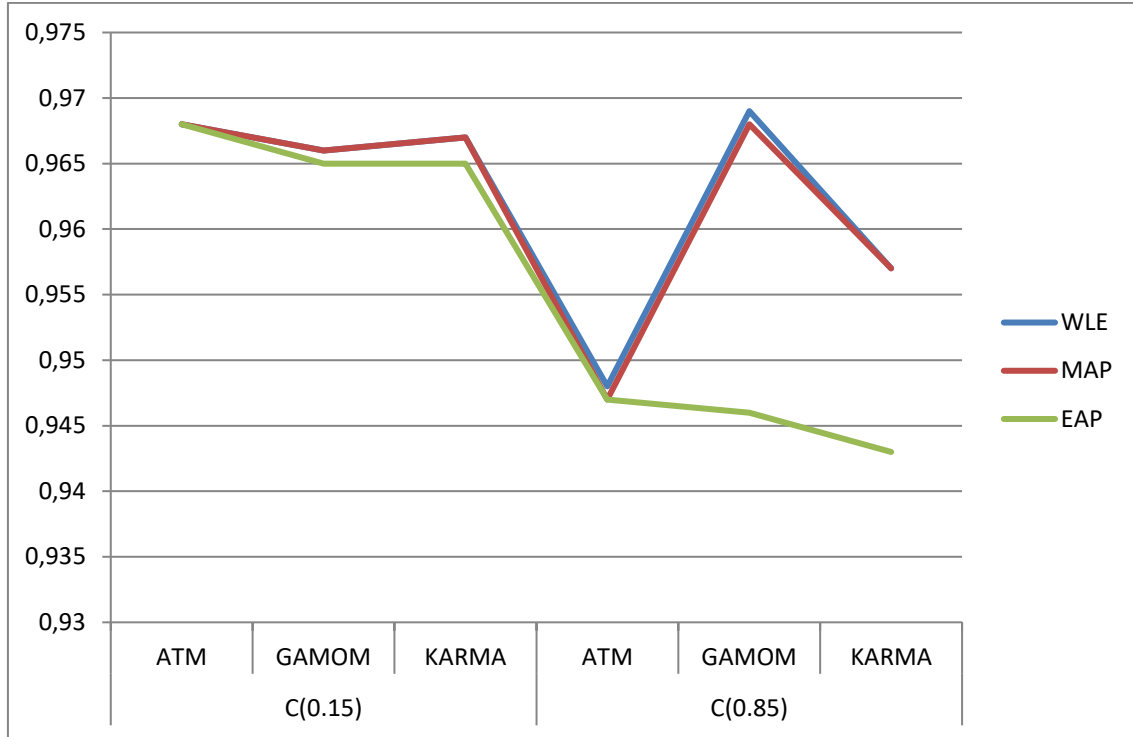
Tablo 9’da yetenek ölçeğinin $c(0,15)$ ve $c(0,85)$ kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilmiş sınıflama tutarlılığı indeksleri yer almaktadır:

Tablo 9

$c(0,15)$ ve $c(0,85)$ Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları

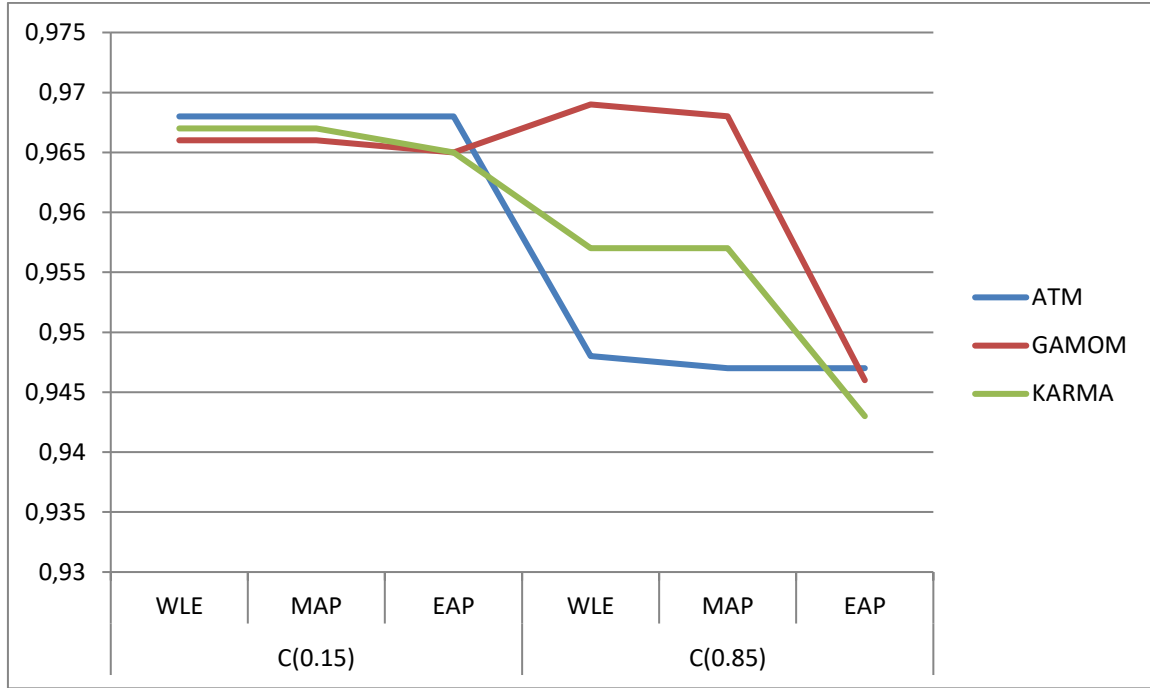
		WLE	MAP	EAP
C(0,15)	ATM	0,968	0,968	0,968
	GAMOM	0,966	0,966	0,965
	KARMA	0,967	0,967	0,965
C(0,85)	ATM	0,948	0,947	0,947
	GAMOM	0,969	0,968	0,946
	KARMA	0,957	0,957	0,943

Tablo 9’daki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama tutarlılıklarına ilişkin değişimi gözlemlemek için grafikler elde edilmiş olup Şekil 8 ve Şekil 9’da sunulmuştur. Şekil 8’de her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 9’da her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişimine ilişkin grafik yer almaktadır:



Şekil 8. $c(0,15)$ ve $c(0,85)$ noktalarında yöntemlerin değişimi - $\hat{\gamma}$

Şekil 8 incelendiğinde $c(0,15)$ kesme puan noktasında modeller altında WLE ve MAP'ın aynı değerlerle sonuçlandığı; EAP'ın ise ATM altında diğer yöntemlerle aynı ancak GAMOM ve karma model altında nispeten daha düşük sınıflama tutarlılığı ile sonuçlandığı gözlenmiştir. Yüksek uçta yer alan $c(0,85)$ noktasında yine bütün modeller altında WLE ve MAP'ın değişiminin benzer olduğu; WLE'nin MAP'a nispeten daha yüksek değerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler benzer değerler ile sonuçlanmasına rağmen, GAMOM ve karma model altında EAP'ın WLE ve MAP'a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 9. $c(0,15)$ ve $c(0,85)$ noktalarında modellerin değişimi - $\hat{\gamma}$

Şekil 9 incelendiğinde $c(0,15)$ kesme puan noktasında ATM'nin bütün yöntemler altında aynı sonuçlar verdiği ve diğer modellerden daha yüksek sınıflama tutarlılıkları sağladığı gözlenmiştir. WLE ve MAP altında karma modelin GAMOM'dan daha yüksek; EAP altında ise GAMOM ve karma modelin aynı sınıflama indeksleriyle sonuçlandığı gözlenmiştir. Yetenek ölçeğinin yüksek ucunda yer alan $c(0,90)$ kesme puan noktasında ise modeller arası farklılığın daha belirgin bir şekilde ortaya çıktığı; WLE ve MAP altında GAMOM'un en yüksek sınıflama tutarlılığı ile sonuçlandığı GAMOM'u ise karma modelin takip ettiği gözlenmiştir. EAP altındaki kestirimlerde GAMOM ve karma modelde düşüş yaşandığı; ATM'ye dayalı elde edilen sonuçların ise yöntemlere göre değişiminin belirgin olmadığı gözlenmiştir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama tutarlılığı üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu görülmüştür ($F_{(1,971, 3312,892)} = 40,572$, $p < 0,01$, $\eta_p^2 = 0,024$). Bu sonuç yöntem ve model arasındaki iki yönlü etkileşim etkilerinin kesme puan noktasının düzeylerine bağlı olarak farklılaştığına işaret etmektedir.

Söz konusu iki yönlü etkileşim etkisinin kesme puan noktasının hangi düzeyinde farklılaştığını ortaya çıkarmak için iki düzeyli olan kesme puan noktasının her bir

düzeyinde yöntem ve model arasında iki yönlü tekrarlı ölçümler ANOVA analizleri yürütülmüştür.

Kesme puan değişkeninin c(0,15) noktasında yöntem ve model değişkenleri için yürütülen iki faktörlü tekrarlı ölçümler ANOVA sonuçlarına ilişkin etkileşim etkisinin anlamlı olduğu ($F_{(2,041, 3431,348)} = 5,143$, $p < 0,01$, $\eta_p^2 = 0,005$) ancak etki büyüklüğünün pratikte anlamsız olduğu görülmüş olup söz konusu etkileşim etkisine yönelik anlamlılık yorumlanmamıştır. Bununla birlikte c(0,15) kesme puan noktasında model değişkeni açısından istatistiksel olarak farklılaşmanın olmadığı gözlenmiştir ($F_{(1,315, 2210,217)} = 4,251$, $p > 0,01$). Yöntem açısından ise istatistiksel olarak anlamlı farklılığın olduğu ($F_{(1,128, 1896,664)} = 16,617$, $\eta_p^2 = 0,010$, $p < 0,01$) ancak bu farklılığın pratikte yorumlanabilir bir etkiye sahip olmadığı gözlenmiştir.

Kesme puanı değişkeninin c(0,85) düzeyinde yöntem ve model arasında yürütülen iki yönlü tekrarlı ölçümler ANOVA sonucunun ise anlamlı çıktığı gözlenmiştir ($F_{(1,899, 3192,052)} = 43,858$, $p < 0,01$, $\eta_p^2 = 0,025$). Yöntem ve model arasındaki etkileşim etkisine ilişkin ana etkiler hem model hem de yöntem düzeyinde incelenmiştir.

İlk olarak model düzeyinde incelemeler yapılmıştır. ATM ($F_{(1,117, 1877,350)} = 135,625$, $p < 0,01$, $\eta_p^2 = 0,075$), GGUM ($F_{(1,021, 1715,815)} = 71,456$, $p < 0,01$, $\eta_p^2 = 0,041$) ve karma model ($F_{(1,023, 1719,020)} = 37,673$, $p < 0,01$, $\eta_p^2 = 0,022$) olmak üzere model türünün bütün düzeylerinde yöntemlerin istatistiksel olarak farklılaştığı sonucuna ulaşılmıştır.

ATM düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin MAP'tan 0,001 farklılık %95 GA [0,0007, 0,0010] ile ($p < 0,01$), WLE'nin EAP'tan 0,001 farklılık %95 GA [0,0008, 0,0012] ile ($p < 0,01$) ve EAP'ın MAP'tan 0,0001 farklılık %95 GA [0,0001, 0,0002] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

GAMOM düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin EAP'tan 0,023 farklılık %95 GA [0,016, 0,029] ile ($p < 0,01$) ve MAP'ın EAP'tan 0,022 farklılık GA [0,016, 0,028] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği, fakat WLE ve MAP arasında ise istatistiksel bir farklılaşmanın olmadığı gözlenmiştir.

Karma model düzeyinde yöntemler arasında yürütülen ikili karşılaştırma testlerine göre WLE'nin EAP'tan 0,014 farklılık %95 GA [0,009, 0,020] ile ($p < 0,01$) ve MAP'ın

EAP'tan 0,014 farklılık GA [0,008, 0,019] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği, WLE ile MAP' ın farklılaşmadığı gözlenmiştir.

Model düzeyinde yöntemler karşılaştırıldıktan sonra yöntem düzeyinde modeller karşılaştırılmıştır. Yöntemin WLE düzeyinde ($F_{(1,469, 2469,493)} = 55,326$, $p < 0,01$, $\eta_p^2 = 0,032$), ve MAP düzeyinde ($F_{(1,440, 2420,858)} = 56,630$, $p < 0,01$, $\eta_p^2 = 0,033$), modellerin birbirinden istatistiksel olarak farklılaştığı ancak EAP düzeyinde ($F_{(1,268, 2131,665)} = 1,222$, $p > 0,01$) modellerin birbirinden istatistiksel olarak farklılaşmadığı gözlenmiştir.

WLE düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,021 farklılık %95 GA [0,015, 0,028] ile ($p < 0,01$), GAMOM'un karma modelden 0,012 farklılık %95 GA [0,008, 0,016] ile ($p < 0,01$), karma modelin ATM'den 0,010 farklılık %95 GA [0,006, 0,014] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

MAP düzeyinde modeller arasında yürütülen ikili karşılaştırma testlerine göre GAMOM'un ATM'den 0,022 farklılık %95 GA [0,015, 0,028] ile ($p < 0,01$), GAMOM'un karma modelden 0,012 farklılık %95 GA [0,008, 0,016] ile ($p < 0,01$), karma modelin ATM'den 0,010 farklılık %95 GA [0,006, 0,014] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Sınıflama Doğruluklarına Yönelik Bulgular

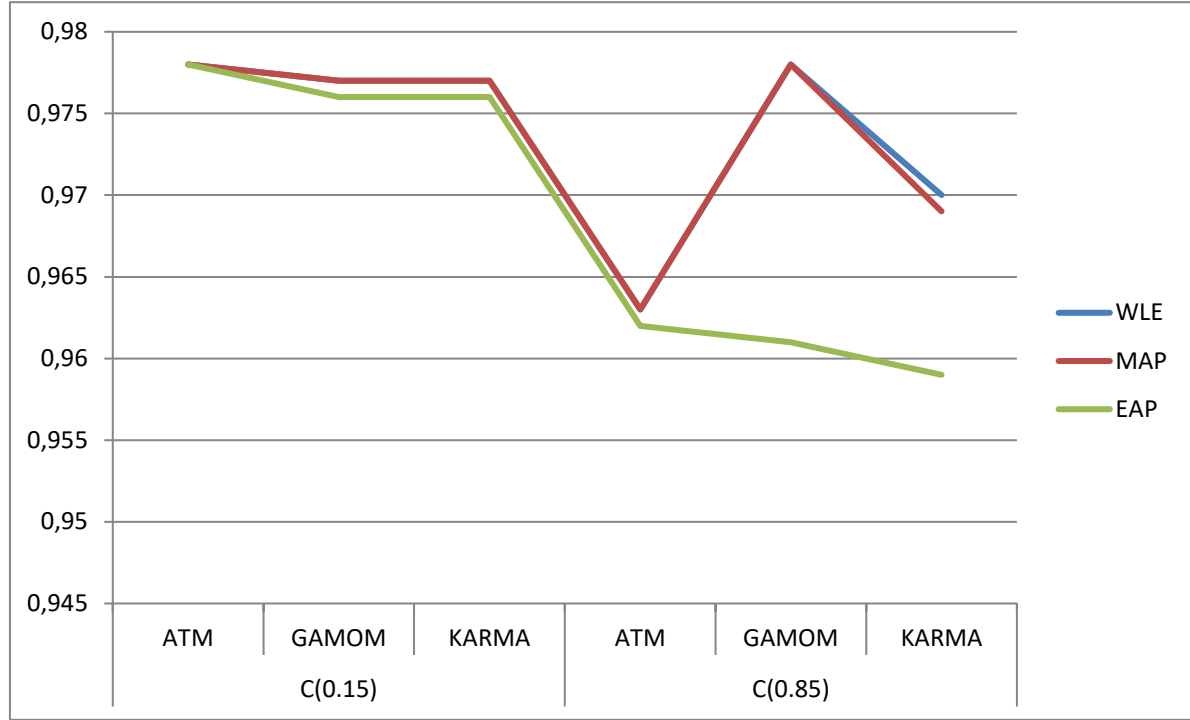
Tablo 10'da yetenek ölçeğinin $c(0,15)$ ve $c(0,85)$ kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilen sınıflama doğruluğu indeksleri yer almaktadır:

Tablo 10

c(0,15) ve c(0,85) Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları

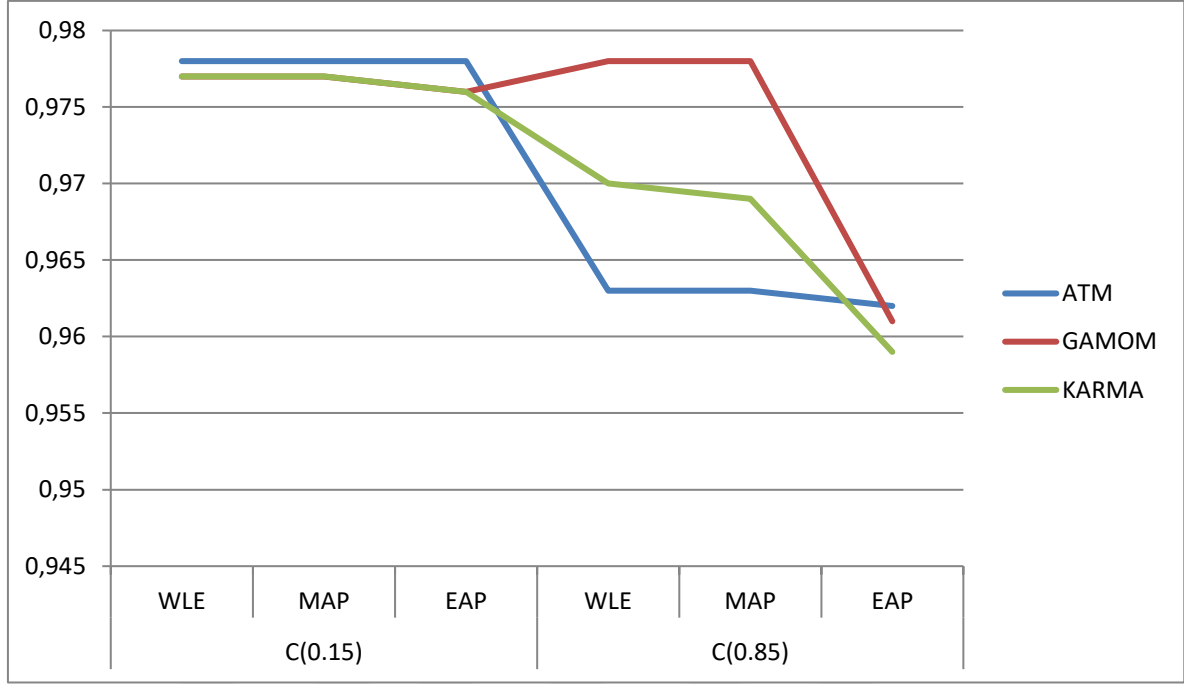
		WLE	MAP	EAP
C(0,15)	ATM	0,978	0,978	0,978
	GAMOM	0,977	0,977	0,976
	KARMA	0,977	0,977	0,976
C(0,85)	ATM	0,963	0,963	0,962
	GAMOM	0,978	0,978	0,961
	KARMA	0,970	0,969	0,959

Tablo 10'daki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama doğruluğuna ilişkin değişimin seyrini gözlemlemek için grafikler elde edilmiş olup Şekil 10 ve Şekil 11'de sunulmuştur. Şekil 10'da her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 11'de her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişimine ilişkin grafik yer almaktadır:



Şekil 10. $c(0,15)$ ve $c(0,85)$ noktalarında yöntemlerin değişimi - $\hat{t}$

Şekil 10 incelendiğinde $c(0,15)$ kesme puan noktasında modeller altında WLE ve MAP'ın aynı değerlerle sonuçlandığı; EAP'ın ise ATM altında diğer yöntemlerle aynı ancak GAMOM ve karma model altında nispeten daha düşük sınıflama doğruluğu ile sonuçlandığı gözlenmiştir. Yüksek uçta yer alan $c(0,85)$ noktasında yine bütün modeller altında WLE ve MAP'ın değişiminin benzer olduğu; WLE'nin MAP'a nispeten karma model altında daha yüksek değerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler benzer değerler ile sonuçlanmasına rağmen, GAMOM ve karma model altında EAP'ın WLE ve MAP'a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 11. c(0,15) ve c(0,85) noktalarında modellerin değişimi - $\hat{\tau}$

Şekil 11 incelendiğinde c(0,15) kesme puan noktasında yöntemler altında GAMOM ve karma modelin aynı sınıflama indeksleriyle sonuçlandığı; ATM'nin ise bu modellere nispeten daha yüksek sonuç verdiği gözlenmiştir. Yetenek ölçeğinin yüksek ucunda yer alan c(0,85) kesme puan noktasında ise modeller arası farklılığın ortaya çıktığı; WLE ve MAP altında GAMOM'un en yüksek sınıflama tutarlılığı ile sonuçlandığı GAMOM'u ise karma modelin takip ettiği gözlenmiştir. EAP altındaki kestirimlerde GAMOM ve karma modelde düşüş yaşandığı; ATM altında elde edilen sonuçların ise yöntemlere göre değişiminin belirgin olmadığı gözlenmiştir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama doğruluğu üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Elde edilen sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu ancak etki büyüklüğünün yorumlanabilir olmadığı görülmüştür ($F_{(2,030, 3412,393)} = 33,069$, $p < 0,01$, $\eta_p^2 = 0,019$). Bu nedenle iki yönlü etkileşim etkileri incelenmiştir. İki yönlü etkileşim etkileri ele alındığında ise kesme puanı ile yöntem, ve model ile yöntem arasında istatistiksel olarak anlamlı farklılığın olduğu ve bu farklılıklara ilişkin etki büyüklüklerinin pratikte yorumlanabilir olduğu gözlenmiştir ($F_{(1,020, 1713,833)} = 53,410$, $p < 0,01$, $\eta_p^2 = 0,031$; $F_{(1,976, 3321,676)} = 35,317$, $p < 0,01$, $\eta_p^2 = 0,021$).

Kesme puan noktası ile yöntem arasındaki etkileşim etkisi için kesme puanının düzeylerinde ikili karşılaştırma testleri ile yöntemler arası farklılıklar incelenmiştir. Kesme puanının $c(0,15)$ düzeyinde WLE'nin EAP'tan 0,001 farklılık %95 GA [0,000, 0,001] ile ($p < 0,01$); MAP'ın EAP'tan 0,0005 farklılık GA [0,0001, 0,0009] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği ve WLE ile MAP arasında istatistiksel bir farklılaşma olmadığı gözlenmiştir. Kesme puanının $c(0,85)$ düzeyinde ise WLE'nin MAP'tan 0,001 farklılık %95 GA [0,000, 0,001] ile ($p < 0,01$), WLE'nin EAP'tan 0,010 farklılık %95 GA [0,007, 0,013] ile ($p < 0,01$) ve MAP'ın EAP'tan 0,009 farklılık GA [0,006, 0,012] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Model ile yöntem arasındaki etkileşim etkisi için hem modelin düzeyinde yöntemler hem de yöntemin düzeyinde modeller için ikili karşılaştırmalar yapılmıştır. İlk olarak her bir model düzeyinde yöntemler karşılaştırılmış ardından her bir yöntem düzeyinde ise modeller karşılaştırılmıştır.

Modellerden ATM altında WLE' nin 0,0005 farklılık %95 GA [0,0004, 0,0006] ile ($p < 0,01$) MAP'tan ve 0,0005 farklılık %95 GA [0,0003, 0,0007] ile ($p < 0,01$) EAP'tan daha iyi sonuç verdiği gözlenmiştir. MAP ve EAP arasında ise istatistiksel olarak farklılaşma gözlenmemiştir. GAMOM altında WLE'nin ve MAP'ın 0,009 farklılık %95 GA [0,006, 0,012] ile ($p < 0,01$) EAP'tan daha iyi sonuç verdiği, WLE ve MAP arasında ise istatistiksel olarak farklılaşmanın olmadığı görülmüştür. Karma model altında WLE ve MAP'ın 0,006 farklılık %95 GA [0,003, 0,008] ile ($p < 0,01$) EAP'tan daha iyi sonuç verdiği, WLE ve MAP arasında ise istatistiksel olarak farklılaşmanın olmadığı görülmüştür.

Model düzeyindeki incelemelerin ardından yöntem düzeyinde incelemeler yapılmıştır. WLE altında GAMOM'un ATM'den 0,007 farklılık %95 GA [0,004, 0,010] ile ($p < 0,01$), karma modelden ise 0,004 farklılık %95 GA [0,002, 0,006] ile ($p < 0,01$) daha iyi sonuç verdiği, ATM ile karma model arasında ise istatistiksel farklılığın olmadığı gözlenmiştir. MAP altında yine GAMOM'un ATM'den 0,007 farklılık %95 GA [0,004, 0,010] ile, karma modelden ise 0,004 farklılık %95 GA [0,002, 0,006] ile daha iyi sonuç verdiği ($p < 0,01$) ve karma modelin ATM'den 0,003 farklılık %95 GA [0,001, 0,005] ile ($p < 0,01$) daha iyi sonuç verdiği gözlenmiştir. Ancak EAP altında modeller arasında istatistiksel olarak anlamlı farklılığın olmadığı gözlenmiştir.

Genel Yorum

Sınıflama doğrulukları ve sınıflama tutarlılıkları incelendiğinde, sınıflama doğruluklarının sınıflama tutarlılıklarından daha yüksek değerler ile sonuçlandığı gözlenmiştir. Kesme puan noktası, yöntem ve model değişkenleri açısından karşılaştırmalar hem sınıflama doğruluğu hem de sınıflama tutarlılığı için yapılmıştır. Sınıflama tutarlılığı için üç yönlü etkileşim etkisi yorumlanabilirken sınıflama doğruluğu için üç yönlü etkileşim etkisinin yorumlanabilir etki büyüklüğüne sahip olmadığı gözlenmiş olup iki yönlü etkileşim etkileri incelenmiştir.

İstatistiksel olarak anlamlı ve yorumlanabilir etki büyüklüğü ile sonuçlanan üç yönlü etkileşim etkisi ile, kullanılan kestirim yöntemi ve model türünün etkileşimine dayalı olarak hesaplanan sınıflama tutarlılığının kesme puan noktasına göre değiştiği sonucu ortaya çıkmıştır. Model ve yöntemin etkileşimine dayalı bu farklılığın kesme puan noktalarının hangi düzeyinde mevcut olduğuna yönelik incelemeler için kesme puan noktasının her bir düzeyinde iki faktörlü tekrarlı ölçümler ANOVA ile incelemeler yapılmıştır. Kesme puan düzeyinin düşük ucunda etkileşim etkisine yönelik pratikte yorumlanabilir istatistiksel bir farklılık elde edilmemiş ancak buna karşın yetenek ölçeğinin yüksek ucunda yöntem ve model türünün etkileşimine dayalı istatistiksel olarak farklılaşmanın olduğu gözlenmiştir. Modeller altında yöntemler karşılaştırıldığında ATM altında bütün yöntemlerin birbirinden istatistiksel olarak farklılaştığı; WLE'nin MAP ve EAP'tan daha yüksek sonuç verdiği MAP ve EAP'ın birbirine daha yakın olduğu gözlenmiştir. GAMOM altında ise WLE ile MAP arasında istatistiksel farklılığın olmadığı ancak bu iki yöntemin de EAP'tan istatistiksel olarak daha yüksek sonuç verdiği görülmüştür. Bununla birlikte WLE-EAP ve MAP-EAP arasındaki ortalama farklılığın ATM altındaki yöntemler arası farklılığa göre daha yüksek olduğu görülmüştür. Modellerin düzeylerinde yöntemler karşılaştırıldıktan sonra yöntemlerin her bir düzeyinde de modeller karşılaştırılmıştır. Modellerin kullanılan kestirim yöntemlerinden WLE ve MAP altında birbirinden anlamlı olarak farklılaştığı ancak EAP altında modeller açısından istatistiksel bir farklılaşma olmadığı görülmüştür. WLE ve MAP altında modellere yönelik farklılığın GAMOM'un lehine olduğu, GAMOM'un her iki modelden daha iyi sonuç verdiği bununla birlikte karma modelin de ATM'den daha iyi sonuç verdiği gözlenmiştir. GAMOM altında WLE'nin her iki yönteme göre; MAP'ın ise EAP'a göre daha iyi sonuç verdiği gözlenmiştir. Karma model altında da GAMOM'da olduğu gibi WLE'nin her iki yönteme göre; MAP'ın ise EAP'a göre daha iyi sonuç verdiği

gözlenmiştir. Sınıflama doğruluğuna ilişkin sonuçlar incelendiğinde iki yönlü etkileşim etkilerinden kesme puanı ile yöntem, ve model ile yöntem arasındaki etkileşim etkisinin anlamlı ve yorumlanabilir olduğu gözlenmiştir. İkili karşılaştırmalar yapıldığında sonuçların sınıflama tutarlılığı ile benzer olduğu gözlenmiştir. Buradan elde edilen sonuçlar da yetenek ölçeğinin üst ucunda GAMOM'un diğer modellere göre daha iyi sonuç verdiğini ve model etkililiğinin yöntemle bağlı olarak farklılaştığını göstermiştir. Yöntemlerden MAP ve WLE, GAMOM ve karma model altında daha iyi sonuç verirken EAP'ın bu modeller altında diğer yöntemlere göre daha düşük sınıflama doğruluğu ile sonuçlandığı gözlenmiştir.

Sınıflama indeksleri genel olarak değerlendirildiğinde yetenek ölçeğinin yüksek ucundaki kesme puan noktasında yöntem ve modellere ilişkin farklılaşmanın ortaya çıktığı görülmektedir. Modellere yönelik yapılan incelemeler GAMOM'un yetenek ölçeğinin üst uçlarında daha iyi sonuç verdiğini göstermektedir. Yöntemlere göre yapılan incelemeler ele alındığında ise WLE'nin bütün modeller altında hem MAP hem de EAP'a göre daha iyi sonuç verdiği ancak EAP yönteminin kullanılan model türüne bağlı olarak etkisinin değiştiği gözlenmiştir.

Dördüncü Araştırma Sorusuna Yönelik Bulgular

Bu araştırma sorusu kapsamında “Yetenek ölçeğinin c(0,20) ve c(0,80) noktalarında oluşturulmuş kesme puanlarına dayalı iki düzeyli sınıflama çalışmalarında ATM, GAMOM ve karma modellerin sınıflama indekslerine olan etkisi her bir kestirim yöntemi altında kesme puan noktalarına bağlı olarak anlamlı olarak farklılaşmakta mıdır?” sorusuna yanıt aranmıştır.

Sınıflama indeksleri sınıflama tutarlılıkları ve sınıflama doğrulukları olmak üzere iki indeks olarak ele alınmış olup her biri için sonuçlar sunulmuştur.

Sınıflama Tutarlılıklarına İlişkin Bulgular

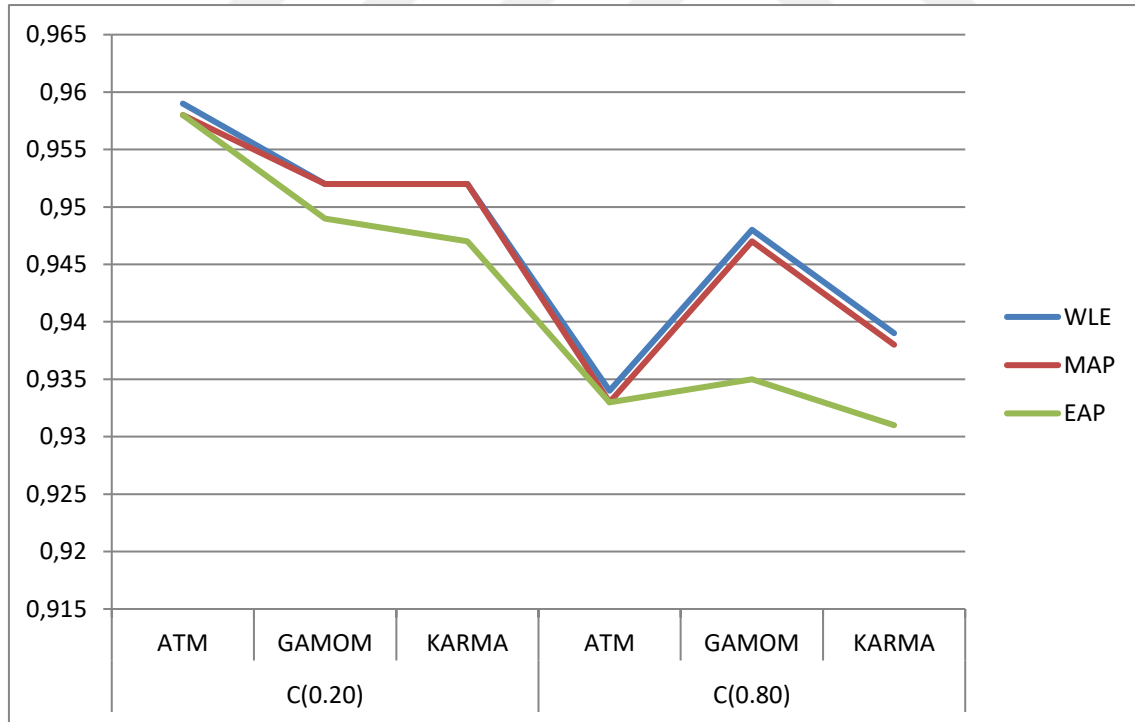
Tablo 11’de c(0,20) ve c(0,80) kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilmiş sınıflama tutarlılığı indeksleri yer almaktadır:

Tablo 11

c(0,20) ve c(0,80) Kesme Puan Noktalarında Elde Edilen Sınıflama Tutarlılıkları

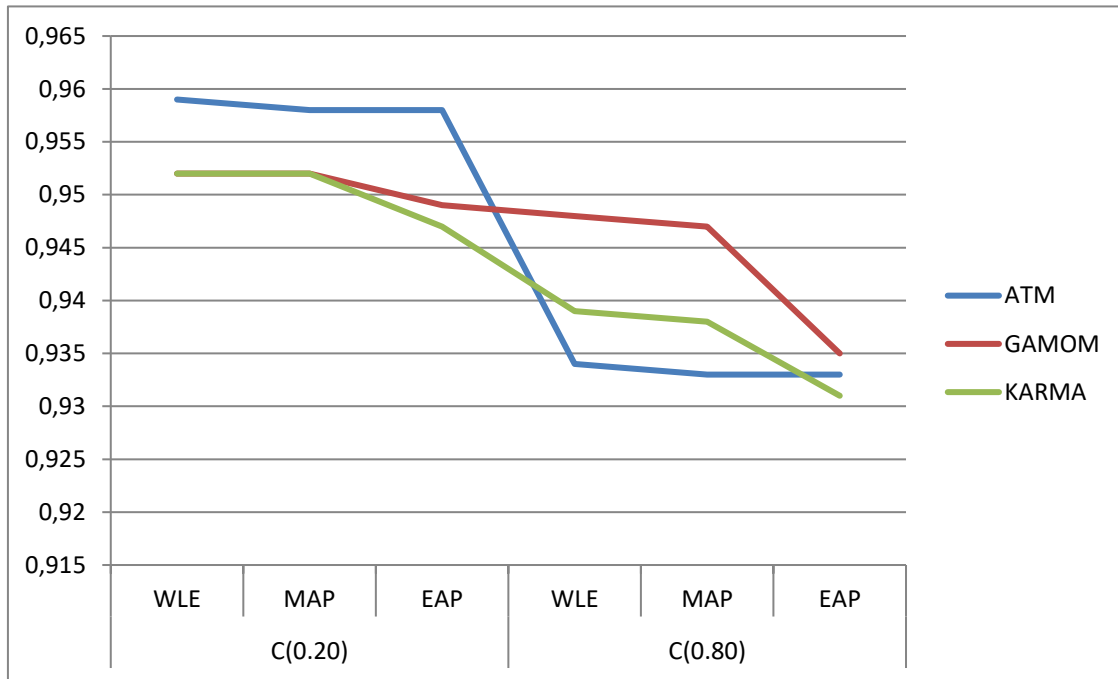
		WLE	MAP	EAP
C(0,20)	ATM	0,959	0,958	0,958
	GAMOM	0,952	0,952	0,949
	KARMA	0,952	0,952	0,947
C(0,80)	ATM	0,934	0,933	0,933
	GAMOM	0,948	0,947	0,935
	KARMA	0,939	0,938	0,931

Tablo 11'deki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama tutarlılıklarına ilişkin değişimi gözlemlemek için grafikler elde edilmiş olup Şekil 12 ve Şekil 13'te sunulmuştur. Şekil 12'de her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 13'te her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişimine ilişkin grafik yer almaktadır:

Şekil 12. c(0,20) ve c(0,80) noktalarında yöntemlerin değişimi - $\hat{\gamma}$

Şekil 12 incelendiğinde c(0,20) kesme puan noktasında bütün modeller altında üç yöntemin de değişiminin benzer olduğu; WLE ve MAP'a dayalı elde edilen sınıflama tutarlılıklarının yaklaşık olarak aynı değerlerle sonuçlandığı ve EAP'a dayalı elde edilen

sınıflama doğruluklarının GAMOM ve karma model altında diğer iki yönteme nispeten daha küçük olduğu görülmektedir. Yüksek uçta yer alan $c(0,80)$ noktasında yine bütün modeller altında WLE ve MAP değişiminin benzer olduğu; WLE'nin MAP'a nispeten daha yüksek değerlerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler benzer değerler ile sonuçlanmasına rağmen, GAMOM ve karma model altında EAP'ın WLE ve MAP'a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 13. $c(0,20)$ ve $c(0,80)$ noktalarında modellerin değişimi - $\hat{y}$

Şekil 13 incelendiğinde $c(0,20)$ kesme puan noktasında bütün yöntemler altında ATM'nin GAMOM ve karma modelden daha yüksek sınıflama tutarlılıkları ile sonuçlandığı görülmektedir. Ancak yetenek ölçeğinin yüksek ucunda yer alan $c(0,80)$ kesme puan noktasında en yüksek sınıflama tutarlılıklarının bütün yöntemler altında GAMOM ile elde edildiği ancak EAP yöntemi altında düşüş olduğu görülmektedir. WLE ve MAP yöntemleri altında GAMOM'u karma modelin takip ettiği gözlenmiştir. Modellerin değişimi genel olarak ele alındığında ATM'nin her bir yöntem altında benzer sonuç verdiği ancak GAMOM ve karma modelin kullanılan kestirim yöntemine bağlı olarak sınıflama tutarlılıkları üzerindeki etkisinin farklılaştığı görülmektedir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama tutarlılığı üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Elde edilen

sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu ($F_{(1,821, 3060,511)} = 15,578$, $p < 0,01$, $\eta_p^2 = 0,009$) ancak bu etkileşim etkisinin pratikte yorumlanabilir bir etki büyüklüğüne sahip olmadığı görülmüştür. Dolayısıyla üç yönlü etkileşim etkisi ihmal edilerek iki yönlü etkileşim etkileri yorumlanmıştır. İki yönlü etkileşim etkileri incelendiğinde kesme puanı ile model arasındaki farklılığa ilişkin anlamlılığın yorumlanabilir etki büyüklüğüne sahip olduğu görülmektedir ($F_{(1,274, 2141,768)} = 45,228$, $p < 0,01$, $\eta_p^2 = 0,026$).

Kesme puan noktası ile model arasındaki etkileşim etkisi için kesme puanının düzeylerinde ikili karşılaştırma testleri ile modeller arası farklılıklar incelenmiştir. Kesme puanının $c(0,20)$ düzeyinde ATM'nin GAMOM'dan 0,007 farklılık %95 GA [0,005, 0,010] ile ($p < 0,01$), ATM'nin karma modelden 0,008 farklılık %95 GA [0,005, 0,011] ile ($p < 0,01$), GAMOM'un karma modelden 0,001 farklılık %95 GA [0,000, 0,002] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir ($p < 0,01$). Kesme puanının $c(0,80)$ düzeyinde ise GAMOM'un ATM'den 0,010 farklılık %95 GA [0,005, 0,015] ile ($p < 0,01$), GAMOM'un karma modelden 0,007 farklılık %95 GA [0,005, 0,010] ile ($p < 0,01$) daha iyi sonuç verdiği gözlenmiş olup ATM ile karma model arasında istatistiksel olarak farklılık gözlenmemiştir.

Sınıflama Doğruluklarına Yönelik Bulgular

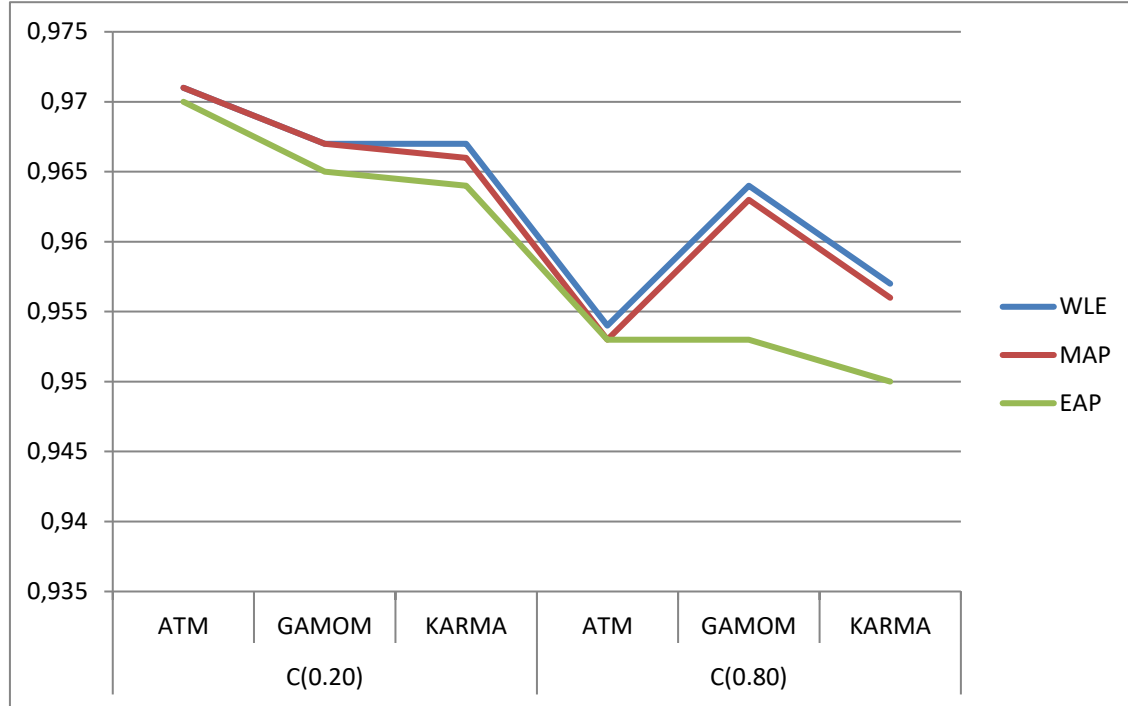
Tablo 12'de yetenek ölçeğinin $c(0,20)$ ve $c(0,80)$ kesme puan noktalarında ATM, GAMOM ve karma model altında WLE, MAP ve EAP kestirim yöntemlerine dayalı olarak elde edilmiş sınıflama doğruluğu indeksleri yer almaktadır.

Tablo 12

c(0,20) ve c(0,80) Kesme Puan Noktalarında Elde Edilen Sınıflama Doğrulukları

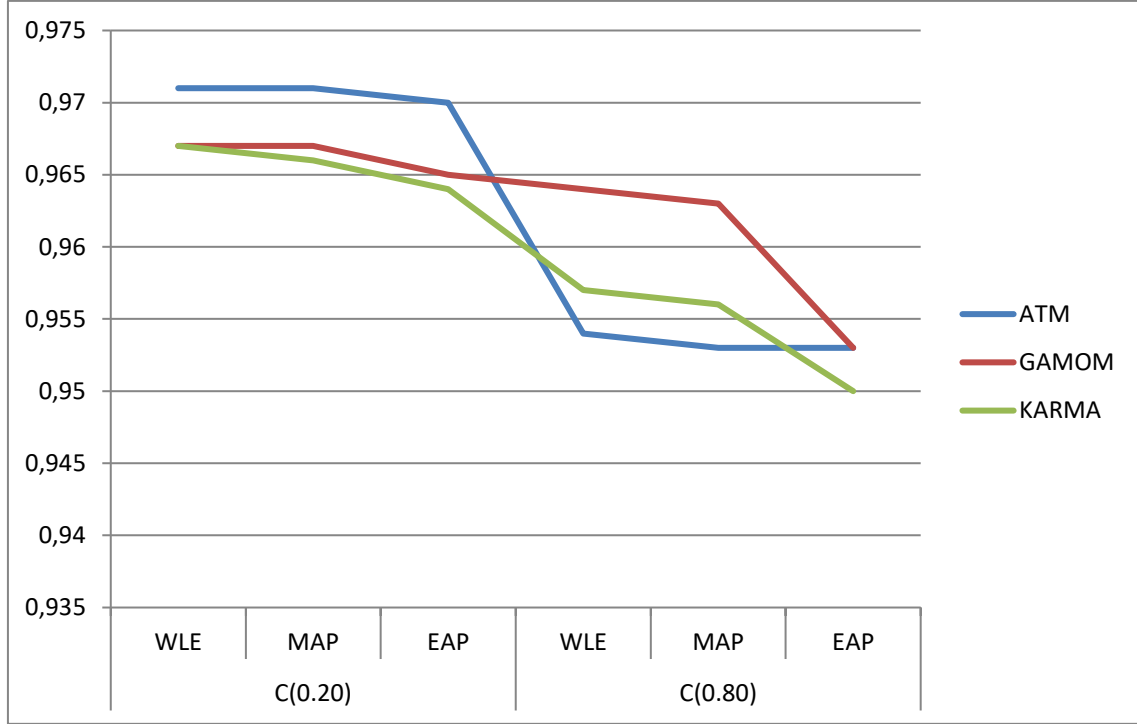
		WLE	MAP	EAP
C(0,20)	ATM	0,971	0,971	0,970
	GAMOM	0,967	0,967	0,965
	KARMA	0,967	0,966	0,964
C(0,80)	ATM	0,954	0,953	0,953
	GAMOM	0,964	0,963	0,953
	KARMA	0,957	0,956	0,950

Tablo 12’deki değerlere bağlı olarak her bir kesme puan noktasında yöntemler ve modeller açısından sınıflama doğruluğuna ilişkin değişimi gözlemlemek için grafikler elde edilmiş olup Şekil 14 ve Şekil 15’te sunulmuştur. Şekil 14’te her bir kesme puan noktasında model türüne bağlı olarak yöntemlerin değişimine ilişkin grafik yer alırken Şekil 15’te her bir kesme puan noktasında yöntem türüne bağlı olarak modellerin değişimine ilişkin grafik yer almaktadır:



Şekil 14. $c(0,20)$ ve $c(0,80)$ noktalarında yöntemlerin değişimi - $\hat{t}$

Şekil 14 incelendiğinde $c(0,20)$ kesme puan noktasında modeller altında üç yöntemin de değişiminin benzer olduğu; WLE ve MAP’a dayalı elde edilen sınıflama doğruluklarının birbirlerine daha yakın değerlerle sonuçlandığı ve EAP’a dayalı elde edilen sınıflama doğruluklarının diğer iki yöntemle nispeten daha küçük olduğu görülmektedir. Yüksek uçta yer alan $c(0,80)$ noktasında yine bütün modeller altında WLE ve MAP değişiminin benzer olduğu; WLE’nin MAP’a nispeten daha yüksek değerlerle sonuçlandığı gözlenmiştir. Ancak ATM altında yöntemler benzer sonuçlanmasına rağmen, GAMOM ve karma model altında EAP’ın WLE ve MAP’a göre belirgin bir düşüş sergilediği gözlenmiştir.



Şekil 15. c(0,20) ve c(0,80) noktalarında modellerin değişimi - $\hat{\tau}$

Şekil 15 incelendiğinde c(0,20) kesme puan noktasında ATM ile en iyi sınıflama doğruluklarının elde edildiği bunu GAMOM ve karma modelin takip ettiği gözlenmiş olup GAMOM ile karma modelin birbirine daha yakın sonuçlar sağladığı görülmektedir. Yetenek ölçeğinin yüksek ucunda yer alan c(0,80) kesme puan noktasında ise modellere ilişkin farklılığın yönünün değiştiği; WLE ve MAP altında GAMOM'un en yüksek sınıflama doğruluğu ile sonuçlandığı gözlenmiştir. GAMOM'u ise karma model takip etmektedir. EAP altındaki kestirimlerde GAMOM ve karma modelde düşüş yaşandığı, ATM altında elde edilen sonuçların ise yöntemlere göre değişiminin belirgin olmadığı gözlenmiştir.

Kesme puanının yeri, kestirim yöntemi ve model türünün sınıflama tutarlılığı üzerindeki etkisini belirlemek için üç yönlü tekrarlı ölçümler ANOVA yürütülmüştür. Elde edilen sonuçlar incelendiğinde kesme puanının yeri, yöntem ve model arasında istatistiksel olarak anlamlı üç yönlü etkileşim etkisinin olduğu ($F_{(1,776, 2985,598)} = 15,615$, $p < 0,01$, $\eta_p^2 = 0,009$) ancak bu etkileşim etkisinin pratikte yorumlanabilir bir etki büyüklüğüne sahip olmadığı görülmüştür. Dolayısıyla üç yönlü etkileşim etkileri ihmal edilerek iki yönlü etkileşim etkileri incelenmiştir. İki yönlü etkileşim etkileri incelendiğinde kesme puanı ile yöntem arasındaki ($F_{(1,004, 1687,012)} = 19,626$, $p < 0,01$, $\eta_p^2 = 0,012$); kesme puanı ile model arasındaki ($F_{(1,288, 2165,505)} = 22,878$, $p < 0,01$, $\eta_p^2 = 0,013$) ve yöntem ve model arasındaki

($F_{(1,580, 2656,590)} = 23,591$, $p < 0,01$, $\eta_p^2 = 0,014$), etkileşim etkilerinin anlamlı olduğu ancak pratikteki etki büyüklüklerinin yorumlanabilir olmadığı görülmektedir. Bu nedenle faktörlere ilişkin ana etkiler incelenmiş olup yöntemler arası farklılığın pratikte yorumlanabilir etki büyüklüğüne sahip olduğu görülmüştür ($F_{(1,002, 1684,699)} = 39,261$, $p < 0,01$, $\eta_p^2 = 0,023$).

Yöntemler arası farklılık incelendiğinde WLE'nin MAP'tan 0,0004 farklılık %95 GA [0,0003, 0,0005] ile ($p < 0,01$) WLE'nin EAP'tan 0,004 farklılık %95 GA [0,003, 0,005] ile ($p < 0,01$); MAP'ın EAP'tan 0,004 farklılık GA [0,002, 0,005] ile ($p < 0,01$) anlamlı olarak daha iyi sonuç verdiği gözlenmiştir.

Genel Yorum

Sınıflama doğruluğu ve sınıflama tutarlılıkları açısından kesme puan noktası, yöntem ve model değişkenleri açısından karşılaştırmalar yapılmıştır. Sınıflama tutarlılığı açısından kesme puanı ve model değişkenleri arasındaki iki yönlü etkileşim etkisini anlamlı olduğu gözlenmiştir. Söz konusu farklılığa ilişkin karşılaştırmalar yapıldığında yetenek ölçeğinin düşük ucundaki kesme puan noktasında ATM'nin GAMOM'dan ve karma modelden daha iyi sonuç verdiği gözlenmiştir. Ancak yetenek ölçeğinin yüksek ucunda yer alan kesme puan noktasında modeller arası farklılaşmanın yetenek ölçeğinin GAMOM'un lehine olduğu, GAMOM'un ATM ve karma modelden daha iyi sonuç verdiği ve modeller arası farklılaşmanın da artış gösterdiği gözlenmiştir. ATM ile karma model arasındaki farklılığın ise istatistiksel olarak anlamsız olduğu görülmüştür. Sınıflama doğruluğu açısından incelemeler yapıldığında ise üç yönlü ve iki yönlü etkileşim etkilerine ilişkin pratikteki etki büyüklüklerinin yorumlanabilir olmadığı gözlenmiştir. Bu nedenle ana etkiler incelenmiş olup ana etkilerden yöntem değişkeni açısından yorumlanabilir anlamlı sonucun olduğu gözlenmiştir. Yöntemlerin değişimine ilişkin ikili karşılaştırma testleri WLE'nin MAP ve EAP'tan; MAP'ın ise EAP'tan daha yüksek değerlerle sonuçlandığını göstermiştir.

BÖLÜM V

SONUÇ VE TARTIŞMA

Bu bölümde araştırmanın bulguları doğrultusunda ulaşılan sonuçlara ve bu sonuçların literatüre dayalı tartışmasına, uygulayıcılara ve araştırmacılara yönelik önerilere yer verilmiştir.

Sonuçlar

Bu araştırmanın amacı madde yanıt süreçlerini baskınlık ve ideal nokta yaklaşımlarına göre modelleyerek uygun model seçimi ve model tanımlaması ile bayesian ve nonbayesian yöntemler çerçevesinde yetenek kestirimlerinin doğruluğu ve farklı kesme puan noktalarında hesaplanan sınıflama indeksleri açısından ele almaktır. Bu doğrultuda araştırmada kümülatif modellerden ATM, ideal nokta modellerinden GAMOM kullanılmıştır. GAMOM altında kalibre edilen madde yanıt örüntülerinden elde edilen delta parametrelerine dayalı olarak monotonik-monotonik olmayan madde spesifikasyonları yapılmış ve karma model tanımlanmıştır. Her bir model kestirimi yapıldıktan sonra ilgili modeller altında WLE, MAP ve EAP yöntemleri kullanılarak yetenek kestirimleri elde edilmiş ve modellere bağlı olarak yetenek kestirimlerinin doğruluğu incelenmiştir. Yetenek kestirimlerinin elde edilmesinin ardından yetenek ölçeğinin düşük ve yüksek ucunda yer alan kesme puan çiftleri üzerinde sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri incelenmiştir. Bu amaç doğrultusunda kesme puan lokasyonunun farklılaşma etkisine bağlı olarak model ve yöntemlerin sınıflama doğruluğu ve sınıflama tutarlılığı üzerindeki etkileri ele alınmıştır.

Yetenek Kestirimlerinin Doğruluğuna İlişkin Sonuçlar

Çalışmanın birinci alt problemi kapsamında kümülatif, ideal nokta ve karma model altında bayesian ve nonbayesian yöntemlere dayalı elde edilen yetenek kestirimlerinin doğruluğu model ve yöntem türü açısından ele alınmıştır. Elde edilen sonuçlara göre yöntem ve model arasında anlamlı etkileşim etkisinin söz konusu olduğu gözlenmiştir. Etkileşim etkisine yönelik farklılıkları ele almak için ilk olarak her bir yöntem düzeyinde modellerin farklılıkları test edilmiştir. Farklılıklara yönelik sonuçlar değerlendirildiğinde MAP ve WLE altında sırasıyla pratikte yüksek ve orta etki ile modellerin farklılaştığı ve her iki yöntem altında değişimin yönünün aynı olduğu gözlenmiştir. Değişimin yönü incelendiğinde ATM'in GAMOM ve karma modelden; karma modelin ise GAMOM'dan daha yüksek standart hata ortalamalarına sahip olduğu görülmüştür. Dolayısıyla yetenek kestirimlerinin sırasıyla en doğru elde edildiği modeller GAMOM, karma model ve ATM olarak ele alınmıştır. Bu sonuç ideal nokta modeli ile yapılan kestirimlerin daha doğru olduğunu göstermektedir. İncelemelerin yürütüldüğü veri seti ele alındığında hem monotonik hem de monotonik olmayan maddelerin mevcut olduğu gözlenmektedir. Çalışmadaki veri setinde monotonik maddeler mevcut olmasına rağmen monotonik olmayan yaklaşımı benimseyen GAMOM ile elde edilen kestirimlerde genel doğruluğun diğer modellere göre daha yüksek olması GAMOM'ın kümülatif modellere göre daha esnek olması ile açıklanabilir (Wilgus & Travis, 2019; Cao vd., 2018). Stark, Chernyshenko, Drasgow ve Williams (2006) çalışmalarında çeşitli ölçekler ile elde edilmiş veri üzerinde ideal nokta ve kümülatif modelleri karşılaştırmalı olarak maddelerin monotonik olma-monotonik olmama durumlarına göre ele almışlardır. Araştırmacılar çalışmalarında elde edilen bulgular doğrultusunda monotonik olmayan maddenin bulunmadığı ölçeklerde/alt boyutlarda ideal nokta ve kümülatif modellere ilişkin uyumun neredeyse benzer olduğunu ortaya çıkarmışlardır. Monotonik olmayan maddelerin gözlemlendiği ölçeklerde/alt boyutlarda ise ideal nokta modellerinin daha iyi uyum sergilediğini göstermişlerdir. Bunun yanı sıra bütün maddelerin monotonik olduğu ölçek/alt boyut üzerinden ideal nokta ve kümülatif modellerden elde edilen yetenek kestirimlerinin neredeyse mükemmel eşleşmesinin söz konusu olduğu gözlenmiştir. Chernyshenko, Stark, Drasgow ve Roberts (2007) ideal nokta sürecine dayalı geliştirilen ölçeğin kümülatif model ile kestirildiğinde; kümülatif sürece dayalı geliştirilen ölçekten elde edilen verinin ise ideal nokta süreci ile kestirildiğinde elde edilen sonuçları 2PLM ve GAMOM altında incelemiştir. Bulgular kümülatif sürece dayalı geliştirilen ölçek verisi

için 2PLM yerine GAMOM kullanıldığında da sonuçların aynı kaldığını ancak ideal nokta sürecine dayalı geliştirilen ölçek verisine dayalı kestirimlerde GAMOM yerine 2PLM'in kullanılmasının uygun olmayacağını göstermiş olup GAMOM'un esnekliğini ortaya çıkarmıştır. Cao, Song ve Tay (2018) da çalışmasında benzer şekilde örneklem büyüklüğü koşulunu da ele alarak kümülatif ve ideal nokta modellerine göre üretilmiş verileri çapraz bir şekilde karşılaştırmıştır. Çalışma sonuçları kümülatif yaklaşıma göre veri üretildiğinde küçük örneklemde GAMOM'un kümülatif yaklaşıma göre daha düşük doğrulukla sonuçlandığını ancak örneklem büyüklüğü orta veya yüksek olduğunda ideal nokta yaklaşımının kümülatif yaklaşım kadar doğru sonuçlar sağladığını göstermiştir. Bu çalışmalar genel olarak değerlendirildiğinde, ölçme aracında bütün maddelerin monotonik olduğu durumda bile ideal nokta modellerinin kümülatif modellerle karşılaştırılabilir olduğunu ancak tam tersi durumun her zaman geçerli olmadığını göstermektedir.

GAMOM ve ATM karşılaştırmasının yanı sıra karma model altında elde edilen yetenek kestirimlerinin doğruluğunun ATM'ye göre daha yüksek olmasına rağmen GAMOM'dan daha düşük olduğu ancak GAMOM'a daha yakın sonuçlar ürettiği görülmektedir. Yani karma model altında monotonik olmayan madde spesifikasyonun yapılması ile sonuçlar iyileşerek GAMOM'a yaklaşmış ancak yine de karma model altında daha düşük doğruluk elde edilmiştir. Bu durum GAMOM altındaki madde parametre kestirim süreci ve ideal nokta modellerinin esnekliği ile ilişkilendirilebilir. Sun, Fraley ve Drasgow (2020) GAMOM altında kestirilen madde yanıt fonksiyonunun maddenin eğimi ve lokasyonu üzerinde değişim gösterebildiğini belirtmişlerdir. Dolayısıyla yer parametresi yüksek olduğunda da GAMOM'un kümülatif veriye de uyum gösterme yeteneğinde olduğu ve kümülatif modellere yakın bir kestirim (approximation) sağlayabileceği ele alınmıştır. Karma modelin ATM'den daha yüksek doğrulukla sonuçlanması ise ölçme aracında yer alan monotonik olmayan maddelerle ilişkilendirilebilir. Çünkü kümülatif modellerin monotonik olmayan modellerin mevcut olduğu durumda, kestirimlerde iyi sonuç veremeyebileceği çalışmalarla gösterilmiştir (Chernyshenko vd., 2007).

Yetenek kestirimlerinin doğruluğu açısından her bir yöntem düzeyinde modeller karşılaştırıldıktan sonra her bir model altında yöntemlerin farklılaşma düzeyleri ele alınmıştır. Elde edilen sonuçlara göre yöntemlere yönelik farklılığın en yüksek GAMOM altında olduğu gözlenmiştir. GAMOM altında yöntemler karşılaştırıldığında en yüksek doğrulukla sonuçlanan yöntemlerin sırasıyla MAP, WLE ve EAP olduğu; MAP ve WLE birbirine daha yakın değerlerle sonuçlanırken EAP'ın yüksek düzeyde standart hatayla

sonuçlandığı gözlenmiştir. Karma model altında da GAMOM'a benzer sonuçlar elde edilmiş ancak pratikteki etki büyüklüğünün GAMOM'dan düşük olduğu gözlenmiştir. Ancak ATM altında yetenek kestirimlerinin doğruluğunun yöntemlere göre değişimi ele alındığında sırasıyla bayesian yöntemlerden MAP ve EAP ile en doğru kestirimlerin elde edildiği; WLE'nin ise diğer yöntemlere göre daha yüksek standart hata ve düşük doğrulukla sonuçlandığı pratikteki etki büyüklüğünün ise yorumlanabilir sınırın altında olduğu görülmüştür. Genel olarak modellere bağlı olarak yöntemlerin değişiminin ve etkililiklerinin farklılaştığı gözlenmiştir. Ancak bütün modeller arasında yöntemlere ilişkin en düşük farklılaşmanın ATM düzeyinde olduğu gözlenmiştir. Wang ve Vispoel (1998) çalışmasında CAT simülasyonlarına dayalı olarak nonbayesian yöntemlerden MLE kestirim yöntemi ile bayesian yöntemlerden MAP, EAP ve OWEN'in yöntemini karşılaştırmıştır. Çalışma sonuçları MLE kestirim yöntemi ile bayesian kestirim yöntemleri arasındaki farklılığın oldukça belirgin olduğunu; MLE yönteminin daha yüksek standart hata değerleriyle sonuçlandığını göstermiştir. Bayesian yöntemler kendi arasında karşılaştırıldığında gerçekçi madde havuzuna dayalı yapılan kestirimlerde MAP'ın EAP'a göre daha düşük standart hata değerleri ile sonuçlandığı görülmektedir. Wang ve Wang (2001) çalışmasında kısmi puan modeli altında kestirim yöntemlerinden WLE, MAP ve EAP yöntemlerini karşılaştırmış ve çalışmalarından elde edilen sonuçlara göre WLE'nin MAP ve EAP'tan daha yüksek standart hata ile sonuçlandığını ancak yöntemler arasındaki ortalama farklılığın istatistiksel olarak anlamlı olmadığını göstermişlerdir. Borgatto, Azevedo, Pinherio ve Andrade (2015) çalışmasında dört farklı test uzunluğunda (15, 30, 45 ve 60 madde) yetenek kestirim yöntemlerinden MAP, EAP, MLE ve WLE kestirim yöntemlerini üç parametrelili lojistik model altında karşılaştırmıştır. Çalışmadan elde edilen bulgular testin uzunluğu arttıkça kestirimlerin doğruluğunun da arttığını; 30 madde ve üzerindeki test uzunluklarında bütün yöntemlerin benzer standart hata ve benzer yanlılıkla sonuçlandığını göstermiştir. Kümülatif modeller altında bayesian ve nonbayesian yöntemlerin parametre kestirimleri açısından karşılaştırıldığı bu çalışmalar incelendiğinde bayesian yöntemlerin nonbayesian yöntemlere göre daha düşük standart hata ile sonuçlandığı ancak belli bir doygunluktaki test uzunluklarında daha küçük standart hata ile sonuçlanmasına rağmen sonuçların bayesian ve nonbayesian yöntemler açısından benzer olduğu görülmektedir. Bu sonuçlar çalışmada elde edilen ATM ve kestirim yöntemleri arasındaki değişim ilişkisine dair sonuçları desteklemektedir. Ancak araştırmada monotonik olmayan maddelerin dahil edildiği GAMOM ve karma modelde yöntemlerin

kümülatif modele göre daha farklı sonuçlandı; MAP ile en doğru sonuçlar sağlanırken, EAP ile elde edilen yetenek kestirimlerinin daha yüksek standart hatayla en düşük doğruluğa neden olduğu gözlenmiştir. Roberts, Donoghue ve Laughlin (2002) EAP yöntemi ile birey parametrelerini kestirerek yöntemin etkililiğini simülasyon çalışması ile ideal olan ve ideal olmayan koşullarda test etmiştir. Çalışma sonuçlarına göre genel olarak EAP ile doğru yetenek kestirimleri elde edilmesine rağmen uç bireylerin olduğu/uç yanıt örüntülerinin olduğu duruma bunun genellenemeyeceğine dikkat çekmiştir. Mevcut araştırmadaki veri seti ele alındığında mükemmel puan örüntüsüne sahip bireylerin olduğu gözlenmiştir. EAP kestirim yönteminin uç yanıt örüntülerinden etkilendiği göz önünde bulundurulursa EAP'ın daha yüksek hata ile sonuçlanması beklenen bir durumdur. Bunun yanı sıra MAP ve EAP kestirim yöntemlerinin altında yer alan algoritmalar incelendiğinde MAP ve EAP ile elde edilen kestirimlerin veri setinin özelliklerine bağlı olarak farklılık göstermesi söz konusu olabilmektedir. Dağılımın merkezine çekilme etkisine temelli olarak MAP dağılımının moduna doğru çekilme etkisi gösterirken EAP dağılımının ortalamasına çekilme etkisi göstermektedir (Han, 2016). Sonsal yetenek dağılımının çarpık olduğu durumlarda modun kullanılması ortalamadan kullanılması göre daha makul bir kestirim olarak değerlendirilmektedir (Williams, 2017). Gerçek veri durumlarında ideal normal dağılımın sağlanmasının pek mümkün olmadığı ve normallikten sapma durumlarıyla karşılaşma olasılığından dolayı MAP ile elde edilen yetenek kestirimleri EAP'tan daha doğru sonuçlar verebilir. Bunun yanı sıra WLE ve MAP iteratif bir süreç (Warm, 1989; Rose, 2010) dayalı olarak yetenek kestirimi yaparken, EAP quadrat noktalarının kullanıldığı non-iteratif bir süreç (Bock & Mislevy, 1982) dayalı olarak yetenek kestirimlerini gerçekleştirmektedir. Ancak ideal nokta kestirim süreci ele alındığında parametrelerin iteratif bir süreç ile elde edildiği gözlenmektedir (Roberts, Donoghue & Laughlin, 1998). Model ve yöntemlerin kestirim süreçlerindeki eşleşmenin EAP ile daha düşük düzeyde olmasından dolayı da EAP'ın yüksek standart hata ve daha düşük doğrulukla sonuçlanmış olması muhtemeldir.

Yetenek kestirimlerinin doğruluğuna yönelik incelemenin ardından GAMOM, ATM ve karma model altında WLE, MAP ve EAP ile elde edilen yetenek kestirimlerine dayalı olarak sınıflama tutarlılığı ve sınıflama doğruluğu indeksleri üç farklı kesme puan çifti üzerinde ayrı ayrı elde edilmiş olup sonuçlar tartışılmıştır.

Sınıflama İndekslerine Yönelik Sonuçlar

Yetenek ölçeğinin düşük ve yüksek ucunda % 5’lik değişimlerle oluşturulan kesme puan noktalarında elde edilen sınıflama indekslerine yönelik sonuçlar kesme puan çiftlerine göre sistematik bir değişimin olduğunu göstermiştir. Bu nedenle sınıflama indekslerinin ele alındığı üç araştırma sorusu kapsamında ortaya çıkan sonuçlara yönelik çıkarımlar karşılaştırmalı olarak ele alınmıştır.

Çalışmanın ikinci araştırma sorusu kapsamında yetenek ölçeğinin düşük ve yüksek olmak üzere $c(0,10)$ ve $c(0,90)$ lokasyonlarında yer alan kesme puan noktalarında model ve yöntemin sınıflama tutarlılıklarına ve sınıflama doğruluklarına olan etkisine yönelik incelemeler yapılmıştır. Elde edilen sonuçlara her iki indeks için de üç yönlü etkileşim etkisi anlamlı olduğu, model ve yöntemlere ilişkin farklılaşmaların da benzer olduğu gözlenmiştir. Bu doğrultuda $c(0,10)$ ve $c(0,90)$ kesme puan noktalarındaki sonuçlar sınıflama tutarlılığı üzerinden sunulmuştur. Üç yönlü etkileşim etkisinin değişimi düşük ve yüksek uçtaki kesme puan noktalarına göre yöntem ve model değişkenleri açısından incelendiğinde yetenek ölçeğinin düşük ucunda farklılaşmanın olmadığı ancak yüksek ucunda yöntemin düzeylerine bağlı olarak modellerin; modellerin düzeylerine bağlı olarak da yöntemlerin farklılaştığı gözlenmiştir. İlk olarak yetenek ölçeğinin üst ucunda yer alan kesme puan noktasında model düzeyinde farklılıkların yönü incelenmiş olup ardından benzer şekilde yöntem düzeyinde farklılaşmalar ele alınmıştır.

Model düzeyinde yöntemlerin farklılıklarına ilişkin sonuçlar ele alındığında GAMOM ve karma model altında yöntemlerin istatistiksel olarak birbirinden farklılaştığı gözlenmiştir. GAMOM altında yöntemler karşılaştırıldığında bütün yöntem çiftleri arasında anlamlı farklılığın meydana geldiği ve en yüksek sınıflama tutarlılıklarının sırasıyla WLE, MAP ve EAP ile elde edildiği gözlenmiştir. Ancak yöntemler arasındaki ortalama farklılık miktarsal olarak incelendiğinde WLE ve MAP’ın birbirine yakın sonuçlar verdiği; EAP ile elde edilen sınıflama tutarlılığında ise diğer iki yöntemle göre belirgin düşüş olduğu gözlenmiştir. EAP’a yönelik elde edilen sonuçlar yetenek kestirimlerinin doğruluğuna ilişkin sonuçlar ile benzerlik göstermektedir. Çünkü GAMOM altında EAP ile daha yüksek standart hata ortalaması ve daha düşük doğrulukta yetenek kestirimleri elde edilmiştir. Yetenek kestirimlerinin standart hatası sınıflama indekslerini etkileyen önemli bir faktör olarak ele alınabilmektedir (Yang vd., 2006). Standart hata ortalamasının yüksek olması itibarıyla EAP’ ın daha düşük sınıflama indeksleri ile sonuçlanması beklenmektedir. ATM altında da benzer şekilde yöntem sıralamasının WLE, MAP ve EAP olarak sonuçlandığı

ancak yöntemler arasındaki miktarsal farklılığın düşük olduğu ve yöntemlerin yaklaşık benzer sonuç verdiği gözlenmiştir.

Çalışmanın üçüncü araştırma sorusu kapsamında $c(0,15)$ ve $c(0,85)$ lokasyonlarında yer alan kesme puan noktalarında model ve yöntemin sınıflama tutarlılıklarına ve sınıflama doğruluklarına olan etkisine yönelik yapılan incelemeler yapılmıştır. Buradan elde edilen sonuçlara göre sınıflama tutarlılığı açısından üç yönlü etkileşim etkisinin anlamlı olduğu, sınıflama doğruluğu açısından ise kesme puanı ile yöntem ve model ile yöntem arasındaki iki yönlü etkileşim etkilerinin anlamlı olduğu gözlenmiştir. Buradaki sonuçların $c(0,10)$ ve $c(0,90)$ kesme puan çiftinin ele alındığı durumdaki sonuçlar ile benzer olduğu gözlenmiştir. Ancak yetenek ölçeğinin orta kısımlarına daha yakın olan $c(0,15)$ ve $c(0,85)$ kesme puan noktalarına dayalı yapılan bu incelemede pratikteki etki büyüklüğünde düşüş olduğu gözlenmiştir. Son olarak yetenek ölçeğinin $c(0,20)$ ve $c(0,80)$ noktalarında elde edilen üç yönlü ve iki yönlü etkileşim etkilerinin daha düşük sonuçlanmasından dolayı yorumlanabilir etkiye sahip olmadığı sadece yöntem ana etkisinin anlamlı olduğu gözlenmiştir. Yönteme ilişkin buradan elde edilen sonuçlar WLE' nin MAP ve EAP'tan; MAP'ın ise EAP'tan daha iyi sonuç verdiğini göstermiştir.

Üç farklı kesme puan setine ilişkin sonuçlar karşılaştırmalı olarak değerlendirildiğinde, kesme puan noktalarının uçlara yakın olduğu durumlarda yetenek ölçeğinin düşük ve yüksek uçlarında yer alan kesme puan noktalarına göre model ve yöntem farklılaşmalarına ilişkin etkileşim etkisinin farklılaştığı gözlenmiştir. Kesme puan lokasyonları orta kısımlara yaklaştıkça üç yönlü etkileşim etkisine ilişkin etki büyüklüğünün düşüş gösterdiği gözlenmiştir. Bu doğrultuda iki yönlü etkileşim etkisine ilişkin sonuçların yorumlandığı durumda ise yine modellerin kesme puan noktasına ve yöntemle göre farklılaştığı, etkileşim etkilerinin anlamlı olmadığı durumlarda ise yöntemlere ilişkin ana etkilerin birbirinden farklılaştığı gözlenmiştir.

Etkileşim etkilerine yönelik sonuçlar incelendiğinde, yöntem ve modellerin sınıflama indeksleri üzerindeki etkililiğinin yetenek ölçeği üzerindeki kesme puan noktasına bağlı olarak farklılaştığı gözlenmiştir. Birçok çalışma (Betebenner vd. 2008; Lathrop & Cheng, 2014; Wyse & Babcock, 2016; Zhang, Du Chen, Xin & Chen, 2017) kesme puanının sınıflama indekslerini etkileyen önemli bir faktör olduğuna işaret etmektedir. Betebenner vd. (2008) koşullu yanlış sınıflama olasılıklarının hem ölçme aracı hem de bireylerin performans düzeylerini tanımlamak için kullanılan kesme puanlarına olan göreceli yerleştirilmelerine yönelik bilgi içerdiğini belirtmektedir. Kesme puan noktasının sınıflama

indekslerine olan etkisi ise yetenek ölçeğinin özelliklerine bağlı olarak da değişebilmektedir. Wyse ve Babcock (2016) kesme puanının yerine göre yetenek dağılımının ortalaması, madde sayısı, çeşitli yetenek (θ) noktalarında sağlanan test bilgi fonksiyonuna ilişkin koşulları manipüle ederek çeşitli senaryolar oluşturmuş ve sınıflama doğruluğu ve tutarlılıklarını bu senaryolarda incelemişlerdir. Çalışmadan elde edilen bulgulara göre kesme puanı yerine başka bir θ noktasında testin bilgisini arttırmanın sınıflama doğruluğu ve sınıflama tutarlılığını arttırma konusunda daha iyi sonuç verdiği sonucuna ulaşılmıştır. Bu durum ise kesme puanının yeri ile yetenek dağılımının etkileşiminin bir sonucu olarak ele alınmıştır. Lathrop ve Cheng (2014) çalışmalarında parametrik yaklaşımlardan Lee (2010) ve Livingston ve Lewis'in (1995) yaklaşımlarını kullanılarak elde edilen sınıflama indeksleri ile nonparametrik yaklaşıma dayalı sınıflama doğruluğu ve tutarlılığı indekslerini çeşitli simülasyon koşullarında test etmişlerdir. Simülasyon koşullarının oluşturulmasında ele alınacak madde tepki kuramı modeli, test uzunluğu ve yetenek dağılımlarına ilişkin faktörler manipüle edilmiştir. Çalışma sonuçlarına göre ise Livingston ve Lewis (1995) yaklaşımının dağılımın normal olmadığı koşullardan etkilendiği ve kesme puanı ile yetenek dağılımı arasındaki etkileşime dayalı olarak yanlış sonuçlar verdiği gözlenmiştir. Çalışmalar genel olarak değerlendirildiğinde kesme puan noktasının sınıflama indekslerini etkileyen önemli bir faktör olduğu ve araştırmadaki diğer koşullar bağlamında yetenek ölçeği üzerindeki lokasyona göre sınıflama indekslerini etkileyebileceği görülmektedir. Bu çalışmada farklı model ve yöntemlerin kullanılmasından dolayı da yetenek kestirimlerinin ve yetenek kestirimlerine koşullu olarak da standart hata değerlerinin farklılaştığı gözlenmiştir. Dolayısıyla yetenek ölçeği üzerinde farklı lokasyonlara konumlandırılmış kesme puan noktalarına dayalı olarak sınıflama indekslerinin farklılaşması pratikte karşılaşılabilecek olası durumlardan biri olarak ele alınabilir.

Yetenek ölçeği üzerinde kesme puan noktasının konumlandırılmasına bağlı olarak değişimin yönü ele alındığında yetenek ölçeğinin düşük uçlarında yer alan kesme puan noktalarında farklılığın anlamsız veya miktarsal olarak düşük olduğu ancak yetenek ölçeğinin yüksek uçlarında istatistiksel olarak anlamlı model ve yöntem farklılaşmalarının ortaya çıktığı gözlenmiştir. Yetenek ölçeğinin üst kısımlarında modeller arasındaki farklılığın yönü incelendiğinde sonuçların GAMOM'un lehine olduğu görülmektedir. Bunu ise karma modelin takip ettiği, ATM'nin ise en düşük sınıflama indeksleri ile sonuçlandığı görülmektedir. Bu durum maddelerin sergiledikleri karakteristik eğrilerinin

yetenek ölçeğinin belli bir kısmına kadar benzer olması ancak belli bir noktadan sonra farklılaşmanın ortaya çıkması ile açıklanabilir. Detaylı olarak ele alınacak olursa monotonik ve monotonik olmayan madde karakteristik eğrileri yetenek ölçeğinin orta kısmına kadar benzer değişime sahipken orta kısımdan sonra modeller arası farklılaşmalar söz konusu olabilmektedir. Çünkü kümülatif model yetenek düzeyine bağlı olarak artış göstermeye devam ederken ideal nokta modeli belli bir noktada (genellikle yetenek ölçeğinin orta kısımlarında) zirve yapıp düşüş sergilemektedir (Roberts, 2008; Speer vd., 2016). Dolayısıyla madde karakteristik eğrilerinin değişiminin ve maddelerin monotonik olmama derecelerine bağlı olarak modeller arası farklılaşmanın yetenek ölçeğinin lokasyonuna bağlı olarak ortaya çıkması olasıdır. Ancak maddelerin monotonik olmama derecesine bağlı olarak hangi noktadan itibaren sınıflama indeksleri açısından modeller arası farklılaşmanın ortaya çıkacağına ilişkin kesme puan noktasının değişimi de buna göre değerlendirilmelidir. Kümülatif ve ideal nokta modellerine dayalı yapılan çeşitli çalışmalarda birey sıralamalarının yetenek ölçeğinin üst kısımlarında farklılaştığı gösterilmiştir (Cao vd., 2014; Roberts vd., 1999; Zampetiks vd., 2015; Stark vd., 2006; Ling vd., 2016). Roberts vd (1999) ideal nokta özelliği gösteren verinin kümülatif modellerle analiz edildiğinde ilgilenilen yetenek açısından yüksek düzeyde olan bir kısım bireylerin yanlış bir şekilde yetenek ölçeğinin orta kısımları ile ilişkilendirilebileceğini hem simülasyon hem de ampirik sonuçlar ile göstermiştir. Zampetiks vd. (2015) tipik davranış ölçümünde GAMOM ve ATM'ye göre karşılaştırma yapmıştır. Çalışmalarından elde edilen sonuçlara göre her iki modelin de veriye uyduğu ancak yetenek kestirimleri arasındaki ilişkinin düşük olduğu belirtilerek iki model arasındaki uyumsuzluğun daha çok üst yetenek düzeylerinde meydana geldiği saçılma grafiği ile gösterilmiştir. Cao vd. (2014) çalışmalarında ölçme aracındaki monotonik olmayan maddelerin monotonik maddelere olan oranına dayalı olarak performans değerlendirmesini yapmayı amaçlamışlardır. Ölçme aracındaki maddelerin %0, %25 ve %50' si monotonik olmayan maddelerden oluşan üç form üzerinde model performanslarına yönelik karşılaştırma yapılmıştır. Çalışmalarından elde edilen sonuçlar monotonik olmayan madde yüzdesinin %50 olduğu durumda, monotonik olmayan madde yüzdesininin %0 ve %25 olduğu durumlara göre $\theta = 1,5$ ve üzerindeki noktalarda belirgin iyileşmelerin olduğu yani üst yetenek düzeyinde GAMOM'un ATM'e göre daha yüksek ölçme doğruluğu ile sonuçlandığı belirtilmiştir. Bu çalışmalardan elde edilen sonuçlar genellikle yetenek kestirimleri açısından modelleri karşılaştırmıştır. Yetenek kestirimlerinin sınıflama indekslerinin önemli bir fonksiyonu

olduğu hususu ele alınacak olursa buradaki sonuçların çalışmadaki sınıflama indeksleri ile benzer olması beklenir. Dolayısıyla bu sonuçlar ile araştırma kapsamındaki sonuçlar desteklenmektedir.

Genel olarak sınıflama indekslerinin elde edilmesinde GAMOM diğer modellere göre üst kısımlardaki farklılık sergilemesine rağmen modeller arası farklılaşmanın pratikteki etkilerinin düşük etkiler olarak ortaya çıktığı görülmüştür. Bu durum ise maddelerin monotonik olmama ve monotonik olma dereceleri ile ilişkilendirilebilir. Monotonik olmayan olarak değerlendirilen maddeler ilgili delta aralığında yer almasına rağmen bazı maddelerin sıfırdan oldukça uzaklaştığı yani monotonik olmama derecesinin düşük olabileceği düşünülebilir. Çünkü monotonik olmayan maddeler yetenek ölçeğinin orta kısımlarında yayılmakta olup maddelerin lokasyonu sıfıra yakın oldukça monotonik olmama derecesi artarken sıfırdan uzaklaştıkça bu derecesi azalmaktadır (Cao vd., 2014). Araştırmada bazı maddeler monotonik olmayan olarak tanımlansa bile bu maddelerin sıfırdan uzaklaşma derecesinin modeller arası farklılıkların gözlenmesinde etki derecesini değiştirdiği söylenebilir. Bunun yanı sıra araştırmada yer alan monotonik maddelerin büyük çoğunluğunun çok uç madde kategorisinde (Ling vd., 2016) yer alması da monotonik olmayan maddeleri baskılamış ve dolayısıyla modeller arasında gözlenen etkiyi düşürmüş olabilir.

Sınıflama Tutarlılığı ve Sınıflama Doğruluğuna İlişkin Karşılaştırmalı Sonuçlar

Sınıflama doğruluğu ve tutarlılığı indeksleri karşılaştırmalı değerlendirildiğinde kesme puanı, yöntem ve modelin birleşiminin sınıflama doğrulukları ve sınıflama tutarlılıklarına olan etkisinin benzer bir şekilde değişim gösterdiği ancak istatistiksel manidarlık sonuçlarındaki pratik etkilerde sınıflama tutarlılıklarında sınıflama doğruluğuna göre artış olduğu gözlenmiştir. Çalışmadan elde edilen bu sonuç sınıflama doğruluğu ve sınıflama tutarlılığının hesaplanmasında ortak kullanılan beklenen olasılıklar matrisi ($\hat{P}$) ile ilişkilendirilebilir. $\hat{P}$ matrisi her bir performans – düzey kategorisindeki puanlamanın beklenen olasılıklarını içerir. Bu olasılıklar yetenek kestirimleri ve standart hata değerlerine dayalı olarak oluşturulmaktadır. Sınıflama doğruluğu $\hat{P}$ matrisi ile bireylerin puanlarına göre performans düzey kategorilerinde sınıflanmasına ilişkin ağırlıklandırma matrisinin (W) çarpımına dayanmaktadır. Ancak sınıflama tutarlılığı $\hat{P}$ matrisinin kendisi ile çarpımına dayanmaktadır (Wyse & Hao, 2012). $\hat{P}$ matrisinde yer alan elementlerin

standart hataya dayalı oluşturulmasından kaynaklı olarak sınıflama doğruluğu ve sınıflama tutarlılığı arasındaki farklılaşma düzeyi ortaya çıkmaktadır. Bu çalışmadaki sınıflama tutarlılığındaki farklılaşma etki düzeyinin sınıflama doğruluğuna nispeten yüksek olması standart hataların yöntem ve model açısından anlamlı olarak farklılaşması ve standart hatanın sınıflama tutarlılığı indeksinde daha baskın olması ile ilişkilendirilebilir. Farklı bir ifadeyle standart hatalarda gözlenen farklılıklar sınıflama tutarlılığında baskın olması itibariyle model ve yöntemler arası farklılaşmayı vurgulamıştır.

Yetenek Kestirimlerinin Doğruluğuna ve Sınıflama İndekslerine İlişkin Karşılaştırmalı Sonuçlar

Çalışmada yetenek kestirimlerinin doğruluğu ve sınıflama indekslerine ilişkin sonuçlar karşılaştırmalı değerlendirildiğinde bayesian ve nonbayesian yöntemlerine ilişkin farklılaşmanın sınıflama indekslerinde yön değiştirdiği gözlenmiştir. Yetenek kestirimlerinde bayesian yöntemler daha düşük standart hata ortalamasına sahip olmasına rağmen sınıflama indekslerinde nonbayesian yöntem öne çıkmıştır. Bu durum yetenek kestirimlerinin yetenek ölçeği üzerindeki yayılımı ile ilişkilendirilebilir. Bayesian ve nonbayesian yöntemlere dayalı yetenek kestirimlerinin ranjı incelendiğinde nonbayesian yöntemlerde çekilme etkisi (Han, 2016) ile ranjda daralma gözlenirken buna karşı nonbayesian yöntemlerden WLE'nin daha geniş bir ranjda yetenek kestirimleri ile sonuçlandığı gözlenmiştir. Çelikten ve Çakan (2019) MLE, WLE, MAP ve EAP'a dayalı elde edilen yetenek kestirimleri üzerinde sınıflama indekslerini karşılaştırmış olup sınıflama indekslerinin yetenek ölçeğinin yayılımına bağlı olarak yöntemlerin değiştiğini; bayesian yöntemlere dayalı elde edilen sınıflama indekslerinde nonbayesian yöntemlere göre düşüş olduğunu göstermişlerdir. Bu durum daha dar bir ranjda yapılan sınıflama çalışmalarında bireylerin kesme puan noktalarında daha yakın mesafelerde konumlandırılması ve dolayısıyla bireylerin yanlış sınıflama olasılığına bağlı artan sınıflama hatası ile ilişkilendirilebilir. Çünkü sınıflama çalışmalarında bireylerin kesme puan noktalarına yaklaşma derecelerine bağlı olarak sınıflama doğrulukları ve tutarlılıkları değişmekte, kesme puan noktasına yaklaşan birey sayısı arttıkça indeksler düşüş göstermektedir (Ercikan & Julian, 2002; Lathrop & Cheng, 2014; Wyse, 2011). Bayesian yöntemlerde ise çekilme etkisinden dolayı bireylerin sistematik bir şekilde yetenek kestirimleri ile kesme puanları arasındaki mesafe azalmaktadır. Dolayısıyla

yapılan sınıflama sonuçlarında bireylerin yanlış gruba atanma ihtimali artmakta ve sınıflama indeksleri azalabilmektedir. Ancak yöntemler karşılaştırmalı olarak değerlendirildiğinde WLE en yüksek sınıflama indeksleri ile sonuçlanmasında rağmen MAP'ın da WLE'ye çok yakın ve hatta bazı durumlarda aynı sonuçlar ürettiği gözlenmiştir. farklı bir ifadeyle belirtilecek olursa sınıflama indekslerinde gözlenen yöntemler arası farklılaşmanın yetenek kestirimlerinde gözlenen farklılığa göre daha az olduğu sonucuna ulaşılmıştır. Bu durum sınıflama indekslerinin çeşitli faktörlerin bir fonksiyonu olması ve ilgili koşul çerçevesinde söz konusu faktörlerin olası etkileşim etkileri (Montgomery & Skorupski, 2012) ile ilişkilendirilebilir. Bununla birlikte sınıflama indekslerinde kümülatif ve ideal nokta modeller fark etmeksizin WLE ve MAP'a dayalı elde edilen sınıflama indekslerinin karşılaştırılabilir olmasına rağmen EAP'ın yetenek ölçeğinin üst kısmındaki sınıflamalarda GAMOM altında yüksek hataya bağlı olarak düşük sınıflama doğruluğu ve tutarlılığı ile sonuçlandığı gözlenmiştir. EAP'a dayalı bu sonuç yetenek kestirim yöntemlerinde tartışılan durumların bir neticesi olarak değerlendirilebilir. Özet olarak çalışmada elde edilen sonuçlara göre yöntem ve modellerin sınıflama indekslerine olan etkisinin yetenek ölçeğinin düşük ve yüksek uçlarında yer alan kesme puan noktalarına bağlı olarak değiştiği gözlenmiştir. Yöntem ve model farklılaşmalarının etkileşimi yetenek ölçeğinin yüksek uçlarında öne çıkarken düşük uça etkileşimin olmadığı gözlenmiştir. Yetenek ölçeğinin yüksek ucunda etkileşime yönelik yapılan incelemeler kapsamında modellerin birbirinden farklılaştığı ve modellerin etkililiğinin ise kullanılan kestirim yöntemine bağlı olarak değiştiği sonuçlarına ulaşılmıştır. En yüksek sınıflama indekslerinin WLE ve MAP altında GAMOM ile elde edildiği gözlenmiş olup EAP'ın sınıflama indekslerinde düşüşe neden olduğu gözlenmiştir. Bununla birlikte yöntemlerden genel olarak WLE'nin düşük farklılıkla yöntemler arasında en yüksek sınıflama indeksleri vermesine karşın MAP'ın WLE'ye çok yakın veya aynı değerler ürettiği gözlenmiştir. Ancak yetenek kestirimlerinin doğruluğuna yönelik incelemelerde en iyi sonuç veren yöntemin MAP olduğu gözlenmiştir. Yani yetenek kestirimlerinin doğruluğuna yönelik incelemede bayesian yöntem öne çıkarken sınıflama indekslerine yönelik incelemede nonbayesian yöntem öne çıkmıştır. Bu durum sınıflama indekslerinin çeşitli faktörlerin bir fonksiyonu olarak elde edilmesinin bir sonucu olarak değerlendirilmiştir. Bu doğrultuda yöntemlerin kestirimlerinin yapıldığı puan varyansının yöntemlerin yön değiştirmesindeki faktörlerden biri olabileceği ele alınmıştır.

Öneriler

Çalışma kapsamında elde edilen sonuçlar doğrultusunda uygulayıcılara ve araştırmacılara yönelik öneriler sunulmuştur.

Uygulayıcılara Yönelik Öneriler

- Çalışma kapsamında elde edilen sonuçlar, bilişsel olmayan yapıların ölçülmesinde ölçme aracında yer alan maddelerin yanıt süreçlerinin MTK çerçevesinde modellenerek buna uygun model seçiminin yapılmasının yetenek kestirimlerinin doğruluğunu arttırdığını göstermiştir. Uygulayıcıların bireylere yönelik yetenek kestirimlerini yapmadan önce kullandıkları ölçme aracına yönelik yapıyı dikkate alması, bu doğrultuda madde yanıt süreçlerini grafiksel ve istatistiksel yaklaşımlarla modelleyerek maddelerin davranışlarını ortaya çıkarması ve bu doğrultuda model seçimini yapması önerilmektedir.
- Çalışma kapsamında ele alınan ATM, GAMOM ve karma modelin yetenek kestirimlerine olan etkisinin kullanılan kestirim yöntemine bağlı olarak farklılaştığı, GAMOM ve karma model kullanımının söz konusu olduğu durumda EAP kestirim yönteminin yüksek hata ile sonuçlandığı bulunmuştur. Bu sonuç ise mevcut veri setinin sahip olduğu özellikler bağlamında ele alınmış olup uç yanıt örüntülerinin ve veri setinin normallikten sapma derecelerinin EAP yönteminin yüksek hatayla sonuçlanmasıyla ilişkilendirilmiştir. Dolayısıyla uygulayıcılara yetenek kestirimlerini yaparken her bir model altında elde ettikleri veri setinin niteliklerini dikkate almaları ve veri setinde tamamı doğru veya tamamı yanlış yanıt örüntülerine sahip bireylerin olduğu durumlarda veya veri setinin normallikten olası sapma durumlarında EAP kestirim yöntemi yerine MAP veya WLE'yi kullanmaları önerilmektedir.
- Çalışmada model seçiminin sınıflama indekslerine olan etkisinin kesme puanının yetenek ölçeği üzerindeki konumuna bağlı olarak değişebileceği gösterilmiştir. Monotonik olmayan maddelerin söz konusu olduğu GAMOM ve karma modelin yetenek ölçeğinin üst kısımlarında ATM'den daha iyi sonuç verdiği; yöntemlerden ise WLE ve MAP'ın öne çıktığı gözlenmiştir. Dolayısıyla sınıflama çalışmalarında kesme puanının yerini gözeterek yetenek ölçeğinin üst kısımlarında GAMOM veya karma modeli tercih etmeleri önerilmektedir.

- Yetenek ölçeğinin üst kısımlarında belirlenen kesme puan noktalarının sırasıyla $c(0,90)$, $c(0,85)$ ve $c(0,80)$ şeklinde orta kısımlara yaklaşması itibariyle modeller arası farklılıkların azaldığı gözlenmiştir. Bu doğrultuda uygulayıcılara sınıflama çalışmalarını yürütmeden önce sınıflamada ele alınacak kesme puan noktasını gözeterek modellerin etkilerini gözlemlemesi ve buna göre model seçiminde bulunmaları önerilmektedir.
- Çalışmada sınıflama indekslerinin her bir faktörün bireysel etkisinden ziyade birden fazla faktörün etkileşimine bağlı olarak farklılaşabildiği bulunmuştur. Özellikle kesme puan noktalarının yetenek ölçeğinin uçlarına yaklaştığı durumda kesme puanı, model ve yöntem faktörlerinin etkileşim etkilerinin söz konusu olduğu gözlenmiştir. Bu doğrultuda uygulayıcıların yürütecekleri sınıflama çalışmalarında farklı yöntem ve model kullanımlarının söz konusu olduğu durumda bireylere yönelik karar verme çalışmasına kullanılacak kesme puanını dikkate alarak bütünsel bir yaklaşımla faktörleri eşzamanlı ele almaları önerilmektedir.

Araştırmacılara Yönelik Öneriler

- Çalışma kapsamında maddelerin kümülatif, ideal nokta ve karma modellerin yetenek kestirimlerinin doğruluğuna olan etkisinin bayesian ve nonbayesian yöntemler çerçevesinde farklılaştığı bulunmuştur. Mevcut veri seti yetenek kestirimleri açısından yaklaşık normal dağılım sergilemektedir. Dolayısıyla araştırmacıların pratikte karşılaşılabilecek farklı dağılımsal özellikler (çarpık, bimodal) açısından da incelemeleri yapması önerilmektedir.
- Çalışma kapsamında model ve yöntemlerin yetenek kestirimlerinin doğruluğuna ve sınıflama indekslerine olan etkisi karşılaştırmalı ele alındığında model arasındaki farklılığa ilişkin etki büyüklüğü yetenek kestirimlerinin doğruluğuna yönelik incelemede daha yüksek olmasına karşın sınıflama indekslerine yönelik incelemelerde etki bütünlüğünün düştüğü gözlenmiştir. Çalışmada incelenen ölçme aracındaki monotonik olmayan maddelerin delta parametrelerinin manipüle edilerek sıfıra yakın farklı derecelerdeki maddeler üzerinde incelemelerin yapılması önerilmektedir.
- Çalışma kapsamında incelenen ölçme aracında yer alan monotonik maddelerin büyük bir kısmının çok uç olduğu bulunmuştur. Bu durumun kestirim sonuçları

üzerinde etkili olduđu düşünöldöğü için monotonik maddelerin sadece uç veya sadece çok uç olduđu her bir durum için kestirimlerin yapıldığı ve madde etkisinin ortaya çıkarıldığı çalışmaların yapılması önerilmektedir.

-



•



KAYNAKLAR

- AERA, APA, & NCME (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Anderson Koenig, J., & Roberts, J. S. (2007). Linking Parameters Estimated With the Generalized Graded Unfolding Model: A Comparison of the Accuracy of Characteristic Curve Methods. *Applied Psychological Measurement*, 31(6), 504-524. <https://doi.org/10.1177/0146621606297315>
- Andrich, D., and Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, 17(3), 253–276. <https://doi.org/10.1177/014662169301700307>
- Andrich, D. (1988). The application of an unfolding model of the PIRT type for the measurement of attitude. *Applied Psychological Measurement*, 12(1), 33–51. <https://doi.org/10.1177/014662168801200105>
- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49(2), 347–365. <https://doi.org/10.1111/j.2044-8317.1996.tb01093.x>
- Benson, M. J., & Campbell, J. P. (2007). To be or not to be linear: An expanded representation of personality and its relationship to leadership performance. *International Journal of Personality and Assessment*, (15)2, 232-249. <https://doi.org/10.1111/j.1468-2389.2007.00384.x>
- Betebenner, D. W., Shang, Y., Xiang, Y., Zhao, Y., & Yue, X. (2008). The impact of performance level on the accuracy and precision of percent at performance level measures. *Journal of Educational Measurement*, 45(2), 119–137. <https://doi.org/10.1111/j.1745-3984.2007.00056.x>

- Bock, R. D., and Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444. doi: 10.1177/014662168200600405
- Borgatto, A. F., Azevedo, C. L. N., Pinheiro, A., & Andrade, D. F. (2015). Comparison of ability estimation methods using irt for test with different degrees of difficulty. *Communications in Statistics – Simulation and Computation*, 44(2), 474-488. <https://doi.org/10.1080/03610918.2013.781630>
- Brennan, R. L., & Wan, L. (2004). *Bootstrap procedures for estimating decision consistency for single administration complex assessments* (CASMA Research Report No. 7). Iowa City: Center for Advanced Studies in Measurement and Assessment, University of Iowa.
- Broadfoat, A.A., (2008). *Comparing the dominance approach to the ideal-point approach in the measurement and predictability of personaity*. Doctoral Dissertation, Graduate College of Bowling Green State University, Ohio.
- Brown, A., and Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 489–493. doi: 10.1111/j.1754-9434.2010.01277.x
- Cao, M., Song, Q. C., and Tay, L. (2018). Detecting curvilinear relationships: a comparison of scoring approaches based on different item response models. *Internationa Journal of Testing*, 18(2), 178–205. <https://doi.org/10.1080/15305058.2017.1345913>
- Cao, M., Drasgow, F., & Cho, S. (2014). Developing ideal intermediate personality items for the ideal point model. *Organizational Research Methods*, 18, 252-275. <https://doi.org/10.1177/1094428114555993>
- Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work satisfaction scale. *Personality and Individual Differences*, 49(7), 743-748. <https://doi.org/10.1016/j.paid.2010.06.019>
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48, 1-29. Doi: 10.18637/jss.v048.i06

- Cheng, Y., Liu, C., & Behrens, J. (2015). Standard error of ability estimates and the classification accuracy and consistency of binary decisions. *Psychometrika*, 80(3), 645-664. <https://link.springer.com/article/10.1007/s11336-014-9407-z> sayfasından erişilmiştir.
- Chernyshenko, O., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562. doi: 10.1207/S15327906MBR3604_03
- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, 19(1), 88–106. doi: 10.1037/1040-3590.19.1.88
- Christensen, K. B., Makransky, G., Horton M. (2016). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, 41(3), 178-194. doi:10.1177/0146621616677520
- Ching-Fung, B. S. (2002). Ability estimation under different item parametrization and scoring models. Doctoral Dissertation, University of North Texas, Denton Texas.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Lawrence Erlbaum Associates Publishers.
- Conn, S., & Rieke, M. L. (Eds.). (1994). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Holt, Rinehart and Winston, Inc.
- Çelikten, S., & Çakan, M. (2019). Bayesian ve nonbayesian kestirim yöntemlerine dayalı olarak sınıflama indekslerinin TIMSS-2015 matematik testi üzerinde incelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi* , 13 (1) , 105-124 . doi: 10.17522/balikesirnef.566446

- Dalal, D. K., & Carter, N. T. (2014). Consequences of ignoring ideal point items for applied decisions and criterion-related validity estimates. *Journal of Business and Psychology*, 30, 483-498. <https://link.springer.com/article/10.1007/s10869-014-9377-2> sayfasından erişilmiştir.
- Davidson, M., Keating, J. L., Eyres, S. (2004). A low back-specific version of the SF-36 Physical Functioning Scale. *Spine*, 29(5), 586-594. doi: 10.1097/01.brs.0000103346.38557.73
- Deng, N. (2011). *Evaluating ,irt- and ctt- based methods of estimating classification consistency and accuracy indices from single administrations*. Doctoral Dissertation, Massachusetts Amherst University, Amherst.
- DeNunzio, M. M. (2019). *The Development and validation of an ideal point measure of work engagement*. Doctoral Dissertation, The City University of New York, New York.
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 Years After Likert: Thurstone Was Right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 465–476. <http://dx.doi.org/10.1111/j.1754-9434.2010.01273.x>
- Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Multivariate applications book series (4th ed.). New Jersey: Lawrence Erlbaum Associates.
- Ercikan, K., & Julian, M. (2002). Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education*, 15(3), 269-294. https://doi.org/10.1207/S15324818AME1503_3
- Erol, M. ve Yıldırım, İ. (2016). Yükseköğrenim yaşam doyumu ölçeğinin geliştirilmesi. *Eğitimde Kuram ve Uygulama*, 12(1), 221-243. <https://dergipark.org.tr/tr/download/article-file/262390> sayfasından erişilmiştir.
- Fan, Y. (2020). *A method for assessing the dimensionality of unfolding responses*. Doctoral Dissertation, University of Georgia, Athens Georgia.
- Ferrando, P. J., & Chico, E. (2007). The external validity of scores based on the two-parameter logistic model: Some comparisons between IRT and CTT. *Psicológica*, 28, 237–257. <https://eric.ed.gov/?id=EJ802623> sayfasından erişilmiştir.

- Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). How to design and evaluate research in education (Eight Edition). New York: McGraw-Hill.
- Gamst, G., Meyers, L., & Guarino, A. (2008). Analysis of Variance Designs: A Conceptual and Computational Approach with SPSS and SAS. Cambridge: Cambridge University Press.
- Gundula, A. M. (2012). Comparing the correctness of classical test theory and item response theory in evaluating the consistency and accuracy of student proficiency classifications. Doctoral Dissertation, University of Alberta, Department of Educational Psychology, Edmonton, Alberta
- Guo, F. (2006). Expected classification accuracy using the latent distribution. *Practical Assessment Research & Evaluation*, 11(6). <https://doi.org/10.7275/bxba-7466>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hambleton, R. K, & Novick, M. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10(3), 159-170. <https://doi.org/10.1111/j.1745-3984.1973.tb00793.x>
- Hambleton, R. K., & Han, N. (2005). Assessing the fit of IRT models to educational and psychological test data: A five step plan and several graphical displays. In W. R. Lenderking & D. Revicki (Eds.), *Advances in health outcomes research methods, measurement, statistical analysis, and clinical applications* (pp. 57-78).
- Han, K. T. (2016). Maximum likelihood score estimation method with fences for short-length tests and computerized adaptive tests. *Applied psychological measurement*, 40(4), 289-301. <https://doi.org/10.1177/0146621616631317>
- Hofstee, W. K. B. (2001). Intelligence and personality: Do they mix? In J. M. Collis & S. Messick (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement* (s. 45–60). Mahwah, NJ: Erlbaum.
- Hojtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika* 55, 641–656. <https://link.springer.com/article/10.1007/BF02294613> sayfasından erişilmiştir.
- Hojtink H. (1991). The Measurement of Latent Traits by Proximity Items. *Applied Psychological Measurement*, 15(2), 153-169. doi:10.1177/014662169101500205

- Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity conditions of repeated measures F tests. *Psychological Bulletin*, 87(3), 479–481. <https://doi.org/10.1037/0033-2909.87.3.479>
- Kim, S. Y. & Lee, W. C. (2019) Classification Consistency and Accuracy for Mixed-Format Tests, *Applied Measurement in Education*, 32(2), 97-115, <https://doi.org/10.1080/08957347.2019.1577246>
- Kirk, R. E. (2008). *Statistics: An Introduction* (5th ed.) Belmont, CA: Thomson Learning.
- Kirk, Roger E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.
- Knupp, T. L. (2009). *Estimating decision indices based on composite scores*. Doctoral Dissertation. University of Iowa, Iowa City.
- Kula Kartal, S. (2018). *Akademik Gdlenme leęi' nin boyutluluk aısından madde tepki kuramı modellerine dayalı olarak incelenmesi*. Doktora Tezi, Ankara niversitesi, Ankara niversitesi Eęitim Bilimleri Enstits, Ankara.
- Laerd Statistics (2015). Three-way repeated measures ANOVA using SPSS Statistics. Statistical tutorials and software guides. <https://statistics.laerd.com/> sayfasından eriřilmiřtir.
- LaFond, L. J. (2014). *Decision consistency and accuracy indices for the Bifactor and Testlet response theory models*. Doctoral Dissertation The University of Iowa. UMI Number: 3638391.
- Lathrop, Q. N., & Cheng, Y. (2013). Two approaches to estimation of classification accuracy under the IRT framework. *Applied Psychological Measurement*, 37(3), 226-241. <https://doi.org/10.1177/0146621612471888>
- Lathrop, Q. N., & Cheng, Y. (2014). A nonparametric approach to estimate classification accuracy and consistency. *Journal of Educational Measurement*, 51(3), 318-334. <https://doi.org/10.1111/jedm.12048>
- LaPalme, M., Tay, L., & Wang, W. (2018). A within-person examination of the ideal-point response process. *Psychological Assessment*, 30(5), 567-581. <https://doi.org/10.1037/pas0000499>

- Lee, W. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement*, 47(1), 1-17. <https://doi.org/10.1111/j.1745-3984.2009.00096.x>
- Lee, W. C. , Hanson B. A., Brennan R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement*, 26(4), 412–432. <https://doi.org/10.1177/014662102237797>
- Lee, W. C., Brennan, R. L., & Wan L. (2009). Classification Consistency and Accuracy for Complex Assessments Under the Compound Multinomial Model. *Applied Psychological Measurement*. 33(5), 374-390. <https://doi.org/10.1177/0146621608321759>
- Lee, W. (2005). *Classification consistency under the compound multinomial model* (CASMA Research Report No. 13). Iowa City: University of Iowa, Center for Advanced Studies in Measurement and Assessment. <http://www.education.uiowa.edu/casma> sayfasından erişilmiştir.
- Levine, M. V. (1984). *An introduction to multilinear formula score theory* (Measurement Series No. 84–4). Arlington, VA: Office of Naval Research, Personnel and Training Research Programs. <https://apps.dtic.mil/sti/pdfs/ADA145696.pdf> sayfasından erişilmiştir.
- Lim, J. X. (2015). *Modelling unfolding response data within the structural equation modelling framework*. Master's thesis, Nanyang Technological University, Singapore.
- Ling, Y., Zhang, M., Locke, K. D., Li, G., & Li, Z. (2016). Examining the process of responding to Circumplex Scales of Interpersonal Values items: should ideal point scoring methods be considered?. *Journal of personality assessment*, 98(3), 310-318. <https://doi.org/10.1080/00223891.2015.1077852>
- Liu, C. W., & Wang, W.C. (2019). A general unfolding IRT model for multiple response styles. *Applied Psychological Measurement*, 43(3), 195–210. <https://doi.org/10.1177/0146621618762743>
- Liu, C. W., & Chalmers, R. P. (2018). Fitting item response unfolding models to Likert-scale data using mirt in R. *PloS one*, 13(5), e0196292. <https://doi.org/10.1371/journal.pone.0196292>

- Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179–197. <https://doi.org/10.1111/j.1745-3984.1995.tb00462.x>
- Lord, F. M., & Wingersky, S. M. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, 8(4), 453-461. <https://doi.org/10.1177/014662168400800409>
- Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability, *Psychometrika*, 48, 233-245. <https://link.springer.com/article/10.1007/BF02294018> sayfasından erişilmiştir.
- Lord, F. M. (1984). *Maximum likelihood and Bayesian parameter estimation in item response theory* (Research Rep. No. RR-84-30-ONR). Princeton, NJ: Educational Testing Service.
- Luo G., Andrich D. (2004) Weighted Likelihood Estimation of Person Locations in an Unfolding Model for Polytomous Responses. In: Banks D., McMorris F.R., Arabie P., Gaul W. (eds) *Classification, Clustering, and Data Mining Applications. Studies in Classification, Data Analysis, and Knowledge Organisation*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17103-1_23
- Maraun, M., Rossi, N. (2011) The Extra-Factor Phenomenon Revisited: Unidimensional Unfolding as Quadratic Factor Analysis, *Applied Psychological Measurement*, 25(1), 88-87. <https://doi.org/10.1177/01466216010251006>
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). New York: Psychology Press.
- McMillan, J. T. (2019). Computerized adaptive testing for ideal point personality assessment: a comparison of test characteristics. Doctoral Dissertation, University of Georgia, Athens.
- Montgomery, M., & Skorupski, W. (2012, April). *Investigation of IRT Parameter Recovery and Classification Accuracy in Mixed Format*. Paper presented at the Annual meeting of the Nation Council of Measurement in Education.

- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159–176. <https://doi.org/10.1177/014662169201600206>
- Ranger, J., & Much, S. (2020). Analyzing the Fit of IRT Models With the Hausman Test. *frontiers in Psychology*, 11(149). doi: 10.3389/fpsyg.2020.00149
- Reise, S. P. (2010). Thurstone might have been right about attitudes, but Drasgow, Chernyshenko, and Stark fail to make the case for personality. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 485–488. <https://doi.org/10.1111/j.1754-9434.2010.01276.x>
- Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.3, <https://cran.r-project.org/web/packages/psych/index.html> sayfasından erişilmiştir.
- Roberts, J. S. (2016). Generalized Graded Unfolding Model In van der Linden, W. J. (Edt.), *Handbook of Item Response Theory*, Volume One: Models (pp.369-390). (Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences). Chapman and Hall/CRC.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, 26(2), 192–207. <https://doi.org/10.1177/01421602026002006>
- Roberts, J. S., Donoghue, J. R., and Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24(1), 3–32. <https://doi.org/10.1177/01466216000241001>
- Roberts, J., & Laughlin, J. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement*, 20(3), 231-255. <https://doi.org/10.1177/014662169602000305>
- Roberts, J. S., Fang, H., Cui, W., & Wang, Y. (2006). GAMOM2004: A Windows-based program to estimate parameters in the generalized graded unfolding model. *Applied Psychological Measurement*, 30(1), 64–65. <https://doi.org/10.1177/0146621605280141>

- Roberts, J., Laughlin, J., & Wedell, D. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, 59(2), 211-233. <https://doi.org/10.1177/00131649921969811>
- Roberts, J. S. (2008). Modified likelihood-based item fit statistics for the generalized graded unfolding model. *Applied Psychological Measurement*, 32(5), 407–423. <http://dx.doi.org/10.1177/0146621607301278>
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1998). The generalized graded unfolding model: A general parametric item response model for unfolding graded responses (Research Rep. RR-98- 32). Princeton NJ: Educational Testing Service. <https://www.ets.org/Media/Research/pdf/RR-98-32.pdf> sayfasından erişilmiştir.
- Roberts, J. S., & Thompson, V. M. (2011). Marginal Maximum A Posteriori Item Parameter Estimation for the Generalized Graded Unfolding Model. *Applied Psychological Measurement*, 35(4), 259-279. <https://doi.org/10.1177/0146621610392565>
- Rose, N. (2010). Maximum Likelihood and Bayes Modal Ability Estimation in Two-Parametric IRT Models: Derivations and Implementation. (*Schriften zur Bildungsf.*)
- Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment Research & Evaluation*, 7(14). <https://doi.org/10.7275/an9m-2035>
- Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment Research & Evaluation*, 10(13). <https://doi.org/10.7275/56a5-6b14>
- Salvucci, S., Walter, E., Conley, V., Fink, S., & Saba, M. (1997). Measurement error studies at the National Center for Education Statistics (NCES). Washington D. C.: U. S. Department of Education.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*, 17(4), 2. doi:10.1002/j.2333-8504.1968.tb00153.x
- Scherbaum, C. A., Sabet, J., Kern, M. J., and Agnello, P. (2013). Examining faking on personality inventories using unfolding item response theory models. *Journal of Personality Assessment*, 95(2), 207–216. doi: 10.1080/00223891.2012.725439

- Sgammato, A. N. (2009). An application of unfolding and cumulative item response theory models for noncognitive scaling: examining the assumptions and applicability of the generalized graded unfolding model. Doktoral Dissertation, University of North Carolina, Chapel Hill
- Speer, A. B., Robie, C., and Christiansen, N. D. (2016). Effects of item type and estimation method on the accuracy of estimated personality trait scores: polytomous item response theory models versus summated scoring. *Personality and Individual Differences*, 102, 41–45. <https://doi.org/10.1016/j.paid.2016.06.058>
- Stark, S., Chernyshenko, O. S., Drasgow, F., and Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25–39. <https://doi.org/10.1037/0021-9010.91.1.25>
- Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, 13(4), 265-276. <https://doi.org/10.1111/j.1745-3984.1976.tb00017.x>
- Sun, T., Fraley, R. C., & Drasgow, F. (2020). Matches made with information: Fitting measurement models to adult attachment data. *Assesment* <https://doi.org/10.1177/1073191120971847>
- Swaminathan H, Hambleton RK, Algina J (1974). Reliability of Criterion-Referenced Tests: A Decision-Theoretic Formulation. *Journal of Educational Measurement*, 11(4), 263–267. <https://doi.org/10.1111/j.1745-3984.1974.tb00998.x>
- Tan, Ş. (2016). *SPSS ve Excel Uygulamalı Temel İstatistik-1*. Ankara: Pegem Akademi.
- Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal? *Journal of Applied Psychology*, 94(5), 1287–1304. <https://doi.org/10.1037/a0015899>
- Tay, L., and Kuykendall, L. (2016). Why self-reports of happiness and sadness may not necessarily contradict bipolarity: a psychometric review and proposal. *Emotion Review*, 9(2), 146–154. doi: 10.1177/1754073916637656

- Tay, L., Ali, U. S., Drasgow, F., and Williams, B. A. (2011). Fitting IRT models to dichotomous and polytomous data: assessing the relative model–data fit of ideal point and dominance models. *Applied. Psychological Measurement*, 35(4), 280–295. doi: 10.1177/0146621610390674
- Tay, L., & Drasgow, F. (2012). Theoretical, statistical, and substantive issues in the assessment of construct dimensionality: Accounting for the item response process. *Organizational Research Methods*, 15(3), 363–384. <http://dx.doi.org/10.1177/1094428112439709>
- Tay, L. & Ng, V. (2018). Ideal Point Modeling of Non-cognitive Constructs: Review and Recommendations for Research. *Frontiers in Psychology*. 9:2423. doi: 10.3389/fpsyg.2018.02423.
- Ten Klooster, P. M., Taal, E., van de Laar, M. A. F. J. (2008). Rasch analysis of the Dutch health assessment questionnaire disability index and the Health Assessment Questionnaire II in patients with rheumatoid arthritis. *Arthritis & Rheumatism*, 59(12), 1721-1728. <https://doi.org/10.1002/art.24065>
- Thompson, V. M. (2014). *Marginal bayesian parameter estimation in the multidimensional generalized graded unfolding model*. Doktoral Dissertation, Georgia Institute of Technology, Atlanta.
- Usami, S. (2011). Generalized graded unfolding model with structural equation for subject parameters. *Japanese Psychological Research*, 53(3), 221–32. <https://doi.org/10.1111/j.1468-5884.2011.00476.x>
- Wan, L., Brennan, R. L., & Lee, W. (2007). *Estimating classification consistency for complex assessments* (CASMA Research Report No. 22). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment. <https://education.uiowa.edu/sites/education.uiowa.edu/files/documents/centers/casma/publications/casma-research-report-22.pdf> sayfasından erişilmiştir.
- Wang, W. (2013). A bayesian markov chain monte carlo approach to the generalized graded unfolding model estimation: the future of non-cognitive measurement. Doctoral Dissertation, Graduate College of the University of Illinois, Urbana-Champaign.

- Wang, W. C., & Wu, S. L. (2016). Confirmatory multidimensional IRT unfolding models for graded response items. *Applied Psychological Measurement*, 40, 56-72. doi: 10.1177/0146621615602855
- Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54(3), 427-450. <https://link.springer.com/content/pdf/10.1007/BF02294627.pdf> sayfasından erişilmiştir.
- Wang, W.C., & Liu, C.W. (2011). Computerized classification testing under the generalized graded unfolding model. *Educational and Psychological Measurement*, 71(1), 114-128. <https://doi.org/10.1177/0013164410391575>
- Wang, J., & Engelhard, G. (2019). Exploring the Impersonal Judgments and Personal Preferences of Raters in Rater-Mediated Assessments With Unfolding Models. *Educational and Psychological Measurement*, 79(4), 773-795. doi:10.1177/0013164419827345
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Education Measurement*, 35(2), 109-135. <https://doi.org/10.1111/j.1745-3984.1998.tb00530.x>
- Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, 25(4), 317-331. <https://doi.org/10.1177/01466210122032163>
- Wainer, H., & Thissen, D. (1987). Estimating ability with the wrong model. *Journal of Educational Statistics*, 12(4), 339-368. https://www.jstor.org/stable/1165054?seq=1#metadata_info_tab_contents sayfasından erişilmiştir.
- Waterson, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika*, 63(3), 639-645. https://www.jstor.org/stable/2335746?seq=1#metadata_info_tab_contents sayfasından erişilmiştir.

- Wilgus S., & Travis J. (2019) A Comparison of Ideal-Point and Dominance Response Processes with a Trust in Science Thurstone Scale. In: Wiberg M., Culpepper S., Janssen R., González J., Molenaar D. (eds) Quantitative Psychology. IMPS 2017. Springer Proceedings in Mathematics & Statistics, (vol 265). Springer, Cham.
- Williams, E. J. (2017). Estimation of parameters in the generalized graded unfolding model using a genetic algorithm. Doctoral Dissertation, School of Psychology, Georgia Institute of Technology, Atlanta.
- Wyse, A. E. (2011). The Potential Impact of Not Being Able to Create Parallel Tests on Expected Classification Accuracy. *Applied Psychological Measurement*, 35(2), 110-126. doi:10.1177/0146621610377449
- Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement*, 36(7), 602-624. <https://doi.org/10.1177/0146621612451522>
- Wyse, A. E., & Babcock, B. (2016). Does maximizing information at the cut score always maximize classification accuracy and consistency? *Journal of Educational Measurement*, 53(1), 23–44. <https://doi.org/10.1111/jedm.12099>
- Yang, X., Poggio, J. C., & Glasnapp, D. R. (2006). Effects of estimation bias on multiple-category classification with an IRT-based adaptive classification procedure. *Educational and Psychological Measurement*, 66(4), 545–564. <https://doi.org/10.1177/0013164405284031>
- Zampetakis, L. A., Lerakis, M., Kafetsios, K. & Moustakis, V. (2015). Using item response theory to investigate the structure of anticipated affect: do self-reports about future affective reactions conform to typical or maximal models?. *Frontiers in Psychology*, 24(6): 1-8. DOI: 10.3389/fpsyg.2015.01438
- Zhang, S., Du, J., Chen, P., Xin, T., & Chen, F. (2017). Using Procedure Based on Item Response Theory to Evaluate Classification Consistency Indices in the Practice of Large-Scale Assessment. *Frontiers in Psychology*, 26(8), 1676. <http://doi.org/10.3389/fpsyg.2017.01676>

Zhang, B., Cao, M., Tay, L., Luo, J., and Drasgow, F. (2019). Examining the item response process to personality measures in high-stakes situations: issues of measurement validity and predictive validity. *Personnel Psychology*. 73(2), 305-332. doi: 10.1111/peps.12353





EKLER



EK 1. Yükseköğrenim Yaşam Doyumu Ölçeği

YÜKSEKÖĞRENİM YAŞAM DOYUMU ÖLÇEĞİ

(YYDÖ)

Sevgili öğrenciler,

Bu ölçek, şu anda öğrenim gördüğünüz üniversite yaşantınızla ilgili düşüncelerinizi öğrenmek amacıyla geliştirilmiştir. Ölçekte 29 madde bulunmaktadır. Lütfen her bir maddeyi dikkatlice okuyunuz. Okuduğunuz maddeye ne derecede katılıp katılmadığınıza karar veriniz. Seçeneklerinden size uygun gelen seçeneği çarpı (X) işareti ile belirtiniz. Bütün maddelere tepkide bulununuz. Ölçek üzerine adınızı ve adresinizi yazmayınız.

Duyarlılığınız için teşekkür ederim.

	Hiç Katılmıyorum	Katılmıyorum	Kısmen Katılıyorum	Katılıyorum	Tamamen Katılıyorum
1. Tekrar tercih yapacak olsam yine bu üniversiteyi seçerim.	()	()	()	()	()
2. Her yönüyle buradaki yaşamımdan hoşnutum.	()	()	()	()	()
3. Bu üniversitede sahip olduğum imkânlardan memnunum.	()	()	()	()	()
4. Bu üniversitede şu ana kadar istediğim şeylere ulaştım.	()	()	()	()	()
5. Bu üniversite birçok yönüyle beklentilerimi karşıladı.	()	()	()	()	()
6. Bu üniversitede şu ana kadar geçirdiğim zamanın boşa gittiğini düşünmekteyim.	()	()	()	()	()
7. Bu üniversitede aldığım eğitimle ideallerimi gerçekleştireceğimden eminim.	()	()	()	()	()
8. Bu üniversitede eğitim aldığım için kendimi şanslı hissetmekteyim.	()	()	()	()	()
9. Bu üniversitede eğitim almaktan keyif almıyorum.	()	()	()	()	()
10. Bu üniversitede verilen akademik eğitimden memnun değilim.	()	()	()	()	()
11. Bu üniversitede kendimi değerli hissetmiyorum.	()	()	()	()	()
12. Bu üniversitenin öğrencisi olmaktan gurur duymaktayım.	()	()	()	()	()
13. Üniversitede olduğum süre içerisinde kendimi iyi hissetmekteyim.	()	()	()	()	()
14. Bu üniversitede kendimi özgür hissetmekteyim.	()	()	()	()	()
15. Öğretim üyelerinin ders işleme tarzından memnunum.	()	()	()	()	()
16. Öğretim üyelerinin öğrencilerle iletişiminden memnunum.	()	()	()	()	()

17. Öğretim üyelerinin öğrencilere değer vermediği kanısındayım.	()	()	()	()	()
18. Öğretim üyelerinin öğrencilerin “görüş ve önerilerini” dikkate almasından memnunum.	()	()	()	()	()
19. Üniversite yönetiminin öğrencilere değer vermediği kanısındayım.	()	()	()	()	()
20. Üniversite yönetiminin öğrencilerin “görüş ve önerilerini” dikkate almasından memnunum.	()	()	()	()	()
21. Üniversite yönetiminin öğrencilerle iletişiminden memnunum.	()	()	()	()	()
22. Barınma (yurt vb.) imkânlarından memnun değilim.	()	()	()	()	()
23. Dershane (derslik, seminer odası vb.) imkânlarından memnun değilim.	()	()	()	()	()
24. Ulaşım imkânlarından memnun değilim.	()	()	()	()	()
25. Burs imkânlarından memnunum.	()	()	()	()	()
26. Sinema/tiyatro imkânlarından memnun değilim.	()	()	()	()	()
27. Yemekhane hizmetlerinden memnunum.	()	()	()	()	()
28. Sağlık hizmetlerinden memnunum.	()	()	()	()	()
29. Psikolojik danışma ve rehberlik hizmetlerinden memnunum.	()	()	()	()	()

EK 2. Gazi Üniversitesi Etik Komisyonu İzni

Evrak Tarih ve Sayısı: 11/01/2019-E.4684



T.C.
GAZİ ÜNİVERSİTESİ
Etik Komisyonu



Sayı : 77082166-302.08.01-
Konu : Bilimsel ve Eğitim Amaçlı

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 06/12/2018 tarihli ve 80287700-302.08.01- 161466 sayılı yazı.

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı **Doktora Öğrencisi Serpil ÇELİKTEN'in, Prof.Dr.Mehtap ÇAKAN'ın** danışmanlığında yürüttüğü *"Bayesian ve Nonbayesian Kestirim Yöntemlerine Dayalı Hesaplanan Sınıflama İndekslerinin Farklı Performans Düzeyleri ve Örneklem Koşullarında İncelenmesi"* adlı tez çalışması ile ilgili konu Komisyonumuzun **08.01.2019** tarih ve **01** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi ve gereğini rica ederim.

e-imzalıdır
Prof. Dr. Alper CEYLAN
Komisyon Başkanı

Araştırma Kod No: 2019-038

Ek:1 Liste



EK 3. Gazi Üniversitesi Eğitim Fakültesi Uygulama İzni

Evrak Tarih ve Sayısı: 18/01/2019-E.8056			
T.C. GAZİ ÜNİVERSİTESİ GAZİ EĞİTİM FAKÜLTESİ Eğitim Bilimleri Bölüm Başkanlığı			
Sayı : 29554917-044- Konu : Anketler			
GAZİ EĞİTİM FAKÜLTESİ DEKANLIĞINA			
İlgi : 17/01/2019 tarihli ve 89377925-044- 7070 sayılı yazı.			
Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı Doktora öğrencisi Serpil ÇELİKTEN'in Öğretim Üyesi Prof. Dr. Mehtap ÇAKAN'ın danışmanlığında yürüttüğü "Bayesian ve Nonbayesian Kestirim Yöntemlerine Dayalı Hesaplanan Sınıflama İndekslerinin Farklı Performans Düzeyleri ve Örneklem Koşullarında İncelenmesi" isimli tez çalışması kapsamında anket uygulama isteği, Bölümümüz tarafından uygun görülmüştür.			
Bilgilerinizi ve gereğini arz ederim.			
e-imzalıdır Prof. Dr. Yücel GELİŞLİ Bölüm Başkanı			

EK 4. GAMOM'a İlişkin Ayırtedicilik, Delta ve Eşik Parametreleri

Tablo 13

GAMOM'a İlişkin Ayırtedicilik, Delta ve Eşik Parametreleri

	a1	b1	t1	t2	t3	t4
M1	1,970	1,234	1,742	1,719	1,192	0,556
M2	1,780	1,289	2,235	1,905	0,726	0,165
M3	2,832	1,119	1,933	1,454	0,802	0,079
M4	2,650	1,121	1,860	1,357	0,541	0,056
M5	3,133	1,117	1,894	1,464	0,703	0,016
M6	0,867	-3,652	4,796	3,063	2,658	2,664
M7	1,984	1,306	2,242	2,251	1,437	0,310
M8	3,299	1,372	2,353	2,093	1,516	0,685
M9	0,633	-4,467	5,958	3,686	3,887	3,546
M10	0,956	-2,503	3,574	1,883	1,404	1,442
M11	0,929	-3,215	4,296	2,747	2,560	2,203
M12	2,607	1,352	2,363	2,257	1,515	0,680
M13	3,067	1,330	2,296	2,001	1,296	0,451
M14	1,424	1,297	1,815	1,844	1,345	0,012
M15	1,448	1,292	2,223	2,299	1,037	0,007
M16	1,399	1,362	2,257	2,484	1,537	0,116
M17	0,513	-5,454	7,474	4,634	4,029	4,069
M18	1,446	1,299	2,210	2,216	1,379	-0,039
M19	0,672	-3,844	5,658	3,591	2,995	2,794
M20	1,411	1,263	2,078	1,883	0,931	-0,184
M21	1,527	1,247	1,948	1,647	0,869	-0,160
M22	0,335	-7,001	8,780	7,180	4,740	6,509
M23	0,467	-4,887	7,224	4,913	4,049	4,090
M24	0,293	-5,700	7,289	3,593	4,295	5,760
M25	0,574	1,065	1,190	1,771	0,683	-0,658
M26	0,366	-4,801	6,749	5,407	4,218	4,368
M27	0,627	1,069	0,948	2,943	1,946	-0,011
M28	0,883	1,129	1,480	2,804	1,411	-0,437
M29	1,194	1,115	1,571	1,971	0,609	-0,094

EK 5. ATM'ye İlişkin Ayırtedicilik ve Eşik Parametreleri

Tablo 14

ATM'ye İlişkin Ayırtedicilik ve Eşik Parametreleri

	a1	d1	d2	d3	d4
M1	2,591	2,228	0,846	-0,835	-2,952
M2	2,167	2,781	0,896	-1,701	-3,815
M3	2,682	2,724	0,549	-1,866	-4,765
M4	2,529	2,331	0,088	-2,558	-4,889
M5	2,906	2,794	0,570	-2,329	-5,502
M6	1,436	2,653	1,465	0,146	-1,886
M7	2,405	3,447	1,973	-0,281	-3,270
M8	3,547	4,507	2,414	-0,217	-3,670
M9	1,199	2,191	0,991	-0,039	-1,901
M10	1,490	2,835	1,668	0,280	-1,835
M11	1,557	2,543	1,197	-0,049	-1,988
M12	3,115	4,322	2,568	-0,092	-3,163
M13	3,369	4,140	2,035	-0,859	-4,343
M14	2,041	2,086	0,780	-0,759	-3,297
M15	1,867	2,856	1,421	-0,882	-3,268
M16	1,876	3,017	1,764	-0,198	-2,856
M17	0,953	2,521	1,357	0,092	-1,902
M18	1,931	2,885	1,455	-0,470	-3,214
M19	1,162	2,187	0,937	-0,343	-2,267
M20	1,945	2,402	0,782	-1,250	-3,719
M21	2,007	2,099	0,423	-1,468	-3,907
M22	0,564	1,822	0,906	-0,375	-1,715
M23	0,810	1,678	0,573	-0,590	-2,256
M24	0,735	1,733	0,939	0,122	-1,419
M25	1,009	1,380	0,342	-0,860	-2,400
M26	0,715	1,271	0,226	-0,853	-2,158
M27	1,084	2,161	1,471	0,304	-1,471
M28	1,347	2,341	1,492	-0,149	-2,405
M29	1,677	1,885	0,680	-1,275	-3,145

EK 6. ATM Altında Yen’ in Q3 İstatistikleri

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29
M1	NA	0,123	-0,104	0,096	0,119	-0,210	-0,105	0,130	-0,256	-0,198	-0,191	0,143	0,159	0,095	-0,138	-0,135	-0,167	-0,138	-0,185	-0,167	-0,150	-0,154	-0,162	-0,133	-0,134	-0,157	-0,144	-0,158	-0,174
M2	102,090	NA	0,131	0,134	0,156	-0,224	-0,136	-0,133	-0,255	-0,227	-0,237	-0,148	0,167	0,112	-0,150	-0,133	-0,174	-0,156	-0,197	-0,150	-0,140	-0,138	-0,168	-0,126	-0,139	-0,146	-0,128	-0,152	-0,151
M3	72,379	115,078	NA	0,160	0,155	-0,293	-0,137	-0,134	-0,332	-0,336	-0,308	-0,131	-0,116	0,112	-0,133	-0,146	-0,227	-0,137	-0,247	-0,149	-0,116	-0,166	-0,188	-0,141	-0,136	-0,190	-0,143	-0,155	-0,150
M4	62,054	120,372	171,560	NA	0,187	-0,295	-0,170	-0,148	-0,309	-0,346	-0,285	-0,143	-0,141	-0,116	-0,136	-0,150	-0,238	-0,130	-0,249	-0,146	-0,116	-0,163	-0,174	-0,143	-0,131	-0,181	-0,144	-0,158	-0,153
M5	94,578	163,901	162,393	235,427	NA	-0,374	-0,161	-0,154	-0,383	-0,441	-0,375	-0,194	-0,174	-0,122	-0,136	-0,140	-0,282	-0,165	-0,286	-0,183	-0,154	-0,186	-0,205	-0,142	-0,128	-0,200	-0,149	-0,161	-0,155
M6	297,319	337,791	577,960	585,924	939,920	NA	-0,283	-0,271	0,271	0,218	0,215	-0,262	-0,294	-0,190	-0,234	-0,204	0,176	-0,210	0,178	-0,268	-0,239	0,157	0,161	-0,142	-0,180	0,137	-0,170	-0,212	-0,231
M7	74,593	124,471	125,784	193,590	173,738	538,611	NA	0,167	-0,283	0,267	-0,254	0,126	-0,162	-0,143	0,124	-0,133	-0,228	-0,131	-0,253	-0,183	-0,181	-0,162	-0,194	-0,134	-0,121	-0,171	-0,144	-0,141	-0,137
M8	112,918	119,603	120,163	146,418	160,479	495,673	187,466	NA	-0,325	-0,264	-0,253	0,200	-0,184	-0,130	-0,130	-0,132	-0,200	-0,138	-0,255	-0,225	-0,193	-0,164	-0,175	-0,142	-0,151	-0,160	-0,142	-0,155	-0,142
M9	442,646	437,513	743,214	641,948	986,810	495,492	539,077	712,399	NA	0,260	0,260	-0,333	-0,358	-0,239	-0,243	-0,191	0,195	-0,217	0,207	-0,260	-0,266	0,167	0,164	0,149	-0,191	0,168	-0,177	-0,241	-0,248
M10	262,444	346,245	761,726	803,612	1308,755	319,124	478,069	470,659	455,928	NA	0,255	-0,257	-0,278	-0,210	0,249	-0,220	0,221	-0,220	0,202	-0,279	-0,273	0,172	0,169	0,148	-0,186	0,158	-0,153	-0,215	-0,250
M11	246,296	377,643	636,557	546,716	947,423	310,620	432,590	432,162	454,756	438,028	NA	-0,260	-0,270	-0,205	-0,199	-0,191	0,212	-0,204	0,213	-0,261	-0,225	0,163	0,169	0,163	-0,181	0,162	-0,174	-0,235	-0,247
M12	138,389	148,292	114,670	138,249	253,898	463,541	105,994	268,877	746,928	443,484	453,522	NA	0,210	-0,132	-0,137	-0,144	-0,233	-0,147	-0,259	-0,221	-0,176	-0,166	-0,192	-0,154	-0,156	-0,164	-0,146	-0,164	-0,166
M13	171,045	188,757	89,853	133,755	203,889	580,280	177,090	228,743	862,082	519,742	491,680	297,866	NA	0,159	-0,137	-0,149	-0,232	-0,151	-0,290	-0,225	-0,165	-0,171	-0,196	-0,134	-0,148	-0,177	-0,147	-0,152	-0,178
M14	60,582	84,842	85,014	90,162	100,309	244,158	136,941	114,261	383,842	296,241	281,956	117,961	170,062	NA	-0,151	-0,136	-0,170	-0,144	-0,208	-0,154	-0,135	-0,137	-0,164	-0,135	-0,145	-0,156	-0,126	-0,143	-0,134
M15	127,489	151,138	119,469	124,872	124,613	368,387	103,400	112,831	397,511	417,551	265,736	125,623	126,125	153,832	NA	0,276	0,272	0,181	-0,218	-0,162	0,147	-0,158	-0,185	-0,135	-0,122	-0,189	-0,123	-0,138	-0,123
M16	122,745	118,477	144,239	150,408	131,684	281,091	119,257	117,417	245,887	324,413	244,387	138,737	150,344	125,227	512,027	NA	0,300	0,225	-0,235	0,173	0,167	-0,135	-0,198	-0,134	-0,123	-0,168	-0,120	-0,132	-0,127
M17	188,149	204,471	345,620	382,626	536,303	209,457	349,475	269,000	254,769	328,513	302,663	365,642	361,045	194,236	498,999	605,645	NA	0,300	0,295	-0,281	-0,273	0,170	0,184	0,178	-0,167	0,170	-0,146	-0,217	-0,232
M18	127,684	163,145	126,855	113,700	183,290	298,112	115,926	127,225	317,738	325,746	279,860	144,659	153,964	138,650	220,321	339,939	607,412	NA	0,259	0,257	0,198	-0,155	-0,188	-0,144	-0,137	-0,180	-0,125	-0,158	0,159
M19	231,489	260,154	410,332	417,425	550,547	213,612	429,947	438,787	288,319	275,473	305,239	451,363	565,713	291,337	320,429	371,505	587,327	452,558	NA	0,333	0,306	0,172	0,181	0,178	-0,176	0,192	-0,182	-0,217	-0,239
M20	188,724	150,593	149,315	142,617	225,574	483,656	225,947	339,775	453,559	523,602	460,003	330,017	341,558	160,338	175,574	202,089	531,719	444,015	746,976	NA	0,304	-0,202	-0,233	-0,162	-0,154	-0,194	-0,135	-0,153	0,163
M21	151,415	131,335	91,269	90,408	160,215	385,190	221,084	251,444	476,859	500,121	339,113	208,598	182,214	122,644	145,048	186,943	499,628	263,282	629,164	620,029	NA	-0,187	-0,217	-0,150	0,148	-0,206	-0,135	-0,148	0,175
M22	158,872	128,652	184,445	178,700	232,206	165,465	175,936	181,691	186,578	198,557	179,456	186,195	197,301	126,033	168,031	122,535	194,018	161,416	199,954	273,531	236,453	NA	0,308	0,206	-0,166	0,201	-0,146	-0,168	-0,182
M23	177,260	189,005	237,160	203,317	282,561	174,303	252,236	205,921	180,984	191,604	193,237	248,705	257,583	181,404	230,153	264,015	227,256	238,040	219,931	366,689	317,657	636,781	NA	0,224	-0,190	0,239	-0,179	-0,203	-0,215
M24	119,651	106,748	133,375	136,688	135,031	134,779	121,243	135,700	149,566	147,566	179,211	159,736	120,513	122,147	122,237	120,121	212,315	139,541	214,081	175,561	150,466	285,877	338,792	NA	-0,192	0,194	0,163	0,154	-0,168
M25	121,060	130,777	125,227	115,623	109,828	217,585	97,974	152,683	246,311	233,744	219,749	164,728	147,909	141,866	100,277	102,570	187,729	126,344	208,070	159,216	148,197	186,413	241,625	248,590	NA	-0,207	0,162	0,179	0,155
M26	166,722	143,075	244,073	219,384	269,025	126,373	196,446	172,071	190,535	168,326	177,183	181,488	210,942	163,290	240,074	189,894	194,110	218,117	247,037	253,250	286,589	270,503	384,595	252,573	289,319	NA	-0,190	-0,201	-0,228
M27	139,533	110,421	136,660	139,836	149,682	194,590	139,014	136,153	211,833	156,533	202,659	142,604	144,540	107,222	102,268	97,209	143,669	104,436	221,875	123,281	122,152	142,489	216,363	179,774	177,160	243,749	NA	0,314	0,216
M28	168,491	156,369	162,227	167,604	173,539	302,858	133,711	161,222	389,195	310,438	371,811	181,790	156,380	138,170	128,246	117,225	317,975	168,147	317,917	157,972	147,251	189,322	277,360	158,557	214,719	271,523	665,061	NA	0,279
M29	203,478	152,523	150,453	157,655	161,528	357,977	125,717	136,290	412,416	420,194	409,662	185,190	212,158	120,597	102,104	109,157	362,269	170,522	385,825	178,445	205,934	221,983	311,074	190,887	160,997	349,424	314,075	523,970	NA

EK 7. Karma Model Altında Yen' in Q3 İstatistikleri

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29
M1	NA	0,171	-0,173	-0,169	-0,179	-0,255	-0,166	-0,191	-0,295	-0,259	-0,242	-0,211	-0,237	-0,149	-0,200	-0,192	-0,207	-0,194	-0,242	-0,230	-0,223	-0,188	-0,201	-0,159	-0,170	-0,190	-0,168	-0,186	-0,233
M2	195,818	NA	-0,198	-0,204	-0,241	-0,270	-0,208	-0,234	-0,297	-0,286	-0,282	-0,257	0,273	-0,159	-0,218	-0,209	-0,204	-0,210	-0,252	-0,213	-0,206	-0,163	-0,206	-0,148	-0,170	-0,168	-0,152	-0,183	-0,198
M3	201,696	263,844	NA	0,196	0,209	-0,318	-0,191	-0,193	-0,354	-0,370	-0,328	-0,211	-0,206	-0,171	-0,199	-0,213	-0,245	-0,213	-0,284	-0,221	-0,184	-0,191	-0,218	-0,163	-0,179	-0,211	-0,166	-0,185	-0,208
M4	191,471	278,682	258,435	NA	0,227	-0,318	-0,270	-0,248	-0,332	-0,386	-0,308	-0,305	-0,347	-0,172	-0,220	-0,228	-0,253	-0,199	-0,284	-0,208	-0,177	-0,191	-0,201	-0,166	-0,171	-0,199	-0,167	-0,189	-0,209
M5	215,563	390,480	294,281	346,927	NA	-0,378	-0,275	-0,407	-0,388	-0,450	-0,372	-0,474	-0,421	-0,181	-0,208	-0,209	-0,286	-0,241	-0,314	-0,266	-0,240	-0,209	-0,226	-0,166	-0,171	-0,215	-0,170	-0,190	-0,213
M6	436,459	490,091	681,428	680,094	960,265	NA	-0,350	-0,326	0,261	0,197	0,192	-0,320	-0,342	-0,227	-0,297	-0,272	0,161	-0,265	0,161	-0,312	-0,288	-0,145	-0,153	-0,146	-0,198	-0,131	-0,193	-0,247	-0,266
M7	185,817	292,173	245,157	489,563	510,203	825,505	NA	0,232	-0,342	-0,322	-0,311	-0,191	-0,245	-0,202	-0,155	-0,173	-0,258	-0,180	-0,326	-0,261	-0,251	-0,192	-0,232	-0,146	-0,150	-0,198	-0,160	-0,165	-0,181
M8	244,923	368,071	251,869	415,202	1112,500	716,169	363,614	NA	-0,372	-0,319	-0,301	0,320	-0,313	-0,202	-0,195	-0,193	-0,233	-0,204	-0,320	-0,339	-0,274	-0,199	-0,216	-0,165	-0,186	-0,193	-0,169	-0,188	-0,195
M9	583,596	592,747	842,411	739,731	1013,738	459,404	786,582	930,870	NA	0,245	0,240	-0,384	-0,405	-0,275	-0,297	-0,248	-0,185	-0,260	-0,206	-0,302	-0,311	-0,167	-0,163	-0,153	-0,211	0,159	-0,202	-0,271	-0,285
M10	449,865	551,527	921,284	1000,561	1363,965	261,686	696,129	682,603	404,724	NA	0,227	-0,313	-0,335	-0,257	-0,299	-0,275	0,198	-0,266	0,189	-0,329	-0,329	-0,156	-0,161	-0,151	-0,209	-0,152	-0,176	-0,249	-0,294
M11	394,165	533,576	725,716	637,608	930,645	247,723	652,628	611,122	388,800	345,767	NA	-0,307	-0,309	-0,242	-0,250	-0,253	0,195	-0,247	0,199	-0,300	-0,266	-0,156	0,151	-0,162	-0,200	0,153	-0,190	-0,265	-0,280
M12	299,038	444,085	300,160	626,367	1510,500	690,314	244,915	688,649	991,918	657,363	633,741	NA	0,300	-0,197	-0,190	-0,209	-0,264	-0,215	-0,322	-0,314	-0,246	-0,199	-0,226	-0,170	-0,192	-0,191	-0,172	-0,198	-0,227
M13	377,341	502,516	286,235	809,369	1191,967	785,592	405,294	659,083	1103,211	756,692	641,066	607,055	NA	-0,220	-0,207	-0,223	-0,262	-0,225	-0,350	-0,339	-0,274	-0,202	-0,231	-0,161	-0,191	-0,208	-0,175	-0,188	-0,252
M14	150,314	170,340	196,910	199,752	219,501	347,079	273,222	274,959	510,553	445,774	393,434	261,492	325,895	NA	-0,207	-0,192	-0,201	-0,197	-0,253	-0,202	-0,182	-0,156	-0,190	-0,149	-0,171	-0,168	-0,145	-0,173	-0,175
M15	269,358	320,251	266,145	324,223	290,752	591,920	160,679	254,782	592,287	601,704	419,331	243,898	287,986	288,860	NA	0,291	-0,296	0,201	-0,268	-0,207	-0,188	-0,181	-0,224	-0,149	-0,148	-0,218	-0,132	-0,158	-0,157
M16	249,237	294,274	303,941	349,585	295,138	497,883	202,297	250,492	414,775	509,312	431,156	295,067	334,636	247,102	569,333	NA	0,318	0,244	-0,294	-0,211	-0,191	-0,163	-0,243	-0,146	-0,145	-0,195	-0,127	-0,149	-0,158
M17	288,424	280,670	404,360	429,994	549,903	175,172	446,322	366,492	231,495	264,626	254,569	470,666	462,948	271,480	587,837	679,631	NA	-0,315	0,262	-0,297	-0,292	0,155	-0,173	-0,172	-0,178	-0,160	-0,162	-0,237	-0,251
M18	251,925	295,648	304,368	265,138	389,805	474,158	218,903	281,148	456,046	474,941	410,531	310,078	340,590	261,037	270,624	398,965	668,020	NA	-0,305	0,283	0,228	-0,181	-0,222	-0,154	-0,155	-0,201	-0,137	-0,183	-0,193
M19	392,506	426,448	541,089	543,321	665,401	174,816	716,949	689,951	284,837	240,719	266,507	696,453	823,322	431,314	481,818	582,228	461,408	623,970	NA	-0,362	-0,334	-0,160	0,158	-0,174	-0,189	0,169	-0,197	-0,245	-0,266
M20	355,469	304,993	328,090	291,239	476,897	654,109	459,619	774,718	614,632	727,981	605,080	664,243	775,267	273,233	289,332	299,489	594,053	539,558	881,618	NA	0,330	-0,220	-0,261	-0,172	-0,176	-0,207	-0,143	-0,175	-0,195
M21	333,462	284,939	228,678	210,481	388,459	557,794	425,343	503,306	651,604	726,246	475,527	405,714	504,751	221,991	237,712	245,589	573,261	349,324	749,180	733,486	NA	-0,206	-0,242	-0,164	0,172	-0,220	-0,145	-0,169	0,206
M22	237,524	178,310	246,456	244,742	292,610	141,931	248,681	265,520	187,094	164,754	164,196	265,310	275,177	163,447	220,269	178,108	161,031	220,083	172,979	326,959	286,684	NA	0,272	0,194	-0,172	0,179	-0,162	-0,182	-0,195
M23	272,002	286,551	319,060	271,150	344,829	157,873	362,532	314,245	178,692	173,800	153,647	343,600	357,748	242,097	336,711	396,524	201,249	331,679	167,086	456,842	394,921	498,398	NA	0,207	-0,198	0,208	-0,193	-0,222	-0,231
M24	169,510	147,698	179,160	185,031	185,024	144,366	143,870	182,063	157,558	153,491	177,607	194,258	175,054	150,027	150,176	144,015	200,177	160,189	204,838	198,281	179,950	252,626	287,796	NA	-0,194	0,184	0,163	0,152	-0,173
M25	195,178	195,271	215,269	196,585	197,475	263,399	152,024	231,909	298,501	293,785	267,821	248,086	245,135	197,432	147,220	142,324	212,029	161,054	239,160	208,601	199,725	198,588	264,830	253,821	NA	-0,213	0,168	0,191	0,176
M26	242,244	189,488	299,823	266,274	311,334	114,773	264,314	250,482	169,316	154,546	158,169	244,543	290,100	189,501	320,674	255,006	171,478	271,716	192,858	289,429	326,491	216,338	292,365	227,595	306,327	NA	-0,199	-0,217	-0,236
M27	190,739	155,304	185,777	187,939	193,435	251,130	172,424	192,971	275,225	208,998	242,236	200,156	206,431	140,882	116,500	108,640	176,672	126,672	259,897	138,041	141,156	176,727	251,032	178,072	189,175	265,619	NA	0,316	0,221
M28	232,807	224,351	230,808	240,328	242,121	411,945	183,197	237,908	495,664	418,261	472,928	264,371	238,740	200,525	167,730	148,656	377,987	224,374	404,301	206,927	192,937	221,855	330,369	155,596	245,043	316,955	673,866	NA	0,291
M29	364,291	264,187	291,934	292,483	304,018	476,989	221,437	256,369	547,871	582,717	528,132	346,321	427,778	205,713	164,953	167,797	424,769	251,541	475,380	256,165	286,788	255,898	358,139	202,057	207,604	375,087	328,438	571,356	NA

EK 8. GAMOM Altında Yen' in Q3 İstatistikleri

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29
M1	NA	0,179	-0,185	-0,180	-0,190	0,263	-0,177	-0,201	0,287	0,268	0,245	-0,221	-0,246	-0,159	-0,208	-0,201	0,199	-0,202	0,228	-0,237	-0,230	0,190	0,204	0,156	-0,178	0,196	-0,176	-0,193	-0,241
M2	214,419	NA	-0,208	-0,210	-0,249	0,284	-0,216	-0,245	0,282	0,317	0,296	-0,268	-0,283	-0,167	-0,223	-0,215	0,199	-0,216	0,240	-0,219	-0,212	0,168	0,210	0,148	-0,177	0,173	-0,158	-0,189	-0,204
M3	230,492	290,707	NA	0,204	0,219	0,342	-0,201	-0,204	0,321	0,435	0,348	-0,223	-0,219	-0,183	-0,207	-0,217	0,236	-0,221	0,270	-0,231	-0,195	0,196	0,222	0,162	-0,189	0,221	-0,175	-0,195	-0,219
M4	217,267	298,104	278,666	NA	0,234	0,351	-0,274	-0,259	0,303	0,461	0,332	-0,313	-0,359	-0,184	-0,224	-0,230	0,245	-0,206	0,271	-0,217	-0,187	0,195	0,204	0,165	-0,181	0,205	-0,177	-0,199	-0,219
M5	241,644	417,630	323,445	367,498	NA	0,411	-0,277	-0,418	0,354	0,542	0,400	-0,478	-0,432	-0,192	-0,216	-0,217	0,275	-0,249	0,303	-0,277	-0,250	0,211	0,230	0,165	-0,182	0,221	-0,180	-0,200	-0,223
M6	465,666	543,475	786,638	830,620	1137,769	NA	0,344	0,334	0,282	0,219	0,218	0,320	0,357	0,235	0,289	0,261	0,175	0,259	0,181	0,319	0,292	0,153	0,164	-0,149	0,194	0,135	0,182	0,236	0,267
M7	209,883	314,326	272,944	504,124	517,608	794,860	NA	0,240	0,314	0,337	0,315	-0,199	-0,254	-0,210	-0,162	-0,180	0,249	-0,187	0,310	-0,266	-0,256	0,194	0,238	0,147	-0,156	0,203	-0,164	-0,170	-0,187
M8	270,817	404,007	280,784	450,092	1177,817	748,930	386,502	NA	0,357	0,346	0,319	0,330	-0,325	-0,212	-0,201	-0,200	0,232	-0,212	0,310	-0,345	-0,278	0,203	0,226	0,173	-0,193	0,202	-0,175	-0,194	-0,202
M9	555,008	533,478	692,140	617,117	843,024	533,339	662,584	856,010	NA	0,264	0,265	0,365	0,381	0,256	0,269	0,223	0,197	0,234	0,215	0,275	0,280	0,168	0,168	0,149	0,198	0,164	0,187	0,245	0,259
M10	482,632	676,453	1271,723	1427,848	1977,735	321,628	765,000	805,977	468,160	NA	0,256	0,336	0,372	0,269	0,317	0,282	0,217	0,276	0,202	0,350	0,357	0,160	0,165	0,152	0,208	0,150	0,168	0,245	0,305
M11	403,841	590,027	814,549	739,410	1074,508	319,030	665,906	682,649	470,846	440,057	NA	0,322	0,333	0,247	0,248	0,245	0,211	0,246	0,217	0,305	0,270	0,163	0,163	0,166	0,195	0,157	0,183	0,257	0,282
M12	327,650	482,778	334,517	660,470	1539,760	688,057	267,689	733,107	896,257	758,631	698,344	NA	0,311	-0,207	-0,198	-0,217	0,258	-0,222	0,313	-0,320	-0,251	0,204	0,235	0,174	-0,199	0,199	-0,178	-0,205	-0,233
M13	406,736	540,061	322,625	867,158	1255,776	857,098	432,479	712,022	977,313	932,961	746,475	650,397	NA	-0,227	-0,213	-0,230	0,256	-0,233	0,340	-0,346	-0,280	0,208	0,240	0,161	-0,199	0,215	-0,181	-0,195	-0,259
M14	169,646	187,494	225,592	227,047	246,854	370,219	296,097	303,411	440,547	487,974	411,173	287,169	347,849	NA	-0,214	-0,199	0,193	-0,204	0,241	-0,207	-0,188	0,162	0,190	0,146	-0,176	0,173	-0,150	-0,178	-0,180
M15	291,454	333,802	287,203	337,047	312,945	563,530	177,318	272,427	485,237	674,742	413,486	263,753	306,421	309,236	NA	0,295	-0,294	0,205	0,258	-0,211	-0,193	0,181	0,225	0,148	-0,153	0,221	-0,135	-0,162	-0,161
M16	270,502	309,832	317,708	356,816	316,631	457,138	217,722	270,435	335,266	536,663	405,113	317,651	356,386	266,477	585,328	NA	-0,322	0,247	0,282	-0,215	-0,196	0,163	0,245	0,148	-0,150	0,199	-0,130	-0,152	-0,163
M17	267,546	265,128	374,499	403,479	509,290	206,062	417,106	363,077	260,130	316,413	298,277	447,789	440,214	251,017	579,890	697,735	NA	-0,318	0,292	0,292	0,285	0,159	0,185	0,180	0,175	0,165	0,152	0,229	0,243
M18	275,456	315,112	329,353	286,744	416,422	452,720	235,238	302,134	368,188	513,624	405,614	331,617	365,147	280,306	283,376	411,572	680,035	NA	0,297	0,287	0,232	0,181	0,224	0,153	-0,158	0,204	-0,141	-0,187	-0,197
M19	348,567	386,364	491,349	495,434	617,083	219,316	644,565	648,281	311,413	273,666	316,594	657,040	775,934	390,309	447,299	534,951	574,667	593,579	NA	-0,359	-0,331	0,167	0,171	0,180	0,186	0,181	0,186	0,232	0,256
M20	378,836	323,356	357,589	318,105	515,629	684,625	475,605	800,290	507,385	825,678	623,867	689,521	803,905	287,974	300,350	309,658	572,272	552,537	866,693	NA	0,333	0,223	0,260	0,165	-0,180	0,211	-0,147	-0,179	-0,198
M21	356,458	302,974	254,693	234,364	420,744	572,931	439,942	518,601	527,050	855,758	489,718	425,387	526,698	237,509	251,444	258,836	546,408	362,889	736,527	746,848	NA	0,207	0,243	0,158	-0,176	0,221	-0,150	-0,174	0,209
M22	242,020	190,234	257,865	256,704	300,955	156,914	252,666	278,361	190,125	171,460	177,842	280,186	291,193	176,973	221,462	178,497	169,979	219,721	188,703	334,515	288,302	NA	0,289	0,203	0,171	0,187	0,159	0,179	0,195
M23	278,898	297,884	332,503	280,911	356,839	179,864	379,582	344,489	189,517	182,065	179,139	370,194	386,162	242,100	340,324	403,375	229,055	336,705	197,846	455,120	398,622	563,283	NA	0,219	0,198	0,224	0,189	0,219	0,230
M24	164,029	146,688	176,900	182,372	182,583	149,915	144,688	200,592	148,848	155,836	185,295	204,404	175,404	143,717	146,789	147,273	219,000	156,505	217,944	182,890	167,914	276,924	322,792	NA	0,189	0,189	-0,163	-0,151	-0,170
M25	213,732	211,598	239,594	220,394	222,961	254,247	163,782	249,994	263,592	290,598	255,290	265,877	265,446	208,602	157,015	150,387	205,507	168,234	233,161	216,926	209,056	196,619	263,276	240,942	NA	0,213	-0,168	0,192	0,178
M26	258,280	200,288	328,378	282,594	329,268	122,428	278,058	275,622	180,897	152,190	165,311	265,610	311,027	200,420	328,174	265,958	183,066	278,732	219,601	299,955	329,716	234,525	337,706	239,233	305,004	NA	0,197	0,216	0,237
M27	208,851	167,383	207,070	210,450	217,291	222,901	181,830	205,704	235,111	189,183	226,356	214,244	221,495	151,808	122,574	113,542	156,460	134,164	232,205	145,636	150,610	170,515	239,479	178,839	190,564	261,108	NA	0,314	0,222
M28	251,863	239,467	254,818	267,037	268,083	374,995	194,880	253,622	404,337	402,476	443,150	281,386	255,374	212,344	175,982	156,057	352,568	234,667	362,904	216,481	203,073	216,091	323,634	154,117	247,002	315,256	664,422	NA	0,291
M29	391,003	280,371	323,053	322,922	335,348	480,891	235,280	273,822	452,323	626,767	533,789	366,600	452,795	219,060	175,403	177,811	397,656	260,947	441,841	264,553	294,599	254,690	356,015	195,175	214,105	376,640	330,254	571,473	NA

EK 9. Kesme Puanları

Tablo 15

Kesme Puanları

			WLE	MAP	EAP
ATM	Düşük	C(0,10)	-1,355	-1,318	-1,327
		C(0,15)	-1,065	-1,038	-1,039
		C(0,20)	-0,763	-0,743	-0,743
	Yüksek	C(0,80)	0,771	0,749	0,747
		C(0,85)	0,942	0,912	0,911
		C(0,90)	1,187	1,146	1,149
GAMOM	Düşük	C(0,10)	-1,121	-1,106	-1,120
		C(0,15)	-0,853	-0,844	-0,845
		C(0,20)	-0,667	-0,657	-0,656
	Yüksek	C(0,80)	0,501	0,491	0,590
		C(0,85)	0,694	0,679	0,888
		C(0,90)	1,342	1,263	1,259
KARMA	Düşük	C(0,10)	-1,134	-1,121	-1,134
		C(0,15)	-0,858	-0,848	-0,852
		C(0,20)	-0,652	-0,641	-0,640
	Yüksek	C(0,80)	0,532	0,520	0,541
		C(0,85)	0,706	0,691	0,744
		C(0,90)	0,975	0,961	1,126



GAZİLİ OLMAK AYRICALIKTIR...