



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

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

SATO TEST KURAMI'NIN KLASİK TEST KURAMI VE MADDE TEPKİ KURAMI  
İLE PSİKOMETRİK AÇIDAN KARŞILAŞTIRILMASI

Sait ÇÜM

Doktora Tezi

Ankara, 2019

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye... En iyiye...*



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

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

SATO TEST KURAMI'NIN KLASİK TEST KURAMI VE MADDE TEPKİ KURAMI  
İLE PSİKOMETRİK AÇIDAN KARŞILAŞTIRILMASI

A PSYCHOMETRIC COMPARISON OF SATO TEST THEORY WITH  
CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY





Sait ÇÜM

Doktora Tezi

Ankara, 2019

### Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,  
Sait ¼M¼n hazırladıđı "SATO TEST KURAMI'NIN KLASİK TEST KURAMI VE  
MADDE TEPKİ KURAMI İLE PSİKOMETRİK AIDAN KARŐILAŐTIRILMASI" baŐlıklı  
bu alıŐma j¼rimiz tarafından Eđitim Bilimleri Ana Bilim Dalı, Eđitimde ¼lme ve  
Deđerlendirme Bilim Dalında Doktora Tezi olarak kabul edilmiŐtir.

J¼ri BaŐkanı	Prof. Dr. Nizamettin KO	İmza 
J¼ri Üyesi (DanıŐman)	Prof. Dr. Selahattin GELBAL	İmza 
J¼ri Üyesi	Prof. Dr. H¼lya KELECİOđLU	İmza 
J¼ri Üyesi	Prof. Dr. Nuri DOđAN	İmza 
J¼ri Üyesi	Do. Dr. Celal Deha DOđAN	İmza 

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, ¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 21 / 12 / 2018 tarihinde uygun g¼r¼lm¼Ő ve Enstit¼ Y¼netim Kurulunca .... / .... / ..... tarihinde kabul edilmiŐtir.

Prof. Dr. Ali Ekber ŐAHİN  
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

## Öz

Bu arařtırmada, Öğrenci Başarılarının Belirlenmesi Sınavına (ÖBBS) katılan öğrencilerin verdikleri yanıtlar üzerinden, Türkçe ve matematik alt testlerinin madde özelliklerinin ve öğrenci başarı puanlarının Sato Test Kuramı (STK),Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) ile belirlenmesi ve elde edilen bulguların karşılaştırılarak incelenmesi amaçlanmıştır. Arařtırma, ÖBBS'ye (2005) katılan 8.sınıf öğrencilerinden oluşan 15461 kiři üzerinde yürütölmüřtür. Yapılan analizler sonrasında incelenen STK indekslerinin küçük örneklemlerde dahi istatistiksel tutarlılık gösterdiği sonucuna ulařılmıştır. STK ile yapılan ve maddelerin işlerliğiyle ilgili bilgi veren sınıflandırmaların farklı örneklemler arasındaki uyumu incelendiğinde, Türkçe testi maddelerinin 100 ve 200'er kişilik örneklemler arasında tamamına yakınının 600'er kişilik örneklemler arasında ise tümünün yüksek uyum yüzdeleriyle sınıflandırıldığı belirlenmiştir. Matematik testi maddeleri için yapılan sınıflandırmaların ise 600'er kişilik örneklemler arasında yüksek uyum gösterdiği fakat daha küçük örneklemlerde benzer bir sonuca ulaşamadığı belirlenmiştir. Söz konusu nedenle Türkçe testi maddeleri için yapılan sınıflandırmaların daha kararlı bir duruş sergilediği sonucuna ulařılmıştır. Birey ve madde özelliklerinin belirlenmesi noktasında ise STK'nın diđer kuramlarla benzer sonuçlar ortaya koyduğu ve sorunlu maddelerin belirlenmesi konusunda KTK ve MTK ile yakın sonuçlar ürettiği görölmüřtür. Bu arařtırmada ulařılan sonuçlar STK'nın, psikometrik açıdan kendi içerisinde tutarlı ve diđer test kuramlarıyla çeliřmeyen kestirimleriyle geçerli ve güvenilir ölçmeler yapılmasına olanak tanıyabilecek alternatif bir test kuramı olarak deđerlendirilebileceği iddiasını desteklemektedir.

**Anahtar sözcükler:** sato test kuramı, öğrenci-madde çizelgesi analizi, klasik test kuramı, madde tepki kuramı, test geliştirme, psikometri

## Abstract

In this study, based on the answers given by the students who participated in the Student Achievement Determination Exam, it is aimed to determine the item characteristics of the mathematics and Turkish sub-tests and student achievement scores with Sato Test Theory (STT), Classical Test Theory (CTT) and Item Response Theory (IRT) and to compare the obtained findings. The research was conducted on 15461 people, 8th-grade students. After the analyzes performed, it has achieved a solid posture that the STT indices examined are statistically consistent even in small samples. When the harmony between the different samples of the STT classifications giving information about the functioning of the items is examined, it was determined that nearly all of the Turkish test items were classified with high compliance percentages among 100 and 200 person samples, and all of them were classified with high compliance percentages among 600 person samples. Classification for Turkish test items has reached a result that has a more stable stance than mathematical test items. At the point of determining the individual and items properties, it is seen that the STT has similar results with other theories and very close results with the CTT and IRT were found especially at the point of identifying the problematic items. The results of this research, it supports the claim that the STT can be evaluated as an alternative test theory, which is psychometric consistent in itself and allows valid and reliable measurements to be made with estimates that do not contradict other test theories.

**Keywords:** sato test theory, student-problem chart analysis, classical test theory, item response theory, test development, psychometry

## Teşekkür

Bu tez, eğitimle ve araştırmayla iç içe geçmiş yılların, karşılaşılan bütün zorluklara rağmen azimle verilen uzun bir mücadelenin ürünüdür. Bu ürünün ortaya çıkmasında, sözünü ettiğim mücadelede bana destek olanların ve yol gösterenlerin de en az benim kadar payı vardır. Bu güne kadar okuduğum her satır ve işittiğim her sözcük bu tezin yapı taşını oluşturan paha biçilemez değerlerdir. Bana bu değerleri katan ve öğrenmeme fırsat veren herkese sonsuz teşekkürlerimi sunarım.

Değerli görüşleriyle yoluma ışık tutan, çalışmamı zenginleştiren ve beni her anlamda cesaretlendiren danışmanım Prof. Dr. Selahattin GELBAL olmasa bu tez de ortaya çıkmazdı. Bu aşamadan sonra da her daim görüşlerine başvurmaya devam edeceğim fakat katılmaktan son derece keyif aldığım derslerini eminim çok özleyeceğim. Öğrencisi olmaktan onur duydum. Kendisine sonsuz teşekkürlerimi sunuyorum.

Yüksek lisans tez danışmanım olan, doktora tez çalışmamda da beni bir an olsun yalnız bırakmayan, akademide ve hayatta varlığından büyük güç aldığım saygıdeğer hocam Prof. Dr. Nizamettin KOÇ'a da önemli katkılarından dolayı teşekkür etmeyi bir borç bilirim. Bu yolu o olmadan yürüyemezdim.

Yaptığım pek çok araştırmada olduğu gibi bu araştırmanın da her aşamasında Prof. Dr. Hülya KELECİOĞLU'ndan izler vardır. Tez izleme komitemde yer aldığı için de kendisine ayrıca bir teşekkür borçluyum. Bana ayırdığı her saniye benim için çok kıymetliydi.

Lisansüstü eğitimim boyunca derslerine katılmaktan ve birlikte çalışmaktan büyük keyif aldığım saygıdeğer hocalarım, Prof. Dr. Nuri DOĞAN, Doç. Dr. Burcu ATAR, Prof. Dr. Halil YURDUGÜL, Prof. Dr. Ezel TAVŞANCIL, Doç. Dr. Ömay ÇOKLUK, Prof. Dr. Nükhet ÇIKRIKÇI, Dr. Öğretim Üyesi Ömer KUTLU ve Doç. Dr. Deha DOĞAN'a da teşekkür etmesem olmazdı. İyi ki kendilerini tanıma fırsatı yakaladım.

Tezimle yakından ilgilenen ve bu konu üzerinde alıřmam iin bana cesaret veren saygıdeęer hocam Do. Dr. İrfan YURDABAKAN'a da teřekkürlerimi sunarım. Kendisiyle Sato Test Kuramı üzerinde alıřma fırsatı bulduęum iin de ayrıca mutluluk duydum.

Sato Test Kuramı üzerinde birlikte alıřma řansı yakaladıęım deęerli dostlarım Dr. Murat Doęan řAHİN, Dr. Ching-Pin TSAİ ve Arř. Gör. Elif Kübra DEMİR'e de teřekkür etmeden gemek olmazdı. Sayelerinde konu üzerindeki bilgilerimi derinleřtirme olanaęı buldum.

Hayatım boyunca varlıklarından ok büyük mutluluk duydugum annem Emine Gül ÜM, babam Bilal ÜM ve kardeřim Bilgesu ÜM'e de sevgileri ve destekleri iin sonsuz teřekkürlerimi sunarım. Onlar olmadan bařaramazdım.

Hayatıma yeni bir anlam kazandıran deęerli eřim řölen řENER ÜM'e de neřesiyle enerjimi oęalttıęı, bana her zaman güvenerek motivasyonumu artırdıęı iin sonsuz teřekkürlerimi sunarım. Sayesinde her gün yenileniyorum ve yeniden alıřabilmek iin ihtiyacım olan gücü buluyorum.

Yeniden ve yeniden...

Teřekkürler.



## İçindekiler

Öz.....	ii
Abstract .....	iii
Teşekkür.....	iv
İçindekiler .....	vi
Tablolar Dizini.....	ix
Şekiller Dizini.....	xii
Simgeler ve Kısaltmalar Dizini.....	xiii
Bölüm 1 Giriş.....	1
Problem Durumu .....	1
Araştırmanın Amacı ve Önemi .....	5
Araştırma Problemi .....	7
Alt Problemler .....	7
Sınırlılıklar .....	8
Tanımlar .....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	10
Klasik Test Kuramı .....	10
KTK'da Madde İstatistikleri.....	11
Madde güçlük indeksi.....	11
Madde varyansı .....	11
Madde ayırıcılık indeksi.....	12
Madde güvenilirlik katsayısı.....	13
KTK'nın Avantajları .....	13
KTK'nın Dezavantajları .....	14
Madde Tepki Kuramı.....	15
MTK'da Madde Parametreleri .....	16
<b>bi</b> parametresi.....	16
<b>ai</b> parametresi .....	16
<b>ci</b> parametresi .....	17
İkili Puanlanan Maddeler için MTK Modelleri .....	18

1 parametrelı lojistik model.....	18
2 parametrelı lojistik model.....	18
3 parametrelı lojistik model.....	18
Madde Bilgi Fonksiyonu .....	19
Test Bilgi Fonksiyonu .....	20
MTK'nın Avantajları.....	20
MTK'nın Dezavantajları.....	22
Sato Test Kuramı .....	23
Madde Uyarı İndeksi .....	24
Öğrenci Uyarı İndeksi.....	25
Öğrenciler için Sınırlandırılmıř Gri İliřki Derecesi .....	27
Maddeler için Sınırlandırılmıř Gri İliřki Derecesi .....	28
STK'nın Avantajları .....	29
STK'nın Dezavantajları .....	30
İlgili Arařtırmalar .....	31
İlgili Arařtırmalar Özet.....	33
<b>Bölüm 3 Yöntem.....</b>	<b>34</b>
Arařtırmanın Türü .....	34
Arařtırmanın Çalıřma Grubu .....	34
Veri Toplama Araçları .....	34
Verilerin Analizi .....	36
Varsayımların Test Edilmesi .....	36
Veri Çözümleme Süreci .....	39
<b>Bölüm 4 Bulgular ve Yorumlar.....</b>	<b>43</b>
Arařtırmanın Birinci Alt Problemine İliřkin Bulgular ve Yorumlar .....	43
Arařtırmanın İkinci Alt Problemine İliřkin Bulgular ve Yorumlar .....	54
Arařtırmanın Üçüncü Alt Problemine İliřkin Bulgular ve Yorumlar .....	55
Arařtırmanın Dördüncü Alt Problemine İliřkin Bulgular ve Yorumlar .....	62
Arařtırmanın Beřinci Alt Problemine İliřkin Bulgular ve Yorumlar .....	64
Arařtırmanın Altıncı Alt Problemine İliřkin Bulgular ve Yorumlar .....	65
Arařtırmanın Yedinci Alt Problemine İliřkin Bulgular ve Yorumlar .....	67
Arařtırmanın Sekizinci Alt Problemine İliřkin Bulgular ve Yorumlar .....	68

Araştırmanın Dokuzuncu Alt Problemine İlişkin Bulgular ve Yorumlar .....	70
Araştırmanın Onuncu Alt Problemine İlişkin Bulgular ve Yorumlar.....	72
Bölüm 5 Sonuç, Tartışma ve Öneriler.....	86
Sonuçlar.....	86
Öneriler .....	89
Araştırma Bulgularına ve Sonuçlarına Dayalı Öneriler.....	89
Benzer Araştırmalar için Öneriler .....	90
Kaynaklar .....	91
EK-A: Türkçe Testi için Maddeler Arası Standardize Edilmiş ki-kare Değerleri	97
EK-B: Matematik Testi için Maddeler Arası Standardize Edilmiş ki-kare Değerleri .....	99
EK-C: Türkçe Testi Madde Bilgi Fonksiyonu Değerleri .....	101
EK-Ç: Matematik Testi Madde Bilgi Fonksiyonu Değerleri.....	102
EK-D: Tez Çalışması Etik Komisyon İzin Muafiyeti Formu.....	103
EK-E: Etik Beyanı.....	104
EK-F: Doktora Tez Çalışması Orijinallik Raporu .....	105
EK-G: Dissertation Originality Report.....	106
EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı .....	107

## Tablolar Dizini

Tablo 1 ÖİSGİD Değerlendirme Ölçütleri .....	28
Tablo 2 Lojistik Modeller için -2log Likelihood Değerleri .....	38
Tablo 3 Farklı Gruplardan Elde Edilen Madde Parametreleri Arasındaki Korelasyonlar .....	39
Tablo 4 Türkçe Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	43
Tablo 5 Türkçe Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	44
Tablo 6 Türkçe Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	45
Tablo 7 Türkçe Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	45
Tablo 8 Türkçe Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	46
Tablo 9 Türkçe Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	46
Tablo 10 Türkçe Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	47
Tablo 11 Türkçe Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	47
Tablo 12 Matematik Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	49
Tablo 13 Matematik Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	49
Tablo 14 Matematik Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	50
Tablo 15 Matematik Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	51
Tablo 16 Matematik Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar .....	51
Tablo 17 Matematik Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular .....	52

Tablo 18 <i>Matematik Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar</i> .....	52
Tablo 19 <i>Matematik Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular</i> .....	53
Tablo 20 <i>Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 50'şer Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	55
Tablo 21 <i>Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 100'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	56
Tablo 22 <i>Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 200'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	57
Tablo 23 <i>Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 600'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	58
Tablo 24 <i>Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 50'şer Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	59
Tablo 25 <i>Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 100'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	60
Tablo 26 <i>Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 200'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	61
Tablo 27 <i>Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 600'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular</i> .....	62
Tablo 28 <i>Türkçe Testi Madde Ayırıcılık İndeksleri ile Madde Uyarı İndeksleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	63
Tablo 29 <i>Matematik Testi Madde Ayırıcılık İndeksleri ile Madde Uyarı İndeksleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	63
Tablo 30 <i>Türkçe Testi Madde Güçlük İndeksleri ile Maddeler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	64
Tablo 31 <i>Matematik Testi Madde Güçlük İndeksleri ile Maddeler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> ..	65
Tablo 32 <i>Türkçe Testi Toplam Puanları ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	66
Tablo 33 <i>Matematik Testi Toplam Puanları ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	66

Tablo 34 <i>Türkçe Testi Madde Uyarı İndeksleri ile a Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	67
Tablo 35 <i>Matematik Testi Madde Uyarı İndeksleri ile a Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	68
Tablo 36 <i>Türkçe Testi Maddeler için Sınırlandırılmış Gri İlişki Dereceleri ile b Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	69
Tablo 37 <i>Matematik Testi Maddeler için Sınırlandırılmış Gri İlişki Dereceleri ile b Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	70
Tablo 38 <i>Türkçe Testini Yanıtlayan Öğrencilerin Yetenek Parametreleri ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	71
Tablo 39 <i>Matematik Testini Yanıtlayan Öğrencilerin Yetenek Parametreleri ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular</i> .....	71
Tablo 40 <i>Türkçe Testi Maddelerinin Ayırıcılık İndeksleri</i> .....	72
Tablo 41 <i>Türkçe Testindeki Sorunlu Maddelerin Klasik Test Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	73
Tablo 42 <i>Türkçe Testi Maddelerinin Atandığı Sınıflar</i> .....	74
Tablo 43 <i>Türkçe Testindeki Sorunlu Maddelerin Sato Test Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	74
Tablo 44 <i>Türkçe Testi Maddelerinin a ve c Parametrelerinin Değerleri</i> .....	77
Tablo 45 <i>Türkçe Testindeki Sorunlu Maddelerin Madde Tepki Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	78
Tablo 46 <i>Matematik Testi Maddelerinin Ayırıcılık İndeksleri</i> .....	79
Tablo 47 <i>Matematik Testindeki Sorunlu Maddelerin Klasik Test Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	80
Tablo 48 <i>Matematik Testi Maddelerinin Atandığı Sınıflar</i> .....	80
Tablo 49 <i>Matematik Testindeki Sorunlu Maddelerin Sato Test Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	81
Tablo 50 <i>Matematik Testi Maddelerinin a ve c Parametrelerinin Değerleri</i> .....	84
Tablo 51 <i>Matematik Testindeki Sorunlu Maddelerin Madde Tepki Kuramı'na göre Belirlenmesine İlişkin Bulgular</i> .....	85

## Şekiller Dizini

Şekil 1. İkili puanlanan bir maddeye ait madde karakteristik eğrisi (üç parametrelî model). .....	16
Şekil 2. Madde bilgi fonksiyonu grafiđi.....	19
Şekil 3. Test bilgi ve standart hata fonksiyonları grafiđi. ....	20
Şekil 4. MUİ ve oran deđerlerine göre maddelerin atanacađı sınıflar. ....	24
Şekil 5. ÖÜİ ve oran deđerlerine göre öğrencilerin atanacađı sınıflar. ....	26
Şekil 6. Türkçe test puanlarının dađılımına ilişkin histogram grafiđi. ....	35
Şekil 7. Matematik test puanlarının dađılımına ilişkin histogram grafiđi. ....	36
Şekil 8. Türkçe testi MİSGİD deđerleri arasındaki korelasyon katsayılarının ortanca deđerlerinin örneklem büyüklüğü bağlamında karşılaştırılması. ....	48
Şekil 9. Matematik testi MİSGİD deđerleri arasındaki korelasyon katsayılarının ortanca deđerlerinin örneklem büyüklüğü bağlamında karşılaştırılması. ....	53
Şekil 10. Türkçe testi madde bilgi fonksiyonları. ....	76
Şekil 11. Matematik testi madde bilgi fonksiyonları. ....	83

## Simgeler ve Kısaltmalar Dizini

**1PLM:** 1Parametrelı Lojistik Model

**2PLM:** 2Parametrelı Lojistik Model

**3PLM:** 3Parametrelı Lojistik Model

**KTK:** Klasik Test Kuramı

**MİSGİD:** Maddeler için Sınırlandırılmıř Gri İliřki Derecesi

**MTK:** Madde Tepki Kuramı

**MUI:** Madde Uyarı İndeksi

**ÖBBS:** Öğrenci Başarılarının Belirlenmesi Sınavı

**ÖİSGİD:** Öğrenciler için Sınırlandırılmıř Gri İliřki Derecesi

**ÖUI:** Öğrenci Uyarı İndeksi

**STK:** Sato Test Kuramı



## Bölüm 1

### Giriş

Bu bölümde araştırmanın problem durumuna, problem cümlesine, alt problemlere, araştırmanın amacı ve önemi başlıklarına yer verilmiştir.

#### Problem Durumu

Modern toplumda eğitimde ve psikolojide kullanılan testler bireylerin hayatında çok önemli bir rol oynamaktadır. Test puanları, bireylerin eğitim görecekleri okulları büyük ölçüde belirlemekte dolayısıyla uzmanlaşacakları mesleklerin belirlenmesi sürecinde kritik bir unsur olarak ortaya çıkmaktadır. Başka bir bakış açısıyla test puanları, kişinin hayatı boyunca kendisini nasıl göreceği ve başkalarının onu nasıl göreceğini belirleyen bir etmen haline gelmiştir. Bu anlamda test puanlarının nasıl elde edileceği ve nasıl yorumlanacağı konuları üzerinde dikkatle durulmalıdır. Eğitimde ve psikolojide test geliştirme ve ölçme süreçleri, karmaşık ve geniş kapsamlı bilgiler bütünü üzerine inşa edilmiş bir bilim dalı olan psikometri çerçevesinde ele alınır. Psikometristler, test maddelerinin nasıl geliştirileceği, nasıl seçileceği, verilen cevapların sayılara nasıl dönüştürüleceği, bunların nasıl yorumlanacağı ve sonuçların nasıl değerlendirileceği gibi sorulara matematiksel ve istatistiksel teorilere dayalı olarak yanıt ararlar. Söz konusu teorilere “test kuramı” ya da “psikometrik kuram” adı verilmektedir (Suen, 1990).

Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) geçmişten günümüze üzerinde en fazla çalışılan başlıca iki test kuramı olarak öne çıkmaktadır. Alanyazında, Klasik Test Kuramı ve Madde Tepki Kuramının bazı üstün ve zayıf yönlerinin tartışıldığı görülmektedir. Hambleton, Swaminathan ve Rogers (1991), Klasik Test Kuramının sınırlılıklarını şu şekilde açıklamaktadırlar: Madde ve test istatistikleri gruba bağlı olarak hesaplanır. Bu sınırlılık madde ve test özelliklerinin birbirinden ayrılamamasıyla ilişkilidir. Eğer bir test zor ise testi yanıtlayan kişi düşük yetenek düzeyinde, eğer kolay ise yüksek yetenek düzeyinde görünebilir. Testi yanıtlayanların yetenek düzeyi ise test maddelerinin güçlük düzeyini belirlemektedir çünkü KTK’da maddelerin ayırt ediciliği ve güçlüğü belli bir grup için geçerlidir. Bu nedenle madde parametreleri uygulamanın yapıldığı gruba bağlı olarak değişecektir. Bu durum paralel formlar oluşturma ve farklı testlere

katılan bireyleri karşılaştırma konularında sorunların yaşanmasına neden olur. Ayrıca KTK'ya göre doğru yanıtlanan madde sayısından elde edilen toplam puan bireyin yetenek düzeyi olarak kabul edilir. Bu tür bir hesaplamada farklı maddelere doğru yanıt vermiş olmalarına rağmen eşit sayıda doğru yanıtı bulunan bireyler aynı yetenek düzeyinde kabul edilmektedir. Örneğin, 20 maddelik bir matematik testini alan ve farklı yedi maddeye doğru yanıt veren iki öğrenci olduğu varsayılırsa, doğru yanıtlanan maddelerin güçlüklerinin ve ayırt ediciliklerinin önemli derecede birbirinden farklı olması durumunda bu iki öğrencinin aynı yetenek düzeyinde kabul edilmesi sonuçların sağlıklı değerlendirilmesi bakımından sorun oluşturabilir. Özellikle, testlerden elde ettikleri sonuçları öğretim süreçlerini düzenlemek (iyileştirmek) için bir dönüt olarak kullanmak isteyen öğretmenlerin eşit puan alan öğrencilerle ilgili yapacakları değerlendirmeleri yalnızca test puanlarına dayalı olarak yapmalarının yetersiz ve alacakları kararlar bakımından yanıltıcı olabileceği göz ardı edilmemelidir. Bu bakımdan, özellikle biçimlendirmeye yönelik değerlendirme (formative assesment) durumlarında KTK'ya dayalı ölçme süreçlerinden istenilen sonuçların alınamayacağı ve etkili bir geribildirim mekanizmasının işletilemeyeceği iddiası ileri sürülebilir.

MTK modelleri farklı yanıt örüntüsüne sahip eşit sayıda doğru yanıtı bulunan bireylerin yetenek düzeyi bakımından birbirlerinden ayırt edilebilmesi konusunda etkili çözümler sunmaktadır. Fakat MTK'ya dayalı parametrelerin daha hatasız kestirilebilmesi için uygulamaların geniş örneklemeler üzerinde yapılması gerekmektedir (de Ayala, 2009; Hambleton ve van der Linden, 1997; Hulin, Lissak ve Drasgow, 1982; Ree ve Jensen, 1983). Okullarda yapılan ölçmeler ise küçük gruplar üzerinde yürütülmekte ve bu durum başlı başına MTK'ya dayalı test uygulamalarını olanaksız kılmaktadır. Üstelik MTK'ya dayalı test geliştirmek ve sonuçları değerlendirmek alan uzmanlığı gerektiren bir süreçtir. Söz konusu nedenlerden dolayı öğretmen yapımı testlerin geliştirilmesi ve test sonuçlarının değerlendirilmesi sürecinde KTK'ya dayalı tekniklerin kullanılmasının daha uygun olduğu görüşü öne sürülebilir. Buna karşın, KTK'nın sınırlılıklarından kaynaklı olarak ölçme sonuçlarının geçerliği, güvenilirliği ve karşılaştırılabilirliği noktasında ortaya çıkan sorunlara çözüm önerebilecek farklı yöntem ve tekniklerin geliştirilmesi ihtiyacı doğmaktadır.

Takahiro Sato 1969'da diğer test kuramlarının sınırlılıklarından doğan ihtiyaçtan hareketle öğrencilerin cevap örüntüleri üzerinden performanslarını ayırt etmeye yönelik küçük örneklemeler üzerinde de etkili sonuçlar üretebilecek bir sistem üzerinde çalışmalarını yürütmüştür. Sato'nun geliştirmiş olduğu Öğrenci-Madde Çizelgesi Analizi (Student-Problem Chart Analysis) ile hem maddelerin kalitesi hem de öğrencilerin bilme ve öğrenme durumları belirlenebilmektedir. Bu anlamda Sato Test Kuramı (STK) modern test kuramları arasında yerini almaktadır (Sheu vd., 2014a; Switzer ve Connell, 1990; Wu, 1999).

Sato (1984) yaptığı araştırma sonrasında, Japon ilk ve ortaokullarında Öğrenci- Madde Çizelgesi Analizi'nin önemli bir ölçme ve değerlendirme yaklaşımı olarak kabul edildiği sonucunu paylaşmıştır. Söz konusu araştırma, Sato Test Kuramı'nın Japon okullarında uzun zaman önce kullanılmaya başlandığı göstermektedir. Buna karşın, kuramın Avrupa ve Amerika'da hâlâ çok fazla ilgi çekmediği bilinmektedir. Bu durum, Öğrenci-Madde Çizelgesi Analizi'yle elde edilen indekslerin, öğrencilerin yetenek düzeyleri ve madde özelliklerine ilişkin sınırlı bilgi verdiği düşüncesinden kaynaklanıyor olabilir. Söz konusu sınırlılık, analiz sonucu elde edilen nitel veri düzeyindeki çıktıların, bireyleri ve maddeleri belirli özelliklere sahip oluş dereceleri bakımından birbirleri ile karşılaştırabilme, sıralayabilme ve bunlara dayalı olarak çeşitli değerlendirmelerde bulunabilme gibi durumlarda yetersiz kalmasıyla ilişkilidir. Ayrıca, nitel veri yapısında alınan sonuç verisinin ileri düzey analizlerde kullanılması durumunda da çeşitli engellerle karşılaşılacağı belirtilebilir. Sayılan nedenlerle, ortaya çıkışından itibaren uzun süre boyunca, kuramın yalnızca biçimlendirmeye yönelik değerlendirme süreçlerine yönelik olarak kullanıldığı görülmektedir. STK'nın tartışmaya konu olan sınırlılığını ortadan kaldırmak amacıyla Nagai (2010) bulanık mantığa (fuzzy logic) dayalı Gri İlişki Analizi'nin Sato Test Kuramı'na dâhil edilmesini önererek Gri Öğrenci-Madde Çizelgesi Analizi'ni (Grey Student-Problem Chart Analysis) geliştirmiştir. Gri Öğrenci-Madde Çizelgesi Analizi'yle öğrenciler ve maddeler ile ilgili elde edilen nitel çıktıların yanı sıra nicel veri düzeyinde ölçme sonuçları da elde edilebilmektedir (Sheu vd., 2014a).

Gri İlişki Analizi'nin dâhil edilmesiyle birlikte, STK'nın düzey belirleme ve buna dayalı olarak bireyler hakkında çeşitli kararlar alma (değer biçmeye yönelik

değerlendirme) amacına yönelik olarak kullanılabilmesinin de olanaklı hale geldiği ifade edilebilir. Ayrıca alanyazın incelemelerinden hareketle, Gri Öğrenci–Madde Çizelgesi Analizi'nin alan uzmanları tarafından kabul görmesinden sonra özellikle Uzakdoğu Ülkeleri'nde STK'ya ilişkin araştırmaların sayısında önemli bir artış olduğu ve bu artışın giderek ivme kazandığı yönünde bir belirleme yapılabilir.

Günümüzde eğitim, psikoloji, sağlık ve endüstri gibi çok sayıda alanda ölçme ve değerlendirme çalışmalarına verilen önem artmakta ve ölçme faaliyetlerinin önemli bir bölümü küçük gruplar üzerinde gerçekleştirilmektedir. STK, hem ikili puanlanan hem de çoklu puanlanan maddeler için uygun modeller sunmaktadır. Bu bakımdan, yetenek, başarı ya da tipik davranışların ölçülmesi amacıyla pek çok farklı alanda kuramdan yararlanılabileceği iddia edilebilir. Bununla birlikte, özellikle eğitimde ölçme ve değerlendirme için STK diğer test kuramlarına önemli bir alternatif olarak ileri sürülebilir. Gri Öğrenci-Madde Çizelgesi Analizi çıktılarının yorumunun Madde Tepki Kuramı çıktıları kadar uzmanlık gerektirmemesi, analizin küçük grup uygulamaları için tutarlı ve etkili sonuçlar üretmesi (Çüm, Gelbal ve Tsai, 2016; Sheu vd., 2014a; Sheu vd., 2014b), nicel ölçme sonuçlarının yanı sıra öğrencilerin öğrenme durumlarına ve madde özelliklerine ilişkin nitel dönütler sağlaması ve kavram yanılgılarına sahip olan öğrencilerin tespit edilebilmesi gibi kurama ilişkin üstünlükler, sınıf içi ölçme ve değerlendirme süreçlerinde karşılaşılan sorunların aşılabilmesi noktasında öğretmenlerin işini kolaylaştırabilir. Türkiye'de Sato Test Kuramı'yla ilişkili çok az sayıda çalışma yapıldığı ve buna bağlı olarak kuramın yerli alanyazında henüz önemli bir yer edinemediği göz önünde bulundurulduğunda, STK'yla ilgili yapılacak olan çalışmaların kurama olan ilgiyi artırması ve kuramın üstünlüklerine ilişkin iddiaları tartışmaya açması bakımından önemli bir rol üstleneceği beklenebilir. Yapılan tartışmalar doğrultusunda bu araştırmanın psikometri alanında yürütülen çalışmalara katkı sağlayabileceği ve araştırma sonuçlarından hareketle mevcut problemlere çözüm önerileri getirilebileceği düşünülmektedir.

## **Araştırmanın Amacı ve Önemi**

Bu araştırmada, ÖBBS'ye (2005) katılan öğrencilerin Türkçe ve matematik alt testlerine verdikleri yanıtlar üzerinden, madde analizlerinin Sato Test Kuramı (STK), Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) ile yapılarak elde edilen madde özellikleri arasındaki ilişki düzeylerinin ve testlerden çıkarılmasına ya da düzeltilmesine karar verilen sorunlu maddelerin belirlenmesi; sözü edilen kuramlara dayalı olarak elde edilen öğrenci başarı puanlarının karşılaştırılması ve katılımcılar arasından seçkisiz olarak seçilen farklı örneklemeler üzerinden Sato Test Kuramına dayalı olarak elde edilen madde özelliklerinin tutarlılığının incelenmesi amaçlanmıştır.

Günümüzde testler, seçme, tanılama, yönlendirme, durum belirleme, araştırma gibi çeşitli amaçlarla yaşamın pek çok evresinde sıklıkla kullanılmaktadır. Test uygulamalarıyla elde edilen puanların değerlendirilmesi sonucu bireyler hakkında önemli kararlar verilmekte ve bu kararlar onların hayatlarının gidişatıyla ilgili belirleyici bir rol oynamaktadır. Bireyler hakkında verilen kararların isabetliliği oranında işini severek yapan, mutlu ve üretken bir toplum yapısına ulaşılabileceği fikri ileri sürülebilir. Test sonuçlarına dayalı olarak verilen bu kararların isabetliliği üzerinde ise o testte yer alan maddelerin geliştirilmesi, doğru maddelerin seçilerek ilgilenilen özelliği geçerli ve güvenilir bir şekilde ölçebilecek yapının oluşturulması, maddelere verilen tepkilerin sayısal değerlere dönüştürülmesi ve elde edilen değerlerin yorumlanması gibi konulara matematiksel ve istatistiksel alt yapıyı sağlayan "test kuramlarının" temelden bir etkisinin olduğu iddia edilebilir.

Günümüze kadar üzerinde çok sayıda çalışma yapılan, test geliştiriciler ve uygulayıcılar tarafından sıklıkla kullanılan iki büyük test kuramından söz edilebilir. Bunlar, Klasik Test Kuramı ve Madde Tepki Kuramı olarak adlandırılmaktadır. Alanyazında her iki kuramın da birbirlerine göre üstün ve zayıf yönlerinin olduğu ileri sürülmektedir. MTK, içerdiği ileri düzey istatistiksel yöntemlerin varsayımlarından kaynaklı olarak büyük (geniş) örneklemelere ihtiyaç duymaktadır. Bu nedenle, küçük grupların bulunduğu okul, klinik gibi kurumlarda yürütülen ölçme ve değerlendirme çalışmalarında KTK'nın kullanımı yaygın bir şekilde

devam etmektedir. Bununla birlikte, KTK'ya dayalı olarak hesaplanan madde istatistiklerinin gruba bağlı olması ve bireylerin başarı/yetenek puanlarının da madde/test özelliklerine bağlı olması pratikte pek çok sorunu beraberinde getirmektedir. KTK ile yapılan kestirimlerin tutarlılığı da tartışma konusudur. Çüm, Gelbal ve Tsai (2016), KTK'ya dayalı olarak elde edilen madde istatistiklerinin aynı evrene ait farklı örneklemeler (30'ar kişilik) arasında büyük farklılıklar gösterdiği bulgusuna ulaşmışlardır. Bu noktada, küçük örneklemelerde daha etkili ölçmeler yapabilmek ve KTK'da karşılaşılan sorunları aşabilmek bakımından yeni test kuramları üzerinde araştırmalar yapılması ihtiyacı doğmaktadır. Alternatif bir modern test kuramı olarak mevcut sorunlara çözüm getirme iddiasında olan Sato Test Kuramı üzerinde yapılan bu araştırmanın, test geliştirme alanında açılacak olan üçüncü bir yola katkı sağlaması bakımından alan uzmanları ve test uygulayıcıları için önem taşıdığı düşünülmektedir.

Sato Test Kuramıyla ilgili çalışmaların son yıllarda özellikle Uzak Doğu Ülkelerinde bir artış gösterdiği ve yapılan İngilizce yayınların, kuramı tanıtmaya (Lin ve Chen, 2006; Wang, Sheu ve Nagai, 2011; Wang ve Chen 2013), kurama ilişkin geliştirilen bir yazılımı tanıtmaya (Pham, Sheu ve Nagai, 2015; Sheu, Pham, Nguyen ve Nguyen, 2013; Sheu vd., 2014a; Sheu vd., 2014c; Wu, 1999), STK aracılığıyla kavram yanılgılarına neden olan maddeleri ve/veya kavram yanılgılarına sahip olan öğrencileri saptamaya (Sheu vd., 2012; Sheu vd., 2013; Tsai vd., 2014) ve STK ve MTK'nın karşılaştırılması (Sheu vd., 2014b; Tatsuoka, 1984) alanlarıyla ilişkili olduğu belirlenmiştir. Bu belirlemeden hareketle, Gri İlişki Analizi'nin kurama eklendiği 2010 yılından sonra dahi Avrupa ve Amerika'da kurama olan ilginin zayıf kaldığına ilişkin bir çıkarım yapılabilir. Bu anlamda, farklı katılımcı grupları üzerinde STK çerçevesinde üretilen indekslerin çeşitli koşullar altında incelenerek tutarlı sonuçlar alınıp alınmadığının ve elde edilen sonuçların diğer test kuramlarıyla elde edilen sonuçlarla karşılaştırılarak benzerlik ve farklılıkların ortaya konulmasının önemli olduğu düşünülmektedir. Başka bir ifadeyle, kuramın güvenilir ve geçerli sonuçlar üretebildiğine ilişkin elde edilecek olan kanıtların, farklı coğrafyalarda da kurama olan ilgiyi artıracak yönünde bir öngöründe bulunulabilir. Sözü edilen nedenlerden dolayı bu çalışma, kuramın hem kendi içerisindeki tutarlılığının hem de diğer kuramlara göre üstün ve zayıf yönlerinin ortaya koyulabilmesi bakımından önemlidir. Çalışmanın özellikle Türkiye'de şu ana

kadar yeterince ilgi çekmemiş olan Sato Test Kuramı'na olan ilgiyi artıracığı ve ardından yapılacak olan çalışmalar için kapı aralayacağı düşünülmektedir. Bu araştırmanın bu anlamda da önemli olduğu ifade edilebilir.

### **Araştırma Problemi**

Sato Test Kuramı'na dayalı olarak elde edilen madde özelliklerinin farklı büyüklükteki örneklemler arasındaki tutarlılık düzeyleri ve üç farklı test kuramına dayalı olarak (STK, KTK, MTK) elde edilen madde özellikleri ve öğrenci başarı puanlarının farklı büyüklükteki örneklemler üzerinden karşılaştırılması sonucu elde edilen korelasyonlar ne düzeydedir?

### **Alt Problemler**

Bu problem cümlesine ilişkin oluşturulan alt problemler aşağıdaki gibidir:

ÖBBS (2005) Türkçe ve matematik alt testlerinde yer alan maddeler için 50'şer kişilik 10, 100'er kişilik 10, 200'er kişilik 10 ve 600'er kişilik 10 farklı örneklem üzerinden hesaplanan,

1. Maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) arasındaki korelasyonlar ne düzeydedir ve MİSGİD'ler arasındaki farklar istatistiksel olarak manidar mıdır?
2. Alt ve üst gruptan elde edilen maddeler için sınırlandırılmış gri ilişki dereceleri(MİSGİD) arasındaki korelasyonlar ne düzeydedir?
3. Madde uyarı indeksleri (MUİ) aracılığı ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?
4. Sato Test Kuramı'nın madde uyarı indeksleri (MUİ) ile Klasik Test Kuramı'nın madde ayırıcılık indeksleri (Çift Serili Korelasyon Katsayıları) arasındaki korelasyonlar ne düzeydedir?
5. Sato Test Kuramı'nın maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) ile Klasik Test Kuramı'nın madde güçlük indeksleri arasındaki korelasyonlar ne düzeydedir?
6. Sato Test Kuramı'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri (ÖİSGİD) ile Klasik Test Kuramı'nın öğrenci başarı puanları (toplam test puanları) arasındaki korelasyonlar ne düzeydedir?

7. Sato Test Kuramı'nın madde uyarı indeksleri (MUI) ile Madde Tepki Kuramı'nın  $a$  parametreleri arasındaki korelasyonlar ne düzeydedir?
8. Sato Test Kuramı'nın maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) ile Madde Tepki Kuramı'nın  $b$  parametreleri arasındaki korelasyonlar ne düzeydedir?
9. Sato Test Kuramı'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri (ÖİSGİD) ile Madde Tepki Kuramı'nın yetenek parametreleri arasındaki korelasyonlar ne düzeydedir?
10. Klasik Test Kuramı, Madde Tepki Kuramı ve Sato Test Kuramı'na dayalı olarak yürütülen madde analizlerinin sonuçlarından hareketle testten çıkarılması ya da düzeltilmesi gerektiğine karar verilen maddeler ne oranda uyum göstermektedir?

### **Sınırlılıklar**

Bu araştırmada, test kuramları arasındaki psikometrik karşılaştırmalar madde ve birey özellikleri ile sınırlandırılmış olup test özellikleri araştırma kapsamına dâhil edilmemiştir. Madde ve birey özellikleri ise yalnızca ikili puanlanan maddeler üzerinden elde edilmiştir.

Bu araştırmada MTK modellerinden araştırma verisine en iyi uyum sağlayan Üç Parametrelili Lojistik Model ve STK yöntemlerinden Gri Öğrenci-Madde Çizelgesi Analizi kullanılmış ve yapılan karşılaştırmalar yalnızca bunlara dayalı analizler ile sınırlandırılmıştır.

Yapılan karşılaştırmalar, araştırma kapsamında ele alınan her bir örneklem büyüklüğü için seçkisiz olarak belirlenen 10 örneklem üzerinden gerçekleştirilmiştir. Karşılaştırılan örneklemelerin bu sayı ile sınırlandırılmasının nedeni analizler için kullanılan yazılımların tekrarlı analizlere olanak tanıması ve her bir veri grubu için yeni baştan bir düzenleme ve işlem sürecine ihtiyaç duyulması olarak gerekçelendirilebilir.



## Tanımlar

**Madde Özellikleri:** Testlerde yer alan maddelerin güçlük, ayırt edicilik, tahminle doğru yanıtlanma olasılığı gibi bir takım özellikleri bulunmaktadır. Bu araştırmada madde özellikleri KTK'da madde güçlüğü ve madde ayırıcılık indeksleri; MTK'da a, b ve c parametreleri, STK'da ise madde uyarı indeksi ve maddeler için sınırlandırılmış gri ilişki dereceleri ile ifade edilmiştir.

**Birey Özellikleri:** Bu araştırmada birey özellikleri, testi alan bireyin tüm test üzerinde gösterdiği performansın nitel ya da nicel bir ifadesi olarak tanımlanmıştır. KTK'da test puanı, MTK'da yetenek parametresi, STK'da ise öğrenci uyarı indeksi ve öğrenciler için sınırlandırılmış gri ilişki dereceleri ile ifade edilmiştir.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde, araştırma kapsamında ele alınan Klasik Test Kuramı, Madde Tepki Kuramı ve Sato Test Kuramı ile ilgili kavramsal ve kuramsal çerçeve belirginleştirilmeye çalışılmıştır.

#### Klasik Test Kuramı

Psikolojik ölçme tarihinin başlangıcından itibaren test geliştirme, analiz ve psikolojik ölçeklerin puanlanmasında daha yaygın olarak kullanılan kuram Klasik Test Kuramı'dır. KTK, gözlenen ile gözlenemeyen özellikler arasındaki ilişkiyi doğrusal bir model aracılığıyla açıklar (Lord ve Novick, 1968). Kuramda, ölçülmek istenen değişkene ait gerçek değer "gerçek puan" olarak adlandırılır ve kurama göre gerçek puan sabit olsa da gözlem sonuçları (gözlenen puan) pratikte her zaman bir miktar hata içerir. Bu üç bileşen arasındaki eşitlik  $X = T + E$  şeklinde kurulur. Söz konusu eşitlikte "X" gözlenen puanı, "T" gerçek puanı ve "E" hata puanını temsil etmektedir (Crocker ve Algina, 1986; Hambleton ve Swaminathan, 1985; Thompson, 1994). Eşitlikte "gerçek puan" olarak ifade edilen değer varsayımsaldır. Bireyin ölçülen herhangi bir özelliği için gerçek puanının belirlenebilmesi sonsuz defa ölçme yapılması ile mümkün olabilecektir. Çünkü ancak aynı özelliğin sonsuz defa ölçülmesi durumunda hata puanları ortalamasının sıfır olacağı varsayılmaktadır. Pratikte bu durumun sağlanması mümkün olmadığı için ölçme yoluyla elde edilemeyen gerçek puan, gözlenen puandan kestirilmektedir (Coaley, 2009).

KTK'ya göre test puanı, bireyin maddelerden almış olduğu puanların toplanmasıyla elde edilir. Test ve madde istatistikleri de bu şekilde elde edilmiş olan toplam puanlar ve madde puanları üzerinden hesaplanmaktadır. Bu durum, bireylerin ölçülen özelliğe ilişkin seviyelerinin testin güçlük düzeyine bağlı olması gibi birtakım sorunları beraberinde getirirse de özellikle ileri düzey istatistik bilgisi olmayan test uygulayıcılarına kolaylık sağlamaktadır.

Klasik Test Kuramı'nda madde güçlük indeksi, madde ayırıcılık indeksi, madde varyansı ve madde güvenilirlik katsayısı en çok kullanılan madde istatistikleri olarak ifade edilebilir.

### **KTK'da Madde İstatistikleri**

**Madde güçlük indeksi.** Madde güçlük indeksi (p), 1 ve 0 olarak puanlanan bir maddenin madde puanlarının ortalamasıdır. Başka bir ifadeyle madde güçlük indeksi, bir maddeyi doğru yanıtlayanların testi alanların tümünün sayısına bölümüdür (Coaley, 2009; Crocker ve Algina, 1986). Bu anlamda madde güçlük ölçüsü yalnızca maksimum performansı ölçen test veya ölçekler (başarı, yetenek gibi) için söz konusudur. Doğru veya yanlış cevabın söz konusu olmadığı tipik davranış ölçeklerinde (kişilik, ilgi, tutum gibi) bu ölçü kullanılmaz (Erkuş, 2003). Madde güçlük indeksinin temel işlevi, sorunun zorluk ya da kolaylık derecesini göstermesidir. İndeksin (p) değeri ne denli büyükse, madde o denli kolay olarak yorumlanır. Fakat KTK'da madde güçlüğü değeri (p) yalnızca maddenin kendi zorluk/kolaylık düzeyiyle ilişkili olarak ortaya çıkan bir değer değil aynı zamanda testi yanıtlayan bireylerin yetenek düzeylerine bağlı olarak da değişen bir değerdir. Bu KTK'nın birey örnekleme bağımlı sonuçlar vermesi olarak da bilinen dezavantajlarından biridir (Fulcher ve Davidson, 2007).

**Madde varyansı.** Madde varyansı madde standart sapmasının karesine eşittir. Maddelerin 1 ve 0 şeklinde puanlandığı durumlarda ise madde varyansı, o maddenin güçlük indeksi (p) ile güçlük indeksinin tersinin (q) çarpımına eşittir (Lord ve Novick, 1968). Bir maddenin varyansı büyüdükçe o madde ile ölçülmek istenilen özellik bakımından bireyler arasındaki farklılıkların ortaya koyulma gücü artar. Varyansın, ölçülen nitelik yönünden bireyler arası farkları ortaya koyan bir istatistik olduğu dikkate alınır, madde güçlük indeksi  $\frac{1}{2}$  veya bu civarda olan maddelerin, ölçtüğü nitelik yönünden cevaplayıcılar arasındaki farkları en iyi ortaya koyan maddeler olduğu söylenebilir (Baykul, 2000). Fakat bu belirleme, tümü orta güçlükteki maddelerden oluşturulan bir testin her zaman cevaplayıcılar arasındaki farkı en iyi şekilde ortaya koyacağı anlamına gelmez. Özellikle uç yetenek düzeylerindeki bireylerin birbirinden ayırt edilebilmesi, kolay ve zor maddelerin de testte yer almasıyla mümkün olabilmektedir (Coaley, 2009). Bununla birlikte Ebel ve Frisbie (1991), birey puanları dağılımının ranjını artıracığı gerekçesiyle daha

ayırt edici ölçmeler yapılabilmesi bakımından başarı testlerinde ağırlıklı olarak orta güçlük düzeyindeki maddelere yer verilmesini önermişlerdir.

**Madde ayırıcılık indeksi.** Bir test ya da ölçekteki maddelerden, madde ile ölçülmesi amaçlanan özelliğe sahip olan ve olmayanları birbirinden ayırt edebilmesi beklenir. Maddelerin bu özelliği maddenin ayırt edicilik gücü olarak adlandırılır. Maddenin ayırt edicilik gücü aynı zamanda maddenin geçerliği olarak da bilinir. Maddelerin ayırt edicilik gücünü belirlemeye yönelik saptamalar çeşitli yöntemlerle yapılabilir ve bu amaçla elde edilen indekse madde ayırıcılık indeksi ( $r_{jx}$ ) adı verilir. Maddenin ayırt ediciliğini belirlemek için bir ölçüte ihtiyaç vardır. Bu amaçla kullanılan ölçüt, o maddenin yer aldığı testin ya da ölçeğin istatistikleri ise buna “iç ölçüt”; ölçülmek istenen özelliği ölçen, güvenilirliği ve geçerliği kanıtlanmış başka bir ölçme aracıyla elde edilen istatistikler ise buna “dış ölçüt” denir (Atılğan, Kan ve Doğan, 2009; Baykul, 2000; Erkuş, 2003). Madde ayırıcılık indeksi hesaplamada iç ölçüt kullanmak daha yaygın ele alınan bir yaklaşımdır. Bir ölçeğin iç ölçütü olarak genellikle, o ölçekten elde edilen toplam puanlar alınır. Toplam puanların, maddenin geçerliği için ölçüt alınmasının temelinde, denenen maddelerin, testle ölçülmek istenen özelliği geçerli olarak ölçtüğü sayılısı yatar (Turgut, 1978). İç ölçüt ile madde ayırt edicilik gücü hesaplamada, “madde toplam test korelasyon katsayıları”, “alt-üst %27’lik grup ortalamalarının bağımsız gruplar t-testiyle karşılaştırılması”, “regresyon analizi” gibi yöntemler kullanılmaktadır (Erkuş, 2003).

Madde ayırıcılık indeksi -1 ve 1 aralığında değişim gösteren bir değerdir. Negatif değerler maddenin düşük yetenekli bireyler tarafından büyük oranda doğru yanıtlanırken yetenekli bireyler tarafından doğru yanıtlanamaması durumlarında ortaya çıkar. Ayırıcılık indeksi negatif olan bir madde ters işleyen sorunlu bir maddedir. İndeksin pozitif olması ve 0.25’in üzerinde bir değer alması beklenir. Ayırt ediciliği düşük olan maddelerin yeniden yazılması ya da testten çıkarılması gerekebilir (Courville, 2004; Henning, 1987). Bununla birlikte, ayırıcılık indeksi örneklem büyüklüğünden ve örneklemde yer alan bireylerin yetenek ranjından çabuk etkilenen bir ölçüdür. Yetenek düzeyleri bakımından homojen ve küçük bir örneklem üzerinden hesaplanan madde ayırıcılık indeksleri yanıltıcı bilgi verebilir (Ebel ve Frisbie, 1991; Fulcher ve Davidson, 2007). Bu bakımdan, küçük gruplar

üzerinde yapılan ölçmelerde yaygın olarak kullanılan Klasik Test Kuramı'na dayalı analizlerde örnekleme hatasından kaynaklı sorunlarla karşılaşılacağı düşünülmemelidir. Küçük bir örneklemden elde edilen verilerle yapılan bir analizde ayırt ediciliği yüksek olarak belirlenen bir madde başka küçük bir örnekleme analiz edildiğinde ayırt ediciliği düşük hatta negatif olarak belirlenebilir (Ebel ve Frisbie, 1991).

**Madde güvenilirlik katsayısı.** Bireylerin yetenek düzeyleri arasındaki farka ilişkin bilgiyi en fazla açıklayan maddeler puanların güvenilirliğine en fazla katkı sunan maddelerdir (Ebel ve Frisbie, 1991). Bu anlamda madde güvenilirlik katsayısı, madde ayırt edicilik indeksi ve madde standart sapmasına bağlı olarak  $r_j = r_{jx} \cdot s_j$  eşitliği ile tanımlanır.

### **KTK'nın Avantajları**

Bu bölümde Klasik Test Kuramı'nın sağladığı düşünülen avantajlar maddeler halinde özetlenmiştir.

- Özellikle, okullarda uygulanan başarı testlerini geliştiren öğretmenler ve yaptıkları araştırmalar kapsamında çeşitli ölçekleri kullanarak bilgi toplayan fakat ölçme alanında üst düzey bilgi sahibi olmayan araştırmacılar için KTK, madde ve test istatistiklerinin hesaplanması bakımından daha tercih edilir bir kuram olarak geçerliğini korumaktadır. Bununla birlikte, karmaşık analiz süreçlerini basitleştiren bilgisayar yazılımları farklı yaklaşımların yaygınlaşmasına olanak tanımaktadır. Bu bakımdan KTK'nın hesaplama ve yorumlama kolaylıklarına ilişkin avantajların zaman içerisinde avantaj olmaktan çıkabileceği iddia edilebilir.

- Klasik Test Kuramı'nın örneklem büyüklüğü konusunda katı bir duruşu yoktur. Madde ayırıcılık indeksi gibi istatistiklerin daha doğru hesaplanabilmesi bakımından küçük ve homojen örneklemeler uygun görülmesine de kuramın matematiksel temelini basitliğinden dolayı çok büyük örneklemelere ihtiyaç duyulmaz.

- Özellikle Madde Tepki Kuramı'nın kullanılmasını olanaksız kılan (MTK varsayımlarını karşılamayan) veri yapılarında KTK başarılı sonuçlar vermektedir

(Zickar ve Broadfoot, 2008). Bunun yanı sıra alanyazında, her iki kuramın özellikle büyük örneklemeler üzerinde karşılaştırdığı durumlarda hem madde hem de birey özellikleri bakımından benzer (tutarlı) sonuçlar verdiği kanıt sunan araştırmalar bulunmaktadır (Fan, 1998; Stage, 2003). Bu doğrultuda düşünüldüğünde, MTK'nın pratik engelleri azaltılana kadar klasik yaklaşımların terk edilmesinin erken olacağı söylenebilir (DeVellis, 2014).

### **KTK'nın Dezavantajları**

Bu bölümde Klasik Test Kuramı için dezavantaj olarak kabul edilebilecek durumlar maddeler halinde özetlenmiştir.

- KTK'ya dayalı geliştirilen testler/ölçeklerde madde istatistikleri testin geliştirdiği gruptan kestirilmekte ve grup değiştiğinde elde edilen istatistik değerleri de değişebilmektedir.

- Madde istatistikleri birey (katılımcı) örneğine bağlı olduğu gibi test puanları da madde örneğine bağlıdır. Eğer test göreceli olarak kolay maddelerden oluşuyorsa bireylerin testten aldığı puanlar da yüksek olacak; test göreceli olarak zor maddelerden oluşuyorsa bireylerin testten aldığı puanlar da düşük olacaktır. Bu bağlamda, KTK'da bireylerin başarı/yetenek düzeylerinin de testte yer alan maddelere bağlı olarak göreceli olduğu söylenebilir (McDonald ve Paunonen 2002). Ayrıca birey ve test istatistiklerinin birbirine bağlı olması Bilgisayar Ortamında Bireyselleştirilmiş Test (BOB test) gibi uygulamalarda teorik zorluklar doğurmaktadır (Fan, 1998).

- Klasik Test Kuramı modellemesinde tüm hatalar tesadüfi hata olarak kabul edilmektedir. Buna karşın kültürel altyapı, etnisite, cinsiyet gibi kaynaklardan ileri gelen sistematik hatalar ölçmeye karışabilir. Test yanlılığına neden olan bu tür hatalar KTK'da genellikle görmezden gelinir (Kiany ve Jalali, 2009).

- KTK'da ölçümlerin güvenilirliği testi yanıtlayan bireylerin özelliklerine, testin özelliklerine ve test uygulayıcısının kararlarına ve yönetimine bağlı olarak değişebilir (Traub ve Rowley, 1991).

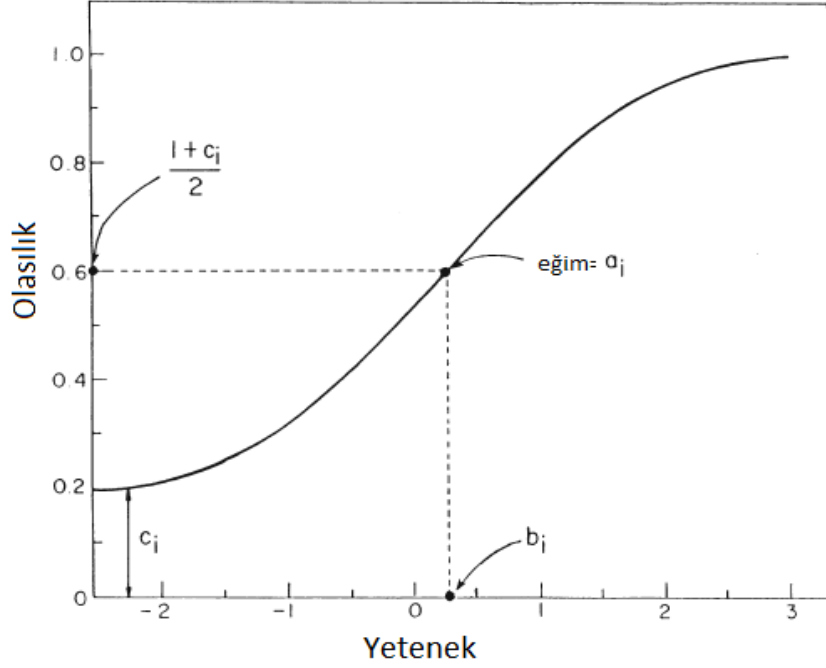
- KTK'da ölçme hataları ve güvenilirlik katsayısı tüm grup için hesaplanır. Farklı yetenek düzeyindeki bireyler bakımından ölçme aracı aynı düzeyde güvenilirliğe sahip olamayacağı için (Nartgün, 2002), bu durum bir dezavantaj olarak öne sürülebilir.

- KTK'ya dayalı uygulamalarda teste katılan bireylere verilen en önemli bilgi, onun tüm test üzerinde gösterdiği performansını yansıtan toplam test puanıdır. Bireyin farklı maddeler üzerinde gösterdiği performans KTK'da detaylı olarak ele alınmaz (Kiany ve Jalali, 2009).

### **Madde Tepki Kuramı**

Klasik Test Kuramının (KTK) sınırlılıklarını ortadan kaldırılabileceği iddia edilen bir model olarak Madde Tepki Kuramı (MTK) 1930'lu yılların sonunda ortaya atılmıştır. Tucker 1940'da kuramın en önemli kavramlarından "madde karakteristik eğrisi" kavramını ilk defa kullanmıştır. Lord'un 1953'te yayımlanan makalesi ve Birnbaum'un 1957 ve 1958 yılları içerisinde Amerika Birleşik Devletleri Hava Kuvvetleri için yazdığı üç teknik rapor MTK'ya ilişkin çalışmaların başlangıcı olarak kabul edilir. Lord ve Novick'in (1968) "Statistical Theories of Mental Test Scores" kitabında Birnbaum tarafından MTK üzerine kaleme alınan dört bölüm kuramın gelişmesinde büyük rol oynamıştır. 1950-1970 yılları arasında, matematiksel işlemlerin karmaşıklığı nedeniyle kuram yavaş gelişmiş, fakat bilgisayar teknolojisindeki gelişmelere paralel olarak kuramın kullanımı ve gelişimi de hızlanmıştır (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985).

MTK modelleri bir ölçme aracı tarafından ölçülen yetenek ya da özellik ( $\Theta$ ) ile bir maddeye verilen cevap arasındaki ilişkiyi gösterir. Söz konusu ilişki madde karakteristik eğrisi (MKE) adı verilen fonksiyonla ifade edilir (Şekil 1). MKE kabaca "S" şeklindedir ve eğrinin farklı kısımları madde parametreleriyle ilgili bilgiler verir. Madde karakteristik eğrisinde X eksenini ölçülen özelliğin gücünü temsil eder ve Y eksenini her bir yetenek düzeyi için maddenin doğru yanıtlanma olasılığını gösterir. Maddeye verilen cevaplar iki kategorili ya da çok kategorili olabilir. Maddelerle ölçülen yapı akademik yeterlik ya da yetenek olabileceği gibi bir tutum ya da inanış da olabilir (DeMars, 2016; DeVellis, 2014).



Şekil 1. İkili puanlanan bir maddeye ait madde karakteristik eğrisi (üç parametrelî model).

Madde karakteristik eğrisinin eğimi  $a_i$ (ayırıcılık) parametresi ile ilgili, y eksenini kestiği nokta  $c_i$ (tahmin) parametresi ile ilgili ve  $(1+c_i)/2$  olasılık değerine karşılık gelen yetenek düzeyi  $b_i$ (güçlük) parametresi ile ilgili bilgi vermektedir.

### MTK'da Madde Parametreleri

**$b_i$  parametresi.** Madde güçlük parametresi olarak tanımlanmaktadır. Maddenin 0.50 olasılıkla ( $c_i = 0$  olduğu durumda) doğru yanıtlanacağı yetenek düzeyine karşılık gelen değerle ifade edilir. Cevaplayıcı grubunun yetenek puanları, ortalaması sıfır ve standart sapması bir olacak şekilde ölçeklendiğinde,  $b_i$  parametresi genellikle -2 ve +2 aralığında değer alır. Parametrenin değeri arttıkça maddeyi 0.50 olasılıkla doğru yanıtlamak için gerekli olan yetenek düzeyi de artar. Bu bakımdan  $b$  parametresinin değeri +2'ye yakın olan maddeler çok zor maddeler; -2'ye yakın olan maddeler ise çok kolay maddeler olarak ifade edilebilir (DeMars, 2016; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991).

**$a_i$  parametresi.** Madde ayırıcılık parametresi olarak ifade edilmektedir. Madde karakteristik eğrisinin eğimine karşılık gelmektedir (Thorndike, 1982). Maddenin, ölçülen özelliğe sahip olan ve olmayan bireylerin ayırt etme gücünü



ortaya koymasý ve bu anlamda bireyin ölçülen yeteneđinin gerçeđe yakın kestirilmesine ne ölçüde katkı sağladığına dair bilgi vermesi bakımından, maddenin geçerliđi ile ilişkilidir. Kuramsal olarak  $-\infty$  ve  $+\infty$  aralıđında deđerleneceđi belirtilse de genellikle 0 ve 2 aralıđında (ya da 0 ile 3) deđer alır (DeMars, 2016; Hambleton ve Swaminathan, 1985).  $a$  parametresi nadiren negatif deđerler de alabilir. Negatif ayırt edicilik tıpký KTK'daki gibi yüksek yetenekli bireylerin, düşük yeteneklilere göre daha az olasılıkla maddeye dođru cevap vermeleri anlamındadır ve bu nedenle  $a$  parametresi negatif deđerli olan maddeler kesinlikle testten çıkarılmalıdır (DeMars, 2016).

Baker (2001), maddeye ait  $a$  parametresi deđerinin 0.01 – 0.34 aralıđında çok düşük, 0.35 – 0.64 aralıđında düşük, 0.65 – 1.34 aralıđında orta, 1.35 – 1.69 aralıđında yüksek ve 1.70'ten büyük olması durumunda çok yüksek ayırt edici olarak yorumlanabileceđini belirtmiřtir.

**$c_i$  parametresi.** Madde karakteristik eđrisinin  $y$  eksenini kestiđi noktadır. Düşük yetenekli bireylerin maddeyi dođru yanıtlanma olasılıklarını ifade etmektedir (Hambleton ve Swaminathan, 1985). Düşük yetenek düzeyine sahip bir bireyin güçlük düzeyi yeteneđinin üzerinde olan bir maddeyi ancak tahmin ederek dođru yanıtlayabileceđi düşünöldüđu için  $c$  parametresi “tahmin” parametresi olarak da adlandırılır. Lord (1974), iyi hazırlanmış standart testlerde  $c$  parametresinin řans ile dođru cevaplandırmaya göre biraz daha düşük deđer aldığını, çünkü iyi hazırlanmış çeldiricilerin düşük yetenekli bireyleri dođru cevaptan uzaklařtırdığını belirtmiřtir. Ancak, maddede düşük yetenekli bireylerin eleyebileceđi zayıf çeldiriciler varsa  $c$  parametresi řanstan daha yüksek bir deđer alabilir (aktaran, DeMars, 2016). Bir maddenin tahminle dođru yanıtlanma olasılıđının yüksek olmasının madde geçerliliđine olumsuz etki edeceđi söylenebilir. Bu bakımdan,  $c$  parametresinin yüksek bir deđer alması istenen bir durum deđildir.

MTK'da sıklıkla kullanılan madde parametrelerinin ardından ikili puanlanan maddeler için en çok tercih edilen 1, 2 ve 3 Parametrelili Lojistik Modeller ařađıda özetlenmiřtir.

## İkili Puanlanan Maddeler için MTK Modelleri

**1 parametrelili lojistik model.** Bu modelde, yanıtlayıcının bir maddeye (i) doğru yanıt verme olasılığı o maddenin güçlük düzeyine ( $b_i$ ) ve yanıtlayıcının yetenek düzeyine ( $\Theta$ ) bağlıdır. 1PLM'de maddeden maddeye değişkenlik gösteren tek parametre güçlük parametresidir. Ayırıcılık parametresi ( $a_i$ ) tüm maddeler için sabit bir değer olarak kabul edilmekte ve maddelerin tahminle doğru yanıtlanma olasılıklarının ( $c_i$  parametresi) sıfır olduğu varsayılmaktadır. Model aşağıdaki eşitlikle ifade edilebilir.

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \quad i=1,2,\dots,n$$

Verilen eşitlikte  $p_i(\Theta)$ , belirli bir yetenek düzeyine sahip bireylerden seçkisiz olarak belirlenmiş bir bireyin "i" maddesini doğru cevaplama olasılığını;  $b_i$ , madde güçlük parametresini; n ise testteki madde sayısını ifade etmektedir.

**2 parametrelili lojistik model.** Bu modelde yanıtlayıcının bir maddeye doğru yanıt verme olasılığının belirlenebilmesi için madde güçlük parametresinin ( $b_i$ ) yanı sıra ayrı bir değişken olarak madde ayırıcılık parametresi ( $a_i$ ) de analize dâhil edilmektedir. Maddelerin tahminle doğru yanıtlanma olasılıklarının ( $c_i$  parametresi) ise sıfır olduğu varsayılmaktadır. Model aşağıdaki eşitlikle ifade edilebilir.

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad i=1,2,3,\dots,n$$

1PLM'den farklı olarak eşitliğe  $a_i$  parametresi ve D ölçekleme faktörü dâhil edilmiştir. Söz konusu ölçekleme faktörü sabit bir değer olarak (1.7) kabul edilmektedir.

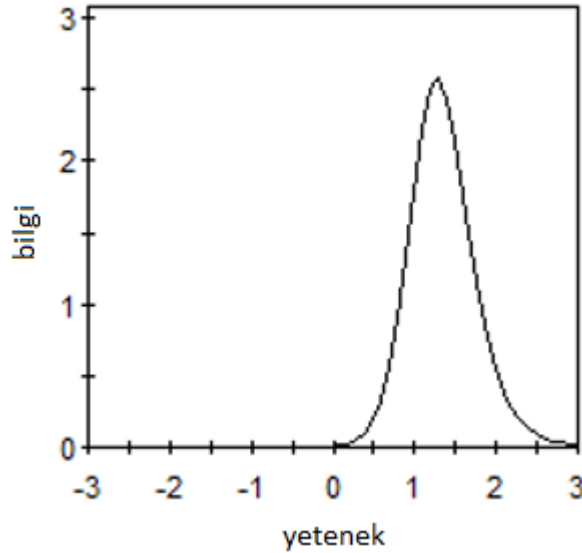
**3 parametrelili lojistik model.** Bu modelde, yanıtlayıcının bir maddeye (i) doğru yanıt verme olasılığı incelenirken  $b_i$  ve  $a_i$  parametrelerinin yanı sıra o maddenin düşük yetenekli bireyler tarafından doğru yanıtlanma olasılığını ifade eden  $c_i$  parametresi de maddeler arası bir değişken olarak dikkate alınmaktadır. Model aşağıdaki eşitlikle ifade edilebilir.

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad i=1,2,3,\dots,n$$

Eşitlik incelendiğinde, 2PLM'den farklı olarak  $c_i$  parametresinin de hesaplamada yer aldığı görülmektedir.

### Madde Bilgi Fonksiyonu

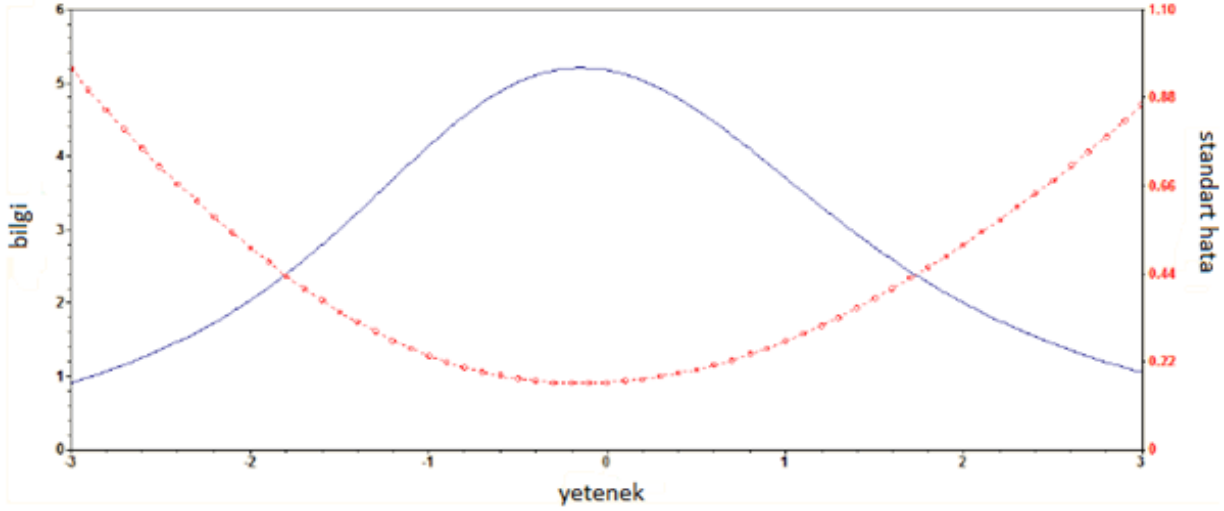
Madde Tepki Kuramı'nda madde bilgisi, bireylerin, testte yer alan herhangi bir madde üzerindeki performanslarının, yetenek düzeylerinin belirlenmesinde ne oranda katkıda bulunduğuyla ilişkili bir kavramdır. Yanıtlanan maddelerin taşıdıkları bilgi ne derece fazlaysa bireyin ölçülen özelliğe ilişkin yetenek düzeyi o derece doğrulukla kestirilir ve yeteneğin gerçek değerine yakın bir değer elde edilmiş olur. Madde bilgi fonksiyonu ise maddenin, her bir yetenek düzeyi için açıkladığı bilgiyi gösterir (Baker, 2001). Ayırt ediciliği yüksek olan maddeler belirli yetenek düzeyleri için yüksek oranda bilgi sağlarlar. Düşük ayırt ediciliğe sahip bir maddeler ise doğru yanıtlanma olasılıklarının yetenek düzeyleri boyunca yavaş yavaş değişmesinden dolayı geniş bir ranj boyunca küçük miktarlarda bilgi sağlarlar. Maddelerin, tahmin parametrelerinin (c parametresi) değeri yükseldikçe açıkladıkları bilgi azalır. Tahmin parametresinin değeri yüksek olan maddelerin özellikle düşük yetenekli bireyler için sağladıkları bilgi çok azdır (DeMars, 2016).



Şekil 2. Madde bilgi fonksiyonu grafiği.

## Test Bilgi Fonksiyonu

Test bilgi fonksiyonu madde bilgi fonksiyonlarının toplamına eşittir. MTK'da ölçmenin standart hatasını ve güvenilirliğini hesaplamak için kullanılır. Bilgi ne kadar yüksek olursa standart hata o kadar düşük ve güvenilirlik de o kadar yüksek olur (DeMars, 2016).



Şekil 3. Test bilgi ve standart hata fonksiyonları grafiği.

Test bilgi fonksiyonu test geliştiriciler için son derece kullanışlı bir madde tepki kuramı çıktısıdır. Temel olarak bu fonksiyon, testin hangi yetenek düzeylerindeki bireyler için yapılan kestirimlerin daha isabetli olduğunun anlaşılmasını sağlar. Kullanılan test, bilginin yüksek olduğu yetenek düzeylerine sahip bireyler için daha geçerli ölçümler elde edilmesine olanak tanımaktadır. Bu anlamda test bilgi fonksiyonu, araştırmacının amacına uygun bir testin geliştirilebilmesi için kolaylık sağlayan bir araç olarak da nitelendirilebilir (Baker, 2001).

## MTK'nın Avantajları

MTK'nın tek boyutluluk, yerel bağımsızlık ve maddelere verilen yanıtlarla bireylerin yetenekleri arasındaki ilişkilerin madde karakteristik eğrisi fonksiyonlarıyla ifade edilebileceği gibi varsayımları bulunmaktadır. Bu varsayımlar ve model-veri uyumu sağlandığı takdirde MTK'yı KTK'dan üstün kılan bazı

avantajların olduğu belirtilmektedir (Dođan ve Tezbařaran, 2003; Fan, 1998; Hambleton ve Swaminathan, 1985).

- Sz konusu avantajların byk ođunluđu MTK'da madde parametrelerinin testin uygulandıđı rneklemden bađımsız yetenek parametresinin de bireyin aldıđı maddelerden bađımsız olarak elde edilebilmesinin sonucu olarak ortaya ıkar. Pratikte bu durum Klasik Test Kuramı iin pek ok sorunu beraberinde getiren test eřitleme ve bireye uyarlanmıř test geliřtirme gibi alanlarda kolaylık sađlar. Ayrıca madde zelliklerinin rneklemden bađımsız oluřu bir testin MTK'ya gre bir kez leklendikten sonra farklı rneklemlerde pek ok kez kullanılabilmesine ve elde edilen sonuların karřılařtırılabilmesine olanak sađlar (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991).

- Eđer kullanılan MTK modeli test verisi iin uygunsa, bireylerin yetenekleri KTK'ya gre daha dođru (sađlam) kestirilir (Birnbaum, 1968; Ferrando ve Cihico, 2007). Bunun en nemli nedeni, KTK'da lme hatalarının tm grup iin hesaplanması, MTK'da ise her birey iin ayrı ayrı kestirilebilmesidir. Yine gvenilirlik katsayısı KTK'da cevaplayıcı grubun puan dađılımı iin tek bir deđer olarak hesaplanırken, MTK'da gvenilirlik her bir madde ve yetenek dzeyi iin madde ve test bilgi fonksiyonları řeklinde hesaplanabilmektedir (Nartgn, 2002).

- MTK'ya dayalı test analizleri maddeler, bireyler ve test hakkında Klasik Test Kuramı ve Sato Test Kuramı analizlerine gre daha fazla bilgi sunar.

- MTK'ya gre hesaplanan yetenek puanlarına her bir maddenin katkısı aynı deđildir; kullanılan modele gre maddenin glğnden, ayırt edicilik gcnden ve tahminle dođru yanıtlanma olasılıđından etkilenen yetenek puanları sz konusudur (elen ve Aybek, 2013).

- Kuramın en nemli avantajlarından bir diđer de, madde glk deđerlerinin ve yetenek dzeylerinin aynı lek dzeyinde yer almasıdır (Spencer, 2004). Bu avantaj, testi almıř ve yetenek dzeyi kestirilmıř olan herhangi bir

bireyin havuzda yer alan herhangi bir maddeyi doğru yanıtlayabilmesinin olasılığının belirlenebilmesine olanak sağlar.

- Klasik Test Kuramı'na göre test geliştirirken deneme uygulaması sonrasında çoğunlukla madde varyansı en yüksek olan orta güçlükteki maddeler nihai teste alınır. Madde Tepki Kuramı'nda ise madde parametrelerinin yanı sıra asıl ölçüt modele uyumluluktur. Bunun yanı sıra her bir maddenin teste olan katkısını her yetenek düzeyi için gösteren madde bilgi fonksiyonu madde seçiminde kriter olarak alınabilmektedir (van der Linden ve Hambleton, 1997). MTK'ya göre geliştirilen bir testin madde havuzunda her yetenek düzeyine uygun madde bulunması ve özellikle bilgisayar ortamında bireye uyarlanmış test uygulamalarında bireylerin karşılaştığı madde sayılarının farklılık gösterebilmesi daha uç yetenek düzeylerinin de açıklanabilmesine olanak sağlar.

- Klasik Test Kuramı'nın "daha uzun testler daha güvenilirdir" kabulünün aksine MTK'ya dayalı olarak az sayıda maddeden oluşan, güvenilirliği yüksek testler geliştirilebilir (Embretson ve Reise, 2000).

### **MTK'nın Dezavantajları**

Bu bölümde Madde Tepki Kuramı için dezavantaj olarak kabul edilebilecek durumlar maddeler halinde özetlenmiştir.

- Kullanılmak istenen MTK modeli ile toplanan veri uyum göstermeyebilir (Stage,2003).

- MTK parametrelerini doğru bir şekilde kestirebilmek için büyük örneklemeler gerekmektedir (de Ayala, 2009; DeMars, 2016; Hulin, Lissak ve Drasgow, 1982; Ree ve Jensen, 1983; van der Linden ve Hambleton, 1997). Ayrıca, MTK'ya dayalı bir test geliştirmek, uygulamalardan elde edilen verileri analiz etmek ve analiz sonuçlarını yorumlamak alan uzmanlığı gerektirmektedir. Söz konusu nedenlerden kaynaklı olarak, kuramın, sınıf içi ölçme ve değerlendirme uygulamalarında kullanılmasının önünde çeşitli engellerin olduğu görüşü ileri sürülebilir.

- Özellikle BOB test uygulamalarında farklı maddelerle ve farklı test uzunluklarıyla karşılaşan bireylerin aynı ölçek üzerinde değerlendirildiklerini, testi alan bireylere ve onların yakınlarına anlatma noktasında güçlük yaşanabilir.

### **Sato Test Kuramı**

Takahiro Sato tarafından 1969'da ileri sürülen kuram, öğrencilerin cevap örüntüleri üzerinden öğrenme durumlarını belirlemeye yönelik olarak normal dağılım varsayımını sağlayamayan küçük örneklerde bile etkili sonuçlar üretmesi beklenen analiz yöntemleri üzerine inşa edilmiştir. STK, ilk başlarda yalnızca Öğrenci Madde Çizelgesi Analizi (Student-Problem Chart Analysis) ile öğrenme durumları bakımından öğrencileri ve işlevlikleri bakımından maddeleri sınıflandırma amaçlarına yönelik olarak kullanılmış olsa da 2010 yılında Nagai'nın önerisiyle Gri İlişki Analizi ile güçlendirilmiş ve öğrencilerin başarı/yetenek düzeylerinin belirlenmesi amacıyla da kullanılabilir hale getirilmiştir. Gri İlişki Analizi Ju-Long Deng tarafından 1982 yılında ortaya konulan Gri Sistem Teorisi içerisinde yer alan analiz yöntemidir. Gri İlişki Analizi, bir derecelendirme, sınıflandırma ve karar verme tekniğidir. Gri Sistem Teorisi'nde renklerin koyuluğu bilginin belirginliğinin derecesini ifade etmektedir. Bu anlamda siyah bilinmeyen bilgileri, beyaz bilinen bilgileri, gri ise kısmen bilinen ve kısmen bilinmeyen bilgileri temsil etmektedir. Gri İlişki Analizi, normal dağılım varsayımını sağlayamayan, belirsizlik nedeniyle modellenemeyen ve yeterli veri içermeyen kısıtlı verilerle karar verme durumlarında istatistiksel çözümler önermektedir.

Günümüzde STK, Gri Öğrenci-Madde Çizelgesi Analizi (Grey S-P Chart Analysis), Kısmi Kredili Öğrenci-Madde Çizelgesi Analizi (Partial Credit S-P Chart Analysis) ve Rasch Gri Öğrenci-Madde Çizelgesi Analizi (Rasch Grey S-P Chart Analysis) modelleri ile hem ikili puanlanan hem de çoklu puanlanan madde türlerine uygulanabilen, test geliştirme ve uygulama alanında mevcut sorunlara çözüm getirebilecek yöntemler sunma çabasıdadır (Pham, Sheu ve Nagai, 2015; Sheu vd., 2014c).

Öğrenci-Madde Çizelgesi Analizi ile öğrenci uyarı indeksi (Student Caution Index) ve madde uyarı indeksi (Problem Caution Index) olmak üzere iki tür indeks üretilmektedir.

## Madde Uyarı İndeksi

Madde uyarı indeksi (MUİ) aracılığıyla maddeler sınıflandırılarak her bir maddenin işlerliği hakkında dönüt elde edilmektedir. MUİ indeksi aşağıdaki formülle hesaplanmaktadır (Sheu vd., 2014a; Lin ve Yih, 2015).

$$MUİ = 1 - \frac{\sum_{n=1}^N (Y_{nm})(Y_n) - (Y_m) \cdot (\mu')}{\sum_{n=1}^N Y_n - (Y_m) \cdot (\mu')}$$

*m*: Maddeler ( $m=1,2,3,\dots,M$ )

*n*: Öğrenciler ( $n=1,2,3,\dots,N$ )

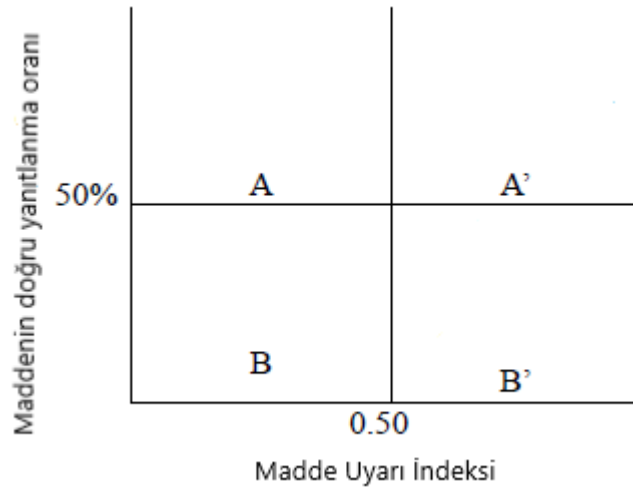
$Y_{nm}$ : *n*. öğrencinin *m*. maddeden aldığı puan (doğru:1, yanlış: 0)

$Y_m$ : Öğrencilerin *m*. maddeden aldıkları puanların toplamı

$Y_n$ : *n*. öğrencinin toplam test puanı

$$\mu' : \frac{\sum_{n=1}^N Y_n}{N}$$

Madde uyarı indeksi (MUİ), maddeleri dört farklı sınıfa atamak için kullanılmaktadır. İndeks 0 ile 1 aralığında değer almakta ve her bir madde için hesaplanan MUİ değerinin 0,50'den daha büyük bir değer alması, ilgili maddede bazı sorunların olabileceğine ilişkin bilgi vermektedir. MUİ değeri maddenin doğru yanıtlanma oranı ile birlikte yorumlanır. Bu iki ölçünün alacağı değerlere göre maddenin hangi sınıfa atanacağına ilişkin görsel Şekil 4'te sunulmuştur.



Şekil 4. MUİ ve oran değerlerine göre maddelerin atanacağı sınıflar.



Şekil 4'te görüldüğü üzere MUI değerinin 0,50'den düşük bir değer aldığı durumda maddenin doğru yanıtlanma oranı %50'den yüksek ise A; düşük ise B sınıfına atandığı ve MUI değerinin 0,50'den büyük bir değer aldığı durumda maddenin doğru yanıtlanma oranı %50'den yüksekse A'; düşük ise B' sınıfına atandığı belirtilebilir. Yapılan bu sınıflandırmalara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

*A: Düzgün çalışan bir madde.*

*A': Kısmi olarak sorunlu bir madde. Tekrar gözden geçirilebilir.*

*B: Zor bir madde.*

*B': Sorunlu bir madde. Ters işliyor olabilir. Testten atılmalı ya da yeniden yazılmalı.*

### Öğrenci Uyarı İndeksi

Öğrenci uyarı indeksi (ÖÜİ) aracılığıyla öğrenciler sınıflandırılarak her bir öğrencinin testle ölçülen konu alanlarını öğrenme ve maddeleri dikkatli bir şekilde yanıtlama durumlarına ilişkin dönüt alınmaktadır. ÖÜİ indeksi aşağıdaki formülle hesaplanmaktadır (Lin ve Yih, 2015; Sheu vd., 2014a).

$$\text{ÖÜİ} = 1 - \frac{\sum_{m=1}^M (Y_{nm})(Y_m) - (Y_n) \cdot (u')}{\sum_{m=1}^M Y_m - (Y_n) \cdot (u')}$$

*m: Maddeler (m=1,2,3,...M)*

*n: Öğrenciler (n=1,2,3,...N)*

*Y<sub>nm</sub>: n. öğrencinin m. maddeden aldığı puan (doğru:1, yanlış: 0)*

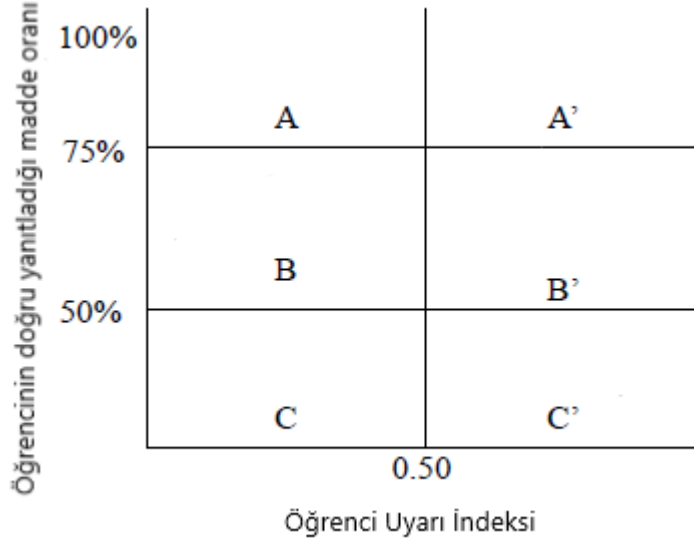
*Y<sub>m</sub>: Öğrencilerin m. maddeden aldıkları puanların toplamı*

*Y<sub>n</sub>: n. öğrencinin toplam test puanı*

*u':  $\frac{\sum_{m=1}^M Y_m}{M}$*

Öğrenci uyarı indeksi (ÖÜİ), öğrencileri altı farklı sınıfa atamak için kullanılmaktadır. İndeks 0 ile 1 aralığında değer almakta ve her bir öğrenci için hesaplanan ÖÜİ değerinin 0,50'den daha büyük bir değer alması, ilgili öğrencinin öğrenme eksikliklerinin olabileceğine ya da dikkatsizlikle kaçırıldığı soruların olabileceğine ilişkin bilgi vermektedir. ÖÜİ değeri öğrencinin doğru yanıt verdiği

madde sayısının testte yer alan madde sayısına oranı ile birlikte yorumlanır. Bu iki ölçünün alacağı değerlere göre öğrencinin(bireyin) hangi sınıfa atanacağına ilişkin görsel Şekil 5'te sunulmuştur.



Şekil 5. ÖUİ ve oran değerlerine göre öğrencilerin atanacağı sınıflar.

Şekil 5'te yer alan değerler doğrultusunda yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

*A: Etkili öğrenme gerçekleşmiş.*

*A': Öğrenme var fakat çok dikkatsiz.*

*B: Genel olarak iyi ancak biraz daha fazla çalışmaya ihtiyacı var.*

*B': Biraz dikkatsiz ve daha fazla çalışmaya ihtiyacı var.*

*C: Öğrenme düzeyi yetersiz.*

*C': Öğrenme gerçekleşmemiş.*

Söz konusu ÖUİ ve MUİ indekslerinin öğrencilerin yetenek düzeyleri ve madde özellikleri hakkında yeterli bilgi vermediği iddia edilebilir. Bunun nedeni, nitel veri düzeyindeki çıktıların, bireyleri ve maddeleri karşılaştırabilme, yapılan değerlendirme sonucu çeşitli kararlar alabilme noktasında sınırlı kalmasıdır.

Öğrenci-Madde Çizelgesi'nin bu sınırlılığını ortadan kaldırmak amacıyla geliştirilen Gri Öğrenci-Madde Çizelgesi Analizi'ne dayalı olarak öğrencilerin test performanslarını belirlemek amacıyla öğrenciler için sınırlandırılmış gri ilişki derecesi –ÖİSGİD- (Localized Grey Relational Grade-Student) ve madde güçlüklerini belirlemek amacıyla maddeler için sınırlandırılmış gri ilişki derecesi - MİSGİD- (Localized Grey Relational Grade-Problem) elde edilmektedir (Sheu vd., 2014a).

### Öğrenciler için Sınırlandırılmış Gri İlişki Derecesi

ÖİSGİD değerleri öğrencilerin yetenek ya da başarı düzeylerine ilişkin bilgi vermektedir. ÖİSGİD hesaplanırken kullanılan vektörler ve formüller aşağıda verilmiştir (Sheu vd., 2014b).

Gri ilişki analizinde karşılaştırılan vektörler:

$$x_0 = (x_0(1), x_0(2), \dots, x_0(k), \dots, x_0(n)) : \text{Referans vektör}$$

$$x_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n)) : \text{İncelenen vektör}$$

Gri öğrenci-madde çizelgesi analizinde ÖİSGİD değeri elde etmek için incelenen vektör:

$S_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n))$  : i. öğrencinin tüm maddelere verdiği yanıtlar vektörü,  $k = 1, 2, 3, \dots, n$

$$\text{ÖİSGİD} = \Gamma_{(x_0, x_i)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{0i}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$\bar{\Delta}_{0i} = (\sum_{k=1}^n (x_0(k) - x_i(k))^p)^{\frac{1}{p}}$  : Referans vektör ve incelenen vektör (öğrenciler için) arasındaki Minkowski uzaklık ölçüsü.

$\bar{\Delta}_{max} = \bar{\Delta}_{0i}$  için maksimum değer.

$\bar{\Delta}_{min} = \bar{\Delta}_{0i}$  için minimum değer.

ÖİSGİD değerlerine dayalı olarak öğrencilerin/bireylerin ölçülen özellikteki performansları aşağıdaki gibi değerlendirilebilir.

Tablo 1

**ÖİSGİD Değerlendirme Ölçütleri**

Bireyin Performansı	ÖİSGİD değeri*
Yüksek	ÖİSGİD $\geq$ 0,75
Ortanın üstü	0,75 > ÖİSGİD $\geq$ 0,50
Orta	0,50 > ÖİSGİD $\geq$ 0,25
Düşük	0,25 > ÖİSGİD

\*ÖİSGİD 0 ile 1 aralığında değer almaktadır.

**Maddeler için Sınırlandırılmış Gri İlişki Derecesi**

Maddeler için sınırlandırılmış gri ilişki derecesi (MİSGİD) maddelerin güçlük düzeylerine ilişkin bilgi vermektedir. MİSGİD 0 ile 1 aralığında değer almakta ve kolay maddeler için yüksek, zor maddeler için düşük değerler elde edilmektedir. Başka bir ifadeyle, MİSGİD değeri ne denli yüksekse madde o denli kolaydır, denilebilir. MİSGİD hesaplanırken kullanılan vektörler ve formüller aşağıda verilmiştir (Sheu vd., 2014b).

Gri öğrenci-madde çizelgesi analizinde MİSGİD değeri elde etmek için incelenen vektör:

$P_k = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_m(k))$  :k. maddeye tüm öğrencilerin verdiği yanıtlar vektörü,  $i = 1, 2, 3, \dots, m$

$$MİSGİD = \Gamma_{(x_j, x_0)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{k0}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$\bar{\Delta}_{k0} = (\sum_{i=1}^m (x_0(k) - x_i(k))^p)^{\frac{1}{p}}$  :Referans vektör ve incelenen vektör (maddeler için) arasındaki Minkowski uzaklık ölçüsü.

$\bar{\Delta}_{max} = \bar{\Delta}_{k0}$  için maksimum değer.

$\bar{\Delta}_{min} = \bar{\Delta}_{k0}$  için minimum değer.

Gri İlişki Analizi'nin dâhil edilmesiyle birlikte Sato Test Kuramı öğrenciler ve öğretim sürecinin etkililiği ile ilgili nitel dönütler sağlamanın yanı sıra düzey

belirlemeye ve buna baęlı olarak bireyler hakkında karar almaya olanak saęlayan, karřılařtırılabilir ve sıralanabilir nicel veriler de saęlamaktadır.

### **STK'nın Avantajları**

Bu bölümde Sato Test Kuramı'nın saęladığı düşünölen avantajlar maddeler halinde özetlenmiştir.

- Kurama dayalı olarak hesaplanan ÖÜİ ve MUİ indeksleri ile kavram yanılgılarına sahip olan öğrencilerin ve kavram yanılgısı içeren maddelerin tespit edilebildięi ileri sürölmektedir (Tsai vd., 2014).

- STK'dan, düzey belirlemeye ve buna dayalı olarak öğrencilerin eğitim durumları ile karar almaya yönelik süreçlerde yararlanılabildięi gibi, öğrencilerin öğrenme durumları ve öğretim sürecinin etkililięi ile ilgili dönüt alabilmek ve alınan dönütlere dayalı olarak öğretimi geliřtirmek amaçlarıyla da yararlanılabilmektedir (Switzer ve Connell; 1990; Yu, 2011).

- Öğrencilerle ilgili verilen nitel dönütler ve öğrencilerin öğrenme durumlarına göre sınıflandırılması, öğretmenlerin dersi bireyselleřtirilmiş öğretime uygun olarak düzenlemesine yardımcı olabilir. Maddelere iliřkin verilen nitel dönütler ise öğretmenleri, kaliteli maddelerden oluřan bir madde havuzu oluřturma konusunda cesaretlendirebilir (Çüm, Demir ve řahin, 2018).

- STK, hem ikili puanlanan hem de çoklu puanlanan maddeler için uygun modeller sunmaktadır. Bu bakımdan geniř bir kullanım alanı vardır. İçerdięi analizler basit ve kullanıřlıdır. Analiz çıktılarının yorumu uzmanlık gerektirmez (McArthur, 1983; Pham, Sheu ve Nagai, 2015).

- Öğrenci-Problem Çizelgesi Analizi'nin yapılabilmesi için veriye (dataya) iliřkin herhangi bir varsayımın saęlanmasına ihtiyaç duyulmamaktadır. Bu bakımdan küçük gruplar üzerinde yürütölen ölçme ve deęerlendirme uygulamaları için uygun ve kullanıřlıdır (Pham, Sheu ve Nagai, 2015).

- Çüm, Gelbal ve Tsai (2016), farklı örneklemlerden (30'ar kişilik) alınan aynı maddelere ait yanıt örüntüleri üzerinden elde ettikleri MİSGİD indekslerinin tutarlı ve MUİ indeksleri aracılığıyla yapılan madde sınıflandırmalarının uyumlu olduğu sonucuna ulaşmışlardır. Bu sonuç, KTK'nın, madde istatistiklerinin testin uygulandığı gruba bağlı olmasıyla ilgili olan dezavantajına karşı STK'nın sağladığı bir avantaj olarak öne sürülebilir.

### **STK'nın Dezavantajları**

Bu bölümde Sato Test Kuramı için dezavantaj olarak kabul edilebilecek durumlar maddeler halinde özetlenmiştir.

- STK'da başarı/yetenek puanı olarak ifade edilen ÖİSGİD indeksinin değeri eşit sayıda maddeyi doğru yanıtlayan öğrenciler arasında farklılık göstermemektedir. Bu anlamda KTK'daki toplam test puanları için yapılan, eşit sayıda maddeyi doğru yanıtlayan bireylerin başarı düzeylerinin aynı olmayabileceği eleştirisi STK'daki ÖİSGİD indekslerine de yöneltilebilir. Buna karşın, öğrencilerin yanıt örüntüleri, içerisinde maddeler ve öğrencilerin başarı düzeyleri ile ilgili çok daha fazla bilgi barındırmaktadır. STK bu bilgileri çözümlyerek öğrencilerine öğrenme durumları ile ilgili dönüt veren öğrenci uyarı indekslerini üretmektedir. Eşit sayıda maddeye doğru yanıt veren öğrencilerin ÖÜİ değerleri ve buna dayalı olarak yapılan sınıflandırmalar farklılaşabilmektedir (Lin ve Chen, 2006). Bu bağlamda, STK'nın test puanı aynı olan öğrencileri ayırt edememe gibi bir dezavantajının olup olmadığı konusunun tartışmaya açık bir konu olduğu belirtilebilir.

- STK'nın maddelerle ilgili açıkladığı bilgi sınırlıdır. MTK'nın sağladığı, bireylerin her bir madde üzerindeki performansının detaylı bir şekilde açıklanması, hangi yetenek düzeyleri için maddenin daha ayırt edici olduğunun ve daha fazla bilgi sağladığının belirlenebilmesi, maddenin tahminle doğru yanıtlanma olasılığının hesaplanabilmesi gibi avantajları STK sağlamamaktadır.

## İlgili Araştırmalar

Wang, Sheu ve Nagai (2011), Tayvan'da 37 mühendislik fakültesi öğrencisine altı hafta boyunca Facebook üzerinden İngilizce eğitimini destekleyici çalışmalar yaptırmışlar ve İngilizce performansındaki gelişimi değerlendirmek amacıyla uyguladıkları öntest ve sontest verilerini Sato Test Kuramı yöntemleriyle incelemişlerdir. Söz konusu çalışmada çevrimiçi ders ortamının öğrenci başarısı üzerinde olumlu yönde ve manidar düzeyde etki yarattığı belirlenmiş ve öğrencilere nitel dönüt sağlayan öğrenci uyarı indekslerinin kullanımı eğitimcilere önerilmiştir.

Wang ve Chen'in (2013), Tayvan'da 100 altıncı sınıf öğrencisi üzerinde yaptıkları çalışmada öğrencilere çevrimiçi testler uygulanmış ve elde edilen sonuçlar Sato Test Kuramı ile değerlendirilerek öğrenci uyarı indeksleri aracılığıyla üretilen nitel değerlendirme sonuçları öğrencilere sunulmuştur. Söz konusu nitel dönütler için öğrenci görüşlerinin alınması amacıyla anket uygulanmış ve yüz yüze görüşmeler yapılmıştır. Öğrencilerin nitel dönütleri olumlu karşıladıkları ve yararlı buldukları sonucuna ulaşılmıştır.

Sheu vd. (2013), Gri Öğrenci-Madde Çizelgesi Analizi ile Rasch Modeli birleştirilerek Rasch Gri Öğrenci-Madde Çizelgesi Modeli adında yeni bir yöntem önermişlerdir. 57 yedinci sınıf öğrencisi üzerinde yürütülen çalışmada söz konusu yöntemle hem öğrencilerdeki kavram yanılgılarının tespit edilebildiği hem de öğretimin iyileştirilebilmesi için kapsamlı analiz çıktılarının alınabildiği sonucuna ulaşmışlardır.

Tsai vd. (2014), Gri Öğrenci-Madde Çizelgesi Analizi ile Pawlak'ın Kaba Kümeler Teorisi'ni bir arada kullanarak, 50 maddeden oluşan bir matematik testine yanıt veren 27 öğrencinin kavram yanılgılarının hangi konularla ne derece ilişkili olduğunu belirlemeye çalışmışlardır.

Sheu vd. (2014d), Lise öğrencilerinden oluşan 33 kişilik bir gruba uyguladıkları 25 maddelik bir matematik testi üzerinden Rasch Gri Öğrenci-Madde Çizelgesi Analizi ile KTK'yı ve yine lise öğrencilerinden oluşan 320 kişilik bir gruba

uyguladıkları 48 maddelik bir İngilizce testi üzerinden Rasch Gri Öğrenci-Madde Çizelgesi Analizi ile Bir Parametrelili Lojistik Modeli (MTK) karşılaştırmışlar ve karşılaştırılan yöntemlerintutarlı sonuçlar üretmekle birlikte RG-ÖMÇA'nın bir takım üstünlüklerinin olduğunu vurgulamışlardır. Araştırmacılar bu çalışma sonrasında hem uygulamasının daha kolay olduğu hem de küçük örneklemeler üzerinde daha iyi sonuçlar elde edildiği gibi gerekçelerle Rasch Gri Öğrenci-Madde Çizelgesi analizinin kullanılmasını önermişlerdir.

Sheu vd. (2014a), Sato Test Kuramı indeksleri ve ROC eğrileriyle kavram yanılgılarının tespit edilebileceğini öne sürmüşler ve bu amaçla geliştirdikleri MATLAB eklentisini tanıtmışlardır.

Sheu vd. (2014b), Tayvan'da 305 lise öğrencisi üzerinde yürüttükleri çalışmada öğrencilere matematik testi uygulamışlar ve elde edilen verileri hem Sato Test Kuramı hem de Madde Tepki Kuramı (1PLM model) yöntemleriyle analiz etmişlerdir. Araştırmacılar, söz konusu çalışmada iki kuramın da çok yakın sonuçlar ürettiğini belirlemişlerdir. Ayrıca STK yöntemlerinin yalnızca küçük örneklemelerde değil büyük örneklemelerde de iyi sonuçlar verdiği sonucuna ulaşmışlar ve uygulamalarının MTK yöntemlerine göre daha kolay olduğunu öne sürerek eğitimcilerle önermişlerdir.

Pham, Sheu ve Nagai (2015), genellikle ikili puanlanan maddeler üzerinden hesaplanan Sato Test Kuramı indekslerini çoklu puanlanan maddeler üzerinden hesaplamışlar ve bu amaçla geliştirdikleri PCSP 1.0 adlı yazılımı tanıtmışlardır.

Acar (2006)'da kuramı tanıtmaya yönelik yaptığı çalışmada, 30 dördüncü sınıf öğrencisine uygulanan sosyal bilgiler dersi izleme testi verilerinden madde uyarı indeksleri ve öğrenci uyarı indeksleri elde ederek Sato Test Kuramı'na dayalı yaptığı analizlerin sonuçlarını incelemiştir.

Bayrak ve Yurdugül (2016), öğrenenlerin kendilerini test edebilmeleri amacıyla hazırlanan web tabanlı öz-değerlendirme sisteminde Sato Test Kuramı indekslerinden öğrenci uyarı indeksini temel alan bir öğrenme analitiği modülü tasarımı geliştirmişlerdir.



Çüm, Gelbal ve Tsai (2016), maddeler için sınırlandırılmış gri ilişki derecelerinin tutarlılığını ve madde uyarı indeksleri aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumu incelemeyi amaçladıkları araştırmada, 20 madde ve bu maddeleri baz alan 1000 adet yanıt örüntüsünü simülasyon yoluyla elde ederek bu yanıt örüntüleri içerisinde seçkisiz olarak seçilmiş 30 yanıt örüntüsü içeren 10 farklı veri seti (örneklem) oluşturulmuşlardır. Elde ettikleri bulgular, Sato Test Kuramında madde güçlüklerinin belirlenebilmesi için üretilen MİSGİD'lerin ve MUİ değerlerine dayalı yapılan sınıflandırmaların tutarlı sonuçlar verdiğini göstermiştir.

Çüm, Demir ve Şahin'in (2018), STK'ya dayalı gerçekleştirilen madde ve öğrenci sınıflamalarına yönelik öğretmen görüşlerini inceledikleri araştırmada 70 öğretmene 12 maddelik bir anket uygulanmış ve katılımcı öğretmenlerin büyük çoğunluğunun STK'yı kullanmaya istekli olduğu, madde ve öğrenciler için yapılan sınıflamaları beğendikleri ve işe yarar buldukları sonucuna ulaşılmıştır. Ayrıca öğretmenlerin, yapılan sınıflamalar aracılığıyla verilen dönütleri madde havuzu oluşturma konusunda teşvik edici buldukları ve öğretimi iyileştirme noktasında işlerini kolaylaştıracağını düşündükleri görülmüştür. Buna karşın, araştırmaya katılan öğretmenlerin %41.4'ü madde analizleri için klasik yöntemleri tercih edip etmeme noktasında kararsızlık yaşadığını ifade etmiştir.

### **İlgili Araştırmalar Özeti**

Alanyazın incelemesi sonucu ulaşılan İngilizce araştırma makalelerinin tümünün Uzakdoğulu araştırmacılar tarafından yürütüldüğü belirlenmiştir. Söz konusu araştırmaların çoğu bilgi üretmeye yönelik kuramsal araştırma niteliğindedir. Araştırmalar, kuramın sağladığı nitel dönütlerin öğretimin iyileştirilmesi üzerindeki etkisi ve sağlanan dönütlere yönelik öğrenci görüşlerinin incelenmesi, öğrencilerin sahip olduğu ve maddelerin içerdiği kavram yanlışlarının belirlenmesi, STK ve MTK'ya dayalı yapılan analizlerin karşılaştırılması gibi amaçlar üzerinden yürütülmüştür. Ayrıca, yapılan incelemeler sonucunda Türkiye'de STK'yla ilişkili olduğu belirlenen dört çalışmadan söz edilebileceği görülmüştür.

## **Bölüm 3**

### **Yöntem**

Bu bölümde, araştırmanın türü, çalışma grubu, veri toplama araçları ve verilerin işlenmesi ve çözümlenmesi başlıklarına yer verilmiştir.

#### **Araştırmanın Türü**

Bu araştırma, ölçme ve değerlendirme çalışmalarında kullanılan mevcut yöntemlerin araştırma verisi üzerinden işleyişlerinin incelenmesi, analiz sonuçlarının karşılaştırılması yoluyla yöntemler arasındaki bağlantıların belirlenmesi ve buradan genel bir tabloya ulaşılması bakımından betimsel bir araştırma niteliği taşımaktadır.

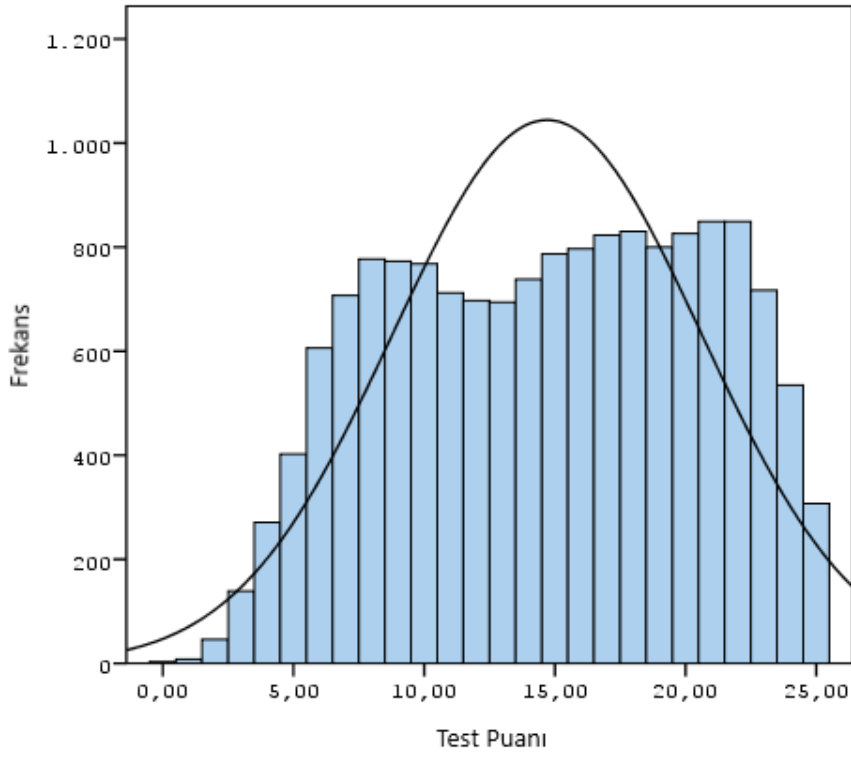
#### **Araştırmanın Çalışma Grubu**

Bu araştırmada, geniş ölçekli bir test uygulamasına katılan öğrenciler üzerinden elde edilen veriler aracılığıyla kuramsal bir inceleme yapılması durumu söz konusudur. Dolayısıyla sonuçların herhangi bir bireyler evrenine genellenmesi amaçlanmamaktadır. Çalışma grubunda, 2005 yılında uygulanan Öğrenci Başarılarının Belirlenmesi Sınavı'na (ÖBBS) katılan 8.sınıf öğrencilerinden oluşan 15461 kişi yer almaktadır. Araştırma verileri, ÖBBS Türkçe ve matematik alt testlerinde yer alan çoktan seçmeli maddelere verilen yanıtların ikili puanlanması (1-0) ile oluşturulmuştur.

#### **Veri Toplama Araçları**

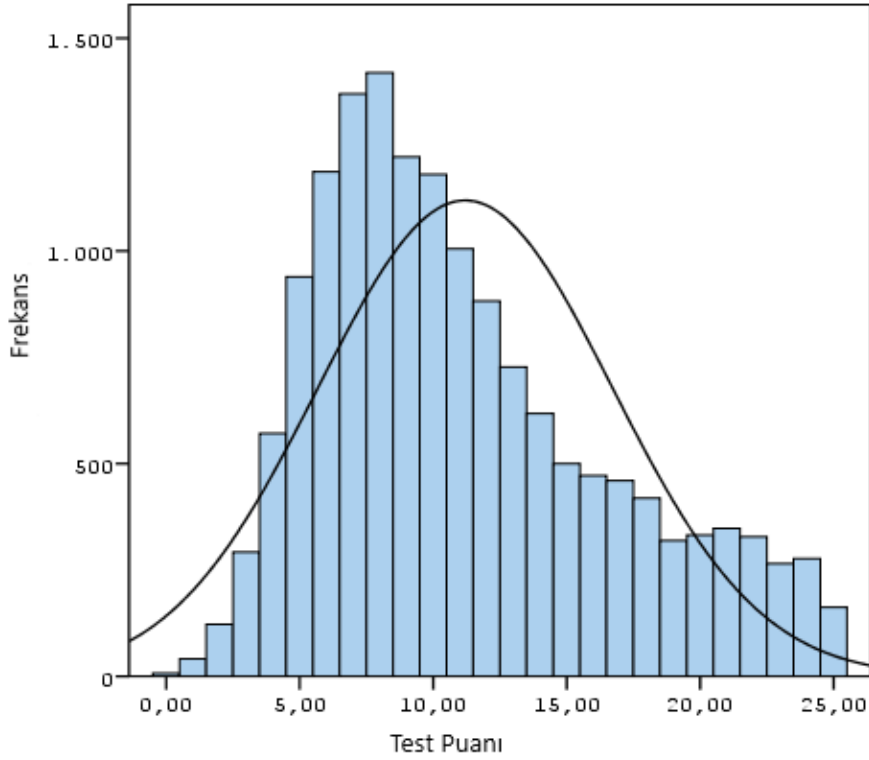
Bu araştırmada veri toplama aracı olarak kullanılan Öğrenci Başarılarının Belirlenmesi Sınavı (ÖBBS), öğrencilerin neler bildiklerinden çok, her sınıf düzeyinde kazandırılması öngörülen bilgi ve becerileri ne düzeyde kazandıklarını ve eksiklerinin neler olduğunu belirlemek amacıyla yapılan bir durum belirleme çalışmasıdır. Bu sınavda, sınıf seviyesine göre becerileri ortaya çıkaracak nitelikte hazırlanmış çoktan seçmeli sorulardan oluşan Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce alanlarında testler bulunmaktadır. Ayrıca sınav kapsamında öğrenci ve öğretmen anketleri de uygulanmaktadır. Sınavın uygulama evreni Türkiye'deki 4-8. sınıflarda okuyan bütün ilk ve ortaokul öğrencileri ve bu okullarda görevli olan öğretmenlerdir. Sınavın katılımcı örnekleme sosyo-ekonomik

durum, bölge, il, taşra ve merkez değişkenleri dikkate alınarak tabakalı örnekleme yöntemiyle e-okul veri tabanından yararlanılarak belirlenmektedir (EARGED, <http://yegitek.meb.gov.tr/earged/arasayfa.php?q=77> Haziran 2016'da erişildi). ÖBBS'nin bu araştırma kapsamında ele alınan Türkçe ve matematik alt testlerinde dört seçenekli 25'er madde yer almaktadır. Sınava katılan öğrencilerin Türkçe testinden aldıkları puanların (doğru yanıtlanan maddeler için 1 yanlış yanıtlanan ve boş bırakılan maddeler için 0 puan verilmiştir) histogram grafiği Şekil 6'da sunulmuştur.



Şekil 6. Türkçe test puanlarının dağılımına ilişkin histogram grafiği.

Şekil 6'da verilen test puanları dağılımının ortalaması 14,703 ve ortancası 15,000 olarak hesaplanmıştır. Dağılımın standart sapması 5,907'dir. Dağılıma ilişkin çarpıklık katsayısı -0,105 ve basıklık katsayısı -1,093'tür. Elde edilen bulgulardan hareketle dağılımın normalden basık olmakla birlikte normale yakın özellikler gösterdiği belirtilebilir. Sınava katılan öğrencilerin matematik testinden aldıkları puanların (doğru yanıtlanan maddeler için 1 yanlış yanıtlanan ve boş bırakılan maddeler için 0 puan verilmiştir) histogram grafiği Şekil 7'de sunulmuştur.



Şekil 7. Matematik test puanlarının dağılımına ilişkin histogram grafiği.

Şekil 7’de verilen test puanları dağılımının ortalaması 11,199 ve ortancası 10,000 olarak hesaplanmıştır. Dağılımın standart sapması 5,513’tür. Dağılıma ilişkin çarpıklık katsayısı 0,697’dir ve test puanları bu anlamda sağa çarpık bir dağılım özelliği göstermektedir. Basıklık katsayısı ise -0,320 olarak hesaplanmıştır.

### Verilerin Analizi

Bu bölümde öncelikle, araştırma verilerinin yapılacak olan analizlerin gerektirdiği varsayımları karşılayıp karşılamadığı incelenmiş daha sonra veri çözümü süreci, araştırma sorularını sırasıyla ele alacak şekilde özetlenmeye çalışılmıştır.

### Varsayımların Test Edilmesi

Bu aşamada ilk olarak hem KTK hem de MTK için gerekli olan tek boyutluluk varsayımı Paralel Analiz yöntemi ile incelenmiştir. Söz konusu yöntem pek çok araştırmacı tarafından faktör (boyut) sayısının belirlenmesinde en iyi yöntem olarak nitelendirilmektedir (Fabrigar ve diğerleri, 1999; Field, 2009;

Hayton, Allen ve Scarpello, 2004; Henson ve Roberts, 2006; Horn, 1965; Thompson, 2004; Zwick ve Velicer, 1986). Bu amaçla, “Factor 10” programı aracılığıyla tetrakorik korelasyon matrisine dayalı faktör analizi yürütülmüş ve faktörleştirme yöntemi olarak Coughlin (2013) tarafından önerilen Ağırlıklandırılmamış En Küçük Kareler Yöntemi (Unweighted Least Squares) seçilmiştir. Korelasyon matrisinin faktörleşebilirliğine ilişkin kanıtlar Bartlett Küresellik Testi (Bartlett’s Test of Sphericity) ve Kaiser-Meyer-Olkin Örneklem Uygunluğu Ölçütü (Kaiser-Meyer-Olkin Test of Sampling Adequacy) ile elde edilmiştir. Türkçe ve matematik alt testleri için Bartlett Küresellik Testi sonuçları korelasyon matrisi ile birim matris arasında istatistiksel olarak manidar bir farklılık olduğuna ( $p= 0.000$ ) işaret etmektedir. Her iki teste ilişkin KMO değerleri (Türkçe= 0.938; matematik=0,901) göz önünde bulundurulduğunda, korelasyon matrislerinin faktörleşebilirlik düzeylerinin “mükemmel” olduğu şeklinde bir değerlendirmede bulunulabilir (Beavers ve diğerleri, 2013). Korelasyon matrisine ilişkin varsayımlar sağlandıktan sonra yapılan Faktör Analizi ve buna dayalı olarak yapılan Paralel Analiz sonuçları hem Türkçe hem de matematik alt testleri için “tek boyut” önermiştir. Elde edilen bulgulara dayalı olarak araştırma verilerinin tek boyutluluk varsayımını karşıladığına karar verilmiştir.

Tek boyutluluk, bir diğer MTK varsayımı olan yerel bağımsızlığın varlığına da kanıt olarak gösterilebilir. Bu kanıtın yanı sıra IRTPRO programı aracılığıyla maddeler arası standardize edilmiş ki-kare değerleri incelenmiş ve elde edilen bulgular EK-A ve EK-B’de sunulmuştur. Söz konusu bulgulardan hareketle, her iki alt test için az sayıda madde arasında yüksek ki-kare değerlerine rastlandığı fakat testler bütünsel olarak ele alındığında maddeler arası ilişkilerin yerel bağımsızlığı ihlal edecek düzeyde olmadığı sonucuna ulaşılmıştır.

Araştırma verisine ilişkin temel varsayımların kontrolünden sonra MTK parametrelerinin daha doğru kestirilebilmesi için veriye en iyi uyum sağlayan MTK modelinin seçilmesi aşamasına geçilmiştir. Bu amaçla, Türkçe ve matematik testlerine ait veriler 1P, 2P ve 3P Lojistik Modellere göre analiz edilmiş ve elde edilen -2log Likelihood değerleri arasındaki farklar incelenmiştir. Söz konusu değerler Tablo 2’de sunulmuştur.

Tablo 2

Lojistik Modeller için  $-2\log$  Likelihood Değerleri

Test	1PLM	2PLM	3PLM
Türkçe	445276,52	440691,74	438268,89
Matematik	471051,75	468898,42	463975,82

Modellerin uyum düzeylerinin karşılaştırılmasında  $\chi^2$  fark testi kullanılmıştır. Türkçe testi için en uygun modelin Üç Parametrelili Lojistik Model olduğu sonucuna ulaşılmıştır. Yapılan hesaplamalar aşağıdaki şekilde ifade edilebilir.

$$\chi^2 = -2 \log Likelihood_{1plm} - (-2 \log Likelihood_{2plm}) = 445276,52 - 440691,74 = 4584,78 > 37,65 (\chi^2_{(df=25 p=;0,05)}) \text{ ve,}$$

$$\chi^2 = -2 \log Likelihood_{2plm} - (-2 \log Likelihood_{3plm}) = 440691,74 - 438268,89 = 2422,85 > 37,65 (\chi^2_{(df=25 p=;0,05)})$$

Matematik testi için de en uygun modelin Üç Parametrelili Lojistik Model olduğu sonucuna ulaşılmıştır. Yapılan hesaplamalar aşağıdaki şekilde ifade edilebilir.

$$\chi^2 = -2 \log Likelihood_{1plm} - (-2 \log Likelihood_{2plm}) = 471051,75 - 468898,42 = 2153,33 > 37,65 (\chi^2_{(df=25 p=;0,05)}) \text{ ve,}$$

$$\chi^2 = -2 \log Likelihood_{2plm} - (-2 \log Likelihood_{3plm}) = 468898,42 - 463975,82 = 4922,60 > 37,65 (\chi^2_{(df=25 p=;0,05)})$$

Seçilen modelin veriye uygunluğuna ilişkin daha fazla kanıt elde edebilmek amacıyla 3PLM ile üretilen madde ve yetenek parametrelerinin değişmezliğine yönelik incelemeler yapılmıştır. Madde parametrelerinin değişmezliğine ilişkin kanıt elde edilmesi amacıyla Türkçe ve matematik testlerine katılan bireylere ait yanıt örüntüleri seçkisiz olarak iki gruba ayrılmış ve her iki gruptan kestirilen madde parametreleri arasındaki ilişki Spearman-Brown Sıra Farkları Korelasyon Katsayıları ile incelenmiştir. Söz konusu bulgular Tablo 3'te sunulmuştur.

Tablo 3

*Farklı Gruplardan Elde Edilen Madde Parametreleri Arasındaki Korelasyonlar*

Test	a	b	c
Türkçe	0,957	0,992	0,875
Matematik	0,926	0,993	0,940

Tablo 3 incelendiğinde, madde parametreleri arasındaki korelasyonların pozitif ve yüksek düzeyde ilişki gösterdiği görülmektedir. Bu bulguya dayalı olarak Türkçe ve matematik testlerinin madde parametrelerinin 3PLM'de değişmezlik özelliği gösterdiği yorumu yapılabilir.

Yetenek parametrelerinin değişmezliğinin incelenmesi amacıyla, Türkçe ve matematik testlerinde yer alan maddeler rastgele bir şekilde ikişer gruba ayrılarak alt testler oluşturulmuş ve bu alt testlerden 3PLM ile yetenek parametreleri kestirilmiştir. Türkçe testi maddelerinden oluşturulan iki ayrı alt testten kestirilen yetenek parametreleri arasındaki korelasyon düzeyi Pearson Momentler Çarpımı Korelasyon Katsayısı tekniği ile incelenmiş ve bu değer 0,783 olarak belirlenmiştir. Matematik alt testlerinden kestirilen yetenek parametreleri arasındaki korelasyon katsayısı ise 0,887 olarak hesaplanmıştır. Elde edilen bulgulara dayalı olarak, incelenen testleri yanıtlayan bireylerin yetenek düzeylerinin 3PLM ile madde örnekleminde bağımsız olarak kestirilebileceği iddiasında bulunulabilir. Yapılan incelemeler sonrasında, araştırma verilerinin çözümlenmeler için gerekli olan varsayımları karşıladığına karar verilmiştir. Bununla birlikte, MTK'ya dayalı çözümlenmelerde araştırma verisine en iyi uyum sağlayan Üç Parametrelili Lojistik Modelin kullanılması ve diğer test kuramlarıyla yapılan karşılaştırmaların bu modele dayalı olarak kestirilen parametrelerle sınırlandırılması kararlaştırılmıştır.

**Veri Çözümleme Süreci**

Öncelikle, ÖBBS'ye (2005) katılan 15461 8.sınıf öğrencisinin yanıt örüntüleri arasından basit seçkisiz örnekleme yöntemiyle her birinden 10'ar adet olmak üzere 50'şer, 100'er, 200'er ve 600'er kişilik örneklemler oluşturularak araştırma verileri araştırma problemini çözümlenmeye uygun hale getirilmiştir.

Araştırmanın birinci alt problemine yanıt aramak amacıyla, oluşturulan örneklemelerin her biri üzerinden ÖBBS (2005) Türkçe ve matematik alt testlerinde yer alan maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) elde edilmiştir. Bu değerler arasındaki ilişkilerin derecesi Spearman'ın Sıra Farkları Korelasyon Katsayısı yöntemiyle hesaplanmış ve değerler arasındaki farkların istatistiksel manidarlığı Kruskal-Wallis-H Testi ile incelenmiştir.

Araştırmanın ikinci alt problemine yanıt aramak amacıyla yanıtlayıcı grubunun tümü en yüksek puanlı öğrenciden en düşük puanlı öğrenciye doğru sıralanmış ve verilerin ilk %25'lik bölümü (üst grup) ile son %25'lik bölümünden (alt grup) iki grup elde edilmiştir. Oluşturulan gruplar üzerinden maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) hesaplanarak değerler arasındaki ilişkiler Spearman'ın Sıra Farkları Korelasyon Katsayısı yöntemiyle incelenmiştir.

Araştırmanın üçüncü alt problemine yanıt aramak amacıyla, oluşturulan örneklemelerin her biri üzerinden Türkçe ve matematik testlerinde yer alan maddelerin madde uyarı indeksleri (MUİ) ve doğru yanıtlanma oranları hesaplanmış ve elde edilen verilerden hareketle maddeler A, A', B ve B' sınıflarına atanmıştır. Yapılan sınıflandırmaların tutarlılığı, yüzde oranlar hesaplanarak incelenmiştir.

Araştırmanın dördüncü alt problemine yanıt aramak amacıyla, oluşturulan tüm örneklemeler üzerinden Türkçe ve matematik testlerinde yer alan maddelerin  $r_{jx}$  ve MUİ değerleri elde edilmiş ve her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen değerler arasındaki korelasyon katsayıları Spearman'ın Sıra Farkları Korelasyon Katsayısı Tekniği hesaplanarak ortancaları alınmıştır.

Araştırmanın beşinci alt problemine yanıt aramak amacıyla, oluşturulan tüm örneklemeler üzerinden Türkçe ve matematik testlerinde yer alan maddelerin  $p$  ve MİSGİD değerleri elde edilmiştir. Söz konusu değerler arasındaki korelasyon katsayıları her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden Spearman'ın Sıra Farkları Korelasyon Katsayısı Tekniği ile hesaplanarak ortancaları alınmıştır.



Araştırmanın altıncı alt problemine yanıt aramak amacıyla, Türkçe ve matematik testlerini yanıtlayan ve oluşturulan örneklem içerisnde yer alan öğrencilerin test toplam puanları ve öğrenciler için sınırlandırılmış gri ilişki dereceleri (ÖİSGİD) elde edilmiştir. Söz konusu değerler arasındaki korelasyon katsayıları her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden Pearson Korelasyon Katsayısı Tekniğı ile hesaplanarak ortancaları alınmıştır.

Araştırmanın yedinci alt problemine yanıt aramak amacıyla, Türkçe ve matematik testlerinde yer alan maddeler için 50, 100, 200 ve 600'er kişilik 10'ar farklı örneklem üzerinden elde edilen MUİ değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $a$  parametreleri arasındaki korelasyon katsayıları Spearman'ın Sıra Farkları Korelasyon Katsayısı Tekniğı ile hesaplanarak ortancaları alınmıştır.

Araştırmanın sekizinci alt problemine yanıt aramak amacıyla, Türkçe ve matematik testlerinde yer alan maddeler için 50, 100, 200 ve 600'er kişilik 10'ar farklı örneklem üzerinden elde edilen MİSGİD değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $b$  parametreleri arasındaki korelasyon katsayıları Spearman'ın Sıra Farkları Korelasyon Katsayısı Tekniğı ile hesaplanarak ortancaları alınmıştır.

Araştırmanın dokuzuncu alt problemine yanıt aramak amacıyla, Türkçe ve matematik testlerini yanıtlayan ve 50, 100, 200 ve 600'er kişilik örneklem içerisnde yer alan öğrencilerin ÖİSGİD değerleri ile aynı öğrencilerin 3PLM ile kestirilen yetenek ( $\theta$ ) parametreleri ( $\theta$  kestirimleri EAP yöntemi ile tüm veri seti üzerinden gerçekleştirilmiştir.) arasındaki korelasyonlar Pearson Korelasyon Katsayısı Tekniğı ile incelenmiştir.

Araştırmanın onuncu alt problemine yanıt aramak amacıyla, Türkçe ve matematik testleri için oluşturulan örneklem üzerinden Çift Serili Korelasyon Katsayısı tekniğı ile madde ayırıcılık indeksleri ( $r_{jx}$ ) elde edilerek sorunlu maddeler KTK bağlamında belirlenmiştir. Daha sonra, aynı örneklemlerden alınan veriler Sato Test Kuramı'na dayalı olarak analiz edilmiş ve sorunlu maddeler MUİ'ler

aracılığı ile yapılan sınıflandırmalar yoluyla belirlenmiştir. Son olarak, Madde Tepki Kuramı'nın  $a$ ,  $c$  parametreleri ve madde bilgi fonksiyonları aracılığı ile sorunlu maddeler belirlenerek her üç kurama göre ulaşılan sonuçlar arasında karşılaştırmalar yapılmıştır.

Verilerin analizi süresince Sato Test Kuramına dayalı kestirimler Sheu, Pham, Nguyen ve Nguyen (2013) tarafından Öğrenci-Madde Çizelgesi Analizi ve Gri Öğrenci-Madde Çizelgesi Analizi için geliştirilen MATLAB Paketi ile Madde Tepki Kuramına dayalı parametreler IRTPRO (V, 4.2) yazılımı ile Klasik Test Kuramına dayalı istatistikler ise TAP (V,14.7.4) yazılımı ile elde edilmiştir. Spearman'ın Sıra Farkları Korelasyon Katsayısı Tekniği, Pearson Korelasyon Katsayısı Tekniği ve Kruskal-Wallis-H Testi ile yapılacak olan karşılaştırmalar için ise SPSS (V,22) yazılımı kullanılmıştır.

## Bölüm 4

### Bulgular ve Yorumlar

Bu bölümde, araştırma problemine ve buna bağlı olarak ortaya koyulan alt problemlere yanıt bulmaya yönelik olarak yürütülen çözümlerden elde edilen bulgular sunulmaktadır. Bulgular ve bulgulara dayalı olarak yapılan yorumlar, alt problemlerin sırasına uygun olarak verilmiş ve bölüm sonunda tartışılarak özetlenmiştir.

#### Araştırmanın Birinci Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın birinci alt problemi olan “ÖBBS (2005) Türkçe ve matematik alt testlerinde yer alan maddeler için 50’şer kişilik 10, 100’er kişilik 10, 200’er kişilik 10 ve 600’er kişilik 10 farklı örneklem üzerinden hesaplanan maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) arasında ne düzeyde korelasyon bulunmaktadır ve MİSGİD’ler arasındaki farklar istatistiksel olarak manidar mıdır?” sorusuna ilişkin bulgular önce Türkçe sonra matematik testi bağlamında incelenmiştir. İlk olarak Türkçe testinde yer alan maddelerin MİSGİD değerleri 50’şer kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 4’te sunulmuştur.

Tablo 4

*Türkçe Testi Maddelerinin 50’şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,868*									
Örn3	0,760*	0,826*								
Örn4	0,788*	0,818*	0,940*							
Örn5	0,847*	0,869*	0,779*	0,730*						
Örn6	0,800*	0,844*	0,699*	0,667*	0,830*					
Örn7	0,745*	0,858*	0,832*	0,806*	0,768*	0,841*				
Örn8	0,771*	0,889*	0,792*	0,827*	0,851*	0,774*	0,821*			
Örn9	0,796*	0,910*	0,818*	0,816*	0,821*	0,774*	0,811*	0,850*		
Örn10	0,685*	0,830*	0,770*	0,774*	0,703*	0,716*	0,823*	0,777*	0,838*	

\* $p < 0,01$

Tablo 4 incelendiğinde, Türkçe testi için 10 farklı örneklemden elde edilen MİSGİD değerleri arasındaki korelasyon katsayılarının 0,667 ile 0,940 değerleri arasında değiştiği ve söz konusu değerlerin 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası 0,816 olarak hesaplanmıştır ve bu değer

katsayılar arasında genel olarak yüksek düzeyde bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgulardan hareketle, Sato Test Kuramı kapsamında hesaplanan ve madde güçlüğü hakkında bilgi veren maddeler için sınırlandırılmış gri ilişki derecelerinin, aynı maddeleri yanıtlayan katılımcılar arasından seçilen farklı örneklerde (50'şer kişilik) tutarlı sonuçlar verdiği iddia edilebilir. Bu iddiayı sağlamlaştırabilmek amacıyla, "maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) arasındaki farklar istatistiksel olarak manidar mıdır?" sorusuna yanıt aranmış ve farklı örneklerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel manidarlığı non-parametrik bir test olan Kruskal Wallis-H testi ile incelenmiştir. Elde edilen bulgular Tablo 5'te sunulmuştur.

Tablo 5

*Türkçe Testi Maddelerinin 50'şer Kişilik Örneklerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	16,118
Sd	9
P	0,064

Tablo 5 incelendiğinde, farklı örneklerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,064>0,050$ ) görülmektedir. Bu bulgunun, 50'şer kişilik örneklerden elde edilen MİSGİD değerlerinin tutarlı sonuçlar verdiğiğine ilişkin ikincil bir kanıt oluşturduğu ifade edilebilir.

Türkçe testinde yer alan maddelerin MİSGİD değerleri 100'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 6'da sunulmuştur.

Tablo 6

*Türkçe Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,901*									
Örn3	0,856*	0,923*								
Örn4	0,914*	0,937*	0,916*							
Örn5	0,882*	0,922*	0,896*	0,950*						
Örn6	0,825*	0,906*	0,840*	0,881*	0,894*					
Örn7	0,876*	0,887*	0,812*	0,894*	0,920*	0,860*				
Örn8	0,908*	0,956*	0,912*	0,925*	0,885*	0,908*	0,855*			
Örn9	0,886*	0,930*	0,921*	0,958*	0,935*	0,886*	0,901*	0,915*		
Örn10	0,865*	0,906*	0,924*	0,908*	0,862*	0,885*	0,835*	0,939*	0,887*	

\* $p < 0,01$

Tablo 6 incelendiğinde, korelasyon katsayılarının 0,812 ile 0,958 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,901 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek düzeyde bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgulardan hareketle, MİSGİD değerlerinin 100'er kişilik farklı örneklemeler arasında tutarlı sonuçlar verdiği yorumu yapılabilir. Ayrıca örneklemelerde yer alan kişi sayısının artmasıyla birlikte MİSGİD değerleri arasındaki korelasyon düzeylerinin de yükseldiği belirlenmiştir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 7'de sunulmuştur.

Tablo 7

*Türkçe Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	5,067
Sd	9
P	0,828

Tablo 7 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,828 > 0,050$ ) görülmektedir. Bu bulgu, 100'er kişilik örneklemelerden elde edilen MİSGİD değerlerinin tutarlı sonuçlar verdiğine ilişkin ikincil bir kanıt oluşturmaktadır.

Türkçe testinde yer alan maddelerin MİSGİD değerleri 200'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 8'de sunulmuştur.

Tablo 8

*Türkçe Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,965*									
Örn3	0,952*	0,934*								
Örn4	0,922*	0,938*	0,904*							
Örn5	0,921*	0,910*	0,928*	0,907*						
Örn6	0,941*	0,964*	0,945*	0,941*	0,935*					
Örn7	0,923*	0,942*	0,892*	0,927*	0,891*	0,937*				
Örn8	0,935*	0,955*	0,938*	0,942*	0,910*	0,955*	0,948*			
Örn9	0,933*	0,926*	0,953*	0,932*	0,959*	0,956*	0,900*	0,907*		
Örn10	0,933*	0,917*	0,940*	0,885*	0,902*	0,928*	0,913*	0,932*	0,926*	

\* $p < 0,01$

Tablo 8 incelendiğinde, korelasyon katsayılarının 0,885 ile 0,965 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,933 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek düzeyde bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgulardan hareketle, 200'er kişilik farklı örneklemeler üzerinden üretilen MİSGİD değerlerinin birbirleri ile tutarlı olduğu yorumu yapılabilir. Örneklemelerde yer alan kişi sayısının artmasıyla birlikte MİSGİD değerleri arasındaki korelasyon düzeylerinin de yükseldiği belirlenmiştir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 9'da sunulmuştur.

Tablo 9

*Türkçe Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	4,430
Sd	9
P	0,881

Tablo 9 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,881 > 0,50$ )

görülmektedir. Bu bulgu, 200'er kişilik farklı örneklemelerden elde edilen MİSGİD değerlerinin tutarlı olduğuna ilişkin diğer bir kanıt olarak ileri sürülebilir.

Türkçe testinde yer alan maddelerin MİSGİD değerleri 600'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 10'da sunulmuştur.

Tablo 10

*Türkçe Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,955*									
Örn3	0,969*	0,979*								
Örn4	0,974*	0,976*	0,985*							
Örn5	0,971*	0,981*	0,981*	0,991*						
Örn6	0,968*	0,973*	0,980*	0,981*	0,982*					
Örn7	0,980*	0,981*	0,990*	0,990*	0,992*	0,975*				
Örn8	0,978*	0,972*	0,978*	0,980*	0,988*	0,983*	0,979*			
Örn9	0,964*	0,970*	0,987*	0,980*	0,979*	0,979*	0,982*	0,976*		
Örn10	0,977*	0,980*	0,989*	0,990*	0,988*	0,976*	0,992*	0,980*	0,974*	

\* $p < 0,01$

Tablo 10 incelendiğinde, korelasyon katsayılarının 0,955 ile 0,992 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,980 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek bir ilişkinin varlığına işaret etmektedir. Söz konusu değerler MİSGİD değerlerinin 600'er kişilik örneklemelerden son derece yüksek bir tutarlılıklar kestirildiğine bir kanıt olarak sunulabilir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 11'de sunulmuştur.

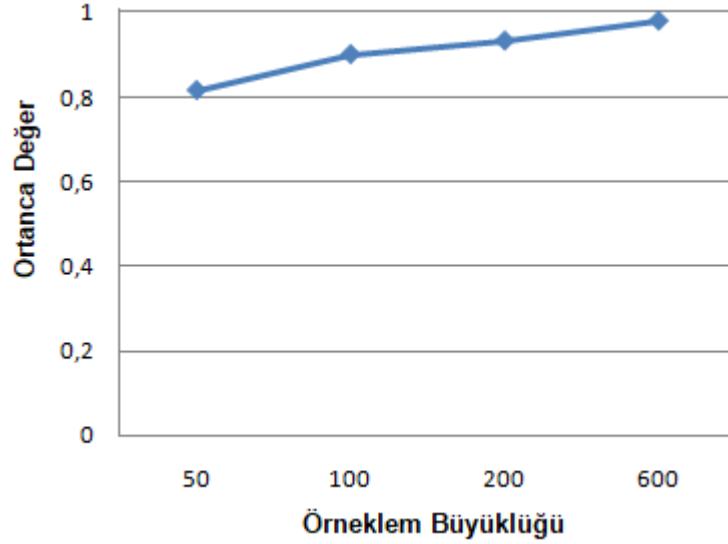
Tablo 11

*Türkçe Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	1,715
Sd	9
P	0,995

Tablo 11 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,995>0,050$ ) görülmektedir. Bu bulgu, 600'er kişilik farklı örneklemelerden elde edilen MİSGİD değerlerinin tutarlı olduğuna ilişkin diğer bir kanıt olarak ileri sürülebilir.

Elde edilen bulgulardan hareketle, Türkçe testinde yer alan maddeler üzerinden hesaplanan MİSGİD değerlerinin araştırma kapsamında ele alınan tüm örneklem büyüklükleri için aynı evrenden seçilen farklı örneklemeler arasında tutarlılık gösterdiği yorumuna gidilebilir. Örneklem büyüklüğüne bağlı olarak korelasyonlarda meydana gelen değişim, ortanca değerler üzerinden görselleştirilerek Şekil 8'de sunulmuştur.



Şekil 8. Türkçe testi MİSGİD değerleri arasındaki korelasyon katsayılarının ortanca değerlerinin örneklem büyüklüğü bağlamında karşılaştırılması.

Şekil 8 incelendiğinde, korelasyon katsayılarının ortancalarının örneklem büyüklüğüne bağlı olarak yaklaşık 0,8 ile 1 değerleri arasında değişim gösterdiği belirlenmiştir. Örneklemelerde bulunan kişi sayısının artmasıyla birlikte MİSGİD değerleri arasındaki korelasyon katsayılarının da artış gösterdiği ve en yüksek korelasyon düzeylerine 600'er kişilik örneklemelerde ulaşıldığı görülmüştür. Bununla birlikte, ele alınan tüm örneklem büyüklükleri için hesaplanan ortanca değerler MİSGİD'ler arasındaki korelasyonların yüksek düzeyde olduğuna işaret



etmektedir. Bu bulgudan hareketle, STK'ya dayalı olarak madde özelliklerinin belirlenebilmesi için çok büyük örneklemelere ihtiyaç olmadığı yorumuna gidilebilir.

Matematik testinde yer alan maddelerin MİSGİD değerleri 50'şer kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 12'de sunulmuştur.

Tablo 12

*Matematik Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,796*									
Örn3	0,715*	0,742*								
Örn4	0,806*	0,683*	0,753*							
Örn5	0,788*	0,647*	0,678*	0,799*						
Örn6	0,857*	0,786*	0,793*	0,788*	0,673*					
Örn7	0,873*	0,878*	0,757*	0,776*	0,771*	0,890*				
Örn8	0,860*	0,652*	0,696*	0,685*	0,720*	0,757*	0,768*			
Örn9	0,813*	0,716*	0,770*	0,809*	0,722*	0,800*	0,765*	0,750*		
Örn10	0,862*	0,706*	0,805*	0,775*	0,704*	0,822*	0,820*	0,820*	0,871*	

\* $p < 0,01$

Tablo 12 incelendiğinde, matematik testi için 50'şer kişilik 10 farklı örneklemde elde edilen MİSGİD değerleri arasındaki korelasyon katsayılarının 0,647 ile 0,890 değerleri arasında değiştiği ve söz konusu değerlerin 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası 0,775 olarak hesaplanmıştır. Elde edilen bulgular doğrultusunda farklı örneklemeler üzerinden hesaplanan MİSGİD değerlerinin birbirleri ile tutarlılık gösterdiği öne sürülebilir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 13'te sunulmuştur.

Tablo 13

*Matematik Testi Maddelerinin 50'şer Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	13,895
sd	9
p	0,126

Tablo 13 incelendiğinde, farklı örneklemlerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,126>0,050$ ) görülmektedir. Bu bulgu, elde edilen MİSGİD değerlerinin 50'şer kişilik örneklemler arasında tutarlılık gösterdiğine ilişkin diğer bir kanıt olarak ileri sürülebilir.

Matematik testinde yer alan maddelerin MİSGİD değerleri 100'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 14'te sunulmuştur.

Tablo 14

*Matematik Testi Maddelerinin 100'er Kişilik Örneklemlerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,878*									
Örn3	0,919*	0,831*								
Örn4	0,814*	0,702*	0,806*							
Örn5	0,747*	0,670*	0,808*	0,676*						
Örn6	0,860*	0,853*	0,882*	0,782*	0,811*					
Örn7	0,804*	0,823*	0,828*	0,658*	0,777*	0,826*				
Örn8	0,810*	0,808*	0,841*	0,824*	0,844*	0,880*	0,866*			
Örn9	0,890*	0,881*	0,913*	0,855*	0,801*	0,926*	0,862*	0,878*		
Örn10	0,825*	0,814*	0,895*	0,784*	0,807*	0,834*	0,789*	0,844*	0,873*	

\* $p<0,01$

Tablo 14 incelendiğinde, korelasyon katsayılarının 0,658 ile 0,926 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,826 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgulardan hareketle, 100'er kişilik farklı örneklemler üzerinden üretilen MİSGİD değerlerinin birbirleri ile tutarlı olduğu yorumu yapılabilir. Örneklemlerde yer alan kişi sayısının artmasıyla birlikte değerler arasındaki korelasyon düzeylerinin de genel olarak arttığı gözlemlenmiştir. MİSGİD değerleri arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 15'te sunulmuştur.

Tablo 15

*Matematik Testi Maddelerinin 100'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	8,975
sd	9
p	0,440

Tablo 15 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,440>0,050$ ) görülmektedir. Bu bulgu, elde edilen MİSGİD değerlerinin 100'er kişilik örneklemeler arasında tutarlılık gösterdiğine ilişkin diğer bir kanıt olarak ileri sürülebilir. Matematik testinde yer alan maddelerin MİSGİD değerleri 200'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 16'da sunulmuştur.

Tablo 16

*Matematik Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,906*									
Örn3	0,902*	0,886*								
Örn4	0,927*	0,894*	0,921*							
Örn5	0,903*	0,873*	0,860*	0,913*						
Örn6	0,902*	0,857*	0,869*	0,879*	0,839*					
Örn7	0,933*	0,909*	0,894*	0,933*	0,872*	0,856*				
Örn8	0,909*	0,916*	0,859*	0,946*	0,953*	0,872*	0,897*			
Örn9	0,893*	0,887*	0,912*	0,855*	0,745*	0,859*	0,885*	0,797*		
Örn10	0,944*	0,904*	0,854*	0,906*	0,914*	0,896*	0,872*	0,910*	0,861*	

\* $p<0,01$

Tablo 16 incelendiğinde, korelasyon katsayılarının 0,745 ile 0,953 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,894 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgulardan hareketle, 200'er kişilik farklı örneklemeler üzerinden üretilen MİSGİD değerlerinin birbirleri ile yüksek düzeyde tutarlılık gösterdiği iddia edilebilir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 17'de sunulmuştur.

Tablo 17

*Matematik Testi Maddelerinin 200'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	2,132
sd	9
p	0,989

Tablo 17 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=0,989>0,050$ ) görülmektedir. Bu bulgu, elde edilen MİSGİD değerlerinin 200'er kişilik örneklemeler arasında tutarlılık gösterdiğine ilişkin diğer bir kanıt olarak ileri sürülebilir. Matematik testinde yer alan maddelerin MİSGİD değerleri 600'er kişilik 10 farklı örneklem üzerinden hesaplanmış ve bunlar arasındaki korelasyon katsayıları Tablo 18'de sunulmuştur.

Tablo 18

*Matematik Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Korelasyonlar*

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	0,970*									
Örn3	0,976*	0,968*								
Örn4	0,976*	0,972*	0,988*							
Örn5	0,985*	0,974*	0,979*	0,982*						
Örn6	0,935*	0,940*	0,937*	0,959*	0,950*					
Örn7	0,973*	0,971*	0,976*	0,993*	0,979*	0,960*				
Örn8	0,967*	0,963*	0,969*	0,981*	0,965*	0,981*	0,980*			
Örn9	0,949*	0,960*	0,962*	0,967*	0,964*	0,955*	0,956*	0,963*		
Örn10	0,972*	0,959*	0,974*	0,969*	0,975*	0,967*	0,963*	0,973*	0,962*	

\* $p<0,01$

Tablo 18 incelendiğinde, korelasyon katsayılarının 0,935 ile 0,993 değerleri arasında değiştiği ve korelasyonların 0,01 düzeyinde manidar olduğu görülmektedir. Katsayıların ortancası ise 0,969 olarak hesaplanmıştır ve bu değer katsayılar arasında genel olarak yüksek bir ilişkinin varlığına işaret etmektedir. Elde edilen bulgular, matematik testi maddeleri için 600'er kişilik örneklemelerden elde edilen MİSGİD değerlerinin yüksek düzeyde tutarlı sonuçlar verdiğine ilişkin bir kanıt olarak sunulabilir. Değerler arasındaki farkların istatistiksel manidarlığına ilişkin bulgular Tablo 19'da sunulmuştur.

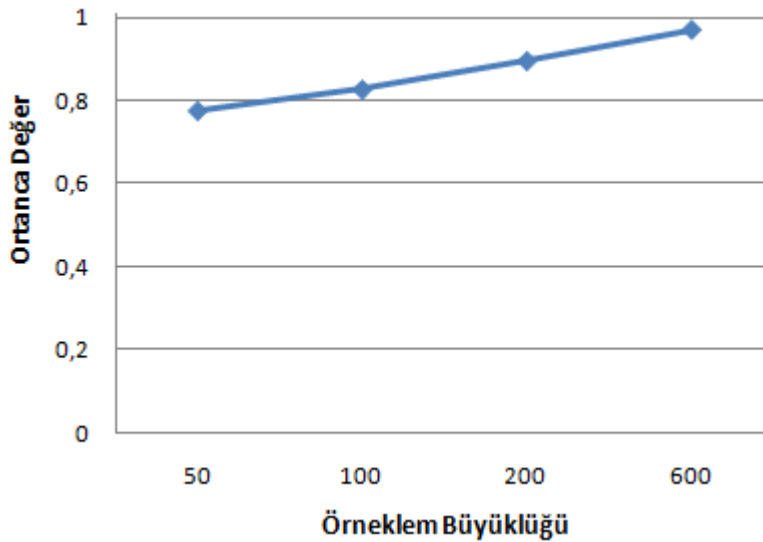
Tablo 19

*Matematik Testi Maddelerinin 600'er Kişilik Örneklemelerden Elde Edilen MİSGİD Değerleri Arasındaki Farkların Manidarlığına İlişkin Bulgular*

	MİSGİD
Ki-Kare	,952
sd	9
$p$	1,000

Tablo 19 incelendiğinde, farklı örneklemelerden elde edilen MİSGİD değerleri arasındaki farkların istatistiksel olarak manidar olmadığı ( $p=1,000>0,050$ ) görülmektedir.

Elde edilen bulgulardan hareketle, matematik testinde yer alan maddeler üzerinden hesaplanan MİSGİD değerlerinin araştırma kapsamında ele alınan tüm örneklem büyüklükleri için aynı evrenden seçilen farklı örneklemeler arasında tutarlılık gösterdiği yorumuna gidilebilir. Örneklem büyüklüğüne bağlı olarak korelasyonlarda meydana gelen değişim, ortanca değerler üzerinden görselleştirilerek Şekil 9'da sunulmuştur.



Şekil 9. Matematik testi MİSGİD değerleri arasındaki korelasyon katsayılarının ortanca değerlerinin örneklem büyüklüğü bağlamında karşılaştırılması.

Şekil 9 incelendiğinde, korelasyon katsayılarının ortancalarının örneklem büyüklüğüne bağlı olarak yaklaşık 0,8 ile 1 değerleri arasında değişim gösterdiği belirlenmiştir. Örneklemelerde bulunan kişi sayısının artmasıyla birlikte MİSGİD

değerleri arasındaki korelasyon katsayılarının da artış gösterdiği ve en yüksek korelasyon düzeylerine 600'er kişilik örneklemelerde ulaşıldığı görülmüştür. Bununla birlikte, ele alınan tüm örneklem büyüklükleri için hesaplanan ortanca değerler MİSGİD'ler arasındaki korelasyonların yüksek düzeyde olduğuna işaret etmektedir. Elde edilen bulgular Türkçe testi üzerinden elde edilen bulgularla benzerlik göstermektedir.

Her iki test için de 50'şer kişilik örneklemelerden elde edilen MİSGİD değerleri arasında beklenenin altında korelasyon değerleriyle karşılaşılsa da katsayı değerleri genel olarak pozitif ve yüksek düzeyde ilişkilere (tutarlılığa) işaret etmektedir. Çüm, Gelbal ve Tsai (2016), simülasyon yoluyla ürettikleri veriler üzerinden yaptıkları çalışmada daha küçük (30'ar kişilik) örneklemeler üzerinden MİSGİD değerleri arasındaki ilişkileri incelemişler ve buna rağmen bu çalışmada incelenen 50'şer kişilik örneklemeler arasındaki korelasyon değerlerinden daha yüksek düzeyde değerler elde etmişlerdir. Söz konusu durumun bu araştırmada kullanılan testlerde yer alan maddelerin güçlük düzeyleri bakımından daha homojen bir dağılım göstermesi nedeniyle ortaya çıktığı düşünülmektedir. Bununla birlikte, farklı örneklemelerden elde edilen MİSGİD değerlerinin tutarlılığına yönelik elde edilen kanıtlar bakımından her iki araştırmadan elde edilen bulguların benzerlik gösterdiği belirtilebilir.

### **Araştırmanın İkinci Alt Problemine İlişkin Bulgular ve Yorumlar**

Bu bölümde araştırmanın ikinci alt problemi olan "Alt ve üst gruptan elde edilen maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) arasındaki korelasyonlar ne düzeydedir?" Sorusuna yanıt aranmıştır. Türkçe testi için alt ve üst gruptan elde edilen MİSGİD değerleri arasında 0,824 düzeyinde; matematik testi için alt ve üst gruptan elde edilen MİSGİD değerleri arasında ise 0,771 düzeyinde korelasyon olduğu belirlenmiştir. Matematik testi için daha düşük korelasyon elde edilmesinin nedeni test puanları dağılımının çarpıklığından (Şekil 7, s.37) kaynaklı olabilir. Bununla birlikte, elde edilen bulguların uç yetenek düzeylerinde yanıtlayıcıların bulunduğu gruptan elde edilen MİSGİD değerlerinin değişimlerinin tutarlılık gösterdiğine işaret ettiği belirtilebilir.

## Araştırmanın Üçüncü Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın üçüncü alt problemi olan “ÖBBS (2005) Türkçe ve matematik alt testlerinde yer alan maddeler için 50’şer kişilik 10, 100’er kişilik 10, 200’er kişilik 10 ve 600’er kişilik 10 farklı örneklem üzerinden hesaplanan madde uyarı indeksleri (MUI) aracılığı ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?” sorusuna ilişkin bulgular, önce Türkçe sonra matematik testi için aşağıda sunulmaktadır. Sınıflandırmalar arasındaki uyum, yüzde oranlar aracılığı ile incelenmiştir. Türkçe test maddeleri için yapılan sınıflandırmaların 50’şer kişilik örneklem arasında gösterdiği uyuma ilişkin bulgular Tablo 20’de sunulmuştur.

Tablo 20

### *Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 50’şer Kişilik Örneklem Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	A'	A'	A	A	A	A	A	A'	A'	%60
2	A'	B	A	A	A	B'	A	A'	A'	B	%40
3	A	A	A	A	A	A	A	A	A	A	%100
4	A	A	A	A	A	A	A'	A	A	A	%90
5	B	B	B'	B'	A	A	A	A'	A	A	%50
6	A	A	A	A	A	B	A	A	A	A	%90
7	A	A	A	A	A	A	A	A	A	A	%100
8	A	A'	A	A	A	A'	A	A'	A	A	%70
9	A	A	A	A	A	A	A	A'	A	A	%90
10	A	A	A	A	A	A	A	A	A	A	%100
11	A'	B	B	A'	A	B	A	A'	A'	A	%40
12	A	B'	A'	B	B	B	A	B	B'	B	%50
13	A	A	A	A	A	B'	A	A'	A	B'	%70
14	A	A	A	A	A	A	A	A	A	A	%100
15	B'	B	B'	B'	B'	B'	B'	B'	B'	B	%80
16	A	B	B	B	A'	B	B	A	B	B	%70
17	B	B	B'	B'	B'	B'	B	B	B	B	%60
18	A	B	B	A	A'	B	B'	B	A'	A	%40
19	A	A	B'	A	A	A	B	B	A	A	%70
20	A	A	A	A	A	A	A	A	A	A	%100
21	A	A	A	A	A	A	A	A	A	A	%100
22	B'	B'	B	B	A	B'	A	B'	B	B'	%50
23	B	A	A	B'	A	A	A	A'	A'	A'	%50
24	A	A	A'	A	A	A'	A	A	A	A'	%70
25	A'	B	A'	A'	A'	B	A	A	A	A	%40

Tablo 20 incelendiğinde, 50’şer kişilik örneklemde Türkçe testi maddeleri için yapılan sınıflandırmaların bazı maddeler için önemli oranda değişkenlik gösterebildiği anlaşılmaktadır. Buna karşın, 3, 7, 10, 14, 20 ve 21. maddelerin tüm örneklemde aynı şekilde sınıflandırıldığı, başka bir ifadeyle %100 uyum gösterdiği ve 4, 6, 8, 9, 13, 15, 16, 19 ve 24. maddeler için yapılan

sınıflandırmaların da yüksek oranda uyum gösterdiği belirlenmiştir (uyum yüzdeleri için modu en yüksek olan sınıf baz alınmıştır).

Türkçe test maddeleri için yapılan sınıflandırmaların 100'er kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 21'de sunulmuştur.

Tablo 21

*Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 100'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	A	A	A'	A'	A'	A'	A	A'	A	%50
2	A	A'	B	A	B	A	A	A	A	A	%70
3	A	A	A	A	A	A	A	A	A	A	%100
4	A	A	A	A	A	A	A'	A	A'	A	%80
5	A	A	B	A	B	B	A'	A	A	B	%50
6	A	A	A	A	A	A	A	A	A	A	%100
7	A	A	A	A	A	A	A	A	A	A	%100
8	A	A	A'	A	A'	A	A	A	A	A'	%70
9	A	A	A	A	A	A	A	A	A	A'	%90
10	A	A	A	A	A	A	A	A	A	A	%100
11	A	A'	A'	A	B'	A	A'	A	A	B	%50
12	B	B	B	B	B	B	B	A	B	B	%90
13	A	A	A	A	A	A	A	A	A	B	%90
14	A	A	A	A	A	A	A	A	A	A	%100
15	B'	B'	B	B'	B	B'	B'	B'	B'	B'	%80
16	B	B	B	B	B'	B	B	B	B	B'	%80
17	B'	B'	B'	B'	B'	B'	B'	B'	A'	B'	%90
18	A'	A	B'	A'	B'	A	B	A	A	B	%40
19	A	A	A	A	A	A'	A	A'	A'	A	%70
20	A	A	A	A	A	A	A	A	A	A	%100
21	A	A	A	A	A	A	A	A	A	A	%100
22	B'	B	B'	B	B'	B	B	B	B	B	%70
23	A	B	A	B	A	A	A	A	A'	A	%70
24	A	A	A	A	A	A	A	A	A	A	%100
25	A	A	A'	A	A'	A	A	A	A	A'	%70

Tablo 21 incelendiğinde, örneklemelerde yer alan kişi sayısının artmasıyla birlikte maddeler için yapılan sınıflandırmalar arasındaki uyumun da önemli oranda arttığı görülmektedir. Bu bağlamda, 3, 6, 7, 10, 14, 20, 21 ve 24. maddelerin tüm örneklemelerde aynı şekilde sınıflandırıldığı (%100 uyum) ve 1, 5, 11 ve 17. maddeler haricindeki tüm maddelerin %70'in üzerinde uyumla sınıflandırıldığı belirlenmiştir. Elde edilen bulgudan hareketle, Türkçe testi maddeleri için yapılan sınıflandırmaların 100'er kişilik örneklemeler arasında genel olarak yüksek bir uyum gösterdiği yorumuna gidilebilir.

Maddelerin 200'er kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 22'de sunulmuştur.



Tablo 22

*Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 200'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	A'	A'	A	A'	A'	A	A	A	A'	%50
2	A	A	A	A	A	A	A	A	A	A	%100
3	A	A	A	A	A'	A	A	A	A	A	%90
4	A'	A'	A'	A	A	A	A	A'	A	A'	%50
5	A	A	A	A	A	A	A	A	A	A	%100
6	A	A	A	A	A	A	A	A	A	A	%100
7	A	A	A	A	A	A	A	A	A	A	%100
8	A	A	A	A	A	A	A	A	A	A	%100
9	A	A	A	A	A	A	A	A	A	A	%100
10	A	A	A	A	A	A	A	A	A	A	%100
11	A	A	A	A	A	A'	A	A	A	A	%90
12	A	B	B'	B	B'	B	A'	B	B	B	%60
13	A'	A	A	A	A	A	A	A	A	A	%90
14	A	A	A	A	A	A	A	A	A	A	%100
15	B	B'	B'	B'	B	B'	B'	B'	B'	B	%70
16	B	B	B	B	B	B'	B	B	B	B	%90
17	B'	B'	B'	B'	B'	B'	B'	A'	B'	B'	%90
18	A	A	A	A	B	A	A	A'	A	A	%80
19	A	A	A	B	A	A	A	A	A'	A	%80
20	A	A	A	A	A	A	A	A	A	A	%100
21	A	A	A	A	A	A	A	A	A	A	%100
22	B	B	B	B'	B	B'	B	B	B	B	%80
23	A	A'	A	A	A	A	A	A	A	A	%90
24	A	A	A	A	A	A	A	A	A	A	%100
25	A'	A	A	A	A	A	A'	A	A	A	%80

Tablo 22 incelendiğinde, 2, 5, 6, 7, 8, 9, 10, 14, 20, 21 ve 24. maddelerin tüm örneklemelerde aynı şekilde sınıflandırıldığı (%100 uyum) belirlenmiştir. Ayrıca üç madde haricinde (1, 4 ve 12) diğer tüm maddeler için yapılan sınıflandırmaların da yüksek oranlarda uyum gösterdiği bulgusuna ulaşılmıştır.

Maddelerin 600'er kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 23'te sunulmuştur.

Tablo 23

*Türkçe Testi Maddeleri için Yapılan Sınıflandırmaların 600'er Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular*

<b>Maddeler</b>	<b>Örn1</b>	<b>Örn2</b>	<b>Örn3</b>	<b>Örn4</b>	<b>Örn5</b>	<b>Örn6</b>	<b>Örn7</b>	<b>Örn8</b>	<b>Örn9</b>	<b>Örn10</b>	<b>Uyum</b>
1	A'	A'	A'	A	A	A'	A'	A'	A'	A'	%80
2	A	A	A	A	A	A	A	A	A	A	%100
3	A	A	A	A	A	A	A	A	A	A	%100
4	A	A'	A	A	A	A	A	A'	A'	A	%70
5	A	A'	A	A	A	A	A	B	A	A	%80
6	A	A	A	A	A	A	A	A	A	A	%100
7	A	A	A	A	A	A	A	A	A	A	%100
8	A	A	A	A	A	A	A	A	A	A	%100
9	A	A	A	A	A	A	A	A	A	A	%100
10	A	A	A	A	A	A	A	A	A	A	%100
11	B	A	A	A	A	A	A'	B	A	A	%70
12	B	A	B	B	B	B	B	B	B	A	%80
13	A	A	A	A	A'	A	A	A	A	A	%90
14	A	A	A	A	A	A	A	A	A	A	%100
15	B'	B'	B	B'	B'	B'	B'	B'	B'	B'	%90
16	B	B	B	B	B	B	B	B	B	B	%100
17	B'	B'	B'	B'	B'	B'	B'	B	B'	B'	%90
18	A	A	A	A	A	A	A	A	A	A	%100
19	A	A	A	A	A	A	A	A	A	A	%100
20	A	A	A	A	A	A	A	A	A	A	%100
21	A	A	A	A	A	A	A	A	A	A	%100
22	B	B	B	B	B	B	B	B	B	B	%100
23	A	A	A	A	A	A	A	A	A	A	%100
24	A	A	A	A	A	A	A	A	A	A	%100
25	A	A	A	A	A	A	A	A'	A	A'	%80

Tablo 23 incelendiğinde, 2, 3, 5, 6, 7, 8, 9, 14, 16, 18, 19, 20, 21, 22, 23 ve 24. maddeler olmak üzere toplamda 16 maddenin tüm örneklemelerde aynı şekilde sınıflandırıldığı (%100 uyum) ve bununla birlikte, testte yer alan tüm maddelerin %70'in üzerinde uyum yüzdeleriyle sınıflandırıldığı belirlenmiştir.

Matematik testi maddeleri için yapılan sınıflandırmaların 50'şer kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 24'te sunulmuştur.

Tablo 24

*Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 50'şer Kişilik Örneklemeler Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	B	A	A	B	A	B	A	A	B	A	%60
2	A	B'	A'	A	B	A	B	A	B	B	%40
3	A	B	B'	A'	B'	B'	B	B	B	B	%50
4	B	B	A	A	B	B	B	B	A	A	%60
5	A	A	A	A	B'	B'	A	B	B	B	%50
6	A	A	A'	A'	A'	A	A'	A	A'	A'	%60
7	A	B'	A'	B	B	A'	A	B'	B'	B'	%40
8	A	A	A	A	A	A	A	B	A	A'	%80
9	A	A	A	A	A	A	A'	A	A'	A	%80
10	B'	B	B'	B	B	B	B	B	A'	B	%70
11	B'	A	B	A	A'	B	B	B	B	B	%60
12	B	B	B	B	B	B'	B	B	B	B	%90
13	B	B'	B	B	B'	B'	B	B	B'	B	%60
14	B'	B	A'	B'	B'	B'	B	B	B	B'	%50
15	A'	A	A	A	A	A'	A	B	A	A	%70
16	B	B'	B	B	B'	B	B	B'	B	B	%70
17	B'	A	B	B'	B	B'	B	B	B	B'	%50
18	B	B	B	B	B	B	B	B	B	B	%100
19	B'	B	B	B	B'	A	B	B'	B	B'	%50
20	A	A	A	A	A	A	A	B	A'	A	%80
21	B	A	B	A	A	A	A	B	A'	B	%50
22	B	A'	B	B'	B'	A	A	B	B	B'	%40
23	B'	B'	B	B'	B'	B	B'	B	B	B'	%60
24	B	B	B	B	B	B	B	B'	B'	B	%80
25	B'	B	B	B	B	B	B'	B'	B'	B'	%50

Tablo 24 incelendiğinde, matematik testinde yer alan 8, 9, 10, 12, 15, 16, 18, 20 ve 24. maddelerin 50'şer kişilik örneklemelerde %70'in üzerinde bir uyum yüzdesiyle sınıflandırıldığı belirlenmiştir. Diğer maddeler için istenilen düzeyde bir sınıflama tutarlılığı elde edilemediği ifade edilebilir. Söz konusu bulgudan hareketle, matematik testi maddeleri için 50'şer kişilik örneklemeler üzerinden yapılan sınıflandırmaların yanıltıcı olabilecek düzeyde değişkenlik gösterdiği yorumuna gidilebilir. Örneğin, yedinci madde farklı örneklemeler üzerinden yapılan sınıflamalarda A, A', B ve B' sınıflarının tümüne atanmıştır. Söz konusu madde için birinci örnekleme ilişkin analizler baz alındığında "Düzgün çalışan bir madde" yorumu yapılabilirken; ikinci örnekleme ilişkin analizler baz alındığında "Sorunlu bir madde" olduğu düşünülebilir.

Maddelerin 100'er kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 25'te sunulmuştur.

Tablo 25

*Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 100'er Kişilik Örneklem Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	A	A	B	B	B	A	A	B	A	%60
2	A	A	B	A	B	A	B	B	B'	B'	%40
3	B	B	B	B	B	B	B	B	B	B	%100
4	A	A	A	B	A'	B'	A'	A'	A	B	%40
5	A	A'	A	A	B	B	B	B'	A'	B	%40
6	A	A	A	A	A'	A	A'	A'	A	A'	%60
7	B'	B	B'	B'	A'	B'	B'	B'	B'	B'	%80
8	A	A	A	A	A	B	A	A	A	A	%90
9	A	A	A	A	A	A	A'	A	A	A	%90
10	B	B	B	B'	B'	B	B'	B'	A	B'	%50
11	A	B	A	B	A	A	B	B'	A	B	%50
12	B	B	B	B	B	B	B	B	B	B	%100
13	B'	B	B	B'	B'	B	B'	B	B'	B'	%60
14	B	B'	B'	B	B	B	B'	B'	B'	B'	%60
15	A'	A	A	A	B	B	B	A	A	A'	%50
16	B'	B	B	B	B'	B	B	B	B	B	%80
17	B	B'	B	A	B	B	B	B	B	B'	%70
18	B	B	B	B	B	B	B	B	B	B	%100
19	B	B	B	B	B	B	B'	B	B	B'	%80
20	B	A	A	A	A	A	B	B	A	B	%60
21	A	A'	A	A	A	A	B	B	A	A	%70
22	B'	B'	B'	B	B'	B	B'	B	B'	B'	%70
23	B	B'	B'	B	B'	B	B'	B'	B'	B'	%70
24	B	B	B	B	B	B	B	B	B	B	%100
25	B	B	B'	B'	B'	B	B'	B'	B	B'	%60

Tablo 25 incelendiğinde, 3, 12, 18 ve 24. maddelerin tüm örneklerde aynı şekilde sınıflandırıldığı (%100 uyum) belirlenmiştir. Söz edilen maddeler de dâhil olmak üzere toplamda 13 madde %70'in üzerinde bir uyum yüzdesiyle sınıflandırılmıştır. Elde edilen bulgudan hareketle, örnekte yer alan kişi sayısının artmasının sınıflandırma tutarlılığına olumlu yönde bir etkisinin olduğu çıkarımı yapılabilir.

Maddelerin 200'er kişilik örnekler arasında gösterdiği uyuma ilişkin bulgular Tablo 26'da sunulmuştur.

Tablo 26

*Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 200'er Kişilik Örneklem Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	B	A	A	A	B	A	B	B	A	%60
2	A	B'	B	A	B	A'	A	A	B	B	%40
3	B	B	B	B	B	B'	B	B	B	B	%90
4	B	A	B'	A'	A	B	B'	A	A'	A	%40
5	B'	B	A	B	A	A	B	B	A	A	%50
6	A	A	A	A	A	A'	A	A'	A'	A	%70
7	A'	B'	B'	B'	B'	B'	B'	B'	A	B'	%80
8	A	A	A	A	A	A	A	A	A	A	%100
9	A	A	A	A	A	A	A	A	A	A	%100
10	B'	B'	B'	B	B'	B	B'	B	B	B'	%60
11	A	A	B	B	A	B	B	A	A	A	%60
12	B	B	B	B	B	B	B	B'	B	B	%90
13	B'	B'	B	B'	B	B'	B'	B'	B'	B'	%80
14	B'	B	B	B	B'	B'	B'	B'	B'	B	%60
15	A	A	A	A	A	A	A	A	A	A	%100
16	B'	B	B'	B	B	B	B	B'	B'	B	%60
17	B'	B	B	B	B	B	B'	B	B	B	%80
18	B	B	B	B	B	B	B	B	B	B	%100
19	B	B	B	B	B	B	B	B	B	B	%100
20	A	B	A	A	B	A	A	A	A	A	%80
21	B	A	A	A	A	B	A	A	A	A	%80
22	B'	B'	B	A'	B'	B	B	B'	B'	B'	%60
23	B'	B'	B'	B'	B'	B	B'	B	B'	B	%70
24	B	B	B	B	B	B	B	B	B	B	%100
25	B'	B'	B	B	B	B'	B	B	B	B'	%60

Tablo 26 incelendiğinde, 8, 9, 15, 18, 19 ve 24. maddeler için yapılan sınıflandırmaların 200'er kişilik örneklemelerin tümünde aynı olduğu (%100 uyum) görülmektedir. Ayrıca, sözü edilen maddeler de dâhil olmak üzere toplamda 15 maddenin %70'in üzerinde uyumla sınıflandırıldığı belirlenmiştir. Örneklemlerde yer alan kişi sayısının artmasıyla birlikte yapılan sınıflandırmaların uyum yüzdelerinin de arttığı ifade edilebilir.

Maddelerin 600'er kişilik örneklemeler arasında gösterdiği uyuma ilişkin bulgular Tablo 27'de sunulmuştur.

Tablo 27

*Matematik Testi Maddeleri için Yapılan Sınıflandırmaların 600'er Kişilik Örneklem Arasında Gösterdiği Uyuma İlişkin Bulgular*

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	A	A	A	A	A	A	A	A	A	A	%100
2	A	A	B	A	B	A	A	A	A	A	%80
3	B	B	B	B	B	B	B	B	B	B	%100
4	B	A	B	A	A'	A	A'	B	A	B	%40
5	A	B	A	B	A	B	B	B	A	B	%60
6	A	A	A'	A	A	A'	A	A	A	A	%80
7	B'	B'	B'	B'	B'	B'	B'	B	B'	B'	%90
8	A	A	A	A	A	A	A	A	A	A	%100
9	A	A	A	A	A	A'	A	A	A	A	%90
10	B'	B'	B'	B'	B'	B'	B	B'	B'	B'	%90
11	B	A	B	B	B	B	B	B	B	B	%90
12	B	B	B	B	B	B	B	B	B	B	%100
13	B	B'	B'	B'	B	B	B'	B'	B	B	%50
14	B	B'	B	B'	B	B	B	B	B'	B	%70
15	A	A	A	A	A	A	A	A	A	A	%100
16	B	B	B	B'	B'	B	B'	B	B'	B'	%50
17	B	B	B	B'	B	B	B	B	B	B	%90
18	B	B	B	B	B	B	B	B	B	B	%100
19	B	B	B	B	B	B	B	B	B	B	%100
20	B	A	A	A	B	A	A	A	A	A	%80
21	A	A	A	B	A	A	B	A	A	A	%80
22	B	B'	B	B	B'	B'	B'	B	B	B	%60
23	B'	B'	B'	B'	B'	B	B'	B	B'	B'	%80
24	B	B	B	B	B	B	B	B	B	B	%100
25	B	B	B	B	B	B	B'	B	B	B	%90

Tablo 27 incelendiğinde, matematik testi maddeleri için 600'er kişilik örneklem üzerinden yapılan sınıflandırmaların genel olarak birbirleriyle yüksek oranda uyum gösterdiği görülmektedir. Testte yer alan 20 madde %70'in üzerinde uyumla sınıflandırılmıştır. Elde edilen bulgulara dayalı olarak hem Türkçe hem de matematik testinde yer alan maddeler için yapılan sınıflandırmalar arası değişkenliğin örneklem büyümesiyle birlikte azaldığı ve en yüksek uyum yüzdelerinin 600'er kişilik örneklem karşılaştırılması sonucu elde edildiği ifade edilebilir.

### **Araştırmanın Dördüncü Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın dördüncü alt problemi olan "Klasik Test Kuramı'nın madde ayırıcılık indeksleri (Çift Serili Korelasyon Katsayıları) ile Sato Test Kuramı'nın madde uyarı indeksleri (MUI) arasında ne düzeyde korelasyon bulunmaktadır?" sorusuna yanıt bulmak amacıyla öncelikle Türkçe testi maddeleri için 50,100,200 ve 600'er kişilik örneklem üzerinden  $r_{jx}$  ve MUI değerleri elde edilmiş ve her bir

örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen değerler arasındaki korelasyon katsayıları hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 28’de sunulmuştur.

Tablo 28

*Türkçe Testi Madde Ayırıcılık İndeksleri ile Madde Uyarı İndeksleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MUI-r<sub>jx</sub>)</sub>	-0,956*	-0,944*	-0,954*	-0,937*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 28 incelendiğinde, ele alınan tüm örneklem büyüklükleri için KTK’nın madde ayırıcılık indeksleri ile STK’nın madde uyarı indeksleri arasında negatif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Korelasyonun negatif yönde olmasının nedeni, maddelerin ayırt ediciliğine ilişkin bilgi veren madde uyarı indeksleri’nin yüksek değerlerinin (madde hakkında yüksek düzeyde uyarı) düşük ayırt ediciliğe işaret etmesinden kaynaklıdır. Elde edilen yüksek korelasyon katsayılarından hareketle, Türkçe testinde yer alan maddelerin ayırt edicilik düzeylerinin belirlenebilmesi bakımından, madde ayırıcılık ve madde uyarı indekslerinin benzer şekilde çalıştığı görüşü öne sürülebilir.

Benzer şekilde matematik testi maddeleri için 50, 100, 200 ve 600’er kişilik örneklem üzerinden  $r_{jx}$  ve MUI değerleri elde edilmiş ve her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen değerler arasındaki korelasyon katsayıları hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 29’da sunulmuştur.

Tablo 29

*Matematik Testi Madde Ayırıcılık İndeksleri ile Madde Uyarı İndeksleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MUI-r<sub>jx</sub>)</sub>	-0,968*	-0,934*	-0,944*	-0,934*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 29 incelendiğinde, ele alınan tüm örneklem büyüklükleri için KTK'nın madde ayırıcılık indeksleri ile STK'nın madde uyarı indeksleri arasında negatif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Elde edilen yüksek korelasyon katsayılarından hareketle, matematik testinde yer alan maddelerin ayırt edicilik düzeylerinin belirlenebilmesi bakımından, madde ayırıcılık ve madde uyarı indekslerinin benzer şekilde çalıştığı görüşü öne sürülebilir.

Tablo 28 ve 29'dan görüldüğü üzere, Türkçe ve matematik testlerinde yer alan maddelerin  $r_{jx}$  ve MUI değerleri arasındaki ilişki düzeylerinin örneklem büyüklüğünden etkilendiğine dair bir bulguya ulaşılmamıştır.

### **Araştırmanın Beşinci Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın beşinci alt problemi olan “Klasik Test Kuramı'nın madde güçlük indeksleri ile Sato Test Kuramı'nın maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) arasında ne düzeyde korelasyon bulunmaktadır?” sorusuna yanıt bulmak amacıyla öncelikle Türkçe testi maddeleri için 50, 100, 200 ve 600'er kişilik örneklemeler üzerinden  $p$  ve MİSGİD değerleri elde edilmiştir. Söz konusu değerler arasındaki korelasyon katsayıları her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 30'da sunulmuştur.

Tablo 30

#### *Türkçe Testi Madde Güçlük İndeksleri ile Maddeler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MİSGİD-p)</sub>	1,000*	0,999*	0,999*	0,999*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elden edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 30 incelendiğinde, ele alınan tüm örneklem büyüklükleri için KTK'nın madde güçlük indeksleri ile STK'nın maddeler için sınırlandırılmış gri ilişki dereceleri arasında pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların örneklem büyüklüğünden etkilendiğine dair bir bulguya ulaşılmamıştır. Elde edilen yüksek korelasyon katsayılarından



hareketle, Türkçe testinde yer alan maddelerin güçlük düzeylerinin belirlenebilmesi bakımından, madde güçlük indekslerinin ve maddeler için sınırlandırılmış gri ilişki derecelerinin benzer şekilde çalıştığı görüşü öne sürülebilir.

Benzer şekilde, matematik testi maddeleri için 50, 100, 200 ve 600'er kişilik örneklemeler üzerinden  $p$  ve MİSGİD değerleri elde edilmiş ve her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen değerler arasındaki korelasyon katsayıları hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 31'de sunulmuştur.

Tablo 31

*Matematik Testi Madde Güçlük İndeksleri ile Maddeler için Sınırlandırılmış Gri İlişki Dereceleeri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MİSGİD-p)</sub>	1,000*	0,998*	0,998*	0,999*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 31 incelendiğinde, ele alınan tüm örneklem büyüklükleri için KTK'nın madde güçlük indeksleri ile STK'nın maddeler için sınırlandırılmış gri ilişki dereceleri arasında pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların örneklem büyüklüğünden etkilendiğine dair bir bulguya ulaşılmamıştır. Elde edilen yüksek korelasyon katsayılarından hareketle, matematik testinde yer alan maddelerin güçlük düzeylerinin belirlenebilmesi bakımından, madde güçlük indekslerinin ve maddeler için sınırlandırılmış gri ilişki derecelerinin benzer şekilde çalıştığı görüşü öne sürülebilir.

### **Araştırmanın Altıncı Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın altıncı alt problemi olan "Klasik Test Kuramı'nın öğrenci başarı puanları (toplam test puanları) ile Sato Test Kuramı'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri (ÖİSGİD) arasında ne düzeyde korelasyon bulunmaktadır?" sorusuna yanıt bulmak amacıyla öncelikle Türkçe testini yanıtlayan ve 50, 100, 200 ve 600'er kişilik örneklemeler içerisinde yer alan

öğrencilerin test toplam puanları ve öğrenciler için sınırlandırılmış gri ilişki dereceleri elde edilmiştir. Söz konusu değerler arasındaki korelasyon katsayıları her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 32’de sunulmuştur.

Tablo 32

*Türkçe Testi Toplam Puanları ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> (ÖİSGİD–TTP)	0,977*	0,978*	0,972*	0,972*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elden edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 32 incelendiğinde, ele alınan tüm örneklem büyüklükleri için test toplam puanları ile öğrenciler için sınırlandırılmış gri ilişki dereceleri arasında pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların örneklem büyüklüğünden etkilendiğine dair bir bulguya ulaşılmamıştır. Elde edilen yüksek korelasyon katsayılarından hareketle, bireylerin Türkçe testinde gösterdikleri performansların belirlenmesi noktasında ÖİSGİD ve test toplam puanlarının benzerlik gösterdiği ifade edilebilir.

Benzer şekilde, matematik testini yanıtlayan ve 50, 100, 200 ve 600’er kişilik örneklem içerisnde yer alan bireylerin test toplam puanları ve öğrenciler için sınırlandırılmış gri ilişki dereceleri elde edilmiştir. Söz konusu değerler arasındaki korelasyon katsayıları her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden hesaplanarak ortancaları alınmıştır. Konuya ilişkin bulgular Tablo 33’te sunulmuştur.

Tablo 33

*Matematik Testi Toplam Puanları ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> (ÖİSGİD–TTP)	0,979*	0,977*	0,978*	0,976*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elden edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 33 incelendiğinde, ele alınan tüm örneklem büyüklükleri için test toplam puanları ile öğrenciler için sınırlandırılmış gri ilişki dereceleri arasında pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların örneklem büyüklüğünden etkilendiğine dair bir bulguya ulaşılmamıştır. Elde edilen yüksek korelasyon katsayılarından hareketle, bireylerin matematik testinde gösterdikleri performansların belirlenmesi noktasında ÖİSGİD ve test toplam puanlarının benzerlik gösterdiği ifade edilebilir. Araştırmanın dördüncü, beşinci ve altıncı alt problemine ilişkin bulgular değerlendirildiğinde STK ve KTK'nın psikometrik özellikler bakımından birbirine yakın sonuçlar ortaya koyduğu görülmüştür.

### **Araştırmanın Yedinci Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın yedinci alt problemi olan “Sato Test Kuramı'nın madde uyarı indeksleri (MUI) ile Madde Tepki Kuramı'nın  $a$  parametreleri arasında ne düzeyde ilişki bulunmaktadır?” Sorusuna yanıt bulmak amacıyla Türkçe testi maddeleri için 50, 100, 200 ve 600'er kişilik 10'ar farklı örneklem üzerinden elde edilen MUI değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $a$  parametreleri arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 34'te sunulmuştur.

Tablo 34

#### *Türkçe Testi Madde Uyarı İndeksleri ile $a$ Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon<sub>(MUI-a)</sub></b>	-0,286*	-0,426*	-0,527*	-0,574*

*\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.*

Tablo 34 incelendiğinde, Türkçe testi maddeleri için elde edilen STK'nın madde uyarı indeksleri ve MTK'nın  $a$  parametreleri arasında örneklem büyüklüğünün 50 olduğu durumda negatif yönde ve düşük düzeyde örneklem büyüklüğünün 100, 200 ve 600 olduğu durumlarda ise negatif yönde ve orta düzeyde bir korelasyonun varlığından söz edilebilir. Korelasyon yönünün negatif olması yüksek madde uyarılarının düşük ayırt ediciliğe işaret etmesinden

kaynaklanmaktadır. Elde edilen bulgular, Türkçe testi maddelerinin ayırt ediciliklerinin belirlenmesi noktasında MUİ ve  $a$  parametrelerinin yüksek düzeyde bir benzerlik göstermediği şeklinde yorumlanabilir. Bununla birlikte, STK'nın Madde Uyarı İndeksleri tek başına maddelerin ayırt edicilik düzeylerinin belirlenmesi amacıyla kullanılmamaktadır. Bu değerler, “maddenin doğru yanıtlanma sayısı/teste katılan kişi sayısı” oranlarıyla birlikte maddelerin işlevliliğiyle ilgili sınıflandırılmasında kullanılmaktadır. Bu araştırma kapsamında, Türkçe testi için MUİ değerlerinin Çift Serili Korelasyon Yöntemi ile hesaplanan madde ayırtıcılık indeksleri ile yüksek düzeyde ilişki gösterdiği (Tablo 28, s.64) fakat  $a$  parametreleri ile benzer düzeyde bir ilişkinin tespit edilemediği ifade edilebilir.

Benzer şekilde, matematik testi maddeleri için 50, 100, 200 ve 600'er kişilik 10'ar farklı örneklem üzerinden elde edilen MUİ değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $a$  parametreleri arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 35'te sunulmuştur.

Tablo 35

*Matematik Testi Madde Uyarı İndeksleri ile  $a$  Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MUİ-a)</sub>	-0,082*	-0,154*	-0,138*	-0,269*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 35 incelendiğinde, matematik testi maddeleri için elde edilen STK'nın madde uyarı indeksleri ve MTK'nın  $a$  parametreleri arasında tüm örneklem büyüklükleri için negatif yönde ve düşük düzeyde bir ilişkinin bulunduğu görülmektedir. Bu bakımdan, matematik testi maddelerinin ayırt ediciliklerinin belirlenmesi noktasında MUİ ve  $a$  parametrelerinin benzer şekilde çalışmadıkları ifade edilebilir.

**Araştırmanın Sekizinci Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın sekizinci alt problemi olan “Sato Test Kuramı'nın maddeler için sınırlandırılmış gri ilişki dereceleri (MİSGİD) ile Madde Tepki Kuramı'nın  $b$

parametreleri arasında ne düzeyde ilişki bulunmaktadır?” Sorusuna yanıt bulmak amacıyla Türkçe testi maddeleri için 50, 100, 200 ve 600’er kişilik 10’ar farklı örneklem üzerinden elde edilen MİSGİD değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $b$  parametreleri arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 36’da sunulmuştur.

Tablo 36

*Türkçe Testi Maddeler için Sınırlandırılmış Gri İlişki Dereceleri ile  $b$  Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MİSGİD–<math>b</math>)</sub>	-0,854*	-0,898*	-0,911*	-0,944*

*\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.*

Tablo 36 incelendiğinde, STK’nın maddeler için sınırlandırılmış gri ilişki dereceleri ile MTK’nın  $b$  parametreleri arasında tüm örneklem büyüklükleri için negatif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Ayrıca, MİSGİD değerlerinin elde edildiği örneklem büyüklükleri arttıkça tüm katılımcıların yanıtları üzerinden elde edilen  $b$  parametreleri ile gösterdikleri korelasyon düzeylerinin de arttığı belirlenmiştir. Korelasyonların negatif yönde olmasının nedeni düşük MİSGİD değerlerinin görece daha zor; yüksek MİSGİD değerlerinin görece daha kolay maddelere işaret etmesinden kaynaklanmaktadır. Elde edilen bulgulardan hareketle, Türkçe testi maddelerinin güçlük düzeylerinin belirlenmesi noktasında MİSGİD ve  $b$  parametrelerinin benzerlik gösterdiği ifade edilebilir.

Benzer şekilde, matematik testi maddeleri için 50, 100, 200 ve 600’er kişilik 10’ar farklı örneklem üzerinden elde edilen MİSGİD değerleri ile teste katılan tüm öğrenciler üzerinden 3PLM ile elde edilen  $b$  parametreleri arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 37’de sunulmuştur.

Tablo 37

*Matematik Testi Maddeler için Sınırlandırılmış Gri İlişki Dereceleri ile b Parametreleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(MİSGİD-b)</sub>	-0,805*	-0,880*	-0,891*	-0,928*

\*Sunulan değerler her bir örneklem büyüklüğü için 10 farklı örneklem üzerinden elde edilen korelasyon katsayılarının ortancaları alınarak elde edilmiştir.

Tablo 37 incelendiğinde, STK'nın maddeler için sınırlandırılmış gri ilişki dereceleri ile MTK'nın  $b$  parametreleri arasında tüm örneklem büyüklükleri için negatif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Ayrıca, MİSGİD değerlerinin elde edildiği örneklem büyüklükleri arttıkça tüm katılımcıların yanıtları üzerinden elde edilen  $b$  parametreleri ile gösterdikleri korelasyon düzeylerinin de arttığı belirlenmiştir. Elde edilen bulgulardan hareketle, matematik testi maddelerinin güçlük düzeylerinin belirlenmesi noktasında MİSGİD ve  $b$  parametrelerinin benzerlik gösterdiği ifade edilebilir.

Bu araştırmada, madde güçlüklerinin belirlenmesi noktasında STK ve MTK ile üretilen değerlerin yüksek düzeyde ilişki gösterdiğinin ortaya koyulmuş olması bakımından elde edilen bulguların Sheu ve diğerlerinin (2014b) yaptıkları araştırmada elde ettikleri bulgularla paralellik gösterdiği ifade edilebilir. Söz konusu araştırmada bu araştırmadan farklı olarak STK ile MTK modellerinden 1PLM karşılaştırılmıştır.

### **Araştırmanın Dokuzuncu Alt Problemine İlişkin Bulgular ve Yorumlar**

Araştırmanın dokuzuncu alt problemi olan "Sato Test Kuramı'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri (ÖİSGİD) ile Madde Tepki Kuramı'nın yetenek parametreleri arasında ne düzeyde korelasyon bulunmaktadır?" Soruna yanıt bulmak amacıyla, Türkçe testini yanıtlayan ve 50, 100, 200 ve 600'er kişilik örneklem içerisnde yer alan öğrencilerin ÖİSGİD değerleri ile aynı öğrencilerin 3PLM ile kestirilen yetenek ( $\theta$ ) parametreleri ( $\theta$  kestirimleri tüm veri seti üzerinden gerçekleştirilmiştir.) arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 38'de sunulmuştur.

Tablo 38

*Türkçe Testini Yanıtlayan Öğrencilerin Yetenek Parametreleri ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(ÖİSGİD-θ)</sub>	0,971	0,973	0,977	0,976

Tablo 38 incelendiğinde, STK'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri ile MTK'nın yetenek parametreleri arasında tüm örneklem büyüklükleri için pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların, ÖİSGİD değerlerinin elde edildiği örneklem büyüklüğünden etkilendiğine dair net bir bulguya ulaşılamamıştır. Elde edilen yüksek korelasyon katsayılarından hareketle, bireylerin Türkçe testinde gösterdikleri performansların belirlenmesi noktasında ÖİSGİD ve  $\theta$  değerlerinin benzerlik gösterdiği ifade edilebilir.

Benzer şekilde, matematik testini yanıtlayan ve 50, 100, 200 ve 600'er kişilik örneklem içerisnde yer alan öğrencilerin ÖİSGİD değerleri ile aynı öğrencilerin 3PLM ile kestirilen yetenek ( $\theta$ ) parametreleri ( $\theta$  kestirimleri tüm veri seti üzerinden gerçekleştirilmiştir.) arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular Tablo 39'da sunulmuştur.

Tablo 39

*Matematik Testini Yanıtlayan Öğrencilerin Yetenek Parametreleri ile Öğrenciler için Sınırlandırılmış Gri İlişki Dereceleri Arasındaki Korelasyonlara İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
<b>Korelasyon</b> <sub>(ÖİSGİD-θ)</sub>	0,936	0,915	0,941	0,949

Tablo 39 incelendiğinde, STK'nın öğrenciler için sınırlandırılmış gri ilişki dereceleri ile MTK'nın yetenek parametreleri arasında tüm örneklem büyüklükleri için pozitif yönde ve yüksek düzeyde korelasyonların olduğu görülmektedir. Söz konusu korelasyonların, ÖİSGİD değerlerinin elde edildiği örneklem büyüklüğünden etkilendiğine dair net bir bulguya ulaşılamamıştır. Elde edilen yüksek korelasyon katsayılarından hareketle, bireylerin matematik testinde

gösterdikleri performansların belirlenmesi noktasında ÖİSGİD ve  $\theta$  değerlerinin benzerlik gösterdiği ifade edilebilir.

### Araştırmanın Onuncu Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın onuncu ve son alt problemi olan “Klasik Test Kuramı, Madde Tepki Kuramı (3PLM) ve Sato Test Kuramı’na dayalı olarak yürütülen madde analizlerinin sonuçlarından hareketle testten çıkarılması ya da düzeltilmesi gerektiğine karar verilen maddeler ne oranda uyum göstermektedir?” Sorusuna yanıt aramak amacıyla öncelikle Türkçe testi için 50, 100, 200 ve 600’er kişilik örneklemeler üzerinden Çift Serili Korelasyon Katsayısı tekniği ile madde ayırıcılık indeksleri ( $r_{jx}$ ) elde edilmiş ve Tablo 40’ta sunulmuştur.

Tablo 40

#### *Türkçe Testi Maddelerinin Ayırıcılık İndeksleri*

Madde	Örneklem	Örneklem	Örneklem	Örneklem
	Büyüklüğü 50	Büyüklüğü 100	Büyüklüğü 200	Büyüklüğü 600
	$r_{jx}$	$r_{jx}$	$r_{jx}$	$r_{jx}$
1	0,55	0,31	0,33	0,35
2	0,34	0,58	0,56	0,48
3	0,67	0,40	0,46	0,42
4	0,47	0,56	0,41	0,35
5	0,53	0,50	0,38	0,49
6	0,57	0,52	0,43	0,53
7	0,49	0,66	0,49	0,47
8	0,23	0,49	0,38	0,36
9	0,43	0,37	0,45	0,44
10	0,69	0,49	0,42	0,49
11	0,50	0,49	0,21	0,42
12	0,47	0,48	0,41	0,41
13	0,35	0,40	0,46	0,45
14	0,71	0,54	0,56	0,54
15	0,13	0,28	0,21	0,25
16	0,48	0,39	0,33	0,46
17	0,03	0,18	0,31	0,27
18	0,38	0,44	0,53	0,43
19	0,44	0,32	0,45	0,41
20	0,46	0,51	0,52	0,55
21	0,59	0,60	0,53	0,49
22	0,17	0,45	0,24	0,41
23	0,41	0,42	0,42	0,45
24	0,22	0,46	0,50	0,52
25	0,46	0,41	0,36	0,43

Tablo 40 incelendiğinde, farklı büyüklükteki örneklemelerden elde edilen  $r_{jx}$  değerlerinin aynı maddeler için önemli düzeyde farklılık gösterebildiği belirlenmiştir. Örneğin, 24. madde örneklem büyüklüğünün 50 olarak belirlendiği



durumda testten çıkarılması gereken bir madde olarak değerlendirilebilecekken ( $r_{jx50} = 0,22$ ) diğer örneklem büyüklükleri üzerinden yürütülen analizlerde daha yüksek ayırt edicilik değerleri elde etmiştir ( $r_{jx100} = 0,46$  ;  $r_{jx200} = 0,50$  ;  $r_{jx600} = 0,52$ ). Söz konusu bulguların, KTK'da madde özelliklerinin gruba bağlı olmasına ilişkin sınırlılığı kanıtlar nitelikte olduğu belirtilebilir.

KTK'ya dayalı olarak elde edilen madde ayırıcılık indeksleri baz alınarak Türkçe testinden çıkarılması ya da düzeltilmesi gerektiğine karar verilen sorunlu maddeler ( $r_{jx} < 0,30$ ) her bir örneklem büyüklüğü için Tablo 41'de sunulmuştur.

Tablo 41

Türkçe Testindeki Sorunlu Maddelerin Klasik Test Kuramı'na göre Belirlenmesine İlişkin Bulgular

Örneklem Büyüklüğü	50	100	200	600
Sorunlu Madde Kümeleri	8, 15, 17, 22, 24	15, 17	11, 15, 22	17

Tablo 41 incelendiğinde, madde ayırıcılık indeksleri baz alınarak belirlenen sorunlu maddelerin farklı örneklem büyüklüklerinde değişiklik gösterdiği belirlenmiştir. Aynı örneklemlerden alınan veriler Sato Test Kuramı'na dayalı olarak analiz edilmiş ve maddeler MUI'ler aracılığı ile sınıflandırılmıştır. Elde edilen bulgular Tablo 42'de sunulmuştur.

Tablo 42

*Türkçe Testi Maddelerinin Atandığı Sınıflar*

Madde	Örneklem Büyüküğü 50	Örneklem Büyüküğü 100	Örneklem Büyüküğü 200	Örneklem Büyüküğü 600
1	A	A'	A'	A'
2	B'	A	A	A
3	A	A	A	A
4	A	A	A	A
5	A	B	A	A
6	B	A	A	A
7	A	A	A	A
8	A'	A	A	A
9	A	A	A	A
10	A	A	A	A
11	B	A	A'	A
12	B	B	B	B
13	B'	A	A	A
14	A	A	A	A
15	B'	B'	B'	B'
16	B	B	B'	B
17	B'	B'	B'	B'
18	B	A	A	A
19	A	A'	A	A
20	A	A	A	A
21	A	A	A	A
22	B'	B	B'	B
23	A	A	A	A
24	A'	A	A	A
25	B	A	A	A

Tablo 42 incelendiğinde, madde uyarı indeksleri ve maddelerin doğru yanıtlanma oranları baz alınarak yapılan sınıflandırmalar üzerinden belirlenen bazı sorunlu maddelerin farklı örneklem büyüklüklerinde değişiklik gösterdiği belirlenmiştir.

STK'ya dayalı olarak yapılan incelemeler sonucu Türkçe testinden çıkarılması ya da düzeltilmesi gerektiğine karar verilen sorunlu maddeler (A' ve B') her bir örneklem büyüklüğü için Tablo 43'te sunulmuştur.

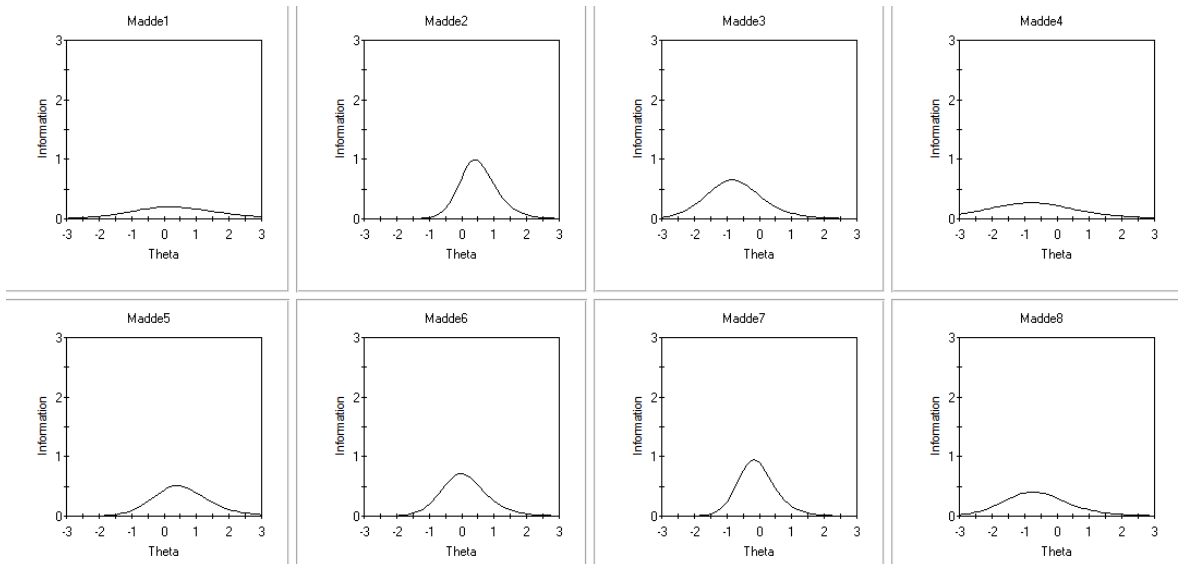
Tablo 43

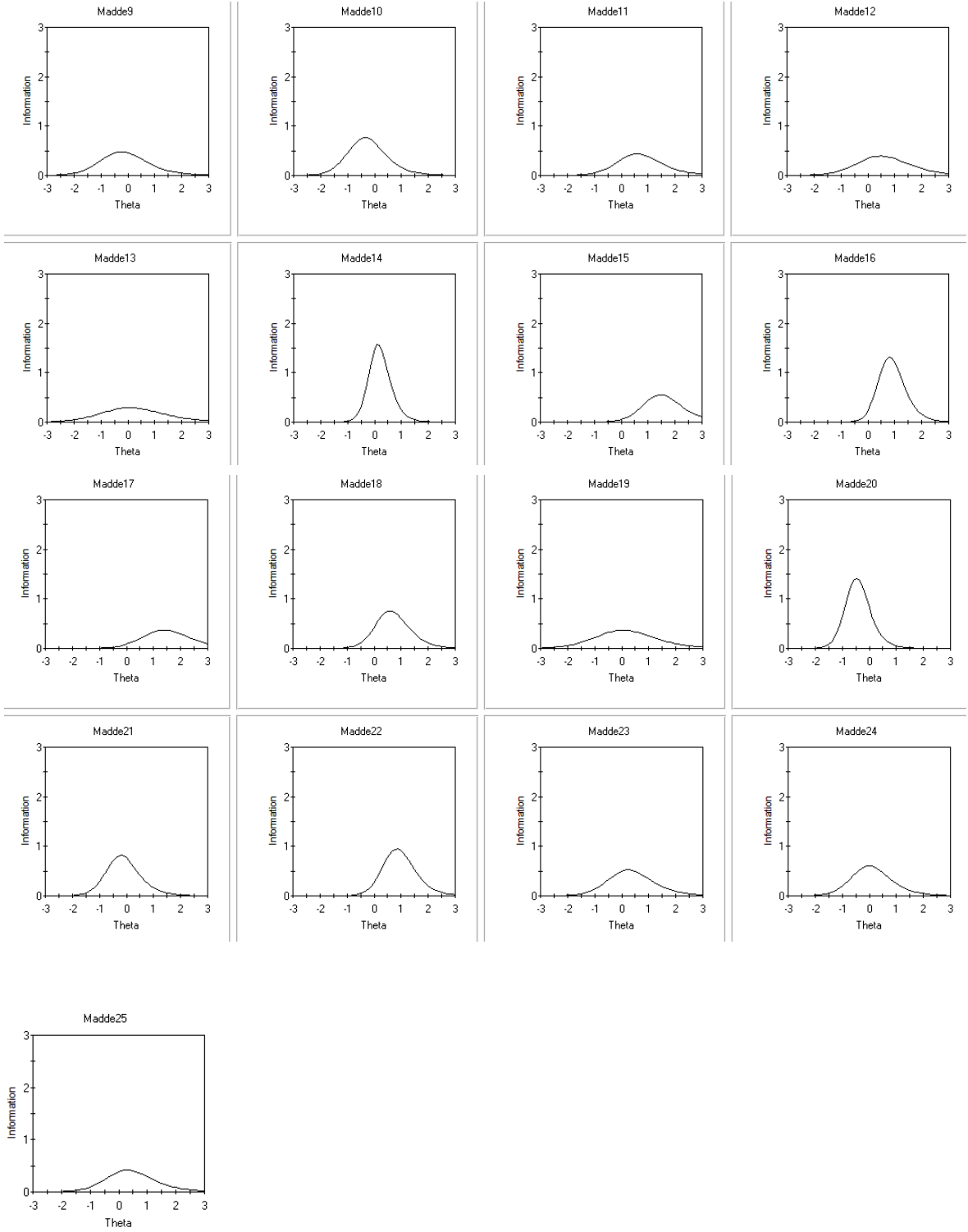
*Türkçe Testindeki Sorunlu Maddelerin Sato Test Kuramı'na göre Belirlenmesine İlişkin Bulgular*

Örneklem Büyüküğü	50	100	200	600
Sorunlu Madde Kümeleri	2, 8, 13, 15, 17, 22, 24	1, 15, 17, 19	1, 11, 15, 16, 17, 22	1, 15, 17

Tablo 43 incelendiğinde, 15 ve 17. maddelerin tüm örneklem büyüklüklerinde ve 1. maddenin 100, 200 ve 600'er kişilik örneklemelerde sorunlu olarak belirlendiği görülmektedir. STK ve KTK'ya dayalı olarak yapılan analizler sonrasında sorunlu olduğu tespit edilen madde kümeleri karşılaştırıldığında, örneklem büyüklüğü 50 iken 8, 15, 17, 22 ve 24. maddelerin; 100 iken 15 ve 17. maddelerin; 200 iken 11, 15 ve 22. maddelerin; 600 iken 17. maddenin ortak eleman olduğu belirlenmiştir. Yapılan karşılaştırmalardan hareketle, her iki kurama dayalı olarak sorunlu maddelerin tespiti amacıyla yürütülen analizlerin benzer sonuçlar ortaya koyduğu iddia edilebilir. Bununla birlikte, STK'ya dayalı olarak belirlenen sorunlu madde kümelerinin eleman sayıları, tüm örneklem büyüklüklerinde, KTK'ya dayalı belirlenen sorunlu madde kümelerinin eleman sayılarından fazladır.

Madde Tepki Kuramı'na göre Türkçe testinden çıkarılması ya da düzeltilmesi gereken sorunlu maddeler  $a$ ,  $c$  parametreleri ve madde bilgi fonksiyonları aracılığı ile belirlenmiştir. Bulgular, tüm katılımcılar üzerinden yürütülen analizler sonucu elde edilmiştir. Madde bilgi fonksiyonları Şekil 10'da sunulmuştur.





Şekil 10. Türkçe testi madde bilgi fonksiyonları.

Şekil 10 incelendiğinde, tüm yetenek (theta) değerleri için 1, 4, 8, 9, 11, 12, 13, 17, 19 ve 25. maddelerin açıkladığı bilgi miktarının tüm yetenek düzeyleri için 0.5'in altında olduğu belirlenmiştir (Türkçe testi için madde bilgi fonksiyonu değerleri EK-C'de sunulmuştur). Bu bakımdan madde bilgi fonksiyonları sözü

edilen maddelerin açıkladığı bilgi miktarının düşük olduğuna (An ve Yung, 2014) işaret etmektedir. Maddelere ilişkin  $a$  parametresi (ayırt edicilik) ve  $c$  parametresi (tahmin) değerleri Tablo 44'te sunulmuştur.

Tablo 44

*Türkçe Testi Maddelerinin  $a$  ve  $c$  Parametrelerinin Değerleri*

Madde	$a_i$	$c_i$
1	1.04	0.15
2	2.49	0.23
3	1.71	0.06
4	1.07	0.03
5	1.67	0.16
6	2.04	0.20
7	2.48	0.25
8	1.42	0.12
9	1.60	0.14
10	1.98	0.13
11	1.64	0.23
12	1.41	0.12
13	1.21	0.11
14	3.42	0.32
15	1.88	0.24
16	2.83	0.22
17	1.53	0.24
18	2.19	0.24
19	1.29	0.07
20	3.00	0.24
21	2.26	0.23
22	2.34	0.19
23	1.69	0.16
24	1.89	0.20
25	1.62	0.24

Tablo 44'te yer alan  $a$  parametrelerine ilişkin değerler dikkate alındığında, Türkçe testinde negatif ya da düşük ayırt edicilik gösteren bir maddenin olmadığı yorumu yapılabilir. Tablodaki  $c$  parametrelerine ilişkin değerler dikkate alındığında ise, 2, 7, 11, 15, 16, 17, 18, 20, 21 ve 25. maddelerin tahminle doğru yanıtlandırılma olasılığının diğer maddelerden görece daha yüksek olduğu ( $c > 0,20$ ); 14. maddenin tahminle doğru yanıtlandırılma olasılığının ise dört seçenekli bir madde için şansla doğru yanıtlandırılma olasılığından daha yüksek bir değer elde ettiği ( $0,32 > 0,25$ ) belirlenmiştir. Söz konusu bulgudan hareketle, Türkçe testinde gösterdiği performans düzeyi düşük olan öğrencilerin 14. maddeyi

normalin üzerinde bir olasılıkla doğru yanıtıma eğiliminde oldukları yönünde bir görüş belirtilebilir.

Madde bilgi fonksiyonları, ayırt edicilik ve tahmin parametreleri dikkate alındığında Türkçe testinden çıkarılması ya da düzeltilmesi gerektiği düşünülen maddeler Tablo 45'te sunulmuştur.

Tablo 45

*Türkçe Testindeki Sorunlu Maddelerin Madde Tepki Kuramı'na göre Belirlenmesine İlişkin Bulgular*

	Ayırt Edicilik Bakımından	Tahmin Olasılığı Bakımından	Açıkladığı Bilgi Bakımından
Sorunlu Madde Kümeleri	-	2, 7, 11, 14, 15, 16, 17, 18, 20, 21, 25	1, 4, 8, 9, 11, 12, 13, 17, 19, 25

Ayırt ediciliği ve taşıdığı bilgi yüksek olan bunun yanı sıra tahminle doğru yanıtlanma olasılığı çok yüksek olmayan maddeler kaliteli olarak nitelendirilebilir. Madde bilgi fonksiyonları maddelerin ayırt edicilik ve tahminle doğru yanıtlanma olasılığı gibi bilgilerinden türetilen bir fonksiyon olmasına karşın bu değerlendirmede ayrı bir kıstas olarak ele alınmış ve parametre değerlerinin bir sağlaması olarak düşünülmüştür.

MTK'ya dayalı yürütülen bu inceleme sonrasında 1, 2, 8, 11, 13, 15, 16, 17 ve 19. maddelerin sorunlu olduğuna ilişkin yapılan belirlemelerin Sato Test Kuramı ile yapılan belirlemelerle benzerlik gösterdiği (Tablo 43, s.75) sonucuna ulaşılmıştır. Bununla birlikte 8, 11, 15 ve 17. maddelerin de Klasik Test Kuramı ile yapılan incelemelerde de sorunlu olarak belirlendiği (Tablo 41, s.74) görülmüştür. STK ve MTK'ya dayalı olarak oluşturulan sorunlu madde kümelerinin ortak eleman sayısı, KTK ve MTK'ya dayalı olarak oluşturulan sorunlu madde kümelerinin ortak eleman sayısından fazladır.

Matematik testi için 50, 100, 200 ve 600'er kişilik örneklemeler üzerinden Çift Serili Korelasyon Katsayısı tekniği ile madde ayıricılık indeksleri ( $r_{jx}$ ) elde edilmiş ve Tablo 46'da sunulmuştur.

Tablo 46

*Matematik Testi Maddelerinin Ayırıcılık İndeksleri*

Madde	Örneklem	Örneklem	Örneklem	Örneklem
	Büyüküğü 50	Büyüküğü 100	Büyüküğü 200	Büyüküğü 600
	$r_{jx}$	$r_{jx}$	$r_{jx}$	$r_{jx}$
1	0,46	0,61	0,53	0,44
2	0,42	0,53	0,39	0,42
3	0,21	0,38	0,44	0,35
4	0,52	0,17	0,33	0,29
5	0,25	0,34	0,42	0,34
6	0,19	0,20	0,31	0,29
7	0,47	0,14	0,30	0,29
8	0,34	0,46	0,52	0,45
9	0,33	0,43	0,36	0,34
10	0,43	0,32	0,21	0,31
11	0,19	0,48	0,43	0,46
12	0,53	0,46	0,60	0,48
13	0,23	0,33	0,45	0,35
14	0,15	0,35	0,30	0,39
15	0,49	0,36	0,47	0,39
16	0,21	0,24	0,40	0,28
17	0,54	0,34	0,46	0,35
18	0,41	0,52	0,50	0,51
19	0,30	0,35	0,50	0,41
20	0,38	0,51	0,45	0,47
21	0,56	0,47	0,38	0,41
22	0,27	0,33	0,17	0,34
23	0,03	0,36	0,28	0,24
24	0,43	0,45	0,51	0,44
25	0,34	0,35	0,39	0,43

Tablo 46 incelendiğinde, matematik testi maddeleri için farklı büyüklükteki örneklemelerden elde edilen madde ayırıcılık indekslerinin bazı maddeler için büyük farklılıklar gösterdiği (3, 4, 7, 10, 11, 13, 16. maddeler) belirlenmiştir. Madde ayırıcılık indeksleri baz alınarak matematik testinden çıkarılması ya da düzeltilmesi gerektiğine karar verilen sorunlu maddeler ( $r_{jx} < 0,30$ ) her bir örneklem büyüklüğü için Tablo 47'de sunulmuştur.

Tablo 47

*Matematik Testindeki Sorunlu Maddelerin Klasik Test Kuramı'na göre Belirlenmesine İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
Sorunlu Madde Kümeleri	3, 5, 6, 11, 13, 14, 16, 22, 23	4, 6, 7, 16	10, 22, 23	4, 6, 7, 16, 23

Tablo 47 incelendiğinde, KTK'ya göre, farklı örneklem büyüklüklerinde en çok kez sorunlu olduğu tespit edilen maddelerin 6, 16 ve 23. maddeler olduğu söylenebilir.

Aynı örneklemelerden alınan veriler Sato Test Kuramı'na dayalı olarak analiz edilmiş ve maddeler MUI'ler aracılığı ile sınıflandırılmıştır. Elde edilen bulgular Tablo 48'de sunulmuştur.

Tablo 48

*Matematik Testi Maddelerinin Atandığı Sınıflar*

Madde	Örneklem Büyüklüğü 50	Örneklem Büyüklüğü 100	Örneklem Büyüklüğü 200	Örneklem Büyüklüğü 600
1	A	B	A	A
2	B	B	B	B
3	B'	B	B	B
4	B	A'	A	A'
5	B'	B	A	A
6	A'	A'	A	A
7	B	A'	B'	B'
8	A	A	A	A
9	A	A	A	A
10	B	B'	B'	B'
11	A'	A	A	B
12	B	B	B	B
13	B'	B'	B	B
14	B'	B	B'	B
15	A	B	A	A
16	B'	B'	B	B'
17	B	B	B	B
18	B	B	B	B
19	B'	B	B	B
20	A	A	B	B
21	A	A	A	A
22	B'	B'	B'	B'
23	B'	B'	B'	B'
24	B	B	B	B
25	B	B'	B	B



Tablo 48 incelendiğinde, madde uyarı indeksleri ve maddelerin doğru yanıtlanma oranları baz alınarak yapılan sınıflandırmalar üzerinden belirlenen bazı sorunlu maddelerin farklı örneklem büyüklüklerinde değişiklik gösterdiği belirlenmiştir.

STK'ya dayalı olarak yapılan incelemeler sonucu matematik testinden çıkarılması ya da düzeltilmesi gerektiğine karar verilen sorunlu maddeler (A' ve B') her bir örneklem büyüklüğü için Tablo 49'da sunulmuştur.

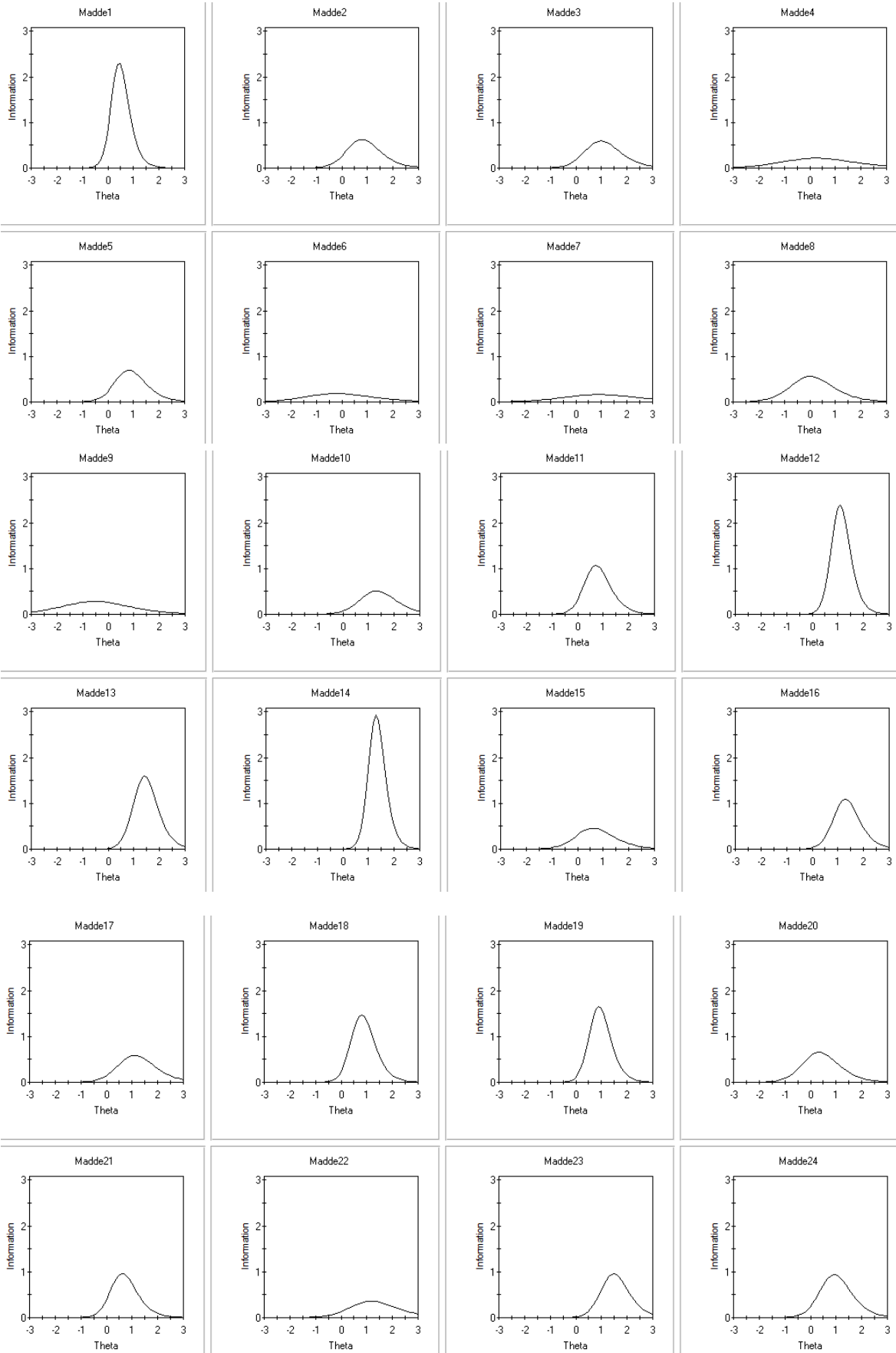
Tablo 49

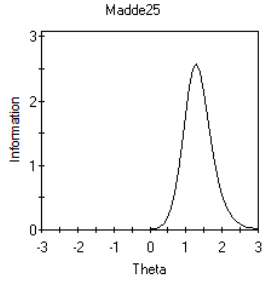
*Matematik Testindeki Sorunlu Maddelerin Sato Test Kuramı'na göre Belirlenmesine İlişkin Bulgular*

Örneklem Büyüklüğü	50	100	200	600
Sorunlu Madde Kümeleri	3, 5, 6, 11, 13, 14, 16, 19, 22, 23	4, 6, 7, 10, 13, 16, 22, 23, 25	7, 10, 14, 22, 23	4, 7, 10, 16, 22, 23

Tablo 49 incelendiğinde, 22. ve 23. maddelerin tüm örneklem büyüklüklerinde; 7. ve 10. maddelerin ise 100, 200 ve 600 kişilik örneklemelerde sorunlu olarak sınıflandırıldığı görülmektedir. STK ve KTK'ya dayalı olarak yapılan analizler sonrasında sorunlu olduğu tespit edilen madde kümeleri karşılaştırıldığında, örneklem büyüklüğü 50 iken 3, 5, 6, 11, 13, 14, 16, 22 ve 23. maddelerin; 100 iken 4, 6, 7 ve 16. maddelerin; 200 iken 10, 22 ve 23. maddelerin; 600 iken 4, 7, 16 ve 23. maddelerin ortak eleman olduğu belirlenmiştir. Yapılan karşılaştırmalardan hareketle, her iki kurama dayalı olarak sorunlu maddelerin tespiti amacıyla yürütülen analizlerin benzer sonuçlar ortaya koyduğu iddia edilebilir. Bununla birlikte, STK'ya dayalı olarak belirlenen sorunlu madde kümelerinin eleman sayıları, tüm örneklem büyüklüklerinde, KTK'ya dayalı belirlenen sorunlu madde kümelerinin eleman sayılarından fazladır.

Madde Tepki Kuramı'na göre matematik testinden çıkarılması ya da düzeltilmesi gereken sorunlu maddeler  $a$ ,  $c$  parametreleri ve madde bilgi fonksiyonları aracılığı ile belirlenmiştir. Bulgular, tüm katılımcılar üzerinden yürütülen analizler sonucu elde edilmiştir. Madde bilgi fonksiyonları Şekil 11'de sunulmuştur.





Şekil 11. Matematik testi madde bilgi fonksiyonları.

Şekil 11’de yer alan ve her bir yetenek düzeyi için maddelerin açıkladığı bilgiyi gösteren grafikler incelendiğinde 4, 6, 7, 9, 15 ve 22. Maddelerin açıkladığı bilgi miktarının tüm yetenek düzeyleri için 0.5’in altında olduğu belirlenmiştir (Matematik testi için madde bilgi fonksiyonu değerleri EK Ç’de sunulmuştur). Söz konusu maddelerin öğrencilerin yetenek kestirimlerine yapacakları katkının düşük olacağı ve kestirimlerin isabetliliğine olumsuz etki edebilecekleri durumları göz önünde bulundurulduğunda bu maddelerin testten çıkarılması ya da düzeltilmesi konusu gündeme getirilebilir. Bu duruma ilişkin diğer bulguların da incelenmesi amacıyla maddelere ilişkin  $a$  parametresi (ayırt edicilik) ve  $c$  parametresi (tahmin) değerleri elde edilerek Tablo 50’de sunulmuştur.

Tablo 50

*Matematik Testi Maddelerinin a ve c Parametrelerinin Değerleri*

Madde	$a_i$	$c_i$
1	3.81	0.24
2	2.04	0.26
3	1.93	0.23
4	0.99	0.06
5	2.18	0.28
6	1.04	0.19
7	0.92	0.13
8	1.64	0.10
9	1.09	0.03
10	1.88	0.28
11	2.63	0.25
12	3.71	0.19
13	2.91	0.15
14	4.18	0.21
15	1.77	0.28
16	2.65	0.25
17	1.86	0.20
18	2.94	0.20
19	3.25	0.24
20	1.90	0.17
21	2.57	0.28
22	1.46	0.21
23	2.48	0.25
24	2.23	0.15
25	3.80	0.18

Tablo 50'de sunulan  $a$  parametrelerine ilişkin değerler incelendiğinde matematik testi içerisinde negatif ya da düşük ayırt ediciliğe sahip herhangi bir maddenin bulunmadığı görülmektedir. Tabloda verilen  $c$  parametrelerine ilişkin değerler incelendiğinde ise 1, 3, 11, 14, 16, 19, 22 ve 23. maddelerin tahminle doğru yanıtlandırılma olasılığının diğer maddelerden görece daha yüksek olduğu ( $c > 0,20$ ) belirlenmiştir. Ayrıca, 2, 5, 10, 15 ve 21. maddelerin tahminle doğru yanıtlandırılma olasılığının dört seçenekli bir madde için şansla doğru yanıtlandırılma olasılığından daha yüksek bir değer elde ettiği ( $c > 0,25$ ) belirlenmiştir. Söz konusu maddelerin matematik performansı düşük düzeyde olan öğrenciler tarafından şanstın daha yüksek bir olasılıkla doğru yanıtlanıyor olması durumu test açısından önemli bir sorun oluşturabilir.

Madde bilgi fonksiyonları, ayırt edicilik ve tahmin parametreleri birlikte değerlendirilerek matematik testinden çıkarılması ya da düzeltilmesi gerektiği düşünülen maddeler belirlenmiş ve elde edilen bulgular Tablo 51’de sunulmuştur.

Tablo 51

*Matematik Testindeki Sorunlu Maddelerin Madde Tepki Kuramı’na göre Belirlenmesine İlişkin Bulgular*

	Ayırt Edicilik Bakımından	Tahmin Olasılığı Bakımından	Açıklanan Bilgi Bakımından
Sorunlu Madde Kümeleri	-	1, 2, 3, 5, 10, 11, 14, 15, 16, 19, 21, 22, 23	4, 6, 7, 9, 15, 22

Ayırt ediciliği ve taşıdığı bilgi yüksek olan bunun yanı sıra tahminle doğru yanıtlanma olasılığı çok yüksek olmayan maddeler kaliteli olarak nitelendirilebilir. Madde bilgi fonksiyonları maddelerin ayırt edicilik ve tahminle doğru yanıtlanma olasılığı gibi bilgilerinden türetilen bir fonksiyon olmasına karşın bu değerlendirmede ayrı bir kıstas olarak ele alınmış ve parametre değerlerinin bir sağlaması olarak düşünülmüştür.

MTK’ya dayalı yürütülen bu inceleme sonrasında sorunlu olduğu belirlenen 3, 4, 5, 6, 7, 10, 11, 14,16, 19, 22 ve 23. maddelerin Sato Test Kuramı ile yapılan belirlemelerle tutarlılık gösterdiği (Tablo 49, s.81) görülmüştür. Bununla birlikte 3, 4, 5, 7, 10, 11, 14, 16, 22 ve 23. maddelerin Klasik Test Kuramı ile yapılan incelemelerde de sorunlu olarak belirlendiği (Tablo 47, s.80) görülmüştür.

Türkçe ve matematik testlerindeki sorunlu maddelerin her üç kurama dayalı olarak belirlenmesi ve karşılaştırılması noktasında elde edilen bulgulardan hareketle STK’nın hem KTK hem de MTK ile tutarlı sonuçlar ürettiği yorumu yapılabilir. Yapılan belirlemelerin farklı büyüklükteki örneklemeler arasındaki tutarlılığı söz konusu olduğunda ise STK’nın KTK’dan daha tutarlı sonuçlar ürettiği gözlemlenmiştir.

## Bölüm 5

### Sonuç, Tartışma ve Öneriler

Bu araştırmada, Sato Test Kuramı (STK), Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) ile elde edilen madde özellikleri arasındaki ilişki düzeylerinin ve testlerden çıkarılmasına ya da düzeltilmesine karar verilen sorunlu maddelerin belirlenmesi; sözü edilen kuramlara dayalı olarak elde edilen öğrenci başarı puanlarının karşılaştırılması ve katılımcılar arasından seçkisiz olarak seçilen farklı örneklemeler üzerinden Sato Test Kuramına dayalı olarak elde edilen madde özelliklerinin tutarlılığının incelenmesi amaçlanmıştır, bu doğrultuda oluşturulan on araştırma sorusuna yanıt aranmıştır. Bu bölümde, yapılan analizlere ve elde edilen bulgulara dayalı olarak ulaşılan sonuçlar benzer çalışmalarla karşılaştırılarak maddeler halinde sunulmuştur. Ayrıca, araştırma sürecinde edinilen bilgiler doğrultusunda, bu alanda çalışma yapmak isteyen araştırmacılara, test geliştiricilere ve uygulayıcılara çeşitli öneriler getirilmiştir.

#### Sonuçlar

- Türkçe ve matematik testlerinde yer alan maddeler için aynı evrenden seçkisiz olarak alınan farklı örneklemeler üzerinden elde edilen MİSGİD değerleri arasındaki korelasyon katsayılarının tüm örneklem büyüklüklerinde ,01 düzeyinde manidar olduğu belirlenmiştir. Her iki test için de 50'şer kişilik örneklemelerden elde edilen MİSGİD değerleri arasında beklenenin altında korelasyon değerleriyle karşılaşılsa da katsayı değerleri genel olarak pozitif ve yüksek düzeyde ilişkilere (tutarlılığa) işaret etmektedir.

- Türkçe ve matematik testlerinde yer alan maddeler için aynı evrenden seçkisiz olarak alınan farklı örneklemelerden elde edilen MİSGİD değerlerinin hiçbir örneklem büyüklüğünde istatistiksel olarak manidar bir farklılık göstermediği belirlenmiştir. Elde edilen bu bulgular STK'da madde güçlük düzeyini ifade eden MİSGİD değerlerinin farklı grupta manidar değişkenlik göstermediğine (tutarlı sonuçlar verdiğine) dair ikinci bir kanıt olarak değerlendirilmiştir.

- Türkçe testi için alt ve üst gruptan elde edilen MİSGİD değerleri arasında 0,824 düzeyinde; matematik testi için alt ve üst gruptan elde edilen MİSGİD

değerleri arasında ise 0,771 düzeyinde korelasyon olduğu belirlenmiştir. Elde edilen bulgular, uç yetenek düzeylerinde yanıtlayıcıların bulunduğu gruptan elde edilen MİSGİD değerlerinin değişimlerinin tutarlılık gösterdiğine ilişkin bir kanıt olarak değerlendirilmiştir.

- Türkçe ve matematik testlerinde yer alan maddelerin işlevliliğiyle ilgili bilgi veren sınıflandırmaların (A, A', B, B') 50'şer kişilik örneklem arasında bazı maddeler için önemli oranda değişkenlik gösterdiği fakat örneklem büyüklüğü arttıkça sınıflandırmalar arasındaki uyumun da arttığı belirlenmiştir. Türkçe testi maddelerinin tamamına yakını 100 ve 200'er kişilik örneklem arasında yüksek uyum oranlarıyla sınıflandırılmıştır. Örneklem büyüklüğü 600 olduğunda ise sınıflandırmalar tüm maddeler için %100'e yakın oranlarda uyum göstermiştir (10 farklı örneklemin her birinde aynı sınıfa atanan maddelerin %100 uyumla sınıflandırıldığı ifade edilmektedir). Matematik testi maddelerinin ise 600'er kişilik örneklemde tamamına yakını yüksek uyumla sınıflandırılmıştır. Elde edilen bulgulardan hareketle Türkçe testi maddeleri için STK ile yapılan sınıflandırmaların daha fazla tutarlılık gösterdiği sonucuna ulaşılabılır. İlk üç araştırma sorusuna yanıt bulmak amacıyla yapılan analizler sonucunda STK'nın Klasik Test Kuramı'nın madde özelliklerinin gruba bağlı olmasına ilişkin sınırlılığını ortadan kaldıracıcağı yönünde bulgulara ulaşıldığı fakat bu iddianın güçlendirilebilmesi bakımından daha fazla kanıt ihtiyacı olduğu belirtilebilir. Bu anlamda, farklı maddeler, gruplar ve koşullar üzerinden yapılacak olan yeni çalışmaların ortaya koyacağı sonuçlar önem taşımakta ve yeterince kanıt ulaşıncaya dek bu iddianın tartışmaya açık kalacağı öngörülmektedir.

- Araştırmanın dördüncü, beşinci ve altıncı alt problemlerinin STK ile KTK'nın karşılaştırılması çerçevesinde ele alındığı belirtilebilir. Bu bağlamda, Türkçe ve matematik testlerinde yer alan maddelerin ayırt edicilik düzeylerinin belirlenebilmesi bakımından madde ayırıcılık (KTK) ve madde uyarı (STK) indekslerinin benzer şekilde çalıştığı sonucuna ulaşılmıştır. Ayrıca söz konusu maddelerin güçlük düzeylerinin belirlenebilmesi bakımından da madde güçlük indeksleri (KTK) ile maddeler için sınırlandırılmış gri ilişki derecelerinin (STK) benzer şekilde çalıştığı belirlenmiştir. Bireylerin Türkçe ve matematik testlerinde gösterdikleri performansların belirlenmesi noktasında da toplam test puanları

(KTK) ile öğrenciler için sınırlandırılmış gri ilişki dereceleri (STK) arasında tüm örneklem büyüklükleri için pozitif yönde ve yüksek düzeyde korelasyonların olduğu tespit edilmiştir. Elde edilen bulgulardan hareketle STK ile KTK'nın psikometrik özellikler bakımından birbirine yakın sonuçlar ürettiği ifade edilebilir.

- Araştırmanın yedinci, sekizinci ve dokuzuncu alt problemlerinin STK ile MTK'nın karşılaştırılması çerçevesinde ele alındığı belirtilebilir. Bu bağlamda, Türkçe ve matematik testlerinde yer alan maddelerin ayırt edicilik düzeylerinin belirlenebilmesi bakımından madde uyarı indeksleri (STK) ile  $a$  parametreleri (MTK) arasında yüksek düzeyde bir benzerliğin bulunmadığı belirlenmiştir. Buna karşın testlerde yer alan maddelerin güçlük düzeylerinin belirlenmesi noktasında maddeler için sınırlandırılmış gri ilişki dereceleri (STK) ile  $b$  parametrelerinin (MTK) benzer şekilde çalıştığı sonucuna ulaşılmıştır. Bireylerin testlerde gösterdikleri performansların belirlenmesi söz konusu olduğunda ise öğrenciler için sınırlandırılmış gri ilişki dereceleri (STK) ile  $\theta$  değerleri (MTK) arasında pozitif yönde ve yüksek düzeyde korelasyonların (ÖİSGİD'lerin elde edildiği tüm örneklem büyüklükleri için) bulunduğu görülmüştür. Elde edilen bulgulardan hareketle STK ile MTK'nın madde ayırt ediciliklerinin belirlenmesi haricinde benzer sonuçlar ürettiği iddia edilebilir. Bununla birlikte, STK'da doğrudan madde ayırt ediciliklerine ilişkin bilgi elde etmeye yönelik olarak kullanılacak bir indeks tanımlanmadığı vurgulanabilir. Maddelerin ayırt edicilik özelliklerine işaret eden madde uyarı indeksleri tek başlarına yorumlanmamakta maddelerin doğru yanıtlanma oranlarıyla birlikte işlevlikleri yönünden sınıflandırılması amacıyla kullanılmaktadır. Araştırma bulguları, STK'nın madde uyarı indeksleriyle KTK'nın madde ayırt edicilik indekslerinin benzer sonuçlar ürettiğini buna karşın MTK'nın  $a$  parametreleri ile bu benzerliğin kurulamadığını göstermektedir.

- Türkçe ve matematik testlerindeki sorunlu maddelerin her üç kurama dayalı olarak belirlenmesi ve karşılaştırılması sürecinde KTK'ya dayalı incelemelerde madde ayırt edicilik indeksleri, STK'ya dayalı incelemelerde madde uyarı indeksleri ve oranlar aracılığı ile yapılan sınıflandırmalar, MTK'ya dayalı incelemelerde ise  $a$ ,  $c$  parametreleri ile madde bilgi fonksiyonları dikkate alınmıştır. STK'ya göre sorunlu olduğu belirlenen maddelerden oluşturulan



kümeler ile KTK ve MTK'ya göre sorunlu olduğu belirlenen maddelerden oluşturulan kümelerin çok sayıda ortak elemanının olduğu belirlenmiştir. Söz konusu bulgudan hareketle, sorunlu maddelerin belirlenmesi konusunda STK'ya dayalı belirlemelerin diğer kuramlara dayalı belirlemelerle paralellik gösterdiği (dış tutarlılık) sonucuna ulaşılmıştır. Ayrıca, yapılan belirlemelerin farklı büyüklükteki örneklemeler arasındaki tutarlılığı söz konusu olduğunda (iç tutarlılık) STK'nın KTK'dan daha tutarlı sonuçlar ürettiği gözlemlenmiştir. Elde edilen bulgular, geçerliliği ve güvenilirliği yüksek ölçmelerin yapılmasına olanak tanıyacak ölçme araçlarının geliştirilmesi sürecinde STK'dan yararlanılabileceği görüşünü destekler niteliktedir.

- Bu araştırma, küçük örneklemeler üzerinden üretilen STK indekslerinin dahi istatistiksel tutarlılık (consistency) gösterdiğini korelasyon analizleri ve fark testleri ile ortaya koymuştur. Ayrıca, birey ve madde özelliklerinin belirlenmesi noktasında STK'nın diğer kuramlarla benzer sonuçlar ortaya koyduğu ve özellikle sorunlu maddelerin belirlenmesi noktasında MTK ile çok yakın sonuçlar elde edildiği görülmüştür. Başka bir ifadeyle bu araştırmada ulaşılan sonuçlar STK'nın, psikometrik açıdan kendi içerisinde tutarlı ve diğer test kuramlarıyla çelişmeyen kestirimleriyle geçerli ve güvenilir ölçmeler yapılmasına olanak tanıyan alternatif bir test kuramı olarak değerlendirilebileceği iddiasını desteklemektedir.

## **Öneriler**

Bu bölümde, ilk aşamada bu araştırmanın bulgularına ve sonuçlarına dayalı olarak alan uzmanlarına ve test uygulayıcılarına; ikinci aşamada ise araştırma süresince edinilen tecrübeler doğrultusunda konuya ilgi duyan araştırmacılara çeşitli öneriler getirilmiştir.

## **Araştırma Bulgularına ve Sonuçlarına Dayalı Öneriler**

- Küçük gruplar üzerinde yürütülen ölçme ve değerlendirme uygulamalarında KTK'nın yaygın olarak kullanıldığı bilinmektedir. Bu araştırmada KTK ve STK'nın psikometrik özellikler bakımından birbirine yakın sonuçlar ürettiği ortaya konulmuş olsa da STK'yı üstün kılan, maddeler ve öğrenciler hakkında nitel dönütler elde edilebilmesi ve aynı puanı alan öğrencilerin sınıflandırmalar

aracılığıyla ayırt edilebilmesi gibi özellikler dikkate alındığında STK, özellikle okullarda yürütülen ölçme ve değerlendirme uygulamaları için öğretmenlere önerilebilir.

- STK'ya dayalı olarak maddelerin işlerliğiyle ilgili sınıflama yapılması söz konusu olduğunda 100 kişilik örneklem büyüklüğünden itibaren yüksek sınıflama tutarlılığına ulaşıldığı görülmüştür. Bu bulgudan hareketle, hazırlayacakları testler için kaliteli maddelerin yer aldığı bir madde havuzu oluşturmak isteyen öğretmenlere bu süreçte en az 100 kişilik gruplar üzerinde çalışmaları önerilebilir.

- Bu araştırmada sorunlu maddelerin belirlenmesi bakımından STK ile MTK'nın benzer sonuçlar ortaya koyduğu görülmüştür. Bununla birlikte madde güçlüklerinin ve öğrenci/birey yeteneklerinin belirlenmesi noktasında da her iki kuramın yakın sonuçlar ürettiği belirlenmiştir. MTK'nın karmaşık yapısından kaynaklı olarak ortaya çıkan uygulama ve yorumlama güçlüğü noktasındaki sorunlar dikkate alındığında STK, test geliştiren, uygulayan kişi, kurum ve kuruluşlara bir alternatif olarak önerilebilir.

- Maddelerin ayırt ediciliklerinin belirlenmesi bakımından STK ve MTK'nın farklılaştığı göz önünde bulundurulduğunda STK, özellikle MTK modellerinden 1PLM'nin tercih edileceği durumlar için alternatif oluşturabilir.

### **Benzer Araştırmalar için Öneriler**

- Bu çalışmada, STK ile araştırma verisine en iyi uyum sağlayan MTK modeli olarak belirlenen 3P Lojistik Model arasında karşılaştırmalar yapılmıştır. Başka araştırmacılar diğer MTK modellerine uygun veriler üzerinden benzer ya da farklı koşullar belirleyerek kuramlar arası karşılaştırmalar yapabilirler.

- Bu çalışmada yapılan karşılaştırmalar yalnızca iki kategorili (dichotomous) puanlanan maddelere uygun STK modeli üzerinden yürütülmüştür. Konuya ilgi duyan araştırmacılara, çok kategorili (polytomous) puanlanan maddelere uygun STK modellerinin diğer test kuramlarıyla psikometrik açıdan karşılaştırılmasını amaçlayan çalışmalar yapmaları önerilebilir.

## Kaynaklar

- Acar, T. (2006). Sato uyarı indeksleri ile madde ve başarı analizleri. [http://www.academia.edu/11390990/Sato Uyarı İndeksleri ile Madde ve Başarı Analizleri](http://www.academia.edu/11390990/Sato_Uyarı_İndeksleri_ile_Madde_ve_Başarı_Analizleri) adresinden erişilmiştir.
- An, X., Yung, Y. F. (2014). Item response theory: What it is and how you can use the irt procedure to apply it. <https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf> adresinden erişilmiştir.
- Atılğan, H., Kan, A. ve Doğan, N. (2009). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı Yayıncılık.
- Baker, F.B. (2001). *The basics of item response theory*. MD: ERIC Clearinghouse on Assessment and Evaluation.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Bayrak, F. ve Yurdugül, H. (2016). Web-tabanlı öz-değerlendirme sisteminde öğrenci uyarı indeksini temel alan öğrenme analitiği modülünün tasarlanması. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(2), 85-99.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. & Novick, M.R. (Eds.), *Statistical Theories of Mental Test Scores*. MA: Addison-Wesley.
- Coaley K. (2009). *Psychological assessment and psychometrics*. California: Sage Publications.
- Coughlin, K.B. (2013). *An analysis of factor extraction strategies: A study of the relative strenghts of principal axis, ordinary least squares and maximum likelihood factor extraction methods in research contexts*(Yayımlanmamış doktora tezi). University of South Florida, Tampa, FL.
- Courville, T. G. (2004). *An empirical comparison of item response theory and classical test theory item/response statistics*. Unpublished Doctoral Dissertation. A & M University, Texas.
- Crocker, L. ve Algina J. (1986). *Introduction to classical and modern test theory*. Orlando: Harcourt Brace Jovanovich Inc.
- Çelen, Ü. ve Aybek, E.C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- Çüm, S., Demir, E.K. ve Şahin, M.D. (2018). Sato test kuramı ile yapılan sınıflamalara yönelik öğretmen görüşlerinin incelenmesi. *Anadolu Journal of Educational Sciences International*, 8(1), 171-194.

- Çüm, S., Gelbal, S. ve Tsai, C.P. (2016). Sato test kuramı yöntemleriyle farklı örneklemelerden elde edilen madde parametrelerinin tutarlılığının incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 170-181.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- DeMars, C. (2016). *Madde tepki kuramı*. Hülya Kelecioğlu (Çev. Ed.). Ankara: Nobel.
- DeVellis, R.F. (2014). *Ölçek geliştirme: Kuram ve uygulamalar*. Tarık Totan (Çev. Ed.). Ankara: Nobel.
- Doğan, N. ve Tezbaşaran, A.A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67.
- EARGED (Eğitimi Araştırma Geliştirme Dairesi) (Tarihsiz). *ÖBBS tanıtımı*. <http://yegitek.meb.gov.tr/earged/arasayfa.php?q=77> adresinden erişilmiştir.
- Ebel, R.L. ve Frisbie D.A. (1991). *Essentials of educational measurement*. New Delhi: Prentice-Hall Inc.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. NJ: Lawrence Erlbaum Associates, Inc.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., ve Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/response statistics, *Educational and Psychological Measurement*, 58(3), 357-381.
- Ferrando, P. J. ve Chico, E. (2007). The external validity of scores based on the two-parameter logistic model: Some comparisons between IRT and CTT. *Psicologica*, 28, 237-257.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications Ltd.
- Fulcher, G. ve Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York: Routledge.
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: Principles and applications*, New York: Springer.

- Hambleton, R. K., Swaminathan, H., ve Rogers, H. (1991). *Fundamentals of item response theory*. CA: Sage.
- Hayton, J.C., Allen, D.G. ve Scarpello, V. (2004). Factor retention decision in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. MA: Newbury House.
- Henson, R.K. ve Roberts, J.K. (2006). Use of exploratory factor analysis in published research: common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66, 393-416.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-85
- Hulin, C. L., Lissak, R. I., ve Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Kiany, G.R. ve Jalali, S. (2009). Theoretical and practical comparison of classical test theory and item response theory. *International Journal of Active Learning*, 12(1), 167-197.
- Lin, Y.H., ve Chen, S.M. (2006). *The investigation of S-P chart analysis on the test evaluations of equality axiom concepts for sixth graders*. Proceedings of the 2<sup>nd</sup> International Conference on Educational Technologies, Romania, Bucharest.
- Lin, Y.H., ve Yih, J.M. (2015). *Application of IIRS in mathematics instruction to promote pupils decimal concept*. The International Conference on Language, Education and Psychology, Taiwan.
- Lord, F. M., ve Novick, M. R. (1968). *Statistical theories of mental test scores*. MA: Addison-Wesley Publishing Company.
- Lord, F. ve Novick R.M. (1968). *Statistical theories of mental test scores*. New York: Addison Wesley Publishing Company.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. NJ: Lawrence Erlbaum Associates.
- MacDonald, P. ve Paunonen, S. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62, 921-943.
- McArthur, D.L. (1983). *Analysis of test score patterns: The student-problem (s-p) technique (Report no: 218)*. Los Angeles: Center for the Study of Evaluation, University of California.

- Nagai, M. (2010). Handouts of Graduate Institute of Educational Measurement and Statistic at NTCU, Taiwan.
- Nartgün, Z. (2002). *Aynı Tutumu Ölçmeye Yönelik Likert Tipi Ölçek ile Metrik Ölçeğin Madde ve Ölçek Özelliklerinin Klasik Test Kuramı ve Örtük Özellikler Kuramına Göre incelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Pham, D.H., Sheu, T.W., ve Nagai, M. (2015). PCSP 1.0 software for partial credit S-P chart analysis. *International Journal of Hybrid Information Technology*, 8(6), 309-322.
- Ree, M. J., ve Jensen, H. E. (1983). *Effects of sample size on linear equating of item characteristic curve parameters: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Sato, T. (1984). *The state of art on S-P analysis activities in Japan*. Tokyo: NEC Corp.
- Sheu, T. W., Nguyen, P. T., Nguyen, P. H., Pham, D. H., Tsai, P. C., ve Nagai, M. (2014a). A MATLAB toolbox for misconceptions analysis based on S-P chart, grey relational analysis and ROC. *Transactions on Machine Learning and Artificial Intelligence*, 2, 72-85.
- Sheu, T.W., Nguyen, P.T., Pham, D.H., Tsai, C.P., Nguyen, P. H. ve Nagai, M. (2014d). A new proposal for applying rasch gsp method in assessment of educational testing. *International Journal of Application or Innovation in Engineering & Management*, 3(7), 152-163.
- Sheu, T.W., Nguyen, P.T., Tsai, C.P., Pham, D.H., Nguyen, P.H. ve Nagai, M. (2014b). Using grey student-problem chart in the evaluation of tests with large data sets. *Education Practise and Innovation*, 1(2), 2372-3106.
- Sheu, T.W., Pham, D.H., Nguyen, P.T., ve Nguyen, P.H. (2013). Amatlab toolbox for student-problem chart and grey student-problem chart and its application. *International Journal of Kansei*, 4(2), 75-86.
- Sheu, T.W., Pham, D.H., Tsai, C.P., Nguyen, P.T., Nguyen, P. H. ve Nagai, M. (2014c). Rasch GSP toolbox for assessing academic achievement. *Journal of Software*, 9(7), 1903-1913.
- Sheu, T.W., Tsai, C.P., Tzeng, J.W., Pham, D.H., Chiang, H.J., Chang, C.L. ve Nagai, M. (2013). An improved teaching strategies proposal based on student's learning misconceptions. *International Journal of Kansei Information*, 4(1), 1-12
- Spencer, G.S. (2004). *The Strength of Multidimensional Item Response Theory in Exporing Consrtuct Space That is Multidimensional and Corralated*. Unpublished doctoral dissertation, Brigam Young University.

- Stage, C. (2003). *Classical test theory or item response theory: the Swedish experience (Report No. 42)*. Umeå: Department of Educational Measurement.
- Suen, H.K. (1990). *Principles of test theories*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Switzer, D. M. ve Connell, M. L. (1990). Practical applications of student response analysis, *Educational Measurement: Issue and Practice*, 9, 15-17.
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, 49(1), 95-110.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurements Yearbook*, 54, 837–847.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thorndike, L. R. (1982). *Applied psychometrics*. Boston: Houghton Mifflin Co.
- Traub, R. ve Rowley, G.(1991). Understanding reliability. *Educational Measurement: Issues and Practice*, 8(1), 8-14.
- Tsai, C.P., Sheu, T.W., Tzeng, J.W., Chen, H.J., Chiang, H.J. ve Nagai, M. (2014). Diagnose learning misconceptions based on rough sets. *International Journal of Applied Mathematics and Statistics*, 52(2), 63-75.
- Turgut, F. (1978). *Test geliştirme teknikleri: Ders notları*. Ankara: Hacettepe Üniversitesi Yayınları.
- van der Linden, W. J. ve Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Wang, B.T., Sheu, T.W., ve Nagai, M. (2011). Evaluating the english-learning of engineering students using the grey S-P chart: a facebook case study in Taiwan. *Global Journal of Engineering Education*, 13(2), 51-56.
- Wang, C.H. ve Chen, C.P. (2013). Employing online S-P diagnostic table for qualitative comments on test results. *The Electronic Journal of e-Learning*, 11(3), 263-271.
- Wu, H. (1999). Software Based on S-P Chart Analysis and Its Applications. *Proceedings of the National Science Council*, 8, 102-107.
- Yu, M. N. (2011). *Educational testing and assessment*. Taiwan: Psychology Publisher.
- Zickar, M.J. ve Broadfoot, A. A. (2008). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C.E. Lance ve

R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends*. New York: Routledge Academic.

Zwick, W. R. ve Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-44.



## EK-A: Türkçe Testi için Maddeler Arası Standardize Edilmiş ki-kare Değerleri

Marginal fit ( $X^2$ ) and Standardized LD  $X^2$  Statistics for Group 1 [\(Back to TOC\)](#)

Item	Label	Marginal $X^2$	1	2	3	4	5	6	7	8	9	10
1	VAR2	0.0										
2	VAR3	0.8	12.8									
3	VAR4	0.0	5.5	6.4								
4	VAR5	0.0	-0.5	1.5	33.4							
5	VAR6	0.3	-0.3	3.4	3.7	3.4						
6	VAR7	0.3	-0.2	1.1	5.7	7.8	12.1					
7	VAR8	0.3	-0.4	0.0	-0.2	-0.3	0.7	9.2				
8	VAR9	0.0	-0.2	5.9	2.9	21.4	-0.5	1.1	6.5			
9	VAR10	0.1	-0.3	0.6	-0.5	3.7	0.2	9.6	2.2	1.2		
10	VAR11	0.1	0.5	0.2	4.7	7.1	-0.4	5.6	-0.4	4.1	18.2	
11	VAR12	0.2	-0.2	0.4	-0.3	-0.3	-0.2	-0.4	3.3	2.6	0.9	3.6
12	VAR13	0.2	0.9	0.2	0.1	0.3	0.9	0.2	0.7	4.3	-0.6	-0.4
13	VAR14	0.1	-0.6	-0.1	-0.2	0.0	1.9	2.3	0.8	0.2	-0.3	0.1
14	VAR15	1.1	0.2	0.5	0.4	0.2	0.1	0.1	4.9	0.4	1.4	0.5
15	VAR16	0.0	-0.2	0.1	-0.7	6.8	1.1	-0.5	-0.5	2.7	-0.6	-0.6
16	VAR17	0.5	-0.1	0.0	5.8	4.5	1.4	-0.0	0.5	4.9	-0.2	4.4
17	VAR18	0.0	0.0	0.8	-0.6	2.4	0.7	-0.2	0.6	3.6	-0.6	0.5
18	VAR19	0.5	0.1	2.8	2.3	1.2	1.2	-0.2	-0.0	0.1	-0.3	-0.3
19	VAR20	0.1	-0.4	-0.1	-0.1	-0.6	0.0	4.0	0.4	-0.5	0.7	-0.5
20	VAR21	0.0	-0.5	2.6	-0.5	3.6	1.0	0.3	-0.1	1.0	-0.5	0.4
21	VAR22	0.2	0.4	-0.0	0.6	-0.2	3.6	1.5	-0.3	-0.3	0.8	-0.0
22	VAR23	0.4	0.4	0.4	-0.2	1.5	0.5	2.0	-0.1	0.3	2.2	0.4
23	VAR24	0.3	0.2	0.7	-0.1	-0.4	-0.2	2.7	-0.4	1.6	-0.3	1.4
24	VAR25	0.3	0.4	-0.0	3.1	0.1	-0.4	4.6	0.1	-0.5	0.5	0.2
25	VAR26	0.2	-0.1	0.1	0.7	-0.4	0.2	-0.4	-0.3	-0.4	1.0	-0.4

Item	Label	Marginal $X^2$	11	12	13	14	15	16	17	18	19	20
11	VAR12	0.2										
12	VAR13	0.2	9.1									
13	VAR14	0.1	0.5	-0.5								
14	VAR15	1.1	0.2	23.8	4.3							
15	VAR16	0.0	-0.4	2.0	0.2	0.2						
16	VAR17	0.5	0.0	0.6	2.5	0.8	2.9					
17	VAR18	0.0	3.0	-0.3	-0.3	0.2	-0.4	2.8				
18	VAR19	0.5	0.3	1.4	-0.3	1.4	-0.2	9.8	21.5			
19	VAR20	0.1	-0.1	-0.5	0.4	0.1	0.7	-0.3	0.0	1.5		
20	VAR21	0.0	-0.5	-0.2	4.1	0.4	-0.2	0.1	-0.2	-0.2	24.8	
21	VAR22	0.2	2.8	37.3	-0.4	0.2	3.6	0.1	0.0	-0.1	4.1	7.7
22	VAR23	0.4	1.5	-0.2	9.2	1.4	2.8	2.4	3.2	1.9	-0.1	3.1
23	VAR24	0.3	-0.3	-0.4	3.2	0.7	2.6	0.0	-0.2	1.2	0.7	0.7
24	VAR25	0.3	1.8	-0.5	17.5	0.2	1.2	10.0	-0.3	-0.0	3.0	5.2
25	VAR26	0.2	-0.4	-0.5	0.5	0.1	-0.3	-0.2	-0.4	0.5	-0.0	0.0

Item	Label	Marginal				
		$\chi^2$	21	22	23	24
21	VAR22	0.2				
22	VAR23	0.4	1.4			
23	VAR24	0.3	0.2	-0.1		
24	VAR25	0.3	1.8	-0.3	30.3	
25	VAR26	0.2	4.1	4.4	10.0	1.0

**Marginal Reliability for Response Pattern Scores: 0.85**

## EK-B: Matematik Testi için Maddeler Arası Standardize Edilmiş ki-kare Değerleri

**Marginal fit ( $X^2$ ) and Standardized LD  $X^2$  Statistics for Group 1** ([Back to TOC](#))

Item	Label	Margin al $X^2$	1	2	3	4	5	6	7	8	9	10
1	VAR2	0.1										
2	VAR3	0.0	24.2									
3	VAR4	0.0	2.7	48.9								
4	VAR5	0.0	-0.2	8.6	27.1							
5	VAR6	0.0	0.7	1.4	-0.5	6.0						
6	VAR7	0.0	9.1	1.9	0.8	0.7	11.7					
7	VAR8	0.0	-0.2	1.6	4.7	-0.7	-0.6	15.4				
8	VAR9	0.0	0.4	-0.6	0.3	1.1	1.0	1.9	4.1			
9	VAR10	0.0	-0.6	-0.7	-0.7	2.4	1.1	5.5	-0.3	8.8		
10	VAR11	0.0	10.5	-0.2	-0.6	-0.7	-0.7	0.6	-0.6	0.2	41.2	
11	VAR12	0.1	-0.6	0.4	-0.4	1.2	5.1	-0.4	-0.4	-0.4	1.3	26.0
12	VAR13	0.0	5.0	0.8	-0.7	-0.2	-0.6	5.8	4.5	8.3	13.0	4.7
13	VAR14	0.0	-0.6	1.5	0.2	-0.1	-0.6	4.5	-0.4	1.7	2.2	-0.4
14	VAR15	0.0	-0.6	0.3	1.9	3.9	-0.5	1.0	5.1	7.2	28.3	-0.7
15	VAR16	0.0	3.9	2.8	9.8	-0.7	-0.1	2.2	-0.6	-0.5	-0.5	-0.6
16	VAR17	0.0	0.3	-0.7	1.5	-0.7	-0.6	-0.6	-0.3	0.1	5.7	0.2
17	VAR18	0.0	0.2	0.0	-0.6	-0.1	-0.6	-0.5	2.8	1.4	2.0	0.6
18	VAR19	0.1	-0.6	-0.5	-0.3	-0.5	4.4	14.6	1.2	2.4	2.8	0.1
19	VAR20	0.1	-0.6	-0.3	-0.3	0.0	0.3	-0.4	-0.6	-0.3	-0.0	1.2
20	VAR21	0.0	0.1	0.8	6.8	-0.6	1.0	0.3	-0.2	0.5	2.0	6.0
21	VAR22	0.1	1.2	-0.3	-0.6	-0.5	-0.6	-0.6	-0.2	-0.4	-0.5	1.2
22	VAR23	0.0	2.3	-0.6	1.6	0.8	-0.5	0.7	-0.3	-0.5	-0.5	0.5
23	VAR24	0.0	-0.3	2.9	1.8	-0.2	-0.6	-0.6	5.2	1.6	1.9	-0.0
24	VAR25	0.0	-0.1	2.8	1.2	1.0	3.7	7.6	-0.7	-0.7	-0.5	-0.7
25	VAR26	0.0	-0.2	-0.4	-0.7	7.5	-0.4	0.4	3.7	5.4	33.7	0.8

Item	Label	Margin al $X^2$	11	12	13	14	15	16	17	18	19	20
11	VAR12	0.1										
12	VAR13	0.0	14.3									
13	VAR14	0.0	0.9	10.5								
14	VAR15	0.0	0.7	-0.1	0.2							
15	VAR16	0.0	-0.6	0.8	-0.3	5.4						
16	VAR17	0.0	-0.1	0.3	-0.0	2.6	22.1					
17	VAR18	0.0	-0.2	1.0	-0.5	1.3	1.5	18.1				
18	VAR19	0.1	-0.4	4.7	-0.5	-0.3	1.1	-0.6	30.2			
19	VAR20	0.1	-0.2	4.8	0.1	1.7	-0.2	1.4	15.7	7.9		
20	VAR21	0.0	0.6	2.6	1.3	2.0	0.3	0.2	-0.4	-0.3	0.5	
21	VAR22	0.1	-0.5	2.8	0.6	-0.2	-0.4	0.3	0.6	1.5	17.8	4.8
22	VAR23	0.0	-0.5	0.7	0.7	0.8	-0.7	-0.0	-0.2	-0.4	6.5	-0.7
23	VAR24	0.0	-0.6	-0.2	0.1	1.9	3.6	0.3	0.5	-0.2	3.4	1.7
24	VAR25	0.0	-0.4	3.8	-0.3	4.2	-0.6	2.1	1.5	3.1	3.6	3.8
25	VAR26	0.0	-0.6	0.8	-0.2	5.0	4.1	-0.5	-0.2	0.8	2.6	-0.5

Item	Label	Marginal				
		$\chi^2$	21	22	23	24
21	VAR22	0.1				
22	VAR23	0.0	8.9			
23	VAR24	0.0	-0.5	-0.4		
24	VAR25	0.0	1.0	-0.4	0.0	
25	VAR26	0.0	-0.5	7.7	8.1	6.8

**Marginal Reliability for Response Pattern Scores: 0.75**

## EK-C: Türkçe Testi Madde Bilgi Fonksiyonu Değerleri

Item Information Function Values for Group 1 at 15 Values of  $\theta$  from -2.8 to 2.8 [\(Back to TOC\)](#)

		$\theta$ :														
Item	Label	-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	Madde1	0.01	0.02	0.04	0.07	0.11	0.15	0.18	0.20	0.20	0.18	0.15	0.12	0.09	0.06	0.04
2	Madde2	0.00	0.00	0.00	0.00	0.01	0.06	0.26	0.69	0.99	0.78	0.41	0.17	0.07	0.03	0.01
3	Madde3	0.04	0.11	0.23	0.42	0.59	0.65	0.56	0.39	0.24	0.13	0.07	0.04	0.02	0.01	0.00
4	Madde4	0.09	0.13	0.18	0.23	0.26	0.27	0.26	0.22	0.18	0.13	0.10	0.07	0.05	0.03	0.02
5	Madde5	0.00	0.00	0.01	0.02	0.06	0.15	0.29	0.44	0.51	0.45	0.32	0.20	0.12	0.06	0.03
6	Madde6	0.00	0.00	0.01	0.03	0.12	0.32	0.59	0.71	0.58	0.36	0.18	0.09	0.04	0.02	0.01
7	Madde7	0.00	0.00	0.00	0.02	0.12	0.43	0.86	0.90	0.56	0.26	0.11	0.04	0.02	0.01	0.00
8	Madde8	0.04	0.08	0.16	0.26	0.36	0.40	0.38	0.30	0.21	0.14	0.08	0.05	0.03	0.02	0.01
9	Madde9	0.01	0.02	0.05	0.11	0.23	0.38	0.47	0.47	0.37	0.25	0.15	0.09	0.05	0.03	0.01
10	Madde10	0.00	0.01	0.04	0.13	0.33	0.60	0.76	0.67	0.44	0.24	0.12	0.06	0.03	0.01	0.01
11	Madde11	0.00	0.00	0.00	0.01	0.03	0.08	0.18	0.32	0.42	0.42	0.33	0.22	0.13	0.07	0.04
12	Madde12	0.00	0.00	0.01	0.03	0.07	0.15	0.25	0.34	0.39	0.37	0.30	0.21	0.14	0.09	0.05
13	Madde13	0.01	0.02	0.05	0.09	0.15	0.21	0.27	0.29	0.28	0.24	0.18	0.13	0.09	0.06	0.04
14	Madde14	0.00	0.00	0.00	0.00	0.01	0.07	0.52	1.48	1.24	0.48	0.14	0.04	0.01	0.00	0.00
15	Madde15	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.15	0.33	0.51	0.55	0.42	0.26	0.14
16	Madde16	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.23	0.82	1.32	0.98	0.44	0.16	0.05	0.02
17	Madde17	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.17	0.28	0.36	0.36	0.29	0.21	0.13
18	Madde18	0.00	0.00	0.00	0.00	0.01	0.04	0.16	0.43	0.71	0.72	0.48	0.25	0.12	0.05	0.02
19	Madde19	0.01	0.03	0.06	0.11	0.18	0.26	0.34	0.37	0.35	0.29	0.21	0.15	0.10	0.06	0.04
20	Madde20	0.00	0.00	0.01	0.05	0.33	1.07	1.39	0.82	0.31	0.10	0.03	0.01	0.00	0.00	0.00
21	Madde21	0.00	0.00	0.01	0.04	0.17	0.46	0.77	0.76	0.49	0.25	0.11	0.05	0.02	0.01	0.00
22	Madde22	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.27	0.66	0.94	0.79	0.45	0.21	0.09	0.04
23	Madde23	0.00	0.00	0.01	0.03	0.09	0.21	0.37	0.50	0.51	0.41	0.27	0.16	0.09	0.05	0.02
24	Madde24	0.00	0.00	0.01	0.04	0.13	0.30	0.52	0.61	0.51	0.34	0.19	0.10	0.05	0.02	0.01
25	Madde25	0.00	0.00	0.01	0.02	0.06	0.15	0.27	0.39	0.41	0.35	0.24	0.15	0.09	0.05	0.02
Test Information:		1.22	1.44	1.88	2.73	4.41	7.44	10.80	12.94	12.72	10.74	7.82	5.18	3.40	2.34	1.73
Expected s.e.:		0.91	0.83	0.73	0.61	0.48	0.37	0.30	0.28	0.28	0.31	0.36	0.44	0.54	0.65	0.76

Marginal Reliability for Response Pattern Scores: 0.85

## EK-Ç: Matematik Testi Madde Bilgi Fonksiyonu Değerleri

Item Information Function Values for Group 1 at 15 Values of  $\theta$  from -2.8 to 2.8 (Back to TOC)

Item	Label	$\theta$ :														
		-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	Madde1	0.00	0.00	0.00	0.00	0.00	0.01	0.09	0.87	2.27	1.46	0.44	0.10	0.02	0.00	0.00
2	Madde2	0.00	0.00	0.00	0.00	0.01	0.03	0.09	0.26	0.51	0.63	0.52	0.32	0.17	0.08	0.04
3	Madde3	0.00	0.00	0.00	0.00	0.00	0.02	0.06	0.18	0.39	0.57	0.57	0.42	0.24	0.13	0.06
4	Madde4	0.02	0.04	0.06	0.09	0.12	0.16	0.19	0.21	0.22	0.20	0.17	0.14	0.11	0.08	0.06
5	Madde5	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.23	0.52	0.69	0.58	0.34	0.17	0.08	0.03
6	Madde6	0.02	0.04	0.06	0.10	0.13	0.17	0.18	0.18	0.17	0.14	0.11	0.08	0.06	0.04	0.03
7	Madde7	0.01	0.01	0.02	0.03	0.05	0.08	0.11	0.13	0.15	0.16	0.16	0.15	0.12	0.10	0.08
8	Madde8	0.00	0.01	0.03	0.08	0.18	0.33	0.49	0.56	0.51	0.37	0.24	0.14	0.08	0.04	0.02
9	Madde9	0.06	0.10	0.15	0.20	0.25	0.28	0.28	0.26	0.22	0.17	0.13	0.09	0.06	0.04	0.03
10	Madde10	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.07	0.19	0.37	0.50	0.47	0.33	0.19	0.10
11	Madde11	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.32	0.83	1.06	0.71	0.33	0.13	0.05	0.02
12	Madde12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.31	1.64	2.29	1.03	0.28	0.07	0.02
13	Madde13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.12	0.57	1.39	1.48	0.81	0.31	0.10
14	Madde14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.71	2.76	1.96	0.52	0.11	0.02
15	Madde15	0.00	0.00	0.00	0.01	0.02	0.06	0.16	0.31	0.44	0.44	0.34	0.22	0.12	0.06	0.03
16	Madde16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.18	0.61	1.06	0.91	0.48	0.19	0.07
17	Madde17	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.15	0.33	0.52	0.58	0.47	0.30	0.17	0.09
18	Madde18	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.24	0.91	1.47	1.05	0.45	0.16	0.05	0.02
19	Madde19	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.11	0.67	1.59	1.30	0.53	0.16	0.05	0.01
20	Madde20	0.00	0.00	0.00	0.01	0.05	0.15	0.35	0.58	0.65	0.53	0.34	0.19	0.09	0.05	0.02
21	Madde21	0.00	0.00	0.00	0.00	0.00	0.02	0.10	0.39	0.85	0.91	0.55	0.25	0.10	0.04	0.01
22	Madde22	0.00	0.00	0.00	0.00	0.01	0.03	0.06	0.13	0.23	0.32	0.36	0.33	0.25	0.17	0.11
23	Madde23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.10	0.37	0.81	0.94	0.63	0.30	0.12
24	Madde24	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.24	0.59	0.91	0.85	0.54	0.27	0.12	0.05
25	Madde25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.10	0.89	2.50	1.77	0.55	0.13	0.03
Test Information:		1.12	1.20	1.32	1.52	1.84	2.38	3.51	6.54	12.50	18.33	21.31	14.63	7.21	3.64	2.17
Expected s. e.:		0.95	0.91	0.87	0.81	0.74	0.65	0.53	0.39	0.28	0.23	0.22	0.26	0.37	0.52	0.68

Marginal Reliability for Response Pattern Scores: 0.75

# EK-D: Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

Form: 40

## Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

23 / 06 / 2016

Hacettepe Üniversitesi  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

<b>Tez Başlığı / Konusu:</b>	SATO TEST KURAMI PARAMETRELERİNİN İNCELENMESİ VE KLASİK TEST KURAMI VE MADDE TEPKİ KURAMI PARAMETRELERİYLE KARŞILAŞTIRILMASI
------------------------------	--

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

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

Gereğini saygılarımla arz ederim.

  
Sait ÇUM  
(Öğrencinin Adı Soyadı, İmzası)

### Öğrenci Bilgileri

<b>Adı Soyadı</b>	Sait ÇUM
<b>Öğrenci No</b>	N12248994
<b>Anabilim Dalı</b>	Eğitim Bilimleri
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme
<b>Statüsü</b>	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

### Danışman Görüşü ve Onayı

Hazır veri kullanıldığından etik kurul izninden muaftır.

  
Prof. Dr. Selahattin GELBAL  
(İmza)  
(Danışmanın Ünvanı, Adı ve Soyadı)

### **EK-E: Etik Beyanı**

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

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

beyan ederim.



10/01/2019

Sait ÇÜM



## EK-F: Doktora Tez Çalışması Orijinallik Raporu

15/01/2019

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

Tez Başlığı: SATO TEST KURAMI'NIN KLASİK TEST KURAMI VE MADDE TEPKİ KURAMI İLE  
PSİKOMETRİK AÇIDAN KARŞILAŞTIRILMASI

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
11/01/2019	107	128786	21/12/2018	%11	1063137910

Uygulanan filtreler:

- 1- Kaynaklar hariç
- 2- Alıntılar dâhil
- 3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

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

Ad Soyadı: Sait ÇÜM  
Öğrenci No.: N12248994  
Ana Bilim Dalı: Eğitim Bilimleri  
Programı: Eğitimde Ölçme ve Değerlendirme  
Statüsü:  Y.Lisans  Doktora  Bütünleşik Dr.

İmza

### DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Selahattin GELBAL

## EK-G: Dissertation Originality Report

15/01/2019

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

**THESIS TITLE: A PSYCHOMETRIC COMPARISON OF SATO TEST THEORY WITH CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY**

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

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
11/01/2019	107	128786	21/12/2018	%11	1063137910

Filtering options applied:

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

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

I respectfully submit this for approval.

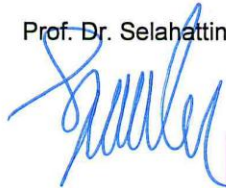
Name Lastname: Sait ÇÜM  
Student No.: N12248994  
Department: Educational Sciences  
Program: Educational Measurement and Evaluation  
Status:  Masters  Ph.D.  Integrated Ph.D.

  
Signature

### ADVISOR APPROVAL

APPROVED

Prof. Dr. Selahattin GELBAL



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

Enstitü tarafından onaylanan lisansüstü tezimin/raporun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivlerime ve aşağıda verilen koşullarla kullanıma açmamla zini Hacettepe Üniversitesi sine verdiğim bildiririm. Bu zini Üniversiteme verilen kullanım hakları dışında tüm fikrî mülkiyet haklarını bana devralacak, tezimin tamamını veya bir bölümünü gelecekteki başka işlerinde (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

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

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

- o Enstitü/Fakülte yönetim kuruluna karar ile tezimin erişime açılması mezu niyet tarihinden itibaren 2 yıl ertelenmiştir.<sup>(1)</sup>
- o Enstitü/Fakülte yönetim kuruluna göre çalılar ile tezimin erişime açılması mezu niyet tarihinden itibaren 3 ay ertelenmiştir.<sup>(2)</sup>
- o Tezimle ilgili lizlilik kararı verilmştir.<sup>(3)</sup>

10/01/2019

Sait ÇÖM

### Yükseköğretim Kurulu'nun "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 7.1. Lisansüstü tezlerin/patent başvurularının yaygın ve yaygın olmayan amaçlarla kullanılmasına ilişkin mezu niyet tarihinden itibaren 2 yıl ertelenmesi, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını ve tezinin elektronik ortamda yaygın ve yaygın olmayan amaçlarla kullanılmasına ilişkin mezu niyet tarihinden itibaren 3 ay ertelenmesi, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını gerektirebilir.
- (2) Madde 7.2. Yaratılmış eserlere/konuların/teknik hariz makalelerdeki özgün ve özgün olmayan bölümlerin elektronik ortamda yaygın ve yaygın olmayan amaçlarla kullanılmasına ilişkin mezu niyet tarihinden itibaren 3 ay ertelenmesi, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını gerektirebilir.
- (3) Madde 7.3. Ulusal kısıtlamalarla yaygın ve yaygın olmayan amaçlarla erişime açılan, yaygın ve yaygın olmayan amaçlarla erişime açılan lisansüstü tezlerin/raporların, Kurum ve kuruluşuna yaygın ve yaygın olmayan amaçlarla erişime açılmasında, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını gerektirebilir. Kurum ve kuruluşuna yaygın ve yaygın olmayan amaçlarla erişime açılan lisansüstü tezlerin/raporların, Kurum ve kuruluşuna yaygın ve yaygın olmayan amaçlarla erişime açılmasında, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını gerektirebilir. Kurum ve kuruluşuna yaygın ve yaygın olmayan amaçlarla erişime açılan lisansüstü tezlerin/raporların, Kurum ve kuruluşuna yaygın ve yaygın olmayan amaçlarla erişime açılmasında, gündüz 09:00-16:00 saatleri arasında Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılmasını gerektirebilir.

<sup>(1)</sup> Tez denizimden itibaren en geç Enstitü ve Fakülte yönetim kuruluna bir yazılı dilekçe ile erişime açılması gerektirebilir.

