



T.C.
EGE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü

**BİLİŞSEL TANI MODELLERİNDE MADDE
PARAMETRELERİ ÜZERİNDEN ÖRTÜK YETENEK
SINIFLARININ EŞİTLENMESİ**

Arzu CANPOLAT
YÜKSEK LİSANS TEZİ
EĞİTİM BİLİMLERİ ANABİLİM DALI

İzmir
2019

T.C.
EGE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü

**BİLİŞSEL TANI MODELLERİNDE MADDE PARAMETRELERİ
ÜZERİNDEN ÖRTÜK YETENEK SINIFLARININ EŞİTLENMESİ**

Arzu CANPOLAT

Tez Danışmanı: Doç. Dr. Tahsin Oğuz BAŞOKÇU

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Yüksek Lisans Programı

İzmir
2019

EGE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ETİKKURALLARA UYGUNLUK BEYANI

EÜ Lisanüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Yüksek Lisans Tezi olarak sunduğum “Bilişsel Tam Modellerinde Madde Parametreleri Üzerinden Örtük Yetenek Sınıflarının Eşitlenmesi” başlıklı bu tezin kendi çalışmam olduğumu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasında yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

Arzu Canpolat





T.C.EGE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



YÜKSEK LİSANS

TEZ SAVUNMA TUTANAĞI

ÖĞRENCİNİN

Adı Soyadı : Arzu CANPOLAT

Numarası : 92140002181

Anabilim Dalı : Eğitimde Ölçme ve Değerlendirme

Tez Başlığı (Türkçe) : Bilişsel Tanı Modellerinden DINA, A-CDM ve G-DINA Modellerinin Önsel Varsayımlarının Test Edilmesi

Tez Başlığı (İngilizce) : Analyzing The Posterior Assumptions of DINA, A-CDM and G-DINA in Cognitive Diagnostic Models

Tez Savunma Tarihi : 26/08/2019

Tez Başlığı Değişikliği Varsa Yeni Başlık: Bilişsel Tanı Modellerinde Madde Parametreleri Üzerinden Örtük Yetenek Sınıflarının Eşitlenmesi

Tez Başlığı Yeni (İngilizce) : Equating The Latent Ability Classes with Item Parameters in Cognitive Diagnostic Models

JÜRİ ÜYELERİ

Jüri Başkanı

Unvan, Adı, Soyadı : Prof. Dr. Tuncay Öğretmen

Karar : Başarılı Başarısız Düzeltme

İmza : *Tuncay Öğretmen*

Jüri Üyesi

Unvan, Adı, Soyadı : Doç. Dr. T. Oğuz BAŞOKÇU (Danışman)

Karar : Başarılı Başarısız Düzeltme

İmza : *T. Oğuz Başokçu*

Jüri Üyesi

Unvan, Adı, Soyadı : Doç. Dr. Duygu Güngör Çulha

Karar : Başarılı Başarısız Düzeltme

İmza : *Duygu Güngör Çulha*

TEZ HAKKINDA JÜRİNİN GENEL GÖRÜŞÜ

(Jüri Başkanı Tarafından Doldurulacaktır)

Tez savunması sonucunda öğrenci tarafından hazırlanan çalışma;

Oybirliğiyle

Oy çokluğuyla

Başarılıdır

Düzeltilmelidir

Başarısızdır

- Bu tutanak üç (3) işgünü içerisinde jüri üyelerinin raporlarıyla beraber Anabilim Dalı Başkanlığı üst yazısıyla Enstitü Müdürlüğüne gönderilmelidir.
- Tezli yüksek lisans programlarında düzeltme alan öğrencinin 3 (üç) ay içerisinde yeniden savunmaya girmesi zorunludur.

ÖNSÖZ

Yüksek lisans öğrenimim boyunca desteğini esirgemeyen, bilgi ve tecrübeleriyle her zaman yardımcı olan, araştırmacı olma yolunda örnek aldığım saygıdeğer tez danışmanım Doç. Dr. T.Oğuz Başokçu'ya çok teşekkür ederim.

Tez jürimde yer alan ve değerli önerilerini sunan Prof. Dr. Tuncay Öğretmen'e ve Doç. Dr. Duygu Güngör Çulha'ya teşekkür ederim.

Proje asistanlığım süresince burs imkanı sağlayan TÜBİTAK'a teşekkürlerimi sunarım.

Desteklerini her daim üzerimde hissettiğim, sağlık sorunları yaşadığım dönemlerde moral ve motivasyon kaynağım olan annem ve ablama tüm kalbimle teşekkür ederim.

Arkadaşlarım Seher Özel'e ve Simge Ceylan'a manevi desteklerinden dolayı teşekkür ederim.

ÖZGEÇMİŞ

Arzu Canpolat, 23.07.1989 yılında Kars'ta doğdu. İlköğrenimini İzmir Osmangazi İlköğretim Okulunda, ortaöğrenimini İzmir Atatürk Lisesinde tamamladı. 2013 yılında Ege Üniversitesi Fen Fakültesi Matematik Bölümünden mezun oldu. 2014 yılında Ege Üniversitesi Eğitimde Ölçme ve Değerlendirme yüksek lisans programına başladı. 2015-2017 yılları arasında tez danışmanı tarafından yürütülen 115K531 No'lu TÜBİTAK projesinde proje bursiyeri olarak görev aldı.



İÇİNDEKİLER

ÖZGEÇMİŞ.....	iv
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ.....	xii
KISALTMALAR LİSTESİ.....	xiii
ÖZET	xiv
ABSTRACT	xvi
BÖLÜM I.....	1
GİRİŞ.....	1
1.1. Problem Cümlesi	4
1.2. Alt Problemler	4
1.3. Sayıtlar	5
1.4. Sınırlılıklar.....	5
1.5. Araştırmanın Amaç ve Önemi.....	5
BÖLÜM II.....	7
İLGİLİ YAYIN VE ARAŞTIRMALAR	7
2.1. Bilişsel Tanı Modelleri Genel Özellikleri	7
2.1.1. Q Matris.....	7
2.1.2. Örtük Yetenek Sınıfı	8
2.2. Başlıca Bilişsel Tanı Modelleri	9
2.2.1. DINA (Deterministic Inputs Noisy And Gate) Model	12
2.2.2. G-DINA (Generalized Deterministic Inputs Noisy And Gate) Model.....	14
2.2.3. DINO (Deterministic Inputs Noisy Or Gate) Model.....	16
2.3. Test Eşitleme	18
2.4. Test Eşitleme Koşulları	19
2.4.1. Simetri Özelliği	19
2.4.2. Aynı Yapıyı Ölçme Özelliği	19

2.4.3. Eşitlik Özelliği.....	19
2.4.4. Eşit Güvenirlik	20
2.4.5. Gruptan Bağımsızlık	20
2.5. Eşitleme Desenleri.....	20
2.5.1. Tek Grup Deseni (The Single Group Design).....	20
2.5.2. Seçkisiz Grup Deseni (Random Group Design).....	21
2.5.3. Denk Olmayan Gruplar Ortak Test Deseni (Nonequivalent Anchor Tests-NEAT).....	22
2.6. Eşitleme Yöntemleri.....	23
2.6.1. Klasik Test Kuramına Dayalı Eşitleme Yöntemleri.....	23
2.6.1.1. Ortalama Eşitleme (Mean Equating)	23
2.6.1.2. Doğrusal Eşitleme (Linear Equating)	24
2.6.1.3. Eşit Yüzdelikli Eşitleme (Equipercentile Equating)	25
2.6.2. Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri.....	25
2.6.2.1. 1 Parametrelili Lojistik Model (1PLM)	25
2.6.2.2. 2 Parametrelili Lojistik Model (2PLM)	26
2.6.2.3. 3 Parametrelili Lojistik Model (3PLM)	26
2.7. Eşitleme Hatası.....	27
2.8. İlgili Araştırmalar	28
BÖLÜM III.....	30
YÖNTEM.....	30
3.1. Araştırmanın Modeli	30
3.2. Araştırmanın Evren ve Örneklemi	30
3.3. Ölçme Aracı	31
3.3.1. Test Geliştirme Aşamaları.....	31
3.3.1.1. Ölçme Aracının Kapsamı.....	32
3.3.1.2. Q Matris Belirleme Süreci	32
3.3.1.3. Soru Yazma Ekibinin Kurulması.....	33

3.3.1.4. Madde Havuzunun Oluşturulması	34
3.3.1.5. Deneme Uygulaması Formlarının Oluşturulması ve Uygulama Deseni.....	34
3.3.1.6. Araştırmada Kullanılan Formlar ve Bu Formlara Ait Q Matrisler .	35
3.3.1.7. Ölçme Aracının Uygulanması.....	36
3.3.1.8. Maddelerin Puanlanması.....	36
3.4. Verilerin Analizi ve Testlerin Psikometrik Özellikleri	37
3.4.1. Betimsel İstatistikler.....	37
3.4.2. Proje Soruları ile PISA TIMSS Sorularının Eşdeğerliği Çalışması	38
3.4.3. Deneme Uygulaması Yapılan Grupların Denkliklerinin Araştırılması.....	38
3.4.4. Proje Sorularının Psikometrik Özelliklerinin İncelenmesi.....	41
3.4.5. Eşitleme Yöntemi	45
BÖLÜM IV	48
BULGULAR	48
4.1. Test 1 ve Test 2'nin Madde Analizi Sonuçlarına İlişkin Bulgular.....	48
4.2. Test 1'e İlişkin Örtük Yetenek Sınıflarının ve Özelliklerin Dağılımına Ait Bulgular.....	52
4.3. Test 2'ye İlişkin Örtük Yetenek Sınıflarının Dağılımına İlişkin Bulgular.....	53
4.4. Test 1'i Alan Öğrencilerin Test 2 İçin Eşitlenmiş Örtük Yetenek Sınıflarındaki Değişime İlişkin Bulgular	54
4.5. Eşitleme Sonrası Örtük Yetenek Sınıflarının Toplam Puanlara Göre Değişimine İlişkin Bulgular	57
4.6. Eşitleme Sonrası Örtük Yetenek Sınıflarının MTK Puanlarına Göre Değişimine İlişkin Bulgular	58
4.7. Ölçülen Özelliklere Göre Örtük Yetenek Sınıflarının Eşitlenmesi Sonrası Öğrencilerin Özelliklere Sahip Olup Olmama Durumlarına İlişkin Bulgular	59
4.7.1. Öğrencilerin İletişim ve İlişkilendirme Özelliğine Sahip Olup Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular	59
4.7.2. Öğrencilerin Matematikleştirme Özelliğine Sahip Olup Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular	59

4.7.3. Öğrencilerin Akıl Yürütme Ve Strateji Geliştirme Özelliğine Sahip Olup Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular	60
4.7.4. Öğrencilerin Sembolik Ve Teknik Dil Kullanımı Özelliğine Sahip Olup Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular.....	61
BÖLÜM V.....	62
TARTIŞMA VE YORUM	62
5.1. Test 1 ve Test 2'ye Ait Madde Analizi Sonuçlarına İlişkin Bulguların Tartışılması ve Yorumu.....	62
5.2. Test 1 ve Test 2'ye İlişkin Örtük Yetenek Sınıflarının ve Özelliklerin Dağılımına İlişkin Bulguların Tartışılması ve Yorumu	62
5.3. Test 1'i Alan Öğrencilerin Test 2 İçin Eşitlenmiş Örtük Yetenek Sınıflarındaki Değişime İlişkin Bulguların Tartışılması ve Yorumu	63
5.4. Eşitleme Sonrası Örtük Yetenek Sınıfların Toplam Puanlarına ve MTK Puanlarına Göre Değişimine İlişkin Bulguların Tartışılması ve Yorumu.....	63
5.5. Ölçülen Özelliklere Göre Örtük Yetenek Sınıflarının Eşitlenmesi Sonrası Öğrencilerin Özelliklere Sahip Olup Olmama Durumlarına İlişkin Bulguların Tartışılması ve Yorumu.....	64
BÖLÜM VI	65
SONUÇ VE ÖNERİLER	65
6.1. Sonuç.....	65
6.2. Öneriler.....	66
KAYNAKÇA	67
EKLER	76
EK1-Proje Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler	76
EK2-TIMSS Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler	77
EK3-PISA Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler	78

TABLolar LİSTESİ

Tablo 1. Q Matris Örneđi	8
Tablo 2. BTM Türleri.....	11
Tablo 3. DINA İçin Örtük Sınıf Olasılıkları	14
Tablo 4. G-DINA İçin Örtük Sınıf Olasılıkları	16
Tablo 5. DINO İçin Örtük Sınıf Olasılıkları	17
Tablo 6. Tek Grup Deseni.....	21
Tablo 7. Dengelenmiş Tek Grup Deseni.....	21
Tablo 8. Seçkisiz Grup Deseni.....	22
Tablo 9. Denk Olmayan Gruplar Ortak Test Deseni.....	22
Tablo 10. Test Formlarının Sınıf Düzeyine ve Cinsiyete Göre Dağılımı	31
Tablo 11. Deneme Uygulaması Deseni.....	35
Tablo 12. C ve D Formlarına Ait Q Matrisler.....	36
Tablo 13. Deneme Uygulamasına Ait Betimsel İstatistikler.....	37
Tablo 14. Grupların Denkliklerine İlişkin ANOVA Sonuçları.....	39
Tablo 15. Ankor Maddelere İlişkin Madde Parametreleri	40
Tablo 16. Deneme Uygulaması Maddelerine İlişkin KTK Parametreleri.....	42
Tablo 17. Deneme Uygulaması Maddelerine İlişkin MTK Parametreleri.....	43
Tablo 18. Deneme Uygulaması Maddelerine İlişkin BTM Parametreleri	44
Tablo 19. Simülasyon Datasında Kullanılan Madde Parametreleri	46
Tablo 20. Simülasyonda Kullanılan Q matris	47
Tablo 21. Test 1 ve Test 2'ye ait Madde Analizi Sonuçları.....	48
Tablo 22. Test 1'e Ait Madde Parametreleri.....	50
Tablo 23. Test 2'ye Ait Madde Parametreleri.....	51
Tablo 24. Testlerin Model Veri Uyumu İstatistikleri.....	51
Tablo 25. Testlerin Madde Veri Uyumu İstatistikleri.....	52
Tablo 26. Test 1'e Ait Örtük Sınıflar ve Sonsal Olasılıkları.....	52
Tablo 27. Test 1'de Özelliklerin Gözlenme Olasılıkları	53
Tablo 28. Test 2'ye Ait Örtük Sınıflar ve Sonsal Olasılıkları.....	53
Tablo 29. Test 2'de Özelliklerin Gözlenme Olasılıkları	54
Tablo 30. Test 1 ve Eşitlenen Teste Ait Örtük Yetenek Sınıfları	55

Tablo 31. Örtük Yetenek Sınıflarının Toplam Puanlara Göre Değişimi	57
Tablo 32. Örtük Yetenek Sınıflarının MTK Puanlarına Göre Değişimi	58
Tablo 33. Öğrencilerin İletişim ve İlişkilendirme Özelliğine Sahip Olup Olmama Düzeyleri	59
Tablo 34. Öğrencilerin Matematikleştirme Özelliğine Sahip Olup Olmama Düzeyleri	60
Tablo 35. Öğrencilerin Akıl Yürütme Ve Strateji Geliştirme Özelliğine Sahip Olup Olmama Düzeyleri	60
Tablo 36. Öğrencilerin Sembolik Ve Teknik Dil Kullanımı Özelliğine Sahip Olup Olmama Düzeyleri	61



ŞEKİLLER LİSTESİ

Şekil 1. Örtük Sınıf Eşitleme Adımları 45



KISALTMALAR LİSTESİ

BTM: Bilişsel Tanı Modelleri

DINA: Deterministic Input Noisy “And”

DINO: Deterministic Input Noisy “Or”

G-DINA: Generalized-Deterministic Input Noisy “And”

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı



ÖZET

BİLİŞSEL TANI MODELLERİNDE MADDE PARAMETRELERİ ÜZERİNDEN ÖRTÜK YETENEK SINIFLARININ EŞİTLENMESİ

CANPOLAT, ARZU

Yüksek Lisans Tezi, Eğitimde Ölçme ve Değerlendirme

Tez Danışmanı: Doç. Dr. Tahsin Oğuz Başokçu

Ağustos, 2019

Bu araştırmada Bilişsel Tanı Modellerine (BTM) uygun olarak geliştirilen iki testten elde edilen örtük yetenek sınıflarının eşitlenmesi amaçlanmıştır.

Araştırma ‘Uluslararası Geniş Ölçekli Sınavlarda Türkiye’nin Matematik Başarısını Arttırabilmek için Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği’ adlı 115K531 No’lu TÜBİTAK projesi kapsamında, proje örnekleme ve ölçme araçları kullanılarak yürütülmüştür. Araştırmanın çalışma grubunu İzmir İl Milli Eğitim Bakanlığına bağlı 12 okulda 6, 7 ve 8. sınıf öğrencilerinden toplam 992 kişi oluşturmaktadır. Bu gruba ortaokul 6. sınıf matematik öğretim programına uygun olarak, alan uzmanları öğretmenler tarafından hazırlanan testler uygulanmıştır. Her bir testte çoktan seçmeli ve açık uçlu olmak üzere toplam 15 madde bulunmaktadır. Testlere ait Q matrisler 4 özellik içermektedir ve uzman görüşlerine dayanarak hazırlanmıştır.

Araştırmada, BTM parametrelerine göre belirlenen örtük yetenek sınıflarına ait cevaplayıcıların diğer testi almış olması durumunda atanacakları örtük yetenek sınıflarının ankor maddeler üzerinden belirlenmesine yönelik bir yöntem önerilmektedir. BTM analizleri DINA model ile yapılmıştır. Veri analizi üç aşamada gerçekleştirilmiştir. Öncelikle Test 1’in uygulandığı cevaplayıcıların cevap örüntüleri ve örtük yetenek sınıfları belirlenmiş, daha sonra Test 1 ve Test 2’den elde edilen madde parametreleri kullanılarak simülasyon data üretilmiştir ve DINA model ile analizi yapılmıştır. Son olarak, ilk test ile aynı örüntüye sahip olan dataların örtük yetenek sınıfları eşleştirilmiştir. Bu durumda Test 1’i alan cevaplayıcıların diğer testi almaları durumunda hangi örtük yetenek sınıflarına atanacakları belirlenmiştir.

Araştırma sonuçları incelendiğinde; ankor madde kullanılarak eşitlenen testlerde iki test örnekleminin aynı sınıflama ölçeğinde değerlendirilmesine yönelik

bulgulara ulařılmıştır. Örtük yetenek sınıflarının eşitlenmesi sonucunda grubun %39'u aynı sınıfa yerleşmiş, %31'i dahil olduđu sınıftan daha üst sınıfta yani daha fazla özelliđe sahip olanların yer aldığı sınıflara atanmıştır. Grubun %22'si ise dahil olduđu sınıftan daha düşük sınıflarda yer almıştır. Arařtırmada örtük yetenek sınıflarının eşitlenmesinin yanı sıra testlerin psikometrik özellikleri, model veri uyumu, sonsal olasılıkları, eşitleme sonrası özelliklere sahip olup olmama durumları incelenmiştir. Arařtırma bulgularına göre BTM çalışmalarında örtük sınıflar için ankor maddeler kullanılarak eşitleme çalışması yapılabileceđine ilişkin istatistiksel kanıtlara ulařılmıştır.

Anahtar Kelimeler: Ankor Madde, Bilişsel Tanı Modelleri, DINA Model, Test Eşitleme, Örtük Sınıf Deđişmezliđi

ABSTRACT

EQUATING THE LATENT ABILITY CLASSES WITH ITEM PARAMETERS IN COGNITIVE DIAGNOSTIC MODELS

CANPOLAT, Arzu

MS, Department of Measurement and Evaluation

Advisor: Assoc. Prof. Dr. Tahsin Oğuz Başokçu

In this study, it is aimed to equalize the latent trait classes obtained from two tests developed with Cognitive Diagnostic Models (CDM).

The study within the scope of “A Recommended Model to Increase Success Level of Turkey in Mathematics in International Wide Scale Exams. Effectiveness of the Cognitive Diagnosis Based Tracking Model” TUBITAK number 115K531 project was conducted with the project’s sample and measurement instruments. The sample of the study consisted of 992 students from 6th, 7th and 8th grades in 12 schools affiliated to İzmir Provincial Ministry of Education. To this group, tests prepared by field teachers in according to 6th grade mathematics curriculum were applied. In each test, there are a total of 15 items, multiple choice and open-ended and, these tests are containing Q matrices that have 4 attributes representing these items based on experts’ opinions.

In the study, a method for determining; in case of the respondents have taken the other test, the latent trait classes to be assigned, with anchor items according to CDM parameters, has been proposed. CDM analyzes were performed with the DINA model. Data analysis was performed in three stages. First of all, the answer patterns and the latent classes of the respondents to whom Test 1 was applied were determined. Then, simulation data were generated by using item parameters obtained from Test 1 and Test 2, and analyzed with the DINA model. Finally, the latent classes of data that have the same pattern as the first test were matched. In this case, it was determined which latent trait classes will be assigned if the respondents taking Test 1, take the other test.

When the research results were examined; in the tests which were equalized by using the anchor items, the results of the two test samples were evaluated at the

same classification scale. As a result of the equalization of latent classes, 39% were placed in the same classes, 31% were in the higher classes than the classes they were included, ie in the class with more attributes. 22% of the group was in the lower classes. In this study, in addition to equalization of latent trait classes, psychometric properties of the tests, model data fit, posterior probabilities, and whether or not the students have post-synchronization attributes were examined. According to the findings of the study, statistical evidence was obtained that an equalization study could be performed by using anchor items for latent classes in CDM studies.

Keywords: Anchor Items, Cognitive Diagnostic Models, DINA Model, Test Equating, Latent Class Invariance

BÖLÜM I

GİRİŞ

Anastasi (1988), bir testi belirli davranışların nesnel ve standartlaştırılmış ölçümü olarak tanımlar. Testler eğitim kademelerinin hemen hemen her basamağında başarıyı ölçmek, kurumlara eleman yerleştirmek, tutum ve ilgi ölçmek ve benzeri sebeplerle çeşitli alanlarda kullanılmaktadır. Testlerin eğitimde kullanılma amacı öğrencilerin öğrenme eksiklerini gidermek, öğrenme düzeylerini belirleyerek başarılarını ortaya koymak ve elde edilen sonuçlara göre aksaklıkları giderici önlemler almaktır (Özçelik, 1997). Testler ve testlerin yapıları birçok kuramın ana konusunu oluşturmaktadır.

Geleneksel değerlendirme yöntemlerinden biri olan Klasik Test Kuramı (KTK), gerçek puanı gözlenen puandan kestirmektedir (Baykul, 2000). KTK’da madde ve test istatistikleri, testin güçlüğü testin uygulandığı örnekleme, cevaplayıcıların yetenekleri ise uygulanan teste bağlıdır (Lord, 1980; Crocker ve Algina, 1986). Cevaplayıcılara testten aldıkları toplam puan rapor edilir fakat eksik ya da güçlü oldukları konular hakkında bilgi verilmez.

KTK’daki sınırlılıklar göz önüne alınarak geliştirilen Madde Tepki Kuramı (MTK), testle ölçüldüğü düşünülen örtük özellikler ile bireyin cevabı arasındaki doğrusal ilişkiyi ortaya koymaya çalışır (Hambleton, 1989). Madde parametreleri kestirimi örneklemeden bağımsızken yetenek parametresi kestirimleri de maddelerden bağımsızdır (Lord ve Novick, 1968). Bu kuramda KTK’nın aksine test istatistikleri gruba bağlı değildir. MTK sonucu elde edilen puan ile her bir öğrenci hakkındaki elde edilen bilgiler sınırlıdır. Ölçülen konu hakkında güçlü ve zayıf olduğu yönler hakkında detaylı bilgi vermemektedir (Snow ve Lohman, 1989).

Özetle, geleneksel değerlendirme cevaplayıcıların öğrenme kapasiteleri hakkında bir fikir vermez (Embretson, 1983), dolayısıyla öğrenci eksiklerini gidermekte eksik kalmaktadır. KTK ve MTK analizleri kişilerin başarıları hakkında bilgi verse de başarı seviyelerini açıklayan özellikler profili hakkında bilgi vermez (Rupp, Mislevy, 2007). Değerlendirmenin etkili ve anlamlı olması için öğretmeyi ve öğrenmeyi desteklemelidir (DiBello ve Stout, 2007). Değerlendirme öğretmenlere, öğrencilere nelerin öğrenilip nelerin öğrenilemediği konusunda bilgi verici olmalıdır

(Tanner ve Jones, 2003). Yeterli dönüt sağlanması durumunda değerlendirmelerin öğrencileri motive etme konusunda yardımcı olduğu ve kendi öğrenmelerinin kontrolünü sağlamaya yönelttiği belirlenmiştir.

KTK ve MTK dışında çeşitli psikometrik modeller geliştirilmiştir. Faktör analizi (Christoffersson, 1975; Mislevy, 1986; McDonald, 1999), Latent trait model (Gustafsson, 1980; Molenaar, 1983), Örtük sınıf modelleri (Bergan, 1983; Haertel, 1984a, 1984b; Macready ve Dayton, 1977; Traub ve Lam, 1985) gibi modeller bazı psikometrik modeller arasındadır. Örtük sınıf modellerinin eğitimde ölçmede kullanımı Macready ve Dayton (1977) zamanında yapılmıştır. Örtük sınıf analizinin sınırlandırılmış hali olan Bilişsel Tanı Modelleri (BTM) eğitim, psikoloji gibi alanlarda önemli bir yere sahiptir.

BTM temel alınarak geliştirilmiş bir testte, bir toplam puan ya da toplam alt ölçek puanları yerine, sınavı alan her bir bireyin hangi becerilere sahip olduğu ve hangileri konusunda eksiklikleri bulunduğu belirlendiği bir ölçüm yapılır. Dolayısıyla geliştirilen bu testler hem değerlendirme sürecine hem de her bir öğrencinin eğitim ihtiyacını belirleme konusuna hizmet eder (Cheng ve Chang, 2007).

BTM'ye olan ihtiyaç ABD'de gerçekleştirilen "Hiçbir Çocuk Geride Kalmasın Reformu" (No Child Left Behind Act) ile artmıştır. Bu reformun sebepleri arasında ABD'de 4. sınıf öğrencilerinin %70'inin ulusal okuma testlerinde oldukça geride kalması, lise son sınıftaki öğrencilerin uluslararası matematik testlerindeki başarılarının oldukça düşük olması, kolejlerdeki birinci sınıf öğrencilerinin çoğunun derslere başlamadan hazırlayıcı eğitim ihtiyacı duyması gibi sebepler gösterilebilir. Bu gibi durumlar göz önüne alınarak öğrencilerin başarılı ve başarısız oldukları yönlerin rapor haline getirilip bu bilişsel tanı raporları ile öğrencilerin, ailelerin ve öğretmenlerin bilgilendirilmeleri sağlanmıştır (The White House, 2003) ve bu durum modellerin gelişmesini hızlandırmıştır.

Ülkemizde kamu görevlerine atama yapmak için KPSS (Kamu Personel Seçme Sınavı), yabancı dil seviyesini ölçmek için YDS (Yabancı Dil Bilgisi Seviye Tespit Sınavı), lisansüstü başvurularında kullanmak, yüksek öğretim kurumlarına atama yapmak için ALES (Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı), yüksek öğretim kurumlarına yerleştirme yapmak için YKS (Yükseköğretim

Kurumları Sınavı) ve birçok farklı alanda çeşitli sebeplerle sınavlar uygulanmaktadır. Uluslararası düzeyde ise PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) gibi büyük ölçekli sınavlar uygulanmaktadır. Bu uygulamalar göz önüne alındığında testlerin, ülkelerin eğitim alanında hangi düzeyde olduğunun belirlenmesi, giderilmesi gereken eksikliklerin ve alınması gereken tedbirlerin belirlenmesi konusunda önemli bir yere sahip olduğu görülmektedir.

Geniş ölçekli ve merkezi sınavlarda güvenliği sağlamak amacıyla öğrencilere her dönem ya da her yıl farklı sorular sorulmaktadır. Bunun için aynı amaçla uygulanan testlerin pek çok formu geliştirilir. Cook ve Eignor (1991), bu formların benzer içerikte olmakla birlikte testlerin güçlüklerinin farklı olabileceğini ve bu durumun bazı cevaplayıcıların daha kolay ya da daha güvenilir bir test olarak avantajlı konuma geçmesine sebep olabileceğini belirtmiştir.

Örneğin bir X sınavındaki başarılarına göre eleman alan bir kuruma iş başvurusunda bulunan iki kişiyi göz önüne alalım. Bu kişiler farklı yıllarda girdikleri sınavlardan elde ettikleri puanlar ile başvuruda bulunmuş olsunlar. A kişinin B kişisinden daha yüksek puan almış olmasını sadece B kişisinden daha fazla bilgiye sahip olması olarak kısıtlayamayız. Bir başka ihtimal A kişinin girdiği testin aynı özellikleri ölçmesine rağmen B'ye uygulanan teste göre daha kolay olmasıdır. İnsan hayatında kritik bir öneme sahip sınavların adil olması beklenmektedir. Bunu sağlamak için aynı özelliği ölçen farklı maddelerden oluşan testlerin aynı ölçek üzerinde gösterilmesi ve cevaplayıcıların buna göre değerlendirilmesi gerekmektedir. Bu işlem, Angoff'un (1971) bir testin birim sistemini başka bir testin birim sistemine dönüştürülmesi olarak adlandırdığı test eşitleme ile sağlanmaktadır. Test eşitleme başarı ile gerçekleştiğinde cevaplayıcıların verilen zaman aralığındaki performanslarındaki değişim tanımlanabilir ya da yılın belirli zamanlarında uygulanan sınavlara giren kişilerin puanları karşılaştırılabilir (Angoff, 1971).

Testin değerlendirilmesi ve test eşitleme süreçleri birlikte değerlendirildiğinde sürekli değişkenlerin aynı ölçek üzerinde değerlendirilmesinin yanı sıra öğrencinin neyi bilip bilmediğini kategorik olarak gösteren örtük yetenek sınıflarının eşitlenmesi de üzerinde durulması gereken bir konudur. İlgili alan yazın incelendiğinde test eşitlemenin KTK ve MTK temelli çeşitli yöntemlerle sürekli

değişkenler üzerinden yapıldığı görülmektedir. Örtük yetenek sınıflarının eşitlenmesine yönelik bir çalışmaya rastlanmamıştır. Bu çalışmada kategorik olarak belirlenen örtük yetenek sınıflarının BTM ile eşitlenmesine yönelik bir yöntem geliştirilmiştir.

1.1. Problem Cümlesi

Bilişsel tanı modelleri temel alınarak ve ortak test desenine uygun olarak hazırlanan testlerin eşitlenmesi çalışmasında öğrencilerin ait oldukları örtük yetenek sınıflarındaki değişim nasıldır?

1.2. Alt Problemler

Araştırma kapsamında problem cümlesi doğrultusunda incelenecek alt problemler şunlardır:

1- Test 1 ve Test 2'nin madde analizi sonuçları nasıldır?

1.1. Testlerin KTK'ya göre psikometrik özellikleri nasıldır?

1.2. Testlerin BTM'ye göre model veri uyumu parametreleri nasıldır?

2- Test 1'e ilişkin örtük yetenek sınıflarının ve özelliklerin dağılımı nasıldır?

3- Test 2'ye ilişkin örtük yetenek sınıflarının ve özelliklerin dağılımı nasıldır?

4- Test 1'i alan öğrencilerin Test 2 için eşitlenmiş örtük yetenek sınıfları nasıl değişmektedir?

5- Eşitleme sonrası örtük yetenek sınıflarının toplam puanlara göre değişimi nasıldır?

6- Eşitleme sonrası örtük yetenek sınıflarının MTK puanlarına göre değişimi nasıldır?

7- Ölçülen özelliklere göre örtük yetenek sınıflarının eşitlenmesi sonrası öğrencilerin özelliklere sahip olup olmama durumları ne düzeyde değişmektedir?

7.1. Öğrencilerin İletişim ve İlişkilendirme özelliğine sahip olup olmama durumu eşitleme sonrasında ne düzeyde değişmektedir?

7.2. Öğrencilerin Matematikleştirme özelliğine sahip olup olmama durumu eşitleme sonrasında ne düzeyde değişmektedir?

7.3. Öğrencilerin Akıl Yürütme ve Strateji Geliştirme özelliğine sahip olup olmama durumu eşitleme sonrasında ne düzeyde değişmektedir?

7.4. Öğrencilerin Sembolik ve Teknik Dil Kullanımı özelliğine sahip olup olmama durumu eşitleme sonrasında ne düzeyde değişmektedir?

1.3. Sayıtlar

1. Araştırma kapsamında uygulanan testlerde öğrenciler gerçek performanslarını göstermişlerdir.
2. Testler için uzman kanılarıyla belirlenen Q matrisler belirlenen özellikleri temsil etmektedir.

1.4. Sınırlılıklar

1. Bu tez çalışması örnekleme ve veri toplama araçları 115K531 No'lu TÜBİTAK projesi kapsamında geliştirilen deneme uygulaması verileri ile sınırlıdır.
2. Deneme uygulaması maddeleri matematik alanında hazırlanmıştır ve 6. sınıf düzeyindedir.
3. DINA modelin kullanılması araştırmanın bir diğer sınırlılığını oluşturmaktadır.

1.5. Araştırmanın Amaç ve Önemi

Eğitimde ve psikolojide kullanılan testlerin değerlendirme aşaması önemlidir. Değerlendirmelerin sınıfın durumu ve öğrenmeleri hakkında bilgi sağlaması beklenir, bunun için de yorumlayıcı, tanılayıcı, bilgi verici ve öngörülü olması gerekmektedir (Pellegrino, Baxter ve Glaser, 1999).

KTK ve MTK gibi kuramlar cevaplayıcıların güçlük yaşadığı ya da iyi olduğu konuları belirlemede yetersiz kalmaktadır. Temelinde örtük sınıf analizi olan BTM'ler, özellikle cevaplayıcıların belirli bir alandaki becerilerinin varlık ya da yokluğunu belirlemek amacıyla geliştirilmiştir. Sonuçların sürekli değil de kategorik olarak verilmesi cevaplayıcılara daha fazla bilgi sağlamaktadır. Bu çalışmada BTM'lerden biri olan DINA model kullanılmıştır. Bu model telafi edici olmayan doğası ile matematik alanında kullanılmaya uygun bir modeldir. Kullanımı ve yorumlanması basit bir model olan DINA model ile yapılan çalışmaların çoğu simülasyon çalışmalarından ya da daha önceden hazırlanmış Q matris kullanılarak ve buna uygun sorular yazılarak yapılan çalışmalardan (Rupp ve Templin, 2008; de la Torre ve Karelitz, 2009; Cheng, 2010; Chen, Xin, Wang ve Chang, 2012) oluşmaktadır. Bazı araştırmalar ise geniş ölçekli sınavların dataları kullanılarak yürütülmüştür (Dogan ve Tatsuoka, 2008; Gierl, Cui ve Zhou, 2009; Birenbaum, Tatsuoka ve Yamada, 2004). Bu çalışmada kullanılan Q matris, BTM test geliştirme mantığına uygun olarak hazırlanmıştır ve maddeler, ölçülen özellikler ve Q matris

aynı anda oluşturulmuştur. Bu araştırmada kullanılan ölçme araçlarının DINA model temel alınarak geliştirilmesi ve gerçek veri üzerinden yürütülmüş olması önem arz etmektedir.

Araştırmada kullanılan testlerde çoktan seçmeli maddelerin yanı sıra doğru yanlış, kısa cevaplı ve açık uçlu maddelere de yer verilmiştir. Maddelerin çoğu rutin olmayan problemlerden oluşmaktadır. Bu problemler matematik programında oldukça önemli bir yer tutmaktadır, öğrencilere öğrendikleri matematiğin gerçek dünya durumlarına uygulamayı öğretmektedir (Verschaffel, De Corte ve Vierstraete, 1999).

Cevaplayıcılara uygulanan testler benzer içerikte olmakla birlikte güçlükleri farklı olabilir, bu durum bazı cevaplayıcıların daha kolay ya da daha güvenilir bir test olarak avantajlı konuma geçmesine sebep olabilir (Cook ve Eignor, 1991). Test eşitleme farklı test formlarını alan bireyler arası yanlılığı önlemeyi ve farklı formlardan alınan puanların aynı ölçek üzerinde rapor edilmesi olarak tanımlanmıştır (Barnard, 1996). İlgili literatür incelendiğinde KTK'ya ve MTK'ya dayalı test eşitleme yöntemleri kullanılarak sürekli değişkenlerin eşitlenmesine yönelik birçok araştırma mevcutken BTM ve test eşitleme birlikte göz önüne alındığında kategorik değişkenlerin eşitlenmesine dair bir araştırmaya rastlanmamaktadır. Bu araştırmada eşitleme çalışması kategorik değişkenler üzerinden gerçekleştirilmiştir.

Yukarıdaki bilgiler ışığında araştırmada BTM üzerinden yetenek kestirimi eşitlemesi ilk kez denenmiş olup ankor maddeler kullanılarak hazırlanan testlerin örtük yetenek sınıflarının aynı sınıflama ölçeğinde değerlendirilmesine yönelik bir çalışma yapılmıştır. Burada istatistiksel süreçler değil yöntemin kendisi test edilmektedir. Bu araştırma ile birlikte farklı BTM modellerinde de kullanılacak bir yöntem, yaklaşım geliştirilmeye çalışılmıştır. Bu alana getirilebilecek yeni bir bakış açısı olarak değerlendirilebilir. Araştırmanın benzerlerine yönelik bir çalışma literatürde bulunmamaktadır ve ilgili modeller için ilk çalışma olması araştırmanın önemini artırmaktadır. Araştırma sonucunda elde edilen bulguların, bu alandaki eksikliğin giderilmesinde önemli bir adım olması ve BTM ile ilgili çalışmalara katkı sağlayabileceği düşünülmektedir.

BÖLÜM II

İLGİLİ YAYIN VE ARAŞTIRMALAR

2.1. Bilişsel Tanı Modelleri Genel Özellikleri

Bilişsel tanılama son zamanlarda eğitimde, psikolojik değerlendirmelerde ve birçok farklı alanlarda önem kazanmaktadır (Tatsuoka, 2009; Rupp, Templin ve Henson, 2010). Temel amacı, cevaplayıcıların maddelere verdikleri yanıtlar, madde parametreleri, madde-özellik ilişkileri gibi madde bilgileri aracılığıyla cevaplayıcıların örtük profillerini belirlemektir.

BTM modelleri cevaplayıcıların yeteneklerini maddeleri doğru cevaplayabilmek için gerekli özelliklere sahip olup olmama durumlarına göre kategorik olarak belirler. Kişilerin performansları hakkında detaylı bilgi verir ve uygun sınıflara atar (de la Torre, 2009).

BTM, cevaplayıcının önceden belirlenmiş beceri ya da yeteneklere sahip olup olmadığına ilişkin bilgi verir. Böylelikle bireyler hakkında daha zengin, anlamlı, yönlendirici bilgiler sağlanabilir. Ayrıca BTM, cevaplayıcıların bir maddeyi çözmeye neden başarılı olamadıklarını da açıklar (Henson, Templin ve Willse, 2009).

Literatür incelendiğinde BTM'ler sınırlandırılmış örtük özellik modelleri (restricted latent class models) (Haertel, 1989), bilişsel tanı modelleri (cognitive diagnosis models) (Nichols, Chipman ve Brennan, 1995), bilişsel psikometrik modeller (cognitive psychometric models) (Rupp, 2007), yapısal madde tepki kuramı modelleri (Rupp ve Misleavy, 2007) gibi isimlerde karşımıza çıkmaktadır. Türkiye'de modelleri ilk kullanan Başokçu (2011), çalışmalarında bilişsel tanı modelleri olarak kullanmaktadır. Bu araştırmada da bu isim kullanılacaktır.

BTM'de analizden önce özellikler ile maddeler arasındaki ilişkilerin belirlenmesi gerekir. Modellerde bu süreç madde özellik matrisi olan Q matrisi ile tanımlanır.

2.1.1. Q Matrisi

BTM'de maddeler ve maddelerin ölçtüğü özellikler arasındaki ilişki Q matrisi ile kurulmaktadır (Embretson, 1984; Tatsuoka, 1985). Q matrisinde her bir satır

maddeleri, her bir sütun ise özellikleri temsil eder. J, hazırlanan testteki toplam madde sayısı; K, bu test ile ölçülen toplam özellik sayısı olmak üzere JxK'lık Q matrisin j. satır, k. sütun elemanı q_{jk} ile gösterilir. Matris elemanlarına değer ataması aşağıdaki gibi yapılır:

$$q_{jk} = \begin{cases} 1, & \text{Eğer } j \text{ maddesini doğru yanıtlamak için } k \text{ özelliği gerekiyorsa} \\ 0, & \text{Diğer durumlar} \end{cases}$$

Eşitlikte de gösterildiği gibi matris elemanları '1' ve '0' lardan oluşur. Bu ikili numaralandırmayı Fischer (1973) ağırlıklandırma olarak adlandırmıştır. Kısaca Q matris, maddelerin doğru cevaplanması için hangi özelliklerin gerekip hangilerinin gerekmediğini tanımlar. Tablo 2'de α_1 , α_2 , ve α_3 olarak adlandırılan 3 özelliğin ölçüldüğü 5 maddelik bir teste ilişkin Q matris örneği verilmiştir.

Tablo 1. *Q Matris Örneği*

Maddeler	Özellikler		
	α_1	α_2	α_3
1	1	0	0
2	0	1	1
3	1	1	1
4	0	0	1
5	1	1	0

Tabloya göre birinci maddenin doğru yanıtlanabilmesi için sadece α_1 özelliğine sahip olmak yeterlidir. Üçüncü maddenin doğru yanıtlanabilmesi için α_1 , α_2 ve α_3 özelliklerinin tamamına sahip olunması gerekmektedir. Q matris cevaplayıcının örtük özellik sınıfı kestirimi sürecini doğrudan etkiler (Henson, 2004), bu nedenle BTM'de madde özellik ilişkilerinin doğru belirlenmesi oldukça önemlidir.

2.1.2. Örtük Yetenek Sınıfı

Cevaplayıcıların test için belirlenen özelliklere sahip olup olmama durumlarının gösterildiği örüntülere örtük yetenek sınıfı adı verilir. Yetenekler; beceri, özellik gibi isimlerle de anılmaktadır (Rupp ve diğ., 2010).

Genel olarak BTM'ler bir test için belirlenen K tane özellik için her bir cevaplayıcıyı ikili özellik sınıflarına atar. I, testi alan toplam kişi sayısı; J testteki toplam madde sayısı; K, test için belirlenen özelliklerin toplam sayısı olmak üzere X_{ij} , i cevaplayıcısının j maddesine verdiği cevaptır. $\alpha_i = \{\alpha_{ik}\}$, i cevaplayıcısının ikili özellik vektörüdür. Burada i cevaplayıcısı k özelliğine sahip olduğu durumlarda $\alpha_{ik} = 1$, k özelliğe sahip olmadığı durumlarda $\alpha_{ik} = 0$ değerini almaktadır.

$$\alpha_{ik} = \begin{cases} 1, & \text{Cevaplayıcı k. özelliğe sahip ise} \\ 0, & \text{Diğer durumda} \end{cases}$$

Cevaplayıcıların belirlenen özelliklere sahip olup olmama durumu olasılıkla temellendirilir. Bu olasılık değerleri, kesme puanı ile karşılaştırılarak sahip olup olmama durumlarına göre ikili puanlara dönüştürülebilir. Genel olarak 0,50 eşiği temel alınır. Olasılık değeri 0,50'den büyük veya eşitse cevaplayıcının o özelliğe sahip olduğu varsayılır ve 1 değeri atanır, eğer 0,50'den küçük ise o özelliğe sahip olmadığı varsayılır ve 0 değeri atanır (de la Torre, Hong ve Deng, 2010; Templin ve Henson, 2006).

Q matriste J madde için tanımlanan K özellik için toplamda 2^K örtük sınıf belirlenir ve cevaplayıcılar bu sınıflara atanırlar. Örneğin 3 özelliğin tanımlandığı bir test için $2^3 = 8$ adet örtük sınıfı belirlenir. Bu sınıflar 000, 001, 010, 100, 110, 101, 011 ve 111 sınıflarıdır. Örneğin 1.cevaplayıcının birinci ve ikinci özelliğe sahip olduğu, üçüncü özelliğe sahip olmadığı durumda bu cevaplayıcının ikili özellik vektörü $\alpha_1 = (\alpha_{11}, \alpha_{12}, \alpha_{13}) = 110$ şeklinde belirlenir.

2.2. Başlıca Bilişsel Tanı Modelleri

Rule space model (Tatsuoka, 1983), İkili Yetenek Modeli (Binary skills model) (Haertel, 1984b; Haertel ve Wiley, 1993), Bayesian inference network (Mislevy, Almond, Yan ve Steinberg, 1999), Noisy Input; Deterministic 'And' Gate (NIDA) model (Maris, 1999), the fusion model (Hartz, 2002; Hartz, Roussos ve Stout, 2002), Deterministic Input Noisy "And" Gate (DINA) model (Macready ve Dayton, 1977; Haertel, 1989; Junker ve Sijtsma, 2001) gibi birçok BTM modeli vardır.

BTM'de cevaplayıcıların sınıflamalarını doğrudan etkilemesi sebebiyle doğru modelin seçilmesi oldukça önemlidir (Lee ve Sawaki, 2009b). BTM'ler telafi edici

modeller ve telafi edici olmayan modeller olarak iki kategoride incelenebilir. Arařtırmacıların kullanacađı modelin telafi edici olup olmadıđını belirlemeleri önemli bir noktadır. Telafi edici modellerde, görevde başarılı bir performans için bir becerideki düşük seviye yeterlik başka bir beceri üzerindeki yüksek seviyedeki yeterlik ile telafi edilebilmektedir. Telafi edici olmayan modellerde ise bir özelliđe yüksek ölçüde sahip olmak başka bir özelliđe düşük derecede sahip olmayı telafi etmez. Örneđin matematikte çözümler genelde işlem adımlarına bölünür. Matematik problemini doğru bir şekilde çözebilmek için bu işlem basamaklarının hepsinde başarılı performans sergilemek gerekir. Dolayısıyla genel olarak matematik testleri için telafi edici olmayan modellerin daha uygun olduđu kabul edilir (Roussos, Templin ve Henson, 2007). Bilişsel yetenekler arası ilişkiler tam olarak bilinmediğinde ise doymuş modellerden biri kullanılabilir. Bu modeller nitelikler arasındaki farklı ilişkileri karşılamak adına daha esnektir ve doğru kestirimler elde etmek için daha büyük örneklem gerektirmektedir (Rojas, de la Torre ve Olea, 2012). Tablo 1’de yaygın olarak kullanılan BTM’ler türlerine göre gösterilmiştir.

Tablo 2. BTM Türleri

BTM Türü	Örnek	Yazar
Telafi Edici	1) DINO (deterministic, input, noisy-or-gate model)	Templin ve Henson (2006)
	2) C-RUM (compensatory reparameterized unified model)	Hartz (2002)
	3) A-CDM (Additive Cognitive Diagnostic Model)	de la Torre (2011)
Telafi Edici Olmayan	1) DINA (deterministic, input, noisy-and-gate model)	Junker ve Sijtsma (2001)
	2) NC-RUM (noncompensatory reparameterized unified model)	Di Bello, Stout ve Roussos (1995)
	3) Rule Space Model	Tatsuoka (1983)
	4) Attribute Hierarchy Model	Leighton, Gierl ve Hunka (2004)
Doymuş	1) GDM (general diagnostic model)	Von Davier (2005)
	2)LCDM (log-linear CDM)	Henson, Templin ve Willse (2009)
	3) G-DINA (generalized DINA)	de la Torre (2011)

BTM'ler genel BTM'ler ve bu modellerin parametrelerinin sınırlandırılması veya yeni parametreler eklenmesi ile elde edilen özel BTM'ler olarak da sınıflandırılabilir (von Davier, 2005; Roussos ve diğ., 2007; Rupp ve Mislevy, 2007; de la Torre, 2011; de la Torre ve Lee, 2013). Özel BTM'lere DINA (Haertel, 1989, Junker ve Sijtsma, 2001), DINO (Templin ve Henson, 2006), NIDA (Junker ve Sijtsma, 2001), R-RUM (Hartz, 2002) ve genel BTM'lere GDM (von Davier, 2005), LCDM (Henson ve diğ., 2009), G-DINA (de la Torre, 2011) örnek verilebilir. Genel BTM modelleri aynı testte hem telafi edici hem de telafi edici olmayan ilişkilere izin

verir. Bir testte her bir madde için tek bir model kullanmak yerine her madde için uygun olan modelin seçilmesini sağlar. Özel BTM'ler ya telafi edici ya da telafi edici olmayan modele izin verir. Bu tez kapsamında DINA, G-DINA ve DINO modelden bahsedilecektir.

2.2.1. DINA (Deterministic Inputs Noisy And Gate) Model

DINA model (Macready ve Dayton, 1977; Haertel, 1989; Junker ve Sijtsma, 2001), BTM'ler arasında yaygın olarak kullanılan modellerden biridir. Bu model, cevaplayıcıların yeteneklerini örtük sınıflara ayırarak belirleyen bir örtük sınıf analizidir (Haertel, 1989). Modelde test için belirlenen K tane özellik için 2^K örtük sınıf belirlenir.

Telafi edici olmayan bu modelde her madde cevaplayıcıları iki farklı sınıfa ayırır. Birincisi o maddeyi doğru yanıtlamak için gerekli tüm özelliklere sahip olanların oluşturduğu tam sınıf, ikincisi ise maddeyi doğru yanıtlamak için gerekli özelliklerden en az birine sahip olmayanların oluşturduğu yokluk sınıfıdır (de la Torre, 2009). Aynı sınıfa ait cevaplayıcıların o maddeyi doğru cevaplama olasılıkları birbirine eşittir.

DINA modelde madde özellik ilişkisi diğer modellerde olduğu gibi Q matris (Embretson, 1984; Tatsuoka, 1985) ile belirlenir. Örtük cevap vektörü olan $\eta_i = \{\eta_{ij}\}$, cevaplayıcının yetenek vektörü ve Q matris tarafından üretilir ve

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$$

olarak ifade edilir. Burada $\alpha_i = \{\alpha_{ik}\}$ i cevaplayıcısının ikili örtük yetenek vektörüdür.

η_{ij} , i cevaplayıcısının j maddesi için gerekli özelliklere sahip olup olmadığını gösterir. $\eta_{ij} = 1$ değerini aldığı anda i cevaplayıcısının j maddesi için gerekli özelliklerden hepsine sahip olduğu, $\eta_{ij} = 0$ değerini aldığı anda ise i cevaplayıcısının j maddesi için gerekli özelliklerden en az birine sahip olmadığı anlaşılır (de la Torre, 2009).

DINA model, göz önüne alınan özelliklerin sayısına bakmaksızın her bir madde için s_j ve g_j olmak üzere iki parametre üretir (Junker ve Sijtsma, 2001). Kaydırma (slip) parametresi olan s_j , i cevaplayıcısının j maddesi için gerekli

özelliklerin tamamına sahip olmasına rağmen o maddeyi yanlış cevaplama olasılığıdır. Tahmin (guess) parametresi g_j ise i cevaplayıcısının j maddesi için gerekli özelliklerden en az birine sahip olmamasına rağmen maddeyi doğru yanıtlama olasılığıdır. s_j ve g_j şu şekilde tanımlanır:

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$$

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$$

Maris (1999), tahmin parametresi olan g_j 'ye alternatif bir yorum getirmiştir ve maddeyi doğru yanıtlayabilmek için gerekli özelliklere sahip olmayanların doğru yanıtlanması dışında o maddeyi doğru yanıtlayabilmek için gerekli özelliklerden farklı bir özellik kullanılarak doğru cevap verilmesi olarak da yorumlanabileceğini ifade etmiştir. Dolayısıyla g_j parametresin yüksek çıkması durumunda maddeyi Q matriste belirlenen özellikler dışında başka bir özellik kullanarak yanıtlama durumu söz konusu olabilir. s_j parametresinin düşük olması o maddeyi doğru yanıtlamak için gerekli tüm özelliklere sahip cevaplayıcıların doğru cevaplama olasılığının yüksek olması demektir.

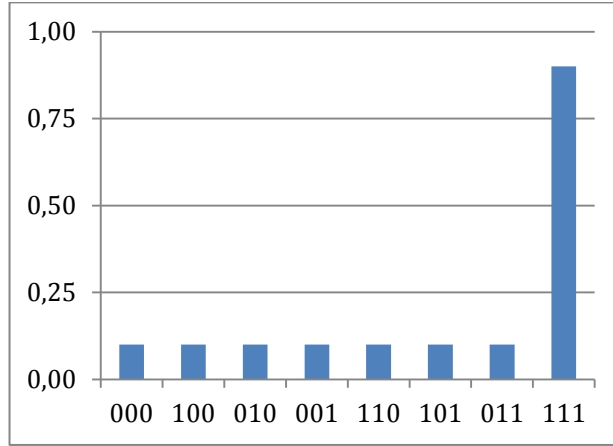
DINA modelde α_i özellik vektörüne sahip i cevaplayıcısının j maddesini doğru cevaplama olasılığını gösteren madde tepki fonksiyonu aşağıdaki gibidir:

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = \begin{cases} 1 - s_j, & \eta_{ij} = 1 \text{ ise} \\ g_j, & \eta_{ij} = 0 \text{ ise} \end{cases}$$

$$= g_j^{1-\eta_{ij}}(1 - s_j)^{\eta_{ij}}$$

Örneğin üç özelliği ölçen, g ve s değerleri 0,1 olan bir j maddesine ait örtük yetenek sınıfları ve bu sınıflardaki bireylerin maddeyi doğru cevaplama olasılıkları Tablo 3'teki gibi olur. Tabloda yatay ekseninde örtük sınıflar, dikey ekseninde maddeyi doğru cevaplama olasılık değerleri yer almaktadır.

Tablo 3. DINA İçin Örtük Sınıf Olasılıkları



DINA modelde konular arasındaki bağımsızlığı gösteren koşullu bağımsızlık

$$L(s, g; \alpha) = \prod_{i=1}^N \prod_{j=1}^J (s_j^{1-y_{ij}} (1-s_j)^{y_{ij}})^{\eta_{ij}} (g_j^{y_{ij}} (1-g_j)^{1-y_{ij}})^{1-\eta_{ij}}$$

olarak ifade edilir.

DINA modelin telafi edici olmayan doğasından dolayı maddenin doğru yanıtlanabilmesi için gerekli işlem adımlarında başarılı bir performans gereken matematik gibi alanlarda kullanılması uygundur (Roussos ve diğ., 2007).

2.2.2. G-DINA (Generalized Deterministic Inputs Noisy And Gate) Model

Doygun bir model olan G-DINA model, DINA modelin geliştirilmiş halidir ve de la Torre (2011) tarafından geliştirilmiştir. DINA modelde maddeyi doğru cevaplamak için gerekli özelliklerden bir veya daha fazlasına sahip olmayanların kaç tane özelliğe sahip olup olmadıklarına bakılmaksızın eşit başarı olasılığına sahiptirler ve yokluk sınıfına atanırlar. G-DINA modelde bu kısıtlama yoktur, bunun yerine her bir örtük sınıfın kendisine ait başarı olasılığı vardır. Cevaplayıcının maddenin gerektirdiği özelliklerden herhangi birine ya da fazlasına sahip olma durumuna göre maddeyi doğru yanıtlama olasılığı özelliğin ağırlıklandırılmasına göre değişmektedir.

DINA modelde olduğu gibi madde özellik ilişkileri Q matris (Tatsuoka, 1983) ile belirlenir. DINA model gerekli özellik sayısına bakmaksızın cevaplayıcıları iki gruba ayırırken G-DINA modelde K sayıda örtük yetenek için 2^{K^*} örtük sınıf belirlenmektedir. Burada K_j^* , j maddesini doğru cevaplayabilmek için gerekli olan özelliklerin sayısıdır ve $K_j^* = \sum_{k=1}^K q_{jk}$ ile gösterilmektedir. Her bir örtük grup, α_{ij}^*

indirgenmiş özellik vektörü ile ifade edilir. Modelde bir maddeyi doğru cevaplamak için K_j^* özellik gerekiyorsa o zaman tüm özelliklerin vektörü olan $\alpha_{lj} = (\alpha_{lj1}, \alpha_{lj2}, \dots, \alpha_{ljK})$ yerine $\alpha_{lj}^* = (\alpha_{lj1}, \alpha_{lj2}, \dots, \alpha_{ljK^*})$ indirgenmiş vektörü göz önüne alınır, böylece 2^K olan örtük sınıf sayısı $2^{K_j^*}$ olur. Özellik vektörü α_{lj}^* olan cevaplayıcıların j maddesini doğru cevaplama olasılığı $P(X_j = 1 | \alpha_{lj}^*)$ ile gösterilir.

G-DINA modelin belirlenen özellik ve bu özelliklerin birbiri ile etkileşimlerinin toplam etkisine göre parçalara ayrılabilen olasılık formülü şu şekildedir:

$$P(\alpha_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

Burada δ_{j0} , j maddesi için kesişim,

δ_{j0} , j maddesini doğru yanıtlamak için gerekli hiçbir özelliğe sahip olmayan cevaplayıcının maddeyi doğru yanıtlama olasılığı,

δ_{jk} , α_k üzerindeki temel etki δ_{jk} bir özelliğe uzman olunması durumunda başarı olasılığındaki değişim,

$\delta_{jkk'}$, α_k ve $\alpha_{k'}$ çoklu etkileşim etkisi ,

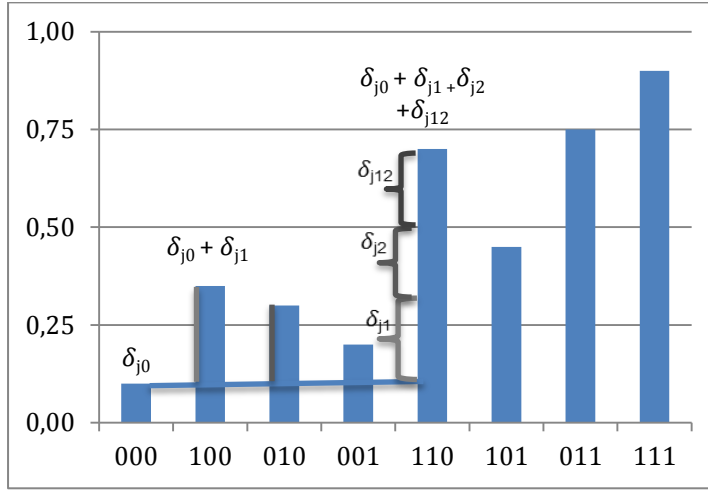
$\delta_{jkk'}$, k ve k' özelliklerine birlikte sahip olunması durumunda başarı olasılığındaki değişime etkisi,

$\delta_{j12\dots K_j^*}$, $\alpha_1 \dots \alpha_{K_j^*}$ etkileşim etkisi,

$\delta_{j12\dots K_j^*}$ maddeyi doğru yanıtlamak için gerekli tüm özelliklere sahip olmanın doğru yanıtlama olasılığındaki değişime etkisidir (de la Torre, 2011).

Örneğin üç özelliği ölçen, g ve s değerleri 0,1 olan bir j maddesine ait örtük yetenek sınıfları ve bu sınıflardaki bireylerin maddeyi doğru cevaplama olasılıkları Tablo 4'teki gibi olur. Tabloda yatay ekseninde örtük sınıflar, dikey ekseninde maddeyi doğru cevaplama olasılık değerleri yer almaktadır.

Tablo 4. *G-DINA İçin Örtük Sınıf Olasılıkları*



Örneğin 110 örtük yetenek sınıfındaki cevaplayıcıların bu maddeyi doğru cevaplama olasılıkları $P(X_j = 1 | \alpha_{ij}^* = (110)) = \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j12}$ ile belirlenir.

G-DINA model parametreleri MMLE kullanılarak kestirilebilir. De la Torre tarafından önerilen algoritma DINA modeldekine benzerdir ve OX Edit programında yazılır.

2.2.3. *DINO (Deterministic Inputs Noisy Or Gate) Model*

DINO model, Templin ve Henson (2006) tarafından geliştirilmiştir. Telafi edici bir model olan bu modelde, bir maddeyi doğru cevaplayabilmek için gerekli özelliklerden en az birine sahip olunması o maddenin doğru cevaplama olasılığının yüksek olması için yeterlidir. DINO model, en az bir özelliğe sahip olunması durumunda doğru cevaplama olasılığının sahip olunan özelliğin türüne ya da sayısına bağlı olmadığını varsayar (Roussos ve diğ., 2007).

DINO model her madde için iki parametre üretir. Tahmin parametresi olan g_j , bir madde için gerekli özelliklerden hiçbirine sahip olmayanların doğru cevaplama olasılığıdır. Kaydırma parametresi s_j , j maddesini doğru yanıtlamak için gerekli tüm özelliklerden en az birine sahip olanların o maddeyi doğru cevaplama olasılığıdır. DINO model için ideal yanıt formülü:

$$\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{g_{jk}}$$

olarak ifade edilir.

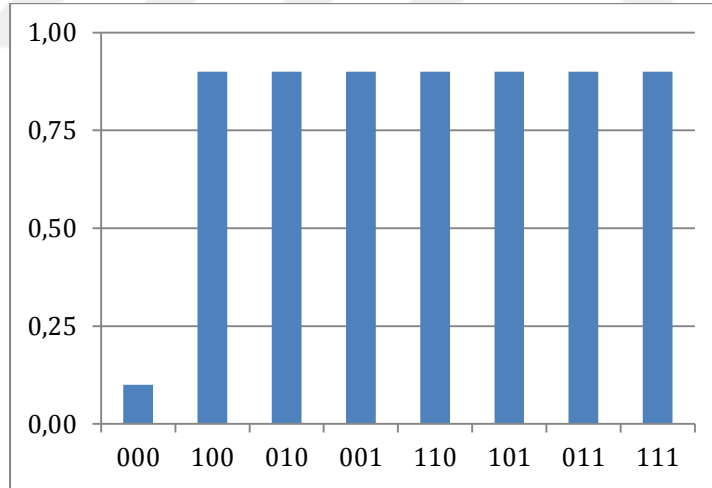
Modele ait madde tepki fonksiyonu aşağıda verilmiştir:

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = \begin{cases} 1 - s_j, & w_{ij} = 1 \text{ ise} \\ g_j, & w_{ij} = 0 \text{ ise} \end{cases}$$

$$= g_j^{1-w_{ij}} (1 - s_j)^{w_{ij}}$$

Herhangi bir j maddesi farklı örtük sınıflarda yer alan cevaplayıcıları gerekli özelliklerden en az birine sahip olanlar ve gerekli özelliklerden hiçbirine sahip olmayanlar olmak üzere iki gruba ayırır. DINA modelin aksine, maddeyi doğru cevaplamak için gerekli hiçbir özelliğe sahip olmayan cevaplayıcılardan oluşan sınıf hariç tüm örtük sınıfların j maddesini doğru cevaplama olasılıkları eşittir (de la Torre, 2011). Örneğin üç özelliği ölçen, g ve s değerleri 0,1 olan bir j maddesine ait örtük yetenek sınıfları ve bu sınıflardaki bireylerin maddeyi doğru cevaplama olasılıkları Tablo 5'teki gibi olur. Tabloda yatay ekseninde örtük sınıflar, dikey ekseninde maddeyi doğru cevaplama olasılık değerleri yer almaktadır.

Tablo 5. DINO İçin Örtük Sınıf Olasılıkları



π_i j maddesi için gerekli özelliklerden en az birine sahip olan cevaplayıcının o maddeye doğru cevap verme olasılığı, r_i j maddesi için gerekli özelliklerden hiçbirine sahip olmayanların o maddeyi doğru cevaplama olasılığı olmak üzere

$$P(X_{ij} = 1 | \alpha_j) = \pi_i^{[1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}]} r_i^{[\prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}]}$$

i cevaplayıcısının j maddesini doğru cevaplama olasılığını vermektedir (Templin ve Henson, 2006).

2.3. Test Eşitleme

Literatürde test eşitlemeye ilişkin birçok tanım yapılmıştır. Angoff (1984) test eşitlemeyi bir test formunun birim sistemini diğer test formunun birim sistemine dönüştürmek olarak tanımlar. Kim ve Hanson (2002), Kolen ve Brennan'a (2004) göre test eşitleme, test formlarının birbiri yerine kullanılmasını sağlamak amaçlı test formlarındaki puanları düzenleyen istatistiksel bir süreçtir. Felan (2002), iki veya daha fazla test formu arasında ilişki kuran istatistiksel bir süreç olarak tanımlamıştır. Braun ve Holland (1982) ise güçlük düzeyleri farklı test formlarından elde edilen puanların birbiri yerine kullanılmasını sağlayan sayısal düzenleme olarak tanımlamaktadır. Test eşitleme farklı test formlarını alan bireyler arası yanlılığı önlemeyi ve farklı formlardan alınan puanların aynı ölçek üzerinde rapor etmeyi amaçlar (Barnard, 1996).

Test eşitleme, yatay eşitleme ve dikey eşitleme olmak üzere iki sınıfa ayrılır (Cook ve Eignor, 1983). Yatay eşitleme test güvenliğini korumak amaçlı, güçlük ve içerik bakımından paralel test formları kullanılacağı zaman uygundur (Cook ve Eignor, 1991). Testin uygulanacağı grubun yetenek dağılımları benzer olmalıdır (Zhu, 1998). Testlerin içeriği, güçlük düzeyi ve güvenilirliği birbirinden çok farklı ise yatay eşitleme uygun olmayacaktır (Kolen, 1988). Dikey eşitleme ise aynı özelliği ölçen farklı güçlükteki test puanlarını eşitlemek için kullanılır. Farklı yetenek seviyelerindeki cevaplayıcıların yetenek dağılımları da farklıdır. Farklı seviyelerdeki formların içerikleri farklı olduğu için ve farklı yetenek düzeyindeki cevaplayıcılara uygulandığı için bu formlardan elde edilen puanlar birbiri yerine kullanılamaz (Barnard, 1996). Kolen (1988) puanların birbiri yerine kullanılmadığı için dikey eşitlemenin eşitleme olarak adlandırılmasına karşı çıkar. Angoff (1971), aynı yapıyı ölçen fakat güçlük veya güvenilirlik bakımından farklılık gösteren testlerin puanlarını bağlamak için kalibrasyon terimini kullanmıştır. Bu tanıma dayanarak dikey eşitlemeyi bir kalibrasyon olarak düşünebiliriz.

Test eşitleme basamakları aşağıda verilmiştir:

1. Eşitlemenin amacına karar verme
2. Benzer içerik ve istatistiksel özelliklere alternatif formlar oluşturma
3. Veri toplama desenini seçme
4. Veri toplama deseninin gerektirdiği şekilde testi uygulama
5. Eşitlemenin bir ya da daha fazla operasyonel tanımını seçme
6. Bir ya da daha fazla istatistiksel kestirim yöntemi seçme
7. Eşitleme sonuçlarını değerlendirme

Test geliştirme süreci testin uygulanması ve istatistiksel süreçler gibi aşamaları kapsar (Kolen ve Brennan, 2004).

2.4. Test Eşitleme Koşulları

Eşitlemenin yapılabilmesi için test formlarının bazı koşulları sağlaması gerekmektedir. Bu koşullar simetriklik, aynı yapıyı ölçme, eşitlik, güvenilirliklerin eşit olması ve gruptan bağımsızlıktır (Dorans ve Holland, 2000).

2.4.1. Simetri Özelliği

Bu özellik, eğer X formundaki ham puan Y formundaki ham puana dönüştürülebiliyorsa Y formundaki ham puanın da X formundaki ham puana dönüştürülmesini gerektirir (Kolen ve Brennan, 2004). Başka bir deyişle, eşitleme puanları, eşitlemenin X formundan Y formuna ya da Y formundan X formuna olmasına bağlı olmaksızın eşittir.

2.4.2. Aynı Yapıyı Ölçme Özelliği

Test formları benzer içerik ve istatistiksel özellikleri ölçecek şekilde hazırlanmalıdır. Aksi takdirde kullanılan istatistiksel süreçlere bakılmaksızın puanlar alternatifli olarak kullanılamazlar (Kolen ve Brennan, 2004). Wolbeck'e (1998) göre eşitlenecek testlerin çok boyutlu olduğu durumlarda eşitleme çalışması alt testlere ait puanlar için ayrı ayrı yürütülmelidir.

2.4.3. Eşitlik Özelliği

Eşitlik özelliği, cevaplayıcının bir testin X formunu ya da Y formunu almasının farklılık oluşturmamasıdır. Lord'un (1980) önerdiği bu özelliğin pratikte sağlanması zordur.

2.4.4. Eşit Güvenirlik

Angoff'a (1971) göre farklı güvenirlige sahip test formlarından elde edilen puanlar birbirleri yerine kullanılamaz ve eşitlenemez. Dorans ve Holland (2000), test eşitlemenin daha iyi sonuç vermesi için testlerin güvenirliklerinin eşit olmasının yanı sıra güvenirliklerinin yüksekliğinin de dikkate alınması gerektiğini belirtmişlerdir.

2.4.5. Gruptan Bağımsızlık

Test eşitleme datadan, eşitleme sürecinin uygulandığı cevaplayıcılardan bağımsız olmalıdır. Eşitleme sonucu elde edilen dönüşüm benzer koşullarda uygulanabilmelidir (Kolen, 1988).

2.5. Eşitleme Desenleri

Test eşitleme koşullarının sağlanmasından sonraki adım test eşitleme desenine karar vermektir. Desen seçilirken dikkat edilecek noktalar uygulanabilirlik ve istatistiksel özelliklerdir (Kolen ve Brennan, 2004). Her birinin avantajı ve dezavantajı olan test eşitleme desenleri üç kısımda incelenebilir. Bu desenler tek grup deseni (single group design), seçkisiz grup deseni (random group design) ve denk olmayan gruplarda ortak madde deseni (nonequivalent anchor tests-NEAT)'dir (Hambleton ve Swaminathan, 1985; Crocker ve Algina, 1986; Kolen ve Brennan, 2004).

2.5.1. Tek Grup Deseni (The Single Group Design)

Tek grup deseni, eşitlenecek olan birden fazla testin aynı gruba uygulandığı test eşitleme desendir. Grubun hedef popülasyonu temsil etme zorunluluğu yoktur (Livingston, 2004). Testler tek bir gruba uygulandığından dolayı ölçme hatası diğer desenlere göre daha düşüktür (Felan, 2002). Puanlar ve puan dağılımlarına ilişkin herhangi bir farklılık, test aynı gruba uygulandığı için testin güçlüğü ile ilişkilendirilir (von Davier, Holland ve Thayer, 2004). Desenin dezavantajı grubun ikinci testteki performansının birinci testteki performansından etkilenmesidir. Bu duruma sıra etkisi denilmektedir. Sıra etkisi ikinci formun uygulanmasında yorgunluk ve aşinalık durumları olarak ortaya çıkabilir. Örneğin X ve Y formlarının uygulandığı bir sınavda yorgunluk kişinin performansını etkilediğinde Y formu X formuna göre daha zor gözükebilir. Aşinalık etkilediğinde ise Y formu X formuna

göre daha kolay gözükabilir (Kolen ve Brennan, 2004). Tek grup deseni Tablo 6’da gösterilmiştir.

Tablo 6. *Tek Grup Deseni*

	X Formu	Y Formu
Grup1	✓	✓

Tek grup deseniindeki sıra etkisini kontrol etmek için dengelenmiş tek grup deseni kullanılabilir. Bu desende kitapçıkların yarısı önce X formu daha sonra Y formu, diğer yarısı ise önce Y formu daha sonra X formunu içerecek şekilde hazırlanır. Cevaplayıcılara uygulanırken ilk kişiye XY sıralamasında, ikinci kişiye YX sıralamasında, üçüncü kişiye XY sıralamasındaki kitapçıklar verilir ve kitapçık dağıtımı benzer şekilde devam eder. Bu süreç, Y formunu önce alan grubun X formunu önce alan grup ile kıyaslanmasını sağlar (Kolen ve Brennan, 2004). Tek grup deseni olduğu gibi bu desende de grupların hedef evreni yansıtmalarına gerek yoktur (Cook ve Eignor, 1991; Livingston, 2004). Dengelenmiş tek grup deseni Tablo 7’de gösterilmiştir.

Tablo 7. *Dengelenmiş Tek Grup Deseni*

	X Formu		Y Formu	
	1. Uygulama	2. Uygulama	1. Uygulama	2. Uygulama
Grup1	✓			✓
Grup2		✓	✓	

2.5.2. *Seçkisiz Grup Deseni (Random Group Design)*

Bu desende eşitlenecek iki test iki farklı gruba ayrı ayrı uygulanır. Kişiler uygulanacak olan testlere random bir şekilde atanır. Sarmal süreç, formları random bir şekilde atarken kullanılabilir bir süreçtir. Sarmal süreçte kitapçıklar dağıtılırken ilk kişi X formunu ikinci kişi Y formunu üçüncü kişi X formunu alır ve dağıtma işlemi benzer şekilde devam eder. Bu süreç X ve Y formunu alan random eşitlenmiş grupların karşılaştırılmasına olanak sağlar. Bu desen kullanıldığında iki form arasındaki grup performans seviyesi farklılığı formlar arası güçlük farkının bir

göstergesidir (Kolen ve Brennan, 2004). İki grup arasındaki eşitlik ilişkileri hedef evrene genellenebilir. Gruplardaki kişiler yalnız bir form alacağı için tek grup deseninde görülen yorgunluk ve aşinalık etkileri giderilmiş olur (Hambleton, Swaminathan ve Rogers, 1991). Desenin dezavantajı ise grubun yetenek dağılımı bakımından benzer olmadığından yanlılık derecesinin belirsizliğidir. Bu sorun daha büyük gruplarla çalışılarak çözülebilir (Zhu, 1998). Seçkisiz grup deseni Tablo 8’de gösterilmiştir.

Tablo 8. *Seçkisiz Grup Deseni*

	X Formu	Y Formu
Grup1	✓	
Grup2		✓

2.5.3. *Denk Olmayan Gruplar Ortak Test Deseni (Nonequivalent Anchor Tests-NEAT)*

NEAT desen, uygulanma tarihlerinde birden fazla testin test güvenliği ya da diğer uygulama sorunlarından dolayı uygulanmadığı durumlarda tercih edilir (Kolen ve Brennan, 2004). Yetenek dağılımı bakımından birbirinden farklı olabilen iki farklı gruba uygulanır. Eşitlenecek test formları ortak maddeler içerir, bu sayede testler arasındaki farklılıklar ortak madde istatistikleri temel alınarak düzeltilebilir (Zhu, 1998). Formlardan elde edilen puanlar arası fark hem grup hem de formlar arası farklılıkları yansıtır. Denk olmayan gruplar ortak test deseni Tablo 9’da gösterilmiştir.

Tablo 9. *Denk Olmayan Gruplar Ortak Test Deseni*

	X Formu	Y Formu	Ortak Form
Grup1	✓		✓
Grup2		✓	✓

Ankor maddeler her iki formda ortak sorulan maddelerdir ve cevaplayıcının toplam puanına etki etme durumuna göre iç ankor ve dış ankor olmak üzere ikiye

ayrılır. İç ankor düzenekte ankor maddeler formların bir parçası olarak gruplara uygulanır ve bu ankor maddelerden alınan puanlar toplam test puanına dahil edilir (Kelecioğlu, 1994). Dış ankor düzenekte iki gruba eşitlenecek testlerle beraber eş zamanlı olarak başka bir ortak test uygulanır ve ortak testten alınan puan toplam test puanına dahil edilmez (Crocker ve Algina, 1986).

Ankor maddelerin doğru bir şekilde seçilmesi önemlidir, aksi halde testlerin karşılaştırılması sonucu elde edilen sonuçlar hatalı olacaktır (Kolen ve Brennan, 2004). Ankor maddeler test yapısı, madde yapısı ve içerik olarak tüm testin küçük bir temsilcisi olmalıdır (Angoff, 1971). İstatistiksel olarak ankor testin uzunluğu arttıkça random hata miktarı azalır (Budescu, 1985). Angoff (1971), ortak madde sayısının X ve Y formundaki madde sayısının %20'sinden az olmaması gerektiğini belirtmiştir. Reckase (1979), örneklem sayısı 300'den fazla olduğu durumlarda ankor madde sayısının en az 5 en çok 15 madde olması gerektiğini belirtmiştir.

2.6. Eşitleme Yöntemleri

Test eşitleme yöntemleri temel aldığı kuramlara göre ikiye ayrılmaktadır. Bu yöntemler geleneksel eşitleme yöntemleri ve madde tepki kuramına dayalı eşitleme yöntemleridir.

2.6.1. Klasik Test Kuramına Dayalı Eşitleme Yöntemleri

Bu kuramda madde güvenilirliği, madde güçlüğü, madde ayırt ediciliği gruba bağlıdır. Cevaplayıcıların yetenekleri testin kolay ya da zor olmasına göre değişir diğer bir deyişle teste bağlıdır (Hambleton ve Swaminathan,1985). Klasik test kuramına dayalı eşitleme yöntemlerinden en sık kullanılan üç yöntem ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitlemedir.

2.6.1.1. Ortalama Eşitleme (Mean Equating)

Ortalama eşitleme yönteminde eşitlenecek olan testlerin güçlükleri arasındaki farklar ölçek boyunca sabittir. Eşitlenecek olan testlerden elde edilen puanların ortalamaya olan uzaklıklarının eşit olduğu kabul edilir (Kolen ve Brennan, 2004).

$$x - \mu(X) = y - \mu(Y) \quad (1)$$

(1) nolu eşitlik düzenlenirse;

$$m_Y(x) = y = x - \mu(X) + \mu(Y)$$

Form X ve Form Y için ortalama eşitleme bağıntısı elde edilir.

$m_Y(x)$: Form X'teki x puanının Form Y'deki ölçüğe ortalama eşitleme kullanılarak dönüştürülmüş hali

x : Form X'teki belirli bir puan

$\mu(X)$: Form X'in ortalaması

$\mu(Y)$: Form Y'nin ortalaması

2.6.1.2. Doğrusal Eşitleme (Linear Equating)

Doğrusal eşitleme, eşitlenecek testlerin ortalama ve standart sapma hariç diğer tüm özelliklerinin eşit olması varsayımına dayanmaktadır (Crocker ve Algina, 1986). Aynı standart puanlara karşılık gelen puanların eşit olduğu kabul edilir (Angoff, 1971). Ortalama eşitlemenin aksine formlar arası güçlük farklılıkları değişkenlik gösterebilir. Düşük seviye başarıya sahip cevaplayıcılar için Form X, Form Y'den daha zor olabilirken yüksek başarı seviyesine sahip cevaplayıcılar için daha kolay olabilir.

Lineer dönüşüm iki test formunun z puanlarının eşit olması ile tanımlanır (Kolen ve Brennan, 2004).

$$\frac{x - \mu(X)}{\sigma(X)} = \frac{y - \mu(Y)}{\sigma(Y)}$$

$$I_Y(x) = y = \sigma(Y) \left[\frac{x - \mu(X)}{\sigma(X)} \right] + \mu(Y)$$

$$I_Y(x) = y = \frac{\sigma(Y)}{\sigma(X)} x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)} \mu(X) \right]$$

$\frac{\sigma(Y)}{\sigma(X)}$ eğimi a , $\mu(Y) - \frac{\sigma(Y)}{\sigma(X)} \mu(X)$ kesişimi b olarak alınırsa $Y = ax + b$ doğru denklemi elde edilir.

$\sigma(X)$: Form X'in standart sapması

$\sigma(Y)$: Form Y'nin standart sapması

$\mu(X)$: Form X'in ortalaması

$\mu(Y)$: Form Y'nin ortalaması

$I_Y(x)$: Form X'teki gözlenen puanların Form Y ölçüğine dönüşümünü sağlayan lineer dönüşüm denklemi

2.6.1.3. Eşit Yüzdellikli Eşitleme (Equipercetile Equating)

Eşit yüzdellikli eşitlemede, Form X ve Form Y'yi alan bireyler yüzdellik sıralarına göre eşitlenir (Kolen, 1988). Eşitlenecek iki formun puan dağılımının yığılmalı frekans dağılımları belirlendikten sonra bu yığılmalı frekanslarda aynı yüzdellik sıraya denk gelen puanlar eşitlenir. Angoff (1984), güvenilirlikleri aynı olan ve aynı özelliği ölçen iki ayrı test formundan elde edilecek puanların aynı yüzdellik sırada olanların eşdeğer olarak düşünölebileceğini belirtmektedir. Bazı durumlarda formların yüzdellik puanları karşılık gelmeyebilir. Bunun için düzeltme çalışmaları yapılmaktadır. Eşitlemeden önce yapılırsa ön düzeltme, eşitlemeden sonra yapılırsa son düzeltme denilmektedir.

2.6.2. Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri

KTK'nın eksik kalan yönleri dikkate alınarak geliştirilen MTK'da madde parametreleri testin uygulandığı gruba bağılı değildir. Cevaplayıcıların yetenekleri testin zorluk düzeyinden etkilenmez yani testten bağımsızdır. Cevaplayıcının aynı yeteneği ölçen testin iki farklı formundan hangisini aldığıın önemi yoktur (Hambleton ve Swaminathan, 1985). Uygulanması için tek boyutluluk, yerel bağımsızlık ve değişmezlik koşulları sağlanmalıdır.

MTK ile eşitleme veri toplama ile başlar ve uygun desen seçilir. Bu yöntemin kullanıldığı eşitlemede daha çok ankor test deseni tercih edilir. Ankor test deseninde testin uygulandığı grubun yetenekleri birbirinden farklı olabilir. Böyle bir durumda geleneksel yöntemler tercih edilmez (Cook ve Eignor, 1983). Desen seçildikten sonra uygun model seçilir. MTK ile eşitlemede kullanılan birçok model vardır. 1-0 şeklinde ikili puanlanabilen testler için yaygın kullanılan modeller 1 Parametrelili Lojistik Model (1PLM), 2 Parametrelili Lojistik Model (2PLM), 3 Parametrelili Lojistik Model (3PLM)'dir. Model veri uyumu sağlandıktan sonra modele ve testi alan kişilere ait parametreler kestirilir (Hambleton ve diğ., 1991). İki ayrı kestirimden elde edilen parametreler ortak bir ölçekte ifade edilir böylece MTK ile eşitleme tamamlanır (Zhu, 1998). 1PLM, 2PLM ve 3PLM'ye kısaca değinilmiştir.

2.6.2.1. 1 Parametrelili Lojistik Model (1PLM)

Yaygın olarak kullanılan 1PLM, Rasch Modeli (Rasch, 1960) olarak da bilinir. Bu modelde cevaplayıcının maddeyi doğru cevaplama olasılığı hesaplanırken

öğrenci yeteneği ve madde güçlüğü dikkate alınır. Yetenek, 0 ve en iyi puan arasındaki olası tüm ham puanlar için kestirilir. Aynı ham puanı alan tüm cevaplayıcıların bu puanı almak için hangi soruları doğru hangilerini yanlış yaptığını bakılmaksızın aynı yetenek düzeyinde oldukları kabul edilir. Bu modelde şans başarısı olarak adlandırılan c_i parametresinin değeri 0'dır ayırt edicilik parametresi a_i ise tüm maddeler için eşittir ve değeri 1'dir.

θ yetenek düzeyindeki cevaplayıcının i maddesini doğru cevaplama olasılığı şu şekildedir:

$$P_i(\theta) = \frac{e^{D_i(\theta-b_i)}}{1 + e^{D_i(\theta-b_i)}}$$

$P_i(\theta)$: θ yetenek seviyesindeki cevaplayıcının i maddesini doğru cevaplama olasılığı

b_i : i maddesinin güçlük indeksi

D : ölçek sabiti (1,7).

2.6.2.2. 2 Parametrelili Lojistik Model (2PLM)

2PLM (Birnbaum, 1968), bir bireyin soruyu doğru cevaplama olasılığını maddenin ayırt ediciliğini de dikkate alarak cevaplayıcının yeteneği ile madde güçlüğü arasındaki bir fonksiyon yardımıyla tanımlar. 1PLM'de olduğu gibi şans başarısı dikkate alınmaz.

θ yetenek düzeyindeki cevaplayıcının i maddesini doğru cevaplama olasılığı şu şekildedir:

$$P_i(\theta) = \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}}$$

$P_i(\theta)$: θ yetenek seviyesindeki cevaplayıcının i maddesini doğru cevaplama olasılığı

a_i : i maddesinin ayırtıcılık gücü indeksi

b_i : i maddesinin güçlük indeksi

D : ölçek sabiti (1,7).

2.6.2.3. 3 Parametrelili Lojistik Model (3PLM)

3PLM'de madde güçlük ve ayırt edicilik parametrelerine ek olarak şansa bağlı tahmin parametresi yer almaktadır. Bu modelin şans başarısının etkili olduğu

testlerde kullanılmasının uygun olduğu belirtilmiştir (Baker, 2001). θ yetenek düzeyindeki cevaplayıcının i maddesini doğru cevaplama olasılığı şu şekildedir:

$$P_i(\theta) = c_i + \frac{(1-c_i)e^{D a_i (\theta-b_i)}}{1+e^{D a_i (\theta-b_i)}}$$

$P_i(\theta)$: θ yetenek seviyesindeki cevaplayıcının i maddesini doğru cevaplama olasılığı

a_i : i maddesinin ayırıcılık gücü indeksi

b_i : i maddesinin güçlük indeksi

c_i : i maddesinin şans parametresi

D : ölçek sabiti (1,7).

2.7. Eşitleme Hatası

Random ve sistematik olmak üzere iki çeşit eşitleme hatası vardır (Kolen ve Brennan, 2004). Random eşitleme hatası eşitleme ilişkisini belirlemede yer alan ortalama, standart sapma ve yüzdeler gibi parametre kestirimleri yapılırken evren yerine örneklem kullanıldığında ortaya çıkar (Barnard, 1996). Örneklem sayısının artırılmasıyla random hata miktarı azalır (Kolen ve Brennan, 2004). Sistematik eşitleme hatasının ortaya çıkma sebebi kullanılan eşitleme yöntemlerinin varsayımları gerçekleşmemesidir (Kolen, 1988). Örneğin tek grup deseninde yorgunluk ve aşinalık durumları kontrol edilmediğinde, seçkisiz grup deseninde spiralleme yöntemi sürecinde sorun olduğunda, NEAT desende yöntem koşulları tam olarak sağlanmadığında ortaya çıkar. NEAT desende ankor maddelerin formlarda farklı davranması, X ve Y formları arası güçlük ve güvenilirlik farkı varsa sistematik hata vardır (Kolen, 1988). Kolen ve Brennan'a (2004) göre sistematik hatayı kontrol edebilmek için desenlerin koşulları sağlanmalı ve uygun istatistiksel yöntemler kullanılmalıdır. Örneklem büyüklüğünü artırmak sistematik hata miktarını azaltmaz.

2.8. İlgili Araştırmalar

İlgili literatür incelendiğinde kategorik değişkenlerin eşitlenmesine yönelik bir araştırmaya rastlanmamıştır. Bu bölümde DINA model ve test eşitlemeye yönelik yapılan bazı araştırmalardan kısaca bahsedilmiştir.

Kelecioğlu (1994), 1990, 1991, 1993 ÖSS Türkçe Matematik Sosyal Bilimler ve Fen Bilimleri alt testleri için doğrusal eşitleme, eşit yüzdelikli eşitleme, bir parametrelili model ve iki parametrelili model eşitleme yöntemlerinden hangisinin daha uygun olacağını araştırmıştır. Testler için iç ankor düzeneği kullanılmıştır. Yöntemlerin uygunluğu ağırlıklandırılmış hata kareleri ortalaması ile hesaplanmış, eşitleme hatası ise eşitlemiş puanlarla test puanlarının ağırlıklandırılmış hata kareleri ortalamaları ile hesaplanmıştır. Ağırlıklandırılmış hata kareleri ortalaması en düşük olan yöntem en uygun model olarak seçilmiştir. Araştırma sonuçlarından elde edilen bulgulara göre; Türkçe alt testi için doğrusal eşitleme, Sosyal Bilimler ve Matematik alt testleri için bir parametrelili model, Fen Bilimleri alt testi için eşit yüzdelikli eşitleme yönteminin daha uygun olduğu kararlaştırılmıştır.

Şahhüseyinoğlu, D. (2005) çalışmasında Hacettepe Üniversitesi İngilizce hazırlık bölümünün uyguladığı 2000, 2001, 2002 yıllarına ait İngilizce yeterlik puanlarının ardışık iki yılına ait testlerin eşitlenmesinde uygun yöntemi belirlemeyi amaçlamıştır. Testler ortak test desenine göre hazırlanmıştır ve üç farklı yıla ait testler üç farklı eşitleme yöntemi ile eşitlenmiştir. Bu yöntemler klasik test kuramına göre eşit yüzdelikli eşitleme ve doğrusal eşitleme, örtük özellikler kuramına göre Rasch parametresinin karakteristik eğri yöntemidir. Araştırma verileri 2000 yılında sınava giren 2233, 2001 yılında sınava giren 1989 ve 2002 yılında sınava giren 2033 öğrenciden elde edilmiştir. Bu sınavda kullanılan testlerin eşitlenmesinde en uygun yöntem olarak hata kareleri ortalaması en düşük olan yöntem önerilmiştir. Bulgular incelendiğinde bu testlerin eşitlenmesinde uygun olan yöntemlerin Rasch yöntemi ve eşit yüzdelikli eşitleme yöntemi olduğu sonucuna ulaşılmıştır.

Lee, Park ve Taylan (2011) araştırmalarında BTM'den biri olan DINA model ile MTK modellerinden olan 1PL, 2PL ve 3PL modeli karşılaştırmışlardır. Veri toplama aracı olarak 2007 TIMSS 4. Sınıf matematik alanına ait maddeleri kullanmışlardır. Seçilen 25 maddeye ait Q matrisi belirlemişlerdir. Araştırmanın çalışma grubu Massachusetts, Minnesota ve Birleşik Devletler olmak üzere üç

bölüme ayrılmıştır ve toplam 823 öğrenciden oluşmaktadır. Test DINA model ve 1PL, 2PL 3PL MTK kullanılarak analiz edilmiştir. Üç ayrı bölgeye ait analizler birbiri ile karşılaştırılmıştır. Çalışma sonucunda elde edilen bulgulara göre DINA modelin model uyumu 1PL, 2PL ve 3PL modellerine kıyasla daha yüksek bulunmuştur.

Başokcu (2012), örneklemin DINA model parametreleri kullanılarak elde edilen madde ayırt edicilik indeksi üzerine etkisini incelemiştir. Bu çalışmada aynı evrenden çekilen aynı büyüklükteki farklı örneklemelerden ve farklı örneklem büyüklüğünden elde edilen ayırt edicilik indekslerinin örneklemden ne derece etkilendiği hesaplanmış daha sonra geleneksel yöntemden elde edilen sonuçlar ile karşılaştırma yapmıştır. Araştırma bulgularına göre evrenden aynı büyüklükte farklı örneklemeler çekildiğinde ve örneklemdeki eleman sayısı azaldığında DINA model daha tutarlı sonuçlar elde edilmiştir.

Demirus, K. B. (2015), Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumlarının test eşitlemeye etkisini incelemiştir. Ortak maddeler cinsiyete göre TB-DMF'li ya da DMF'siz olduğunda MTK temel alınarak yapılan eşitleme yöntemlerinin performansı karşılaştırılmıştır. Araştırmada DMF analizlerinde Matel-Haenszel ve lojistik regresyon yöntemleri kullanılmıştır. Test eşitlemede ise lineer ölçek dönüştürme yöntemlerinden ortalama-ortalama ile ortalama- sigma; karakteristik eğrisi dönüştürme yöntemlerinden Haebara ile Stocking-Lord kullanılmıştır. Eşitleme yöntemlerinin performansları RMSD eşitleme hataları hesaplanarak değerlendirilmiştir. Çalışma grubunu 1350 8.Sınıf öğrencisi oluşturmaktadır. Sonuç olarak ortak maddeler erkekler lehine TB-DMF'li olduğunda en büyük RMSD eşitleme hatasını ortalama-ortalama yönteminde, en küçük hata ise ortalama-sigma yönteminde elde edildiği gözlemlenmiştir. Ortak maddelerin DMF'siz olduğu durumda ise en büyük hata ortalama-sigma yönteminde bulunmuştur. En küçük RMSD eşitleme hatası karakteristik eğrisi yöntemlerinde birbirine eşit bulunmuştur.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın türü, araştırmanın evren ve örnekleme, ölçme aracına ilişkin bilgiler ve verilerin analizi yer almaktadır.

3.1. Araştırmanın Modeli

Bu araştırmada, BTM'ye uygun olarak geliştirilen ortak ve farklı maddeleri içeren iki test formunun psikometrik özellikleri belirlenmiştir. Aynı zamanda bu test formlarının ayrı ayrı uygulandığı iki farklı cevaplayıcı grubundan elde edilen veriler DINA model ile analiz edilmiştir ve madde parametreleri kullanılarak örtük sınıfların eşitlemesi yapılmıştır. Araştırma sonuçlarının var olan kurama katkı sağlayacak özellikte olmasından dolayı temel araştırma niteliği taşımaktadır. Temel araştırma, bilgiyi bilgi için sağlama üzerine kuruludur, kuramsal sorunları test etmeye odaklıdır ve teorilerle, modellerle ilişkilendirilir. Temel araştırma sonuçlarının geniş kapsamlı ve uygulanabilir olması beklenir (Whitley, 2002).

Araştırmanın gerçek uygulama verisi üzerinden kurgulanan bir eşitleme çalışması olması ve sonuçlarının bir proje çalışmasının temel bulguları arasında yer alması sebebiyle uygulama araştırması niteliği de taşımaktadır. Uygulama araştırması, bazı yönleri ile toplumu etkileyen problemlere çözüm bulma amaçlıdır ve teorilerin rehberliğinde olma zorunluluğu yoktur. Bir ya da birkaç durumdaki davranışa odaklıdır. İlgili konu üzerindeki probleme etkisi olan ya da belirli davranışı tahminlemek için kullanılabilecek değişkenler belirlenir (Whitley, 2002).

3.2. Araştırmanın Evren ve Örnekleme

Araştırma verileri okul uygulaması sonucunda elde edilmiştir. Araştırma evrenini, İzmir İl Milli Eğitim Müdürlüğü'ne bağlı 12 okulda 6, 7 ve 8. sınıflardan oluşan 2958 kişilik öğrenci grubu oluşturmaktadır. Bu gruba A,B,C,D,E ve F olarak adlandırılan test formları uygulanmıştır. Bu formlar, 'Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek için Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği' adlı 115K531 No'lu TÜBİTAK projesi kapsamında hazırlanan deneme uygulamasından elde edilmiştir.

Araştırma analizleri ve eşitleme çalışması C formunun uygulandığı 501 kişi ve D formunun uygulandığı 491 kişi olmak üzere toplam 992 öğrenci üzerinde gerçekleştirilmiştir. Uygulamanın yapılacağı okullar tabakalama yöntemiyle seçilmiştir. Cevaplayıcılar formlara random olarak atanmıştır ve her form her sınıftaki öğrencilere eşit sayıda uygulanmıştır. Bu durum, grupların denkleştirilmesi konusunda örnekleme yöntemine bir avantaj sağlamıştır. Örnekleme randomizasyon sınıf düzeyinde değil öğrenci düzeyinde gerçekleştirilmiştir. Tablo 10’da bu formların uygulandığı öğrencilerin sınıf düzeylerine ve cinsiyetlerine göre dağılımları verilmiştir.

Tablo 10. Test Formlarının Sınıf Düzeyine ve Cinsiyete Göre Dağılımı

Formlar	6.sınıf		7.sınıf		8.sınıf		Toplam
	Erkek	Kız	Erkek	Kız	Erkek	Kız	
Form A	125	171	102	112	3	7	520
Form B	125	160	92	127	3	6	513
Form C	119	158	83	126	5	10	501
Form D	146	133	109	95	2	6	491
Form E	132	142	91	110	1	8	484
Form F	122	136	80	103	5	3	449

3.3. Ölçme Aracı

Araştırmada kullanılan ölçme araçları ortak ve farklı maddelerden oluşan, açık uçlu ve çoktan seçmeli maddelerin yer aldığı 15 maddelik test formlarıdır. 6. sınıf matematik dersi öğrenme alanlarını ölçecek şekilde hazırlanan bu formlar, ‘Uluslararası Geniş Ölçekli Sınavlarda Türkiye’nin Matematik Başarısını Artırabilmek için Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği’ adlı 115K531 No’lu TÜBİTAK projesi kapsamında hazırlanan deneme uygulamasından elde edilmiştir.

3.3.1. Test Geliştirme Aşamaları

Araştırmada öncelikle proje amacına uygun olarak testin konusu, kapsamı ve Q matris ile ölçülecek özellikler belirlenmiştir. Testteki maddelerin yazılabilmesi amacıyla 6. sınıf dersine giren 112 matematik öğretmenine soru yazma eğitimi verilmiştir. Eğitim sonunda 77 öğretmen soru yazma ekibine dâhil edilmiştir ve soru

yazma aşaması başlamıştır. Süreç sonunda yazılan 324 soru, belirlenen özelliklere uygunluklarına göre bir değerlendirme sürecine tabi tutulmuşur ve değerlendirme sonucunda hakem puanları temel alınarak en yüksek puana sahip maddeler seçilmiştir. Proje sorularına ek olarak aralarında PISA ve TIMSS maddelerinin yer aldığı toplam 48 madde ile ankor test desenine uygun 6 form oluşturulmuştur. Sorunsuz bir şekilde yapılmış uygulama sonuçları bu uygulamaya özel geliştirilen optik formlara kodlanmış ve okutulmuştur. Araştırmada kullanılmak üzere iki form seçilmiştir. Bu aşamalar ayrıntılı bir şekilde anlatılmıştır.

3.3.1.1. Ölçme Aracının Kapsamı

Testlerin içeriğini oluşturan maddeler, Milli Eğitim Bakanlığı ortaokul matematik dersi öğretim programında 6.sınıflara yönelik belirlenen ‘Sayılar ve İşlemler, Cebir, Geometri ve Ölçme, Veri İşleme’ öğrenme alanlarına göre hazırlanmıştır (MEB, 2013). Testler aynı zamanda projenin amaçlarına uygun olarak yukarıda belirtilen konu alanları için üst düzey düşünme becerilerini de ölçecek şekilde geliştirilmiştir.

3.3.1.2. Q Matris Belirleme Süreci

Araştırmada kullanılan Q matriste yer alan özellikler matematiksel yeterliklere göre oluşturulmuştur. Matematiksel yeterliklerin belirlenmesinde Niss ve arkadaşlarının (Niss, 2003; Niss ve Jensen, 2002; Niss ve Højgaard, 2011) belirlediği 8 tür matematiksel yeterlik üzerinde durduğu çalışması temel alınmıştır. PISA 2012’de de temel alınan bu matematiksel yeterlikler iletişim, matematikleştirme, gösterim, akıl yürütme ve kanıt gösterme, problem çözme stratejisi tasarlama, sembolik formel ve teknik dil ve işlemleri kullanma, matematiksel araçları kullanma olarak ele alınmıştır (OECD, 2013).

Bu araştırmada kullanılacak olan yeterlikler; proje araştırmacıları, danışmanlar ve alan uzmanı öğretmenlerle birlikte madde türleri, konu ve sınıf düzeyi göz önünde bulundurularak dört başlık altında toplanmıştır. ‘İletişim ve ilişkilendirme, matematikleştirme, akıl yürütme ve strateji geliştirme, sembolik ve teknik dil kullanımı’ olarak belirlenen bu dört yeterliğe ilişkin tanımlamalar kısaca açıklanmıştır (Basokcu, 2019).

İletişim ve İlişkilendirme: İletişim becerisi; matematiksel dili, günlük dil ve sembollerle ilişkilendirme, matematiksel düşüncelerin doğruluğunu ve anlamını

yorumlama gibi özellikleri kapsar. İlişkilendirme becerisi; matematik kavramlarının kendi aralarında, diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

Matematikleştirme: Bir problemi kesin matematiksel formda ifade etmeye yönelik modelleme, yapısal gösterim, varsayımlarla tanımlama, formülleştirme, kurulan bir yapı ya da modele yönelik matematiksel çıktılarının elde edilmesi ve yorumlanması gibi belirli eylemleri ifade etmektedir.

Akıl Yürütme ve Strateji Geliştirme: Akıl yürütme, eldeki bilgilerden hareketle matematiğin kendine özgü araç (semboller, tanımlar, ilişkiler, vb.) ve düşünme tekniklerini (tümevarım, tümdengelim, karşılaştırma, genelleme, vb.) kullanarak yeni bilgiler elde etme sürecidir. Strateji geliştirme, problem çözmede matematiksel bilgi ve becerilerin kullanılmasına yönelik bir plan ya da stratejinin seçilmesi ya da tasarlanmasını ifade etmektedir.

Sembolik ve Teknik Dil Kullanımı: Matematik okuryazarlığı açısından sembolik ve teknik dilin kullanılması yeteneği, matematiksel kurallarla ifade edilen bir matematiksel içeriğe yönelik olarak sembol gösterimlerini anlama ve yorumlama davranışlarının bir bütünüdür.

3.3.1.3. Soru Yazma Ekibinin Kurulması

Testin amacı ve kapsamı belirlendikten sonra soru yazma ekibi kurulmasına temel oluşturması amacıyla soru yazma eğitimi verilmiştir. Ekibin kurulması için İzmir İl Milli Eğitim Müdürlüğü desteği ile ülke genelinde, 6. sınıf matematik dersine giren öğretmenlere yönelik bir eğitim çağrısı yapılmıştır. Yapılan eleme sonucunda 31 farklı il ve 63 farklı ilçeden 112 öğretmen seçilmiştir. Eğitimin başlığı “Ortaokul Matematik Öğretmenleri İçin Üst Düzey Düşünme Becerilerine Yönelik Soru Yazma Eğitimi” olarak belirlenmiştir ve 9-10 Şubat 2016 tarihlerinde gerçekleştirilmiştir. Eğitim içeriği öğretmenlere 6. sınıf düzeyinde PISA ve TIMSS testlerinde kullanılan maddeler gibi özellikle üst düzey düşünme becerilerine yönelik sorular yazabilme becerisi kazandıracak şekilde geliştirilmiştir. Eğitimde öğretim görevlileri tarafından program geliştirme, ölçme değerlendirme ve matematik eğitimine yönelik kuramsal seminerler verilmiştir ve katılımcıların üst düzey düşünmeye odaklı soru yazma becerileri geliştirmeye yönelik çalıştaylar düzenlenmiştir. Eğitim sonunda projede soru yazarı olarak çalışmayı kabul eden 77

öğretmene proje kapsamında kullanılacak soruların nasıl yazılacağına ilişkin dokümanlar gönderilmiştir. Öğretmenlerden birden fazla özelliği aynı anda ölçebilen maddeler yazmaları istemiştir. Bu süreçte aynı zamanda proje amaçlarına uygun olarak sorunun hangi özellikleri ölçtüğü, hangi kazanımlar ile tanımlanabileceği ve bilişsel süreç becerileri ya da üst düzey düşünme becerileri ile nasıl ilişkilendirilebileceğine dair bilgileri de tanımlamışlardır.

3.3.1.4. Madde Havuzunun Oluşturulması

Soru yazma ekibinde görev alan 77 öğretmen, proje amaçları doğrultusunda iki aylık süreçte aşamalı olarak 324 soruyu proje web sitesi üzerinden sisteme yüklemişlerdir. Projede soruları değerlendirmek amacıyla bir hakem grubu belirlenmiştir. Bu süreçte her soru üç hakem tarafından değerlendirilmiştir. Hakemler soruları ön ve son değerlendirme olarak iki kere puanlamıştır. Ön değerlendirme hakemin soru yazarı ile görüşmesinden önce, son değerlendirme yazar ile hakem arasındaki görüşmeler sona erdiğinde yapılmıştır.

Projede soruları değerlendirebilmek için 15 maddeden oluşan 5'li puanlama anahtarı geliştirilmiştir. Bu yöntemle 324 proje sorusu 28 hakem tarafından proje web sitesinde çevrimiçi değerlendirilmiştir. Daha sonra hakemler tarafından puanlanan maddelerden en yüksek puanı alan maddeler belirlenmiş aralarından 34 tanesi deneme uygulaması için seçilmiştir. Bu maddelere ek olarak 6. sınıf matematik öğretim programına ve araştırma için belirlenen konulara uygun olan PISA ve TIMSS'te yer alan maddelerden de deneme uygulamasında kullanılmak üzere seçilmiştir.

3.3.1.5. Deneme Uygulaması Formlarının Oluşturulması ve Uygulama Deseni

Deneme uygulamasında bütün maddeleri tek seferde büyük bir gruba uygulayabilmek için ortak ve farklı maddelerin yer aldığı 6 farklı test formu oluşturulmuştur. PISA, TIMSS ve proje sorularından oluşan deneme uygulaması deseni Tablo 11'de verilmiştir.

Tablo 11. *Deneme Uygulaması Deseni*

	Form A	Form B	Form C	Form D	Form E	Form F
PISA	p1	p1	p1	p1	p1	p1
TIMSS	t1	t1	t1	t1	t1	t1
Proje	s2	s2	s2	s2	s2	s2
Proje	s9	s9	s9	s9	s9	s9
PISA	p5	p2	p3	p3	p4	p7
PISA	p6	p5	p2	p4	p7	p6
TIMSS	t3	t2	t2	t5	t7	t7
TIMSS	t4	t4	t5	t6	t6	t3
Proje	s22	s22	s21	s19	s18	s14
Proje	s23	s23	s4	s20	s16	s15
Proje	s12	s21	s19	s18	s14	s27
Proje	s13	s4	s20	s16	s15	s29
Proje	s32	s7	s24	s3	s26	s25
Proje	s1	s28	s33	s31	s6	s34
Proje	s5	s30	s8	s10	s11	s17

Tabloda p ile gösterilen maddeler PISA, t ile gösterilen maddeler TIMSS sınavlarının yayınlanan maddelerinden seçilen maddeleri; s ile gösterilen maddeler ise proje kapsamında yazılan maddeleri temsil etmektedir. Formlarda yer alan ilk 4 madde tüm formlarda ortaktır. Bu 4 madde dışında yan yana iki form 4 ortak maddeye daha sahiptir. Desende PISA'ya ait 7, TIMSS'e ait 7 ve projeye ait 34 olmak üzere toplam 48 farklı madde bulunmaktadır. Bu 48 maddenin 25 tanesi çoktan seçmeli, 23 tanesi açık uçlu maddelerden oluşmaktadır. Açık uçlu maddeler PISA ve TIMSS içeriklerine benzer şekilde kısa cevaplı, uygulamalı ve doğru yanlış türündedir. Her formda 3 PISA, 3 TIMSS ve 9 proje sorusu olmak üzere toplam 15 madde yer almaktadır.

3.3.1.6. Araştırmada Kullanılan Formlar ve Bu Formlara Ait Q Matrisler

Ortalama madde gücü göz önüne alınarak ve Q matris uyumuna bakılarak araştırmada kullanılmak üzere C ve D formları seçilmiştir. Q matrisleri aynı kitapçıkların seçilmesinin sebebi eşitleme çalışmasında Q matris etkisini ortadan kaldırmaktır. Her iki formda da 9 çoktan seçmeli 6 açık uçlu madde bulunmaktadır. Bu testlerde yer alan maddelerden 3 tanesi ölçtüğü matematiksel yeterliklerle birlikte Ek'te yer verilmiştir. C ve D formlarına ait Q matrisler Tablo 12'de gösterilmektedir.

Tablo 12. *C ve D Formlarına Ait Q Matrisler*

Form C Madde Kodu	Özellikler				Form D Madde Kodu	Özellikler			
	α_1	α_2	α_3	α_4		α_1	α_2	α_3	α_4
0502_1	0	1	1	0	204	0	1	1	0
6471	0	0	1	0	506	0	0	1	0
5752_1	1	0	1	0	9342	1	0	1	0
4427	0	1	1	0	8867	0	1	1	0
3019	1	1	1	0	6323_1	1	1	1	0
202	1	1	1	1	1979	1	1	1	1
6188	0	1	0	1	9074	0	1	0	1
203	0	0	0	1	203	0	0	0	1
201	0	1	0	0	201	0	1	0	0
501	0	1	0	1	501	0	1	0	1
505	1	1	0	1	505	1	1	0	1
6251	1	1	1	0	6251	1	1	1	0
6370	1	0	1	0	6370	1	0	1	0
1119	1	0	1	0	1119	1	0	1	0
6050	1	1	1	0	6050	1	1	1	0

Tablo 12’de α_1 , α_2 , α_3 ve α_4 ile gösterilen özellikler sırasıyla iletişim ve ilişkilendirme, matematikleştirme, akıl yürütme ve strateji geliştirme, sembolik ve teknik dil kullanımı özelliklerini temsil etmektedir. Tablodaki veriler göz önüne alındığında her iki formda da bir özelliği ölçen 3 madde; iki özelliği ölçen 7 madde; üç özelliği ölçen 4 madde ve dört özelliği birden ölçen 1 madde olduğu görülmektedir. Testlerde yer alan son 8 madde her iki formda ortaktır.

3.3.1.7. Ölçme Aracının Uygulanması

Testler, önceden belirlenmiş gün ve saatlerde 6-8 Haziran 2016 tarihleri arasında öğretmen gözetiminde uygulanmıştır. Sınav öncesi öğretmenlere sınav ve kodlama hakkında bilgilendirme ve test yönergesi gönderilmiştir. Sınav süresi olan 60 dakika testin tamamlanması için yeterli görülmüştür.

3.3.1.8. Maddelerin Puanlanması

Ölçme araçlarında farklı madde türleri yer almaktadır. Puanlamanın yapılabilmesi amacıyla açık uçlu maddelerin ve çoktan seçmeli maddelerin beraber okutulmasını sağlayan optik form geliştirilmiştir. Açık uçlu maddeler soru okuma ekibi tarafından daha önce belirlenen puanlama anahtarına göre okunmuştur ve

öğrencinin açık uçlu cevabına ilişkin kod öğrencinin optik formuna kodlanmıştır. Bu yöntem ile hem açık uçlu sorulara verilen cevapların tamamının dijital olarak kaydedilmiştir hem de optik okuyucu tarafından tek seferde okunması sağlanmıştır. Optik form bu şekilde okutulduktan sonra Microsoft Excel’de doğru yanıtlanmış maddeler ‘1’, yanlış yanıtlanan maddeler, boş bırakılan maddeler ve kısmi puanlanmış maddeler ‘0’ olarak kodlanmıştır ve DINA model ile analize uygun hale getirilmiştir.

3.4. Verilerin Analizi ve Testlerin Psikometrik Özellikleri

Öncelikle deneme uygulamasına ait betimsel istatistikler, grupların denkliliği, proje sorularının PISA TIMSS soruları ile eşdeğerliği çalışması, maddelerin psikometrik özellikleri hesaplanmıştır. Daha sonra eşitleme yöntemi anlatılmıştır. Klasik test teorisi madde parametreleri için ITEMAN 4.2 paket programı, Madde Tepki Kuramı için IRT pro, DINA model parametre kestirimleri için Ox Edit programı kullanılmıştır.

3.4.1. Betimsel İstatistikler

Tablo 13’te deneme uygulamasına ait betimsel istatistikler yer almaktadır.

Tablo 13. Deneme Uygulamasına Ait Betimsel İstatistikler

	Form A	Form B	Form C	Form D	Form E	Form F
Madde Sayısı	15	15	15	15	15	15
Kişi Sayısı	520	513	501	491	484	449
Ortalama	6,28	6,52	7,69	7,22	8,88	7,83
Medyan	6	7	8	7	9	8
En Düşük Puan	0	0	0	0	0	0
En Yüksek Puan	14	13	15	14	15	15
Varyans	6,69	6,1	9,03	9,94	12,91	10,4
Standart Sapma	2,59	2,47	3,05	3,15	3,59	3,23
Çarpıklık Katsayısı	0,23	0,01	-0,03	0,09	-0,23	-0,13
Basıklık Katsayısı	-0,28	-0,37	-0,63	-0,73	-0,82	-0,69
Güvenirlilik(Alpha)	0,61	0,6	0,71	0,72	0,8	0,76
Ortalama Madde Güçlüğü	0,42	0,44	0,51	0,48	0,59	0,52

Test formları en az 449, en çok 520 kişiye uygulanmıştır. Basıklık ve çarpıklık katsayıları incelendiğinde her formun normallik varsayımını karşıladığı

belirlenmiştir. Formların madde güçlüğü ortalamaları 0,42 ile 0,59 arasında değişmektedir. Güvenirlik katsayıları 0,60 ile 0,80 arasında değer almaktadır.

3.4.2. Proje Soruları ile PISA TIMSS Sorularının Eşdeğerliği Çalışması

Projede geliştirilen soruların PISA ve TIMSS soruları ile eşdeğer olup olmadığını belirlemek için tekrarlı ölçümler için ANOVA analizi yapılmıştır. Analiz sonuçlarına göre PISA, TIMSS ve proje sorularının ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmamıştır ($F_{(896-2)} = 2,358, p>0,05$). Dağılımların varyanslarının eşit olup olmadıkları Mauchly Küresellik testi ile incelenmiştir. Analiz sonuçlarına göre küresellik varsayımının bozulmadığı bir başka deyişle dağılımların varyanslarının eşit olduğu görülmüştür, ($\chi^2(2)=4,881, p>0,05$). Bu durum proje kapsamında üretilen soruların PISA ve TIMSS sorularıyla eşdeğer olduğunun bir göstergesidir. Aynı zamanda eşdeğerlik kapsam ve ölçüt geçerliği yönünden de incelenmiştir. Ölçüt geçerliğine göre test puanlarının birbirleriyle istatistiksel olarak anlamlı ve yüksek bir ilişki verdiği görülmüştür, ($r=0,52, p<0,01$). Kapsam geçerliği çalışması en az doçentlik seviyesinde olan, iki ölçme ve değerlendirme, iki matematik eğitimi ve bir program geliştirme alanında uzman proje araştırmacı ve danışmanlarından oluşan 5 kişilik ekip tarafından gerçekleştirilmiştir. Yapılan çalışma sonucunda projede üretilen soruların PISA ve TIMSS kapsamıyla uyumlu olduğu görülmüştür.

3.4.3. Deneme Uygulaması Yapılan Grupların Denkliklerinin Araştırılması

Deneme uygulamasında yer alan grupların denkliklerini sağlayarak proje soruların psikometrik özelliklerinin farklı örneklemelerden gelmesine rağmen araştırılabilmesi amacıyla testlerde ankor maddeler kullanılmıştır. Tüm formlarda yer alan 4 madde için formlara göre ortalamalar arasında fark olup olmadığı tek yönlü varyans analizi (ANOVA) ile incelenmiştir. Analizler SPSS 23 programı ile yapılmıştır. Tablo 14 grupların denkliklerine ilişkin ANOVA sonuçlarını vermektedir.

Tablo 14. Grupların Denkliklerine İlişkin ANOVA Sonuçları

Betimsel İstatistikler					ANOVA Sonuçları					
		N	X	s.s	Varyans Kaynağı	Kareler toplamı	Kareler Ort.	S.D.	F	P
s201 PISA	FormA	520	0,87	0,34	Gruplar arası Gruplar içi Toplam	1,056	0,211	5	1,987	0,08
	FormB	513	0,9	0,3						
	FormC	501	0,9	0,3						
	FormD	491	0,86	0,35						
	FormE	484	0,86	0,35						
	Form F	449	0,88	0,33						
	Toplam	2958	0,88	0,33						
					314,67		2957			
s501 TIMSS	FormA	520	0,51	0,5	Gruplar arası Gruplar içi Toplam	2,205	0,441	5	1,781	0,11
	FormB	513	0,45	0,5						
	FormC	501	0,46	0,5						
	FormD	491	0,44	0,5						
	FormE	484	0,43	0,5						
	Form F	449	0,43	0,5						
	Toplam	2958	0,45	0,5						
					733,25		2957			
s1119 PROJE	FormA	520	0,49	0,5	Gruplar arası Gruplar içi Toplam	0,229	0,046	5	0,183	0,97
	FormB	513	0,48	0,5						
	FormC	501	0,49	0,5						
	FormD	491	0,51	0,5						
	FormE	484	0,49	0,5						
	Form F	449	0,5	0,5						
	Toplam	2958	0,49	0,5						
					739,31		2957			
s6251 PROJE	FormA	520	0,32	0,47	Gruplar arası Gruplar içi Toplam	2,358	0,472	5	2,132	0,06
	FormB	513	0,38	0,49						
	FormC	501	0,34	0,47						
	FormD	491	0,29	0,45						
	FormE	484	0,33	0,47						
	Form F	449	0,34	0,47						
	Toplam	2958	0,33	0,47						
					655,32		2957			

Tablo 14’te görüldüğü gibi formların türlerine göre soruların ortalamaları arasında istatistiksel olarak anlamlı bir fark yoktur. Bu durum grup denkliklerinin bir göstergesidir. Grup denkliklerine ilişkin bir diğer analiz maddelerin ayırt edicilik parametrelerinin gruplar arası değişimidir. Ankor maddelere ilişkin ayırt edicilik parametreleri Tablo 15’te verilmiştir.

Tablo 15. *Ankor Maddelere İlişkin Madde Parametreleri*

Form	Soru Kodu	r_{jx}	Alt Grup	Üst Grup	Nokta çift serili r	Çift serili r
Form A	201	0,238	0,713	0,951	0,305	0,486
	501	0,665	0,14	0,805	0,54	0,677
	1119	0,535	0,203	0,738	0,442	0,555
	6251	0,378	0,147	0,524	0,349	0,456
Form B	201	0,219	0,775	0,972	0,379	0,652
	501	0,623	0,145	0,768	0,554	0,697
	1119	0,448	0,243	0,691	0,413	0,518
	6251	0,373	0,197	0,569	0,313	0,4
Form C	201	0,224	0,783	0,986	0,367	0,634
	501	0,637	0,152	0,789	0,541	0,679
	1119	0,491	0,25	0,741	0,436	0,546
	6251	0,381	0,163	0,544	0,328	0,425
Form D	201	0,296	0,652	0,949	0,391	0,586
	501	0,612	0,11	0,722	0,545	0,686
	1119	0,441	0,28	0,722	0,413	0,517
	6251	0,241	0,134	0,375	0,309	0,399
Form E	201	0,27	0,686	0,956	0,412	0,643
	501	0,694	0,137	0,831	0,555	0,7
	1119	0,466	0,291	0,757	0,411	0,516
	6251	0,33	0,177	0,507	0,317	0,386
Form F	201	0,348	0,645	0,992	0,453	0,731
	501	0,683	0,091	0,773	0,536	0,675
	1119	0,552	0,198	0,75	0,419	0,525
	6251	0,391	0,132	0,523	0,36	0,466

Tablo 15’te görüldüğü gibi soruların hem ayırt edicilik değerleri hem de madde test korelasyonları yeterlidir ve gruplar arası manidar farklılık göstermemektedir. Özellikle maddelerin nokta çift serili ve çift serili korelasyon katsayıları bütün maddelerin yeterli ayırt edicilik değerine sahip olduğuna işaret

etmektedir. Ayırt edicilik değerlerinin formlar arasında farklılık gösterip göstermediği Friedman Testi ile incelenmiştir. Her bir ankor maddenin formlar için değerleri tekrarlı ölçüm olarak kabul edilmiş ve non-parametrik Friedman testi sonucuna göre soruların formlara göre ayırt edicilik değerleri arasında anlamlı bir fark olmadığı görülmüştür ($\chi^2 = 1,331$, $p > 0,05$). Elde edilen bulgular sonucunda deneme uygulaması için belirlenen grupların denk oldukları ve bu gruplardan gelen veriler ile proje için geliştirilen soruların psikometrik özelliklerinin incelenebileceği anlaşılmaktadır.

3.4.4. Proje Sorularının Psikometrik Özelliklerinin İncelenmesi

Proje kapsamında geliştirilen soruların psikometrik özellikleri farklı yöntemler ile incelenmiştir. Öncelikle KTK'ya göre madde ayırt edicilik indeksleri ve madde test korelasyonları belirlenmiş daha sonra MTK parametrelerine göre sorulara ilişkin bulgulara ulaşılmıştır. Son olarak projenin temel istatistik modelini oluşturan BTM analizleri incelenmiştir. KTK analizlerinde maddelerin güçlük değerleri (P_j), ayırt edicilik parametreleri (r_{jx}) ve nokta çift serili (r_{pbis}) ve çift serili (r_{bis}) korelasyon katsayıları hesaplanmıştır (Haladyna ve Rodriguez, 2013; McDonald, 1999; Murphy ve Davidshofer, 2005). Tablo 16'da deneme uygulamasına ait KTK parametreleri verilmiştir.

Tablo 16. *Deneme Uygulaması Maddelerine İlişkin KTK Parametreleri*

No	P _j	r _{jx}	r _{pbis}	r _{bis}	No	P _j	r _{jx}	r _{pbis}	r _{bis}
1	%87,6	0,25	0,38	0,62	25	%45,5	0,69	0,59	0,74
2	%10,1	0,10	0,18	0,31	26	%48,1	0,60	0,47	0,59
3	%76,0	0,41	0,46	0,63	27	%19,4	0,30	0,36	0,51
4	%67,1	0,46	0,43	0,56	28	%79,0	0,47	0,54	0,76
5	%45,3	0,43	0,37	0,46	29	%61,4	0,61	0,55	0,70
6	%41,8	0,54	0,47	0,60	30	%45,9	0,71	0,61	0,77
7	%63,1	0,57	0,55	0,71	31	%23,4	0,39	0,42	0,58
8	%45,3	0,64	0,54	0,68	32	%43,0	0,56	0,48	0,61
9	%73,3	0,41	0,42	0,56	33	%35,7	0,54	0,50	0,64
10	%70,0	0,40	0,41	0,54	34	%50,5	0,51	0,45	0,56
11	%77,3	0,41	0,41	0,57	35	%67,3	0,44	0,45	0,58
12	%50,1	0,54	0,45	0,56	36	%9,6	0,21	0,37	0,65
13	%35,8	0,70	0,63	0,80	37	%0,5	0,01	0,11	0,54
14	%72,3	0,53	0,53	0,70	38	%30,5	0,40	0,39	0,51
15	%68,1	0,63	0,60	0,78	39	%30,1	0,46	0,46	0,60
16	%54,8	0,42	0,39	0,49	40	%54,8	0,70	0,57	0,72
17	%49,2	0,47	0,42	0,53	41	%10,2	0,28	0,41	0,70
18	%20,2	0,06	0,05	0,07	42	%5,8	0,16	0,38	0,77
19	%52,2	0,52	0,46	0,57	43	%66,6	0,54	0,49	0,63
20	%43,8	0,55	0,44	0,56	44	%53,8	0,45	0,41	0,51
21	%65,3	0,48	0,45	0,57	45	%56,0	0,62	0,55	0,69
22	%63,5	0,57	0,54	0,69	46	%20,2	0,35	0,39	0,56
23	%32,5	0,35	0,32	0,42	47	%53,5	0,58	0,52	0,65
24	%23,2	0,39	0,46	0,64	48	%45,0	0,55	0,50	0,63

Daha sonraki aşamada maddelerin MTK analizleri ile madde parametreleri ve madde karakteristik eğrileri (MKE) belirlenmiştir (Embretson ve Reise, 2013; Hambleton ve Swaminathan, 2013; Lord, 1980; Reckase, 2006). Analizler IRTPro 2.1 ve MPlus 5 programı kullanılarak yapılmıştır. Tablo 17’de deneme uygulamasında kullanılan soruların MTK’da 3 parametrelilik lojistik modele göre (3PLM) belirlenmiş A (ayırıcılık-eğim), B (güçlük-madde konumu) ve C (şans başarısı- düşük asimptot) parametreleri ve standart hataları verilmiştir.

Tablo 17. Deneme Uygulaması Maddelerine İlişkin MTK Parametreleri

No	A	s.e.	C	s.e.	B	s.e.	No	A	s.e.	C	s.e.	B	s.e.
1	0,48	0,12	-0,18	0,09	0,37	0,21	25	0,47	0,13	-1,14	0,11	2,42	0,64
2	0,94	0,16	-0,09	0,1	0,1	0,11	26	1,18	0,18	-1,5	0,14	1,27	0,17
3	0,7	0,16	-1,56	0,13	2,24	0,47	27	1,71	0,23	-0,49	0,13	0,28	0,08
4	1,45	0,3	-2,99	0,31	2,06	0,28	28	1,31	0,18	-0,68	0,12	0,52	0,1
5	2,14	1	-7,77	1,97	3,62	0,91	29	0,76	0,13	0,07	0,1	-0,09	0,13
6	0,86	0,17	-1,57	0,14	1,82	0,31	30	1,47	0,2	0,35	0,12	-0,24	0,08
7	0,43	0,12	-0,22	0,09	0,5	0,25	31	1,22	0,21	2,29	0,2	-1,87	0,24
8	0,81	0,15	0	0,1	-0,01	0,12	32	0,83	0,14	0,98	0,12	-1,17	0,2
9	0,82	0,15	0,72	0,11	-0,89	0,18	33	1,47	0,2	1,46	0,17	-1	0,12
10	0,33	0,12	-0,5	0,09	1,54	0,6	34	2,03	0,27	1,25	0,19	-0,61	0,08
11	0,95	0,17	0,95	0,12	-1	0,18	35	0,73	0,12	-0,04	0,1	0,06	0,14
12	1,22	0,64	-5,87	0,85	4,8	2,06	36	1,68	0,21	-0,25	0,14	0,15	0,08
13	2,53	0,65	-4,88	0,83	1,92	0,21	37	1,8	0,26	1,92	0,22	-1,07	0,11
14	0,7	0,14	0,17	0,1	-0,24	0,14	38	1,93	0,24	0,02	0,15	-0,01	0,08
15	1,24	0,19	1,53	0,15	-1,23	0,16	39	1,45	0,18	0,28	0,13	-0,19	0,09
16	0,7	0,14	1,18	0,12	-1,68	0,32	40	1,53	0,27	2,69	0,27	-1,76	0,21
17	2,26	0,34	-0,96	0,18	0,42	0,07	41	1,45	0,2	0,75	0,14	-0,52	0,09
18	0,74	0,13	-0,06	0,1	0,08	0,13	42	0,52	0,13	-0,72	0,1	1,4	0,37
19	0,83	0,14	0,17	0,1	-0,21	0,12	43	1,77	0,28	2,1	0,23	-1,19	0,12
20	0,97	0,15	0,78	0,11	-0,8	0,14	44	1,42	0,2	0,66	0,13	-0,47	0,09
21	1,22	0,17	0,18	0,11	-0,15	0,09	45	0,98	0,17	-1,41	0,14	1,44	0,23
22	1,17	0,2	2,03	0,17	-1,74	0,24	46	1	0,16	-1,01	0,12	1,02	0,17
23	1,94	0,27	-0,99	0,16	0,51	0,08	47	1,87	0,36	-3,27	0,37	1,75	0,2
24	-0,26	0,13	-1,4	0,11	-5,36	2,68	48	1,11	0,17	0,86	0,13	-0,78	0,13

Deneme uygulaması verileri BTM'ye göre DINA model kullanılarak analiz edilmiştir. BTM analizlerine göre elde edilen bulgular Tablo 18'de verilmiştir.

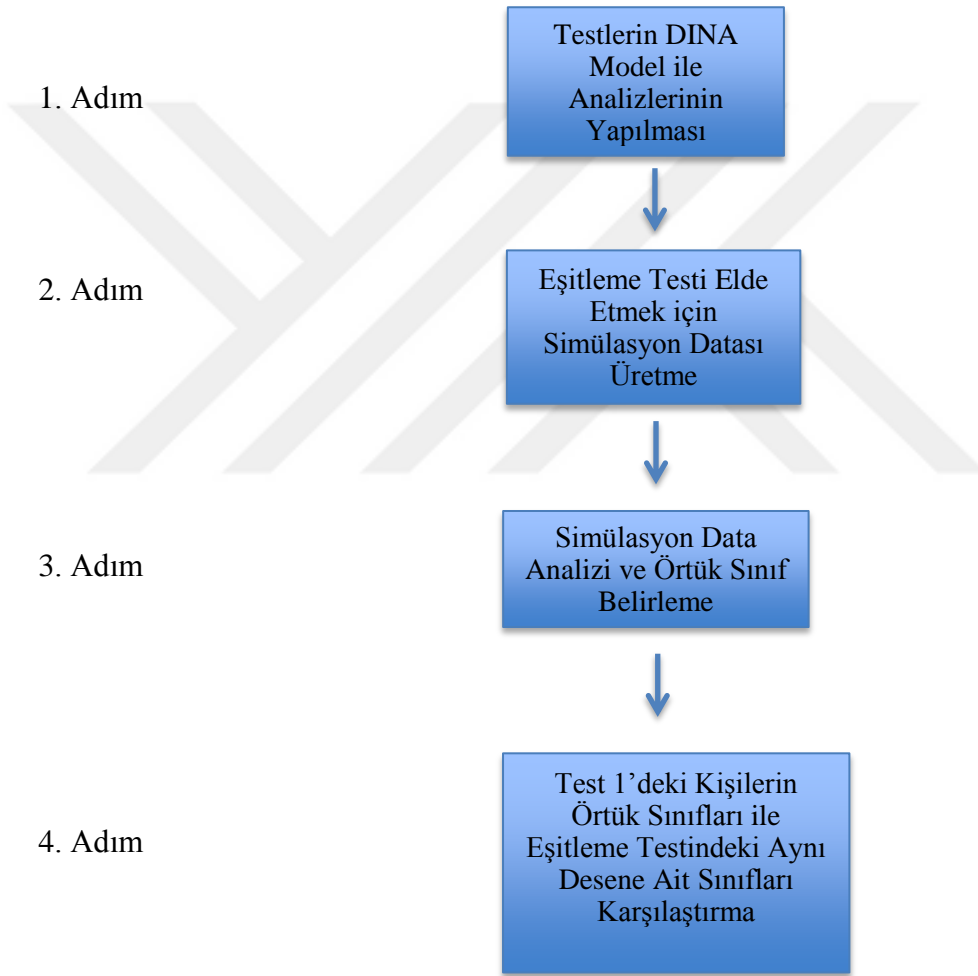
Tablo 18. *Deneme Uygulaması Maddelerine İlişkin BTM Parametreleri*

No	G	s.e.	S	s.e	No	G	s.e.	S	s.e
1	0,39	0,04	0,34	0,03	25	0,54	0,03	0,27	0,02
2	0,16	0,04	0,24	0,03	26	0,32	0,04	0,28	0,04
3	0,35	0,03	0,43	0,04	27	0,22	0,03	0,82	0,03
4	0,02	0,01	0,73	0,04	28	0,08	0,02	0,61	0,03
5	0,29	0,03	0,28	0,04	29	0,24	0,03	0,29	0,04
6	0,11	0,02	0,69	0,03	30	0,35	0,03	0,29	0,03
7	0,09	0,02	0,67	0,03	31	0,32	0,03	0,10	0,03
8	0,36	0,03	0,17	0,05	32	0,50	0,04	0,16	0,02
9	0,18	0,03	0,24	0,04	33	0,55	0,04	0,17	0,03
10	0,35	0,03	0,30	0,04	34	0,39	0,04	0,07	0,02
11	0,38	0,03	0,34	0,04	35	0,16	0,03	0,27	0,03
12	0,50	0,03	0,08	0,03	36	0,07	0,04	0,01	0,04
13	0,01	0,02	0,99	0,01	37	0,31	0,03	0,23	0,03
14	0,00	0,00	0,85	0,03	38	0,39	0,05	0,35	0,03
15	0,43	0,03	0,18	0,05	39	0,53	0,05	0,43	0,03
16	0,09	0,02	0,87	0,02	40	0,36	0,06	0,36	0,03
17	0,61	0,03	0,05	0,02	41	0,02	0,09	0,53	0,03
18	0,63	0,03	0,14	0,03	42	0,23	0,04	0,03	0,06
19	0,42	0,03	0,08	0,03	43	0,18	0,06	0,40	0,03
20	0,19	0,03	0,46	0,04	44	0,16	0,04	0,24	0,03
21	0,48	0,03	0,17	0,03	45	0,34	0,06	0,41	0,03
22	0,19	0,03	0,57	0,04	46	0,36	0,05	0,25	0,03
23	0,33	0,03	0,20	0,03	47	0,46	0,06	0,39	0,03
24	0,10	0,02	0,36	0,04	48	0,17	0,04	0,29	0,03

3.4.5. Eşitleme Yöntemi

Araştırmada kullanılmak üzere seçilen Form C ve Form D sırasıyla Test 1 ve Test 2 olarak adlandırılmıştır. Araştırmanın temel problemini oluşturan, Test 1'in uygulandığı öğrencilere Test 2'nin uygulanmış olması durumunda atanacakları örtük yetenek sınıflarının belirlenmesine yönelik bir takım analizler gerçekleştirilmiştir. Şekil 1'de analizlere yönelik işlem adımları gösterilmiştir.

Şekil 1. Örtük Sınıf Eşitleme Adımları



1. Adımda Test 1 ve Test 2'nin dataları ilk 7 maddeleri birbirinden farklı son 8 maddesi ise ankor maddeler olacak şekilde düzenlenmiştir. Testlerin DINA model ile analizleri yapılmış, maddelere ait g ve s parametreleri belirlenmiştir. Test 1'den elde edilen 501 kişiye ait cevap örüntüleri ve bu cevap örüntülerine karşılık gelen örtük yetenek sınıfları belirlenmiştir.

2. Adımda simülasyon aracılığıyla eşitleme datası elde edebilmek için Ox Edit programı kullanılarak 4 özelliğin ölçüldüğü 15 maddenin yer aldığı 50 000 kişilik data üretilmiştir. Kişi sayısının yüksek tutulmasının sebebi Test 1'den elde edilen cevap örüntülerinin aynısını bulmak amacıyla mümkün olduğunca fazla cevap örüntüsü elde etmektir. Parametre olarak Tablo 19'da gösterildiği gibi ilk 7 madde için Test 2'nin ilk 7 maddesinden elde edilen g ve s parametreleri, son 8 madde için ise Test 1'in ankor maddelerinden elde edilen g ve s parametreleri kullanılmıştır.

Tablo 19. Simülasyon Datasında Kullanılan Madde Parametreleri

Madde No.	g	s
1	gd1	Sd1
2	gd2	Sd2
3	gd3	Sd3
4	gd4	Sd4
5	gd5	Sd5
6	gd6	Sd6
7	gd7	Sd7
8	gc8	Sc8
9	gc9	Sc9
10	gc10	Sc10
11	gc11	Sc11
12	gc12	Sc12
13	gc13	Sc13
14	gc14	Sc14
15	gc15	Sc15

Test 1'deki ankor madde parametreleri

Simülasyon datanın analizinde kullanılan Q matris Tablo 20'de gösterilmektedir. α_1 , α_2 , α_3 ve α_4 sırasıyla iletişim ve ilişkilendirme, matematikleştirme, akıl yürütme ve strateji geliştirme, sembolik ve teknik dil kullanımı özelliklerini temsil etmektedir.

Tablo 20. Simülasyonda Kullanılan Q matris

Madde No.	Özellikler			
	α_1	α_2	α_3	α_4
1	0	1	1	0
2	0	0	1	0
3	1	0	1	0
4	0	1	1	0
5	1	1	1	0
6	1	1	1	1
7	0	1	0	1
8	0	0	0	1
9	0	1	0	0
10	0	1	0	1
11	1	1	0	1
12	1	1	1	0
13	1	0	1	0
14	1	0	1	0
15	1	1	1	0

3. Adımda bir önceki adımda üretilen 50 000 kişilik data için DINA model ile analizler yapılmıştır. 4 özelliğin ölçüldüğü 15 maddelik testte 50 000 kişinin cevap örüntüleri ve ait oldukları örtük yetenek sınıfları belirlenmiştir.

4. Adımda 50 000 kişilik eşitleme datasında Test 2'den elde edilen 491 kişinin cevap örüntüleri Microsoft Excel'de makro kodu kullanılarak bulunmuştur. Eşitleme datasından ve Test 1'den elde edilen aynı cevap örüntülerine karşılık gelen örtük yetenek sınıfları karşılaştırılmıştır. Eşitleme datasından elde edilen örtük yetenek sınıfı, Test 1'e giren öğrencilere Test 2 uygulaması durumunda ait olacakları örtük yetenek sınıflarını belirtmektedir.

BÖLÜM IV

BULGULAR

Bu bölümde araştırmanın ana problemi çerçevesinde öğrencilerin testlere verdikleri yanıtlara dayanılarak yapılan istatistiksel analizlerden elde edilen bulgular ve açıklamaları yer almaktadır.

4.1. Test 1 ve Test 2'nin Madde Analizi Sonuçlarına İlişkin Bulgular

Araştırma kapsamında kullanılan testlerin önce KTK'ya daha sonra BTM'ye göre hesaplanan madde analizi sonuçlarına ilişkin bulgular aşağıda verilmiştir.

4.1.1. Testlerin KTK'ya Göre Psikometrik Özelliklerine İlişkin Bulgular

Tablo 21'de testlere ait madde güçlükleri ve madde ayırt edicilik indeksleri sonuçları gösterilmektedir.

Tablo 21. Test 1 ve Test 2'ye ait Madde Analizi Sonuçları

Test 1			Test 2		
Madde No.	P	Rpbis	Madde No.	P	Rpbis
1	0,74	0,37	1	0,66	0,38
2	0,66	0,46	2	0,70	0,52
3	0,31	0,39	3	0,42	0,47
4	0,53	0,52	4	0,42	0,59
5	0,54	0,46	5	0,23	0,46
6	0,11	0,18	6	0,20	0,05
7	0,63	0,54	7	0,56	0,55
8	0,77	0,46	8	0,75	0,46
9	0,90	0,37	9	0,84	0,39
10	0,46	0,54	10	0,44	0,55
11	0,37	0,63	11	0,35	0,62
12	0,34	0,33	12	0,25	0,29
13	0,50	0,45	13	0,52	0,45
14	0,49	0,44	14	0,51	0,41
15	0,34	0,45	15	0,37	0,54
Ortalama	0,51	0,44	Ortalama	0,48	0,45

Tablo 21'de verilen madde analizi sonuçları değerlendirildiğinde, her iki testten elde edilen madde güçlüğü ve madde ayırt edicilik indekslerinin ortalamalarının birbirine yakın olduğu görülmektedir. Testlerin madde güçlükleri

ortalamları 0,50 civarındadır. Test 1'e ait maddelerin madde güçlükleri 0,11 ile 0,90 arasında değerler almaktadır. Madde güçlüğü 0,30 ve 0,30'un altında 1 madde, 0,71'in üzerinde 3 madde ve orta güçlükte 11 madde bulunmaktadır. Test 2 için madde güçlük değerlerinin 0,20 ile 0,84 arasında değiştiği görülmektedir. Madde güçlüğü 0,30 ve 0,30'un altında 3 madde, 0,71'in üzerinde 2 madde ve orta güçlükte 10 madde bulunmaktadır.

Madde ayırt edicilik indeksleri incelendiğinde Test 1'deki maddelerin ayırt ediciliklerinin 0,18 ile 0,63 arasında değerler aldığı görülmektedir ve ortalaması 0,44 olarak hesaplanmıştır. Test 2 incelendiğinde ayırt ediciliklerinin 0,05 ile 0,62 arasında değerler aldığı ve ortalamasının 0,45 olduğu görülmektedir.

4.1.2. Testlerin BTM'ye Göre Model Veri Uyumu Parametrelerine İlişkin Bulgular

DINA model ile madde düzeyinde belirlenen parametreler g ve s parametreleri olmak üzere iki tanedir. Tahmin parametresi olarak da bilinen g parametresi, ilgili maddeyi doğru cevaplayabilmek için gerekli özelliklerden en az birine sahip olmayanların o maddeyi doğru cevaplama olasılığıdır. Kaydırma parametresi olan s parametresi, ilgili maddeyi doğru cevaplayabilmek için gerekli özelliklerin tamamına sahip olan bireylerin o maddeyi yanlış cevaplama olasılığıdır (de la Torre ve Douglas, 2008). g ve s parametrelerinin düşük olması maddenin o maddeyi çözmek için gerekli özelliklere sahip olanlar tarafından doğru çözülebilmek olasılığının yüksek olması anlamına gelmektedir (Zhang, 2006). Tablo 22'de Test 1'den elde edilen g ve s parametreleri ile bu parametrelere ait standart hatalar verilmiştir.

Tablo 22. Test 1'e Ait Madde Parametreleri

Madde No.	g	se(g)	s	se(s)
1	0,66	0,03	0,08	0,03
2	0,46	0,05	0,17	0,04
3	0,20	0,03	0,45	0,05
4	0,34	0,03	0,05	0,04
5	0,45	0,03	0,13	0,05
6	0,09	0,01	0,78	0,05
7	0,47	0,03	0,02	0,03
8	0,50	0,05	0,01	0,02
9	0,80	0,03	0,01	0,01
10	0,30	0,03	0,21	0,04
11	0,20	0,02	0,01	0,04
12	0,30	0,02	0,51	0,06
13	0,38	0,03	0,23	0,05
14	0,36	0,03	0,22	0,05
15	0,26	0,02	0,34	0,06
Ortalama	0,39	0,03	0,22	0,04

Test 1'de yer alan maddelere ait g parametreleri 0,09-0,80 arasında değer almaktadır ve ortalaması 0,39'dur. Maddelere ait s parametreleri incelendiğinde 0,01-0,78 arasında değerler aldığı ve ortalamasının 0,22 olduğu görülmektedir. Test 1'deki orta güçlükte maddelerin g parametresi ortalaması 0,47, kolay maddelerin 0,80'dir. Zor olan tek bir maddenin g değeri 0,11'dir. Tablo 23'te Test 2'den elde edilen g ve s parametreleri ile bu parametrelere ait standart hatalar verilmiştir.

Tablo 23. *Test 2'ye Ait Madde Parametreleri*

Madde No.	g	se(g)	s	se(s)
1	0,56	0,03	0,13	0,03
2	0,44	0,04	0,04	0,03
3	0,25	0,03	0,16	0,05
4	0,19	0,03	0,1	0,04
5	0,14	0,02	0,38	0,06
6	0,23	0,02	0,94	0,03
7	0,38	0,03	0,08	0,04
8	0,55	0,05	0,07	0,03
9	0,69	0,04	0,04	0,02
10	0,27	0,03	0,21	0,04
11	0,19	0,02	0,03	0,05
12	0,22	0,02	0,6	0,06
13	0,43	0,03	0,26	0,05
14	0,41	0,03	0,26	0,05
15	0,27	0,02	0,22	0,05
Ortalama	0,35	0,03	0,24	0,04

Test 2'de yer alan maddelere ait g parametreleri 0,14-0,69 arasında değerler almaktadır ve ortalaması 0,35'tir. Maddelere ait s parametreleri incelendiğinde 0,03-0,94 arasında değer aldığı ve ortalamasının 0,24 olduğu görülmektedir. Zor maddelerin g parametresi ortalaması 0,23; orta güçlükteki maddelerin 0,50 ve kolay maddelerin 0,80 olarak hesaplanmıştır. Testlere ait veri uyum istatistiklerine ilişkin veriler Tablo 24'te verilmiştir.

Tablo 24. *Testlerin Model Veri Uyumu İstatistikleri*

	Test 1	Test 2
-2LL	8657,7662	8549,8781
AIC	8747,7662	8639,8781
BIC	8937,5135	8828,7181

Testlere ait madde veri uyum istatistiklerine ilişkin istatistikler Tablo 25'te verilmiştir.

Tablo 25. *Testlerin Madde Veri Uyumu İstatistikleri*

	Prop	Test 1		Prop	Test 2	
		Z	Log		Z	Log
Ortalama Sapma	0,0189	0,0621	0,3245	0,0128	0,0696	0,3398
Maksimum Sapma	0,0593	0,2313	0,9545	0,0320	0,2387	1,0880
Standart Hata	0,0169	0,0448	0,3118	0,0219	0,0453	0,2201

Model veri uyum istatistiklerine göre Test 2'nin uyum düzeyi daha fazladır. İki test arasındaki uyum farkı Test 2'nin g parametreleri ortalamasının Test 1'e göre daha düşük olması ile açıklanabilir.

4.2. Test 1'e İlişkin Örtük Yetenek Sınıflarının ve Özelliklerin Dağılımına Ait

Bulgular

BTM'de öğrenciler uygulanan test için belirlenen örtük yeteneklere sahip olup olmama durumlarına göre örtük sınıflara atanmaktadır. Araştırmada kullanılan DINA modelin öğrencileri örtük sınıflara atama olasılıkları ve örtük yeteneklerin gözlenme olasılıkları, testi alan gruplar ve uygulanan testler göz önüne alınarak hesaplanmıştır. Tablo 26, Test 1'e ait örtük sınıfları ve bu sınıflara ait sonsal olasılıkları göstermektedir.

Tablo 26. *Test 1'e Ait Örtük Sınıflar ve Sonsal Olasılıkları*

Örtük Sınıflar	Olasılık	Örtük Sınıflar	Olasılık
0000	0,0659	0110	0,0482
1000	0,0659	0101	0,0541
0100	0,0602	0011	0,0554
0010	0,0588	1110	0,0498
0001	0,0556	1101	0,0563
1100	0,0602	1011	0,048
1010	0,0482	0111	0,0658
1001	0,0556	1111	0,1521

Örtük sınıfların sonsal olasılıkları incelendiğinde gözlenme olasılığı en fazla olması beklenen örtük yetenek sınıfının 1111 sınıfı olduğu belirlenmiştir ve gözlenme olasılık değeri 0,1521'dir. Gözlenme olasılığı en düşük olması beklenen

örtük yetenek sınıfı ise 1011 sınıfıdır ve gözlenme olasılık değeri 0,0480 olarak hesaplanmıştır. Tablo 27, araştırma kapsamında belirlenen 4 özelliğin gözlenme olasılıkları göstermektedir.

Tablo 27. *Test 1’de Özelliklerin Gözlenme Olasılıkları*

Özellikler	Gözlenme Olasılığı
İletişim ve İlişkilendirme (α_1)	0,5360
Matematikleştirme (α_2)	0,5467
Akıl Yürütme ve Strateji Geliştirme (α_3)	0,5263
Sembolik ve Teknik Dil Kullanımı (α_4)	0,5427

Özelliklerin gözlenme olasılıkları incelendiğinde gözlenme olasılığı en fazla olan özellik ‘Matematikleştirme (α_2)’ özelliğidir ve değeri 0,5467 olarak hesaplanmıştır. Gözlenme olasılığı en az olan özellik ‘Akıl Yürütme ve Strateji Geliştirme (α_3)’ özelliğidir ve gözleme olasılığı 0,5263’tür.

4.3. Test 2’ye İlişkin Örtük Yetenek Sınıflarının Dağılımına İlişkin Bulgular

Tablo 28, Test 2’ye ait örtük sınıflar ve bu sınıflara ait sonsal olasılıkları göstermektedir.

Tablo 28. *Test 2’ye Ait Örtük Sınıflar ve Sonsal Olasılıkları*

Örtük Sınıflar	Olasılık	Örtük Sınıflar	Olasılık
0000	0,0703	0110	0,0502
1000	0,0703	0101	0,0580
0100	0,0619	0011	0,0585
0010	0,0570	1110	0,0472
0001	0,0564	1101	0,0470
1100	0,0619	1011	0,0440
1010	0,0387	0111	0,0697
1001	0,0564	1111	0,1527

Örtük sınıfların sonsal olasılıkları incelendiğinde gözlenme olasılığı en fazla olması beklenen örtük yetenek sınıfının 1111 sınıfı olduğu belirlenmiştir ve gözlenme olasılık değeri 0,1527’dir. Gözlenme olasılığı en düşük olması beklenen örtük yetenek sınıfı 1010 sınıfıdır ve gözlenme olasılık değeri 0,0387 olarak

hesaplanmıştır. Tablo 29, araştırma kapsamında belirlenen 4 özelliğin gözlenme olasılıklarını göstermektedir.

Tablo 29. *Test 2’de Özelliklerin Gözlenme Olasılıkları*

Özellikler	Gözlenme Olasılığı
İletişim ve İlişkilendirme (α_1)	0,5181
Matematikleştirme (α_2)	0,5484
Akıl Yürütme ve Strateji Geliştirme (α_3)	0,5179
Sembolik ve Teknik Dil Kullanımı (α_4)	0,5427

Özelliklerin gözlenme olasılıkları incelendiğinde gözlenme olasılığı en fazla olan özellik ‘Matematikleştirme (α_2)’ özelliğidir ve değeri 0,5484 olarak hesaplanmıştır. Gözlenme olasılığı en az olan özellik ‘Akıl Yürütme ve Strateji Geliştirme (α_3)’ özelliğidir ve gözleme olasılığı 0,5179’dur.

4.4. Test 1’i Alan Öğrencilerin Test 2 İçin Eşitlenmiş Örtük Yetenek

Sınıflarındaki Değişime İlişkin Bulgular

Test 1’e ve eşitlenen teste ait örtük yetenek sınıfları ve bu sınıflara dahil olan cevaplayıcı sayıları Tablo 30’da verilmiştir.

Tablo 30. Test 1 ve Eşitlenen Teste Ait Örtük Yetenek Sınıfları

		Eşitlenen Teste Ait Örtük Yetenek Sınıfları					
		0 0 0 0	0 0 0 1	0 0 1 0	0 1 0 0	0 0 1 1	0 1 0 1
Test 1	0 0 0 0	26	22	5	0	4	1
	0 0 0 1	0	42	0	0	15	0
	0 0 1 0	1	0	3	0	3	0
	0 1 0 0	19	1	5	1	0	1
	1 0 0 0	0	0	3	0	3	0
	0 0 1 1	0	3	0	0	5	0
	0 1 0 1	0	9	0	0	2	11
	0 1 1 0	0	0	3	0	5	0
	1 0 0 1	0	1	0	0	8	0
	1 0 1 0	0	0	3	0	4	0
	1 1 0 0	0	0	0	0	0	0
	0 1 1 1	0	0	0	0	6	0
	1 0 1 1	0	0	0	0	9	0
	1 1 0 1	0	4	0	0	4	9
	1 1 1 0	0	0	3	0	0	0
	1 1 1 1	0	0	0	0	2	0

Tablo 30. Test 1 ve Eşitlenen Teste Ait Örtük Yetenek Sınıfları (Devam)

		Eşitlenen Teste Ait Örtük Yetenek Sınıfları					
		0 1 1 0	1 0 1 0	0 1 1 1	1 0 1 1	1 1 1 0	1 1 1 1
Test 1	0 0 0 0	0	0	3	0	1	1
	0 0 0 1	1	0	2	3	5	6
	0 0 1 0	0	0	0	0	0	1
	0 1 0 0	0	0	0	0	1	0
	1 0 0 0	0	1	0	1	2	0
	0 0 1 1	0	0	1	0	0	3
	0 1 0 1	0	0	8	0	0	2
	0 1 1 0	3	0	10	0	2	3
	1 0 0 1	0	1	0	1	2	2
	1 0 1 0	0	3	0	3	1	1
	1 1 0 0	0	0	0	0	1	0
	0 1 1 1	0	0	18	1	0	10
	1 0 1 1	0	1	1	13	2	4
	1 1 0 1	0	0	5	1	0	9
	1 1 1 0	1	0	2	1	7	10
	1 1 1 1	0	0	19	8	5	62

Test 1'in uygulandığı 501 kişilik cevaplayıcı grubuna ait veriler üzerinden DINA model aracılığıyla yapılan analizler sonucunda 452 farklı cevap örüntüsü ve 16 örtük yetenek sınıfı elde edilmiştir. Buradan elde edilen 452 cevap örüntüsü, eşitlenen testte incelendiğinde ise bu örüntülere karşılık gelen toplam 12 örtük yetenek sınıfı olduğu belirlenmiştir. Test 1'in uygulandığı grubun yaklaşık %39'u eşitleme sonrası yine aynı örtük sınıflara atanmıştır. Grubun yaklaşık %31'i olduğundan daha yüksek bir örtük yetenek sınıfına ve yaklaşık %22'si ise olduğundan daha düşük bir sınıfa atanmıştır. Kalan 42 kişi aynı sayıda özelliğe sahip fakat farklı bir sınıfta yer almaktadır. Test 1'de 0000 sınıfına dahil 37 kişi eşitlenen testte daha yüksek sınıfa atanmıştır. Bunlardan 27 tanesi 1 özellik, 5 tanesi 2 özellik, 4 tanesi 3 özellik, 1 tanesi 4 özellik içeren sınıflara atanmıştır. Test 1'de 2 veya daha altında özelliğe sahip sınıflara atanan 19 kişinin eşitleme sonrasında 1111 sınıfında yer aldığı görülmektedir.

4.5. Eşitleme Sonrası Örtük Yetenek Sınıflarının Toplam Puanlara Göre

Değişimine İlişkin Bulgular

Örtük yetenek sınıflarının toplam puana göre değişimi Tablo 31’de gösterilmektedir. Eşitlenmiş testte bazı örtük sınıflara atanan cevaplayıcı yoktur dolayısıyla tabloda bu sınıflara ait değerler yer almamaktadır. Tablonun en sağındaki iki sütunda sırasıyla Test 1’in ve eşitlenmiş testteki sınıfların özellik sayılarına göre ortalama puan değerleri verilmiştir.

Tablo 31. Örtük Yetenek Sınıflarının Toplam Puanlara Göre Değişimi

		Test 1			Eşitlenmiş Test				
		N	X	ss	N	X	ss		
1 Özellik	0 0 0 0	63	4,06	1,86	46	2,91	1,26	4,06	2,91
	0 0 0 1	74	5,81	1,80	82	4,72	1,42		
	0 0 1 0	8	4,75	1,67	25	5,72	1,34	5,15	5,15
	0 1 0 0	28	3,75	1,62	1	5,00			
	1 0 0 0	10	6,30	1,83					
2 Özellik	0 0 1 1	12	6,33	1,97	70	7,00	1,43		
	0 1 1 0	26	7,81	1,36	5	7,40	1,14		
	1 1 0 0	1	8,00					7,28	7,26
	1 0 0 1	15	7,33	1,68					
	1 0 1 0	15	7,07	1,16	6	7,33	1,03		
3 Özellik	0 1 0 1	32	7,16	1,14	22	7,32	1,04		
	0 1 1 1	35	9,17	1,40	69	9,14	1,24		
	1 0 1 1	30	8,33	1,24	32	8,53	1,92	9,07	9,05
	1 1 0 1	32	8,78	1,29					
	1 1 1 0	24	10,00	1,59	29	9,48	2,03		
	1 1 1 1	96	11,81	1,23	114	11,14	1,74	11,81	11,14

N: Kişi Sayısı, X: Ortalama, ss: Standart Sapma

Tablo değerlerine göre, eşitlenmiş testin ortalamaları Test 1’e kıyasla genel olarak daha düşüktür. 1111 sınıfına doğru ortalamalar birbirine yaklaşma eğilimi göstermiştir. Önemli miktarda farklılıklar göze çarpmamaktadır. Test 1’de en çok kişinin atandığı sınıf 1111, en az kişinin atandığı sınıf 1100 sınıfıdır. Eşitlenmiş testte ise en çok kişi 1111, en az sayıda kişi 0100 sınıfına atanmıştır. Eşitlenmiş testte 1000, 1001, 1100 ve 1101 sınıflarına dahil edilen cevaplayıcı gözlemlenmemiştir.

Sınıfların özellik sayısı arttıkça Test 1'e ait sınıfların ortalamasının eşitlenen teste kıyasla daha yüksek olma eğiliminde olduğu görülmektedir.

4.6. Eşitleme Sonrası Örtük Yetenek Sınıflarının MTK Puanlarına Göre

Değişimine İlişkin Bulgular

Örtük yetenek sınıflarının MTK puanlarına göre değişimi Tablo 32'de gösterilmektedir. Eşitlenmiş testte bazı örtük sınıflara atanan cevaplayıcı yoktur dolayısıyla tabloda bu sınıflara ait değerler yer almamaktadır. Tablonun en sağındaki iki sütunda özellik sayılarına göre sırasıyla Test 1'in ve eşitlenmiş testin ortalama değerleri verilmiştir.

Tablo 32. Örtük Yetenek Sınıflarının MTK Puanlarına Göre Değişimi

		Test 1			Eşitlenmiş Test				
		N	X	ss	N	X	ss		
1 Özellik	0 0 0 0	63	-,98	,56	46	-1,30	,41	-,98	-1,30
	0 0 0 1	74	-,52	,46	82	-,71	,46		
	0 0 1 0	8	-,94	,52	25	-,62	,46	-,75	-,71
	0 1 0 0	28	-1,05	,41	1	-,80			
	1 0 0 0	10	-,51	,55					
2 Özellik	0 0 1 1	12	-,43	,47	70	-,17	,45		
	0 1 1 0	26	,06	,40	5	-,16	,33		
	1 1 0 0	1	-,23						
	1 0 0 1	15	-,07	,48					
	1 0 1 0	15	-,31	,32	6	-,31	,24	-,17	-,17
3 Özellik	0 1 0 1	32	-,07	,21	22	,12	,38		
	0 1 1 1	35	,36	,31	69	,49	,42		
	1 0 1 1	30	,08	,36	32	,16	,59	,34	,32
	1 1 0 1	32	,51	,26					
	1 1 1 0	24	,42	,39	29	,30	,63		
	1 1 1 1	96	1,22	,33	114	,91	,57	1,22	,91

Test 1'de ortalaması en yüksek ve en düşük sınıflar sırasıyla 1111 ve 0100 sınıflarıdır. Eşitlenen testte en yüksek ortalamaya sahip sınıf 1111 iken en düşük 0000 sınıfıdır. Sınıfların özellik sayısı arttıkça Test 1'e ait sınıfların ortalamasının eşitlenen teste kıyasla daha yüksek olma eğiliminde olduğu görülmektedir.

4.7. Ölçülen Özelliklere Göre Örtük Yetenek Sınıflarının Eşitlenmesi Sonrası Öğrencilerin Özelliklere Sahip Olup Olmama Durumlarına İlişkin Bulgular

Eşitleme sonrası öğrencilerin araştırma kapsamında belirlenen 4 özelliğe sahip olup olmama durumları incelenmiştir.

4.7.1. Öğrencilerin İletişim ve İlişkilendirme Özelliğine Sahip Olup Olmama

Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular

Tablo 33'te Q matris için belirlenen özelliklerden birincisi olan iletişim ve ilişkilendirme özelliğinin Test 1'deki ve eşitlenen testteki durumu gösterilmektedir.

Tablo 33. Öğrencilerin İletişim ve İlişkilendirme Özelliğine Sahip Olup Olmama Düzeyleri

		Eşitlenen Test		Toplam
		0	1	
Test 1	0	239	39	278
	1	81	142	223
Toplam		320	181	501

Test 1'in uygulandığı cevaplayıcılar arasında 'İletişim ve İlişkilendirme' özelliğine sahip olmayan dolayısıyla 0 sınıfına atanan 278 kişi vardır. Eşitlenen testte, Test 1'de 0 sınıfına atanmış kişilerden 239 tanesinin 0 sınıfına atandığı görülmektedir. Test 1'de bu özelliğe sahip olanların yer aldığı 1 sınıfında 223 kişi vardır, eşitlenen testte ise bu kişilerden 81 tanesi 1 sınıfına atanmıştır. Testlerin bu özelliğe sahip olup olmama konusunda verdiği kararlar incelendiğinde öğrencilerin %76'sının her iki testte de aynı sınıfa atandığı belirlenmiştir.

4.7.2. Öğrencilerin Matematikleştirme Özelliğine Sahip Olup Olmama Durumunun

Eşitleme Sonrası Değişimine İlişkin Bulgular

Tablo 34'te Q matris için belirlenen özelliklerden ikincisi olan Matematikleştirme özelliğinin Test 1'deki ve eşitlenen testteki durumu gösterilmektedir.

Tablo 34. Öğrencilerin Matematikleştirme Özelliğine Sahip Olup Olmama Düzeyleri

		Eşitlenen Test		Toplam
		0	1	
Test 1	0	187	40	227
	1	74	200	274
Toplam		261	240	501

Test 1'in uygulandığı cevaplayıcılar arasında 'Matematikleştirme' özelliğine sahip olmayan dolayısıyla 0 sınıfına atanan 227 kişi vardır. Eşitlenen testte ise Test 1'de 0 sınıfına atanmış kişilerden 187 tanesinin 0 sınıfına atandığı belirlenmiştir. Test 1'de bu özelliğe sahip olanların yer aldığı 1 sınıfta 274 kişi vardır. Eşitlenen testte bu kişilerden 200 tanesi 1 sınıfına atanmıştır. Testlerin bu özelliğe sahip olup olmama konusunda verdiği kararlar incelendiğinde öğrencilerin %77'sinin her iki testte de aynı sınıfa atandığı belirlenmiştir.

4.7.3. Öğrencilerin Akıl Yürütme Ve Strateji Geliştirme Özelliğine Sahip Olup

Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular

Tablo 35'te Q matris için belirlenen özelliklerden üçüncüsü olan Akıl Yürütme ve Strateji Geliştirme özelliğinin Test 1'deki ve eşitlenen testteki durumu gösterilmektedir.

Tablo 35. Öğrencilerin Akıl Yürütme Ve Strateji Geliştirme Özelliğine Sahip Olup Olmama Düzeyleri

		Eşitlenen Test		Toplam
		0	1	
Test 1	0	147	108	255
	1	4	242	246
Toplam		151	350	501

Test 1'in uygulandığı cevaplayıcılar arasında 'Akıl Yürütme ve Strateji Geliştirme' özelliğine sahip olmayan dolayısıyla 0 sınıfına atanan 255 kişi vardır. Eşitlenen testte ise Test 1'de 0 sınıfına atanmış kişilerden 147 tanesinin 0 sınıfına atandığı görülmektedir. Test 1'de bu özelliğe sahip olanların yer aldığı 1 sınıfta 246 kişi vardır, eşitlenen testte ise bu kişilerden 242 tanesi 1 sınıfına atanmıştır.

Testlerin bu özelliğe sahip olup olmama konusunda verdiği kararlar incelendiğinde öğrencilerin %78'inin her iki testte de aynı sınıfa atandığı belirlenmiştir.

4.7.4. Öğrencilerin Sembolik Ve Teknik Dil Kullanımı Özelliğine Sahip Olup

Olmama Durumunun Eşitleme Sonrası Değişimine İlişkin Bulgular

Tablo 36'da Q matris için belirlenen özelliklerden dördüncüsü olan Sembolik ve Teknik Dil Kullanımı özelliğinin Test 1'deki ve eşitlenen testteki durumu gösterilmektedir.

Tablo 36. Öğrencilerin Sembolik Ve Teknik Dil Kullanımı Özelliğine Sahip Olup Olmama Düzeyleri

		Eşitlenen Test		Toplam
		0	1	
Test 1	0	95	80	175
	1	17	309	326
Toplam		112	389	501

Test 1'in uygulandığı cevaplayıcılar arasında 'Sembolik ve Teknik Dil Kullanımı' özelliğine sahip olmayan dolayısıyla 0 sınıfına atanan 175 kişi vardır. Eşitlenen testte ise Test 1'de 0 sınıfına atanmış kişilerden 95 tanesinin 0 sınıfına atandığı görülmektedir. Test 1'de bu özelliğe sahip olanların yer aldığı 1 sınıfında 326 kişi vardır, eşitlenen testte ise bu kişilerden 309 tanesi 1 sınıfına atanmıştır. Testlerin bu özelliğe sahip olup olmama konusunda verdiği kararlar incelendiğinde öğrencilerin %81'inin her iki testte de aynı sınıfa atandığı belirlenmiştir.

BÖLÜM V

TARTIŞMA VE YORUM

Bu bölümde araştırmada kullanılan testlere ilişkin elde edilen bulgular tartışılmış ve yorumlanmıştır.

5.1. Test 1 ve Test 2'ye Ait Madde Analizi Sonuçlarına İlişkin Bulguların

Tartışılması ve Yorumu

Birinci alt probleme ilişkin bulgular incelendiğinde testlerin gerek KTK gerekse BTM açısından yeterli uyum düzeyine sahip olduğu görülmüştür. Testlerin madde güçlüklerinin ve ayırt ediciliklerinin birbirine yakın olduğu, önemli düzeyde farklılıklar olmadığı gözlemlenmiştir. g parametresi ile madde güçlüğü arasında yüksek düzeyde ilişki olduğu göze çarpmaktadır. Test 1'e ait maddelerin madde güçlükleri ve g parametreleri arası korelasyon 0,95, Test 2'nin 0,92 olarak hesaplanmıştır. Dolayısıyla aralarındaki ilişki pozitif yönlü olduğundan madde cevaplanma oranı arttıkça maddeye ait g parametresi de yükselmektedir. s parametresi ile ayırt edicilik arası ters yönlü ve güçlü bir ilişki vardır. s parametreleri ve ayırt edicilikler arası korelasyon Test 1 için -0,74, Test 2 için -0,84 olarak hesaplanmıştır. Bu durum madde test korelasyonunun s parametresi ile ilişkili olduğunu göstermektedir. Bu ilişkinin yapısının madde test korelasyonu yöntemlerinin tek boyutlu yapısından kaynaklandığı da söylenebilir. BTM'de maddenin ilişkili olduğu özellik sayısı arttıkça (madde doygunluğu) s parametresinin fit verme düzeyinin buna bağlı olarak düştüğü söylenebilir.

5.2. Test 1 ve Test 2'ye İlişkin Örtük Yetenek Sınıflarının ve Özelliklerin

Dağılımına İlişkin Bulguların Tartışılması ve Yorumu

İkinci ve üçüncü alt probleme ilişkin bulgular birlikte göz önüne alınmıştır ve örtük yeteneklerin dağılımına ilişkin bulgular incelendiğinde iki testte de gözlenme olasılığı en yüksek örtük yetenek sınıfı 1111 olarak belirlenmiştir. En düşük olasılığa sahip sınıflar Test 1'de 1011, Test 2'de 1010 sınıfıdır. Diğer sınıflara ait olasılık değerleri birbirine yakındır. Bu durum iki testin büyük örneklemelere uygulandığında birbirine yakın sınıflama yapacağının bir göstergesidir.

Özellikler birbirine yakın değerler almakla birlikte Test 1’de gözlenen olasılık değerlerinin daha yüksek olduğu görülmektedir. Bu durum aynı özellikleri ölçen farklı maddelerden oluşan iki ayrı testin cevaplayıcıları atayacakları sınıflarda da farklılaşmalar oluşacağına bir göstergesidir. Her iki formda da gözlenme sıklığı en az olan özellik Akıl Yürütme ve Strateji Geliştirme (α_3), en fazla olan özellik ‘Matematikleştirme (α_2)’ olarak belirlenmiştir. İki test için belirlenen sonsal olasılıkların sıralamalarının değişmemesi ancak ağırlıklarının değişmesi testlerin aynı özellikleri farklı düzeylerde ölçtüğünün bir göstergesi olarak kabul edilebilir.

5.3. Test 1’i Alan Öğrencilerin Test 2 İçin Eşitlenmiş Örtük Yetenek

Sınıflarındaki Değişime İlişkin Bulguların Tartışılması ve Yorumu

Dördüncü alt problem olan Test 1 uygulanan cevaplayıcıların Test 2’yi almaları durumunda atanacakları örtük yetenek sınıflarının belirlenmesi üzerine bulgular incelenmiştir. Sınıfların değişimi gözlemlendiğinde eşitleme sonrası cevaplayıcıların genel olarak daha düşük sınıflara doğru atandıkları belirlenmiştir. Öğrencilerin ilk atandıkları sınıflar ile eşitleme sonrası belirlenen sınıfları arasında 2 veya daha az özelliğe sahiptir kararı verilenlerin oranı %14 düzeyindedir. Bu durum eşitleme sonrasında öğrencilerin ilk sınıflamalara göre daha düşük sınıflara atanma eğiliminde olduğunun bir göstergesi olarak görülebilir. Eşitlenen testin yani Test 2’nin madde güçlük düzeyi ve g parametrelerinin daha düşük olması sebebi ile Test 1’i alan öğrencilerin eşitlenmiş testte atandıkları sınıfların daha düşük çıkması beklenen bir durumu gösterdiği söylenebilir.

5.4. Eşitleme Sonrası Örtük Yetenek Sınıfların Toplam Puanlarına ve MTK

Puanlarına Göre Değişimine İlişkin Bulguların Tartışılması ve Yorumu

Öğrencilerin atandıkları sınıflara ait test puanlarınının değişimi KTK ve MTK ile incelenmiş ve sonuçlar birbirine yakın bulunmuştur. Karşılaştırmaların istatistiksel olarak yapılamamasının nedeni eşitleme sonrası gözlenen sınıf dağılımındaki azalmadan kaynaklanmaktadır. Test 1’de 16 sınıf gözlenirken eşitlenen testte 12 sınıf gözlenmiştir. Takip eden analizler ile incelendiğinde öğrencilerin önceki ve sonraki sınıflarına karşılık gelen KTK ve MTK puanları arasında 0,95 ve 0,94 spearman korelasyon hesaplanmıştır. Bu durum ilk sınıflamalar ve eşitleme sonrasında elde edilen sınıflar arasında gerek MTK gerekse KTK

puanları açısından uyumlu olduđuna işaret etmektedir. Eşitleme işlemi sonrasında öğrencilerin yetenek düzeyleri ile atandıkları sınıflar arasında yüksek düzeyde bir uyum gözlenmektedir.

5.5. Ölçülen Özelliklere Göre Örtük Yetenek Sınıflarının Eşitlenmesi Sonrası Öğrencilerin Özelliklere Sahip Olup Olmama Durumlarına İlişkin Bulguların

Tartışılması ve Yorumu

Ölçülen özelliklere göre örtük yetenek sınıflarının eşitlenmesi sonrası öğrencilerin bu özelliklere sahip olup olmama durumlarına ilişkin bulgular incelenmiştir. Uyum düzeyi arttıkça sınıflarda değişimler gözlenmekle birlikte iki modelin uyum düzeyleri %80 civarındadır. Eşitleme testi ve Test 1, cevaplayıcıların özelliklere sahip olup olmama konusunda birbirine yakın kararlar vermiştir. Eşitleme öncesi ve sonrası özelliklere sahip olan öğrenci sayılarındaki uyumun yüksek olması ile birlikte yaklaşık %20'lik bir grup için iki analiz sonucunda farklı kararlar verildiđi de görülmektedir. Bu farkların iki testin ölçtüđü özelliklerin aynı olmasının yanı sıra maddelerin özellikleri temsil etme düzeylerinden kaynaklandığı düşünülebilir.

BÖLÜM VI

SONUÇ VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen bulgular bağlamında ulaşılan sonuçlar belirtilmiş ve bu sonuçlara ilişkin olarak geliştirilen öneriler sunulmuştur.

6.1. Sonuç

Analiz sonuçlarına göre uygulanan yöntemin BTM için kullanılabilir olduğuna ilişkin bazı bulgulara rastlanmıştır. Bununla birlikte yöntemin geliştirilmesi ve denenmesi aşamalarında bu modellerin gerçek uygulama durumlarına ilişkin bulgular da gözlenmiştir. Modelin farklı gruplara uygulanan aynı maddelerin öğrencilerin örtük sınıflarını belirleme konusunda etkilerinin farklılaştığı görülmektedir. Diğer yandan madde parametrelerinin yapılarına ilişkin bazı dikkat çekici bulgulara da ulaşılmıştır. Test madde ilişkisi içinde madde güçlüğü ve madde ayırtediciliği değerlerinin g ve s parametrelerini yüksek düzeyde etkilemesi BTM çalışmalarında modellerin kendi uyum indeksleri dışında başka ek analizlerden yararlanılabileceğine de işaret etmektedir.

BTM çalışmalarının uygulama alanları düşünüldüğünde sadece test cevaplayıcıların hangi sınıfa atandıklarının değil aynı zamanda bu örtük sınıflara ilişkin test sonuçlarına göre ulaşılan diğer bilgilerin de modellerin uygulama alanlarına katkı sağlayacağı söylenebilir. BTM çalışmalarının istatistiksel perspektifi dışında bu modeller kullanılarak elde edilen sonuçların nasıl kullanılacağı da alan açısından önemli bir nitelik taşımaktadır. Bu nedenle bu araştırmada elde edilen bulguların modellerin istatistiksel sağlamlığı kadar bu modeller ile ulaşılan sonuçların değerlendirme süreçlerindeki kullanılabilirliğini de ilgilendirdiği söylenebilir.

6.2. Öneriler

Bu çalışmanın BTM kullanılarak kategorik değişkenlerin eşitlenmesi ile ilgili yapılacak çalışmalar için bir başlangıç niteliğinde olabileceği düşünülmektedir. Araştırmada kullanılan yöntemin ilk kez uygulanmasından ve elde edilen bulgulardan hareketle daha sonraki yapılacak çalışmalar için bazı öneriler sunulmuştur. Bu öneriler simülasyon çalışmaları için ve gerçek uygulama çalışmaları için olmak üzere iki başlık altında incelenmektedir.

Simülasyon çalışmaları için;

Bu çalışmada önerilen yöntem yeni olduğu için simülasyon çalışmaları ile zengileştirilmesi gerekmektedir. Ankor madde içeren ve madde sayısı daha fazla olan testler üzerinde uygulanabilir.

Çalışmada 15'er maddelik testler için kullanılan ankor madde sayısı 8'dir. Bu yönetime yönelik ideal ankor madde sayısının belirlenmesi amacıyla çalışmalar yapılabilir.

Bu çalışmanın analiz aşamasında telafi edici olmayan BTM modellerinden biri olan DINA model kullanılmıştır. Yöntemin diğer BTM modellerindeki etkisini görmek amacı ile telafi edici olmayan modellerle telafi edici modellerin karşılaştırılması yapılabilir.

Gerçek uygulama çalışmaları için;

Bu araştırma İzmir ilinde bulunan tabakalama yöntemi ile seçilen bazı ortaokullardaki gruplara uygulanmıştır. Benzer çalışmaların daha büyük örneklemeler üzerinde yapılması daha genel bir sonuca varılması açısından katkı sağlayabilir.

Araştırmada kullanılan testler 6.sınıf matematik öğretim programı temel alınarak hazırlanmıştır. Aynı yöntem başka alanlarda uygulanabilir.

Bu çalışmada aynı özellikleri ölçen farklı testler kullanılmıştır ve testin uygulandığı gruplar farklı ve birbirine denktir. Daha sonraki çalışmalarda aynı gruba benzer özelliklerin ölçüldüğü farklı iki test uygulanarak bu iki testin eşitleme çalışması yapılabilir.

KAYNAKÇA

- Anastasi, A. (1988). *Psychological testing*. 6th Edition, Macmillan Publishing Company, New York.
- Angoff, W. H. (1971). *Scales, norms and equivalent scores*. In R. L. Thorndike (Ed.) *Educational measurement* (s.508–600). Washington, DC: American Council on Education.
- Angoff, W. H. (1984). *Scales, norms and equivalent scores*. New Jersey: Educational Testing Service.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
- Barnard, J. J. (1996). *In search for equity in educational measurement: traditional versus modern equating methods*. Paper presented at ASEESA's National Conference, Pretoria South Africa.
- Başokçu, T. O. (2011). *Bağıl ve Mutlak Değerlendirme ile DINA Modele Göre Yapılan Sınıflamaların Geçerliğinin Karşılaştırılması*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Başokçu, T. O. (2012). *DINA Model Parametreleri Kullanılarak Tahminlenen Madde Ayırıcılık İndekslerinin İncelenmesi*. *Eğitim ve Bilim*, 37(163), 310-321..
- Basokcu, T. O. (2019). *A Recommended Model to Increase Success Level of Turkey in Mathematics in International Wide Scale Exams. Effectiveness of the Cognitive Diagnosis Based Tracking Model*. Izmir: TUBITAK 115K531.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Bergan, J. R. (1983). *Latent-Class Models in Educational Research*, s. 305–360 in E. W. Gordon, ed., *Review of Research on Education*, Washington, DC: American Educational Research Association.
- Birenbaum, M., Tatsuoka, C. ve Yamada, T. (2004). *Diagnostic assessment in TIMSS-R: Between countries and within-country comparisons of eighth graders' mathematics performance*. *Studies in Educational Evaluation*, 30, 151-173.
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores* (bölüm 17-20). Reading, MA: Addison-Wesley.

- Braun, H. I., ve Holland, P. W. (1982). Observed- score test equating: A mathematical analysis of some ETS equating procedures. In P. W.Holland ve D.B. Rubin (Ed.), Test equating (s. 9-49). New York: Academic Press.
- Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. *Journal of Educational Measurement*, 22, 13–20.
- Chen, P., Xin, T., Wang, C. ve Chang, H. H. (2012). Online Calibration Methods for the DINA Model with Independent Attributes in CD-CAT. *Psychometrika*, 77(2), 201-222.
- Cheng Y., ve Chang H. (2007). *The modified maximum global discrimination index method for cognitive diagnostic computerized adaptive testing*. Presented at the CAT and Cognitive Structure Paper Session.
- Cheng, Y. (2010). Improving Cognitive Diagnostic Computerized Adaptive Testing by Balancing Attribute Coverage: The Modified Maximum Global Discrimination Index Method. *Educational and Psychological Measurement*, 70(6), 902-913.
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40(1), 5-32.
- Cook, L. L., ve Eignor, D. R. (1983). Practical considerations regarding the use of item response theory to equate tests. In R. K Hambleton (Ed.), Applications of item response theory (s. 175-195). Vancouver, BC: Educational Research Institute of British Columbia.
- Cook, L. L., ve Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice*, 10, 37-45.
- Crocker, L. ve Algina, J. (1986). Introduction to classical & modern test theory. New York: Harcourt Brace Jovanovich College Publishers.
- de la Torre, J., ve Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, 73(4), 595-624.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115–130.
- de la Torre, J., ve Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement*, 46(4), 45.
- de la Torre, J., Hong, Y. ve Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47, 227-249.

- de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, 76 (2), 179-199.
- de la Torre, J., ve Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355-373.
- Demirus, K. B. (2015). Ortak maddelerin deęişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- DiBello, L. V., Stout, W. ve Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. Chipman, & R. Brennan (Ed.), *Cognitively diagnostic assessment* (s. 361-390).
- DiBello, L. V., ve Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 44, 285-291.
- Dogan, E., ve Tatsuoka, K. K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, 68(3), 263-272.
- Dorans, J. N., ve Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281-306.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Embretson, S. E., ve Reise, S. P. (2013). *Item response theory*: Taylor & Francis.
- Felan, G. (2002). Test equating: Mean, linear, equipercentile, and item response theory. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin.
- Fischer, G. H. (1973). The linear logistic model as an instrument in educational research. *Acta Psychologica* 37, 359-374.
- Gierl, M., Cui, Y. ve Zhou, J. (2009). Reliability and attribute-based scoring in cognitive diagnostic assessment. *Journal of Educational Measurement*, 46(3), 293-313.

- Gustafsson, J. E. (1980). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, 40(2), 377-385.
- Haertel, E. H. (1984a). An application of latent class models to assessment data. *Applied Psychological Measurement*, 8, 333-346.
- Haertel, E. H. (1984b). Detection of a skill dichotomy using standardized achievement test items. *Journal of Educational Measurement*, 21, 59-72.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 333-352.
- Haertel, E. H., ve Wiley, D. E. (1993). *Representations of ability structures: Implications for testing*. In N. Frederiksen, R. J. Mislevy, & I. Bejar (Ed.), *Test Theory for a New Generation of Tests* (s. 359-384).
- Haladyna, T. M., ve Rodriguez, M. C. (2013). *Developing and Validating Test Items*: Taylor & Francis.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (s. 147-200). New York, NY, England: Macmillan Publishing Co, Inc; : American Council on Education.
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: Principles and applications*. Baston: Kluwer-Nijhoff Publisihing.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hambleton, R. K., ve Swaminathan, H. (2013). *Item response theory: Principles and Applications*: Springer Netherlands.
- Hartz, S. (2002). A Bayesian Framework for the Unified Model for Assessing Cognitive Abilities: Blending Theory with Practicality. Unpublished Doctoral Dissertation, University of Illinois at Urbana-Champaign.
- Hartz, S., Roussos, L. ve Stout, W. (2002). Skills diagnosis: Theory and practice. Unpublished manuscript. University of Illinois at Urbana Champaign.
- Henson, R. (2004). *Test discrimination and test construction for cognitive diagnostic models*. Yayınlanmamış doktora tezi, University of Illinois.
- Henson, R., Templin, J. ve Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191- 210.

- Junker, B. W., ve Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Keleciođlu, H. (1994). Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Kim, J. S., ve Hanson, B. A. (2002). Test equating under the multiple-choice model. *Applied Psychological Measurement*, 26(3), 255-270.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29-36.
- Kolen, M. J., ve Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd. ed.). New York: Springer.
- Lee, Y. W., ve Sawaki, Y. (2009b). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, 6, 172-189.
- Lee, Y. S., Park, Y. S. ve Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the U.S. national sample using the TIMSS 2007. *International Journal of Testing*, 11, 144-177.
- Leighton, J. P., Gierl, M. J. ve Hunka, S. M. (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement*, 41(3), 205-237.
- Livingston, S. A. (2004). Equating test scores (Without IRT). Educational Testing Service.
- Lord, F. M., ve Novick, M. R. (1968). *Statistical theories of mental test scores*. New York: Addison- Wesley Publishing Company.
- Lord, M. F. (1980). *Application of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Macready, G. B., ve Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2, 99-120.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*: Taylor & Francis.

- Milli Eğitim Bakanlığı (MEB), (2013). Ortaokul matematik dersi (5, 6, 7 ve 8. Sınıflar) öğretim programı. Ankara: Devlet Kitapları Müdürlüğü Basımevi.
- Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical data. *Journal of Educational Statistics*, 11, 3-31.
- Mislevy, R. J., Almond, R. G., Yan, D. ve Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K .B. Laskey & H. Prade (Ed.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (s. 437-446).
- Molenaar, I. W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometrika*, 48, 49-72.
- Murphy, K. R., ve Davidshofer, C. O. (2005). *Psychological testing: principles and applications*: Pearson/Prentice Hall.
- Nichols, P. D., Chipman, S. F. ve Brennan, R. L. (Ed.). (1995). Cognitively diagnostic assessment. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Niss, M. (1999). Kompetencer og uddannelsesbeskrivelse, *Uddannelse* 9, 21-29.
- Niss, M., ve Jensen, T. H. (2002). Kompetencer og matematiklærning (competencies and mathematical learning). *Uddannelsesstyrelsens temaafteserie* nr. 18-2002, Undervisningsministeriet.
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Ed.), *Proceedings of the Third Mediterranean Conference on Mathematics Education* (s. 115-124). Athens, Hellenic Republic.
- OECD (2013), PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing.
- Özçelik, D. A. (1997). Test Hazırlama Kılavuzu (Genişletilmiş Üçüncü Baskı). Ankara: ÖSYM Eğitim Yayınları 8.
- Pellegrino, J. W., Baxter, G. P. ve Glaser, R. (1999). *Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice*. In A. Iran-Nejad & P. D. Pearson (Ed.), *Review of research in education* (vol. 24, s. 307-353). Washington, DC: American Educational Research Association.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4(3), 207-230.
- Reckase, M. D. (2006). Multidimensional Item Response Theory. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26, s. 607-642): Elsevier.
- Rojas, G., de la Torre, J. ve Olea, J. (2012, April). Choosing between general and specific cognitive diagnosis models when the sample size is small. Paper presented at the annual meeting of the National Council of Measurement in Education, Vancouver, British Columbia, Canada.
- Roussos, L., Templin, J. L. ve Henson, R. A. (2007). Skills Diagnosis Using IRT-Based Latent Class Models. *Journal of Educational Measurement*, 44(4), 293-311.
- Rupp, A. A. (2007). The answer is in the question: A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*. 7 (2), 95- 125.
- Rupp, A. A., ve Mislevy, R. J. (2007). Cognitive foundations of structured item response models. In J. P. Leighton & M. J. Gierl (Ed.), *Cognitive diagnostic assessment for education: Theory and applications* (s. 205-241). New York, NY, US: Cambridge University Press.
- Rupp, A. A. ve Templin, J. (2008). The Effects of Q-Matrix Misspecification on Parameter Estimates and Classification Accuracy in the DINA Model. *Educational and Psychological Measurement*, 68(1), 78-96.
- Rupp, A. A., Templin, J. ve Henson, R. J. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.
- Snow, R. E., ve Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (s. 263-331).
- Şahhüseyinoğlu, D. (2005). İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü. Ankara.
- Tanner, H., ve Jones, S. (2003). Self-efficacy in mathematics and students' use of Self- Regulated learning strategies during assessment events. In N. A. Pateman, B. J. Doherty, & J. Zilliox (Ed.), *Proceedings at the 27th Conference of the Int. Group for the Psychology of Mathematics Education* (4. baskı) (s. 275-282). Honolulu, USA: PME

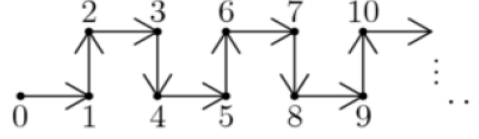
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345–354.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12, 55–73.
- Tatsuoka, K. K. (1990). Toward an Integration of Item-Response Theory and Cognitive Error Diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Ed.), *Diagnostic Monitoring Skills and Knowledge Acquisition* (s. 453-488). Hillsdale, NJ: Erlbaum.
- Tatsuoka, K. K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. New York: Routledge Academic.
- Templin, J. L. ve Henson, R.A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods* 11(3), 287-305.
- The White House [W.H.]. (2003). *An Overview of No Child Left Behind*.
- Traub, R. E., and Lam, Y. R. (1985). Latent structure and item sampling models for testing. *Annual Review of Psychology*, 36,19-48.
- Verschaffel, L., De Corte, E. ve Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving non-standard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education*, 30, 265–285.
- von Davier, A. A., Holland, P. W. ve Thayer, D. T. (2004). *Statistics for social science and public policy: The kernel method of test equating*. New York: Springer-Verlag.
- von Davier, M. (2005). A general diagnostic model applied to language testing data. ETS Research Report.No. RR-05-16. Princeton, NJ: Educational Testing Service.
- Whitley, B. E. (2002). *Principles of Research in Behavioral Science*. 2nd edn. Boston: McGraw Hill.
- Woldbeck, T. (1998). Basic concepts in modern methods of test equating. Paper presented at the annual meeting of the Southwest Psychological Association, New Orleans.
- Zhang, W. (2006). *Detecting Differential Item Functioning Using the DINA Model*. Yayınlanmamış doktora tezi. The University of North Carolina at Greensboro.

Zhu, W. (1998). Test equating: What, why and how? *Research Quarterly for Exercises and Sport*, 69(1), 11–23.

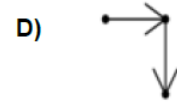
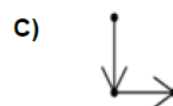
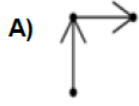


EKLER

EK1-Proje Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler



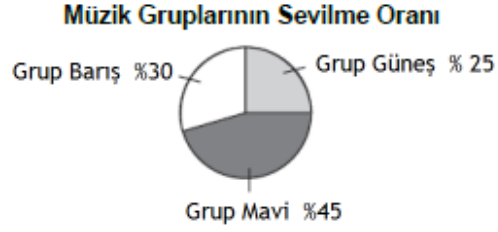
Yukarıdaki şekil örüntüsü devam ettirildiğinde 301 numaralı noktadan 303 numaralı noktaya giden bölüm aşağıdakilerden hangisidir?



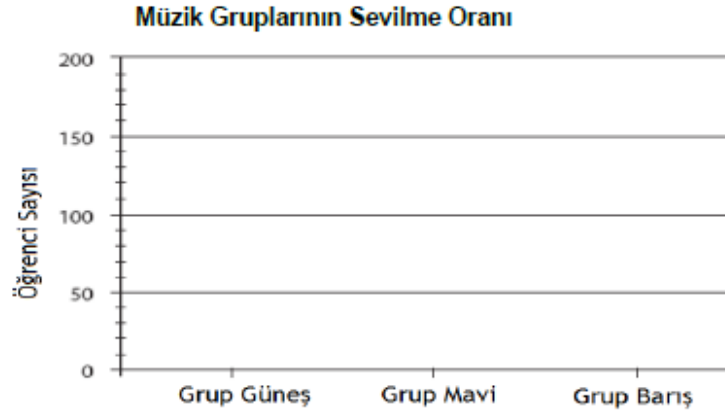
İletişim ve İlişkilendirme, Matematikleştirme ve Akıl Yürütme ve Strateji Geliştirme

EK2-TIMSS Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler

Aşağıdaki daire grafiğinde, 200 öğrenci üzerinde yapılan bir araştırmanın sonuçları görülmektedir.



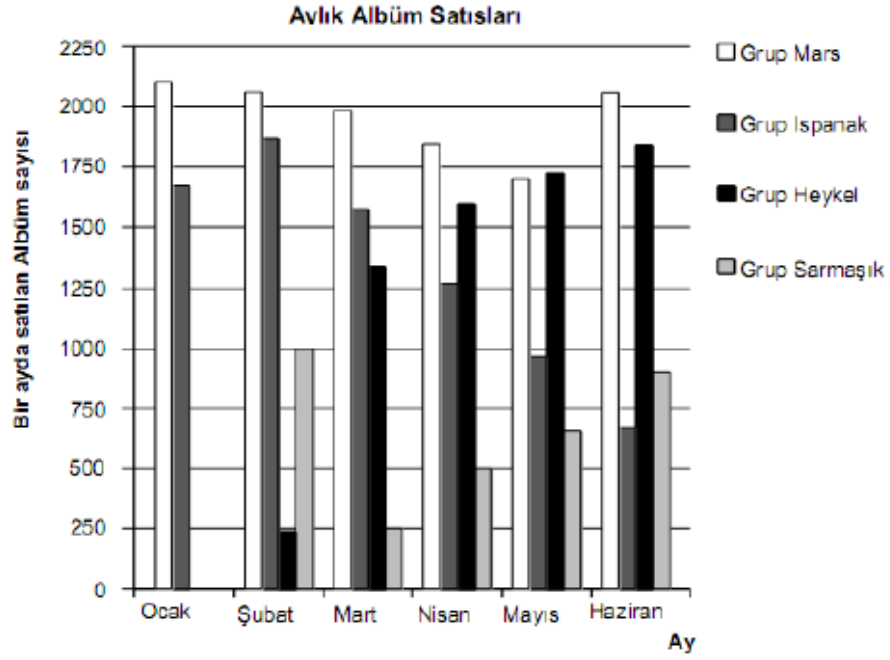
Aşağıda ayrılan yere, daire grafiğinin her bölümündeki öğrenci sayısını gösteren bir sütun grafiği çiziniz.



İletişim ve İlişkilendirme Matematikleştirme Sembolik ve Teknik Dil Kullanımı

EK3-PISA Örnek Soru ve Ölçtüğü Matematiksel Yeterlikler

Müzik gruplarından Grup Mars ve Grup İspanak'ın yeni albümleri Ocak ayında çıkacaktır. Bu albümleri Şubat ayında Grup Heykel ve Grup Sarmaşık'ın albümleri takip edecektir. Aşağıdaki grafik müzik gruplarının Ocak ayından Haziran ayına kadarki albüm satışlarını göstermektedir.



Grup İspanak'ın menajeri, grubun albüm satışları şubat ayından Haziran ayına kadar düşüş gösterdiğinden dolayı endişe etmektedir.

Bu olumsuz gidişat aynı şekilde devam ederse, grubun Temmuz ayı albüm satışı tahmini olarak ne kadar olur?

- A) 70 albüm B) 370 albüm C) 670 albüm D) 1340 albüm

Matematikleştirme ve Akıl Yürütme ve Strateji Geliştirme