



**T.C.**

**EGE ÜNİVERSİTESİ**

**Eğitim Bilimleri Enstitüsü**

**BİLİŞSEL TANI MODELLERİNDEN DINA VE FUSION  
MODELLERİNİN MADDE PARAMETRELERİ, OLASILIK  
DAĞILIMLARI VE ÖĞRENCİ ÖRTÜK SINIF KESTİRİMLERİ  
ÜZERİNDEN KARŞILAŞTIRILMASI**

**GAMZE AKBABA**

**YÜKSEK LİSANS TEZİ**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM  
DALI**

**İzmir**

**Eylül-2019**

**T.C.**

**EGE ÜNİVERSİTESİ**

**Eğitim Bilimleri Enstitüsü**

**BİLİŞSEL TANI MODELLERİNDEN DINA VE FUSION  
MODELLERİNİN MADDE PARAMETRELERİ, OLASILIK  
DAĞILIMLARI VE ÖĞRENCİ ÖRTÜK SINIF KESTİRİMLERİ  
ÜZERİNDEN KARŞILAŞTIRILMASI**

**COMPARISON OF DINA AND FUSION MODELS ON ITEM  
PARAMETERS, POSTERIOR PROBABILITIES AND LATENT  
CLASS PREDICTIONS IN COGNITIVE DIAGNOSTIC MODELS**

**GAMZE AKBABA**

**YÜKSEK LİSANS TEZİ**

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM  
DALI**

**Tez Danışmanı**

**Doç. Dr. Tahsin Oğuz BAŞOKÇU**

**Eylül-2019**

**İZMİR**



T.C.EGE ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ



YÜKSEK LİSANS  
TEZ SAVUNMA TUTANAĞI

**ÖĞRENCİNİN**

Adı Soyadı :

Numarası :

Anabilim Dalı :

Tez Başlığı (Türkçe) :

Tez Başlığı (İngilizce) :

Tez Savunma Tarihi :

Tez Başlığı Değişikliği Varsa Yeni Başlık:

**JÜRİ ÜYELERİ**

**Jüri Başkanı**

Unvan, Adı, Soyadı : .....

Karar :  Başarılı  Başarısız  Düzeltme

İmza : .....

**Jüri Üyesi**

Unvan, Adı, Soyadı : .....

Karar :  Başarılı  Başarısız  Düzeltme

İmza : .....

**Jüri Üyesi**

Unvan, Adı, Soyadı : .....

Karar :  Başarılı  Başarısız  Düzeltme

İmza : .....

**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.

## İÇİNDEKİLER

<b>TABLolar LİSTESİ</b> .....	<b>iv</b>
<b>ŞEKİLLER LİSTESİ</b> .....	<b>vi</b>
<b>DENKLEMLER LİSTESİ</b> .....	<b>vii</b>
<b>KISALTMALAR</b> .....	<b>viii</b>
<b>ÖZET</b> .....	<b>1</b>
<b>EXTENDED ABSTRACT</b> .....	<b>3</b>
<b>BÖLÜM I</b> .....	<b>6</b>
<b>GİRİŞ</b> .....	<b>6</b>
1.1. Problem Durumu .....	6
1.2. Problem Cümlesi.....	9
1.2.1. Alt Problemler.....	9
1.3. Denenceler, Sınırlılık ve Sayıtlar .....	10
1.4. Araştırmanın Önemi.....	10
<b>BÖLÜM II</b> .....	<b>12</b>
<b>LİTERATÜR</b> .....	<b>12</b>
2.1. Bilişsel Tanı Modelleri.....	12
2.1.1. Bilişsel Tanı Modellerinin Genel Özellikleri.....	12
2.1.2. Q Matris .....	14
2.1.3. BTM Hata Kaynakları.....	15
2.2. DINA Model .....	17
2.2.1. DINA Model Parametreleri.....	21
2.3. Fusion Model .....	23
2.4. Model Parametreleri İçin Kestirim Metodu .....	30
2.4.1. Beceri Parametreleri İçin Bayes Yapısı .....	31

2.4.2. Madde Parametreleri İçin Bayes Yapısı.....	32
2.4.3. Markov Chain Monte Carlo (MCMC) .....	33
2.5. Model Kontrolü Yöntemi .....	35
2.5.1. Yakınsama Kontrolü .....	35
2.5.2. Model Parametre Ölçümlerinin Yorumlanması .....	39
2.6. Puan Raporlama İstatistikleri .....	55
2.6.1. ppm ve $\eta$ .....	56
2.6.2. Puanları Beceri Profillerine İlişkilendirme: Yeterlilik Ölçekleme.....	57
2.6.3. Beceriye Sahip Olma Kestirimleri İçin Alt Puanların Kullanımı .....	58
<b>BÖLÜM III.....</b>	<b>64</b>
<b>YÖNTEM.....</b>	<b>64</b>
3.1. Araştırmanın Türü.....	64
3.2. Örneklem.....	64
3.3. Veri Toplama Aracı.....	67
3.3.1. Testin Geliştirilmesi .....	67
3.3.2. İzleme-1 Testi Gerçek Uygulaması.....	70
3.4. Verilerin Analizi.....	72
<b>BÖLÜM IV .....</b>	<b>74</b>
<b>BULGULAR .....</b>	<b>74</b>
4.1. Birinci Alt Probleme İlişkin Bulgular .....	74
4.2. İkinci Alt Probleme İlişkin Bulgular .....	76
4.3. Üçüncü Alt Probleme İlişkin Bulgular .....	77
4.4. Dördüncü Alt Probleme İlişkin Bulgular .....	78
4.5. Beşinci Alt Probleme İlişkin Bulgular .....	79
4.6. Altıncı Alt Probleme İlişkin Bulgular .....	80

4.6.1. İletişim ve İlişkilendirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir? .....	80
4.6.2. Matematikleştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir? .....	81
4.6.3. Akıl Yürütme ve Strateji Geliştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir? .....	81
4.6.4. Sembolik ve Teknik Dil Kullanımı Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir? .....	82
<b>BÖLÜM V</b> .....	<b>83</b>
<b>TARTIŞMA</b> .....	<b>83</b>
<b>BÖLÜM VI</b> .....	<b>87</b>
<b>SONUÇ VE ÖNERİLER</b> .....	<b>87</b>
<b>KAYNAKÇA</b> .....	<b>89</b>

## TABLolar LİSTESİ

Tablo 1. Örnek Q Matrisi .....	14
Tablo 2. Örnek Q Matris .....	18
Tablo 3. Araştırma Örneklemine Oluşturan İlçeler ve Öğrenci Sayıları .....	64
Tablo 4. Araştırma Örneklemi Deney ve Kontrol Grupları .....	65
Tablo 5. Matematiksel Yeterlikler ve Kodları .....	66
Tablo 6. Matematiksel Yeterlikler Matrisi .....	68
Tablo 7. İzleme-1 Testi Pilot Uygulaması Madde Parametreleri.....	69
Tablo 8. İzleme-1 Testi Betimsel İstatistikleri .....	69
Tablo 9. İzleme-1 Testi Madde Parametreleri.....	70
Tablo 10. DINA Madde Parametreleri.....	73
Tablo 11. DINA Model ile Belirlenen Örtük Sınıf Posterior Olasılıkları.....	74
Tablo 12. DINA Modelde Öğrencilerin Örtük Özelliklere Sahip Olma Yaygınlıkları...74	
Tablo 13. Fusion Madde Parametreleri .....	75
Tablo 14. Fusion Model ile Belirlenen Örtük Sınıf Posterior Olasılıkları.....	76
Tablo 15. Fusion Modelde Öğrencilerin Örtük Özelliklere Sahip Olma Yaygınlıkları..76	
Tablo 16. Fusion Model $p_i^*$ Parametresi ile DINA $g$ ve $s$ Parametrelerinin Toplam Değerinin Karşılaştırılması için Veri Seti .....	77
Tablo 17. DINA ve Fusion Model Posterior Olasılıkları Mutlak Farkları.....	78
Tablo 18. DINA ve Fusion Model ile Atanan Sınıf Üyeliklerine Göre Sınıflardaki Birey Sayıları .....	78
Tablo 19. İletişim ve İlişkilendirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi .....	79
Tablo 20. Matematikleştirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi .....	80
Tablo 21. Akıl Yürütme ve Strateji Geliştirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi.....	80
Tablo 22. Sembolik ve Teknik Dil Kullanımı Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi.....	81

Tablo 23. Fusion Modelde, DINA'ya Göre Daha Fazla Yeteneğe Sahip Bulunan Öğrenci Sayısı.....	84
Tablo 24. DINA Modelde, Fusion'a Göre Daha Fazla Yeteneğe Sahip Bulunan Öğrenci Sayısı.....	84





## ŞEKİLLER LİSTESİ

- Şekil 1. Zincir uzunluğunun madde parametreleri kestirim yakınsamasındaki etkisi (Jang, 2005; (Jang, 2005; Roussos, DiBello, Henson, Jang, & Templin, 2010).....37
- Şekil 2. Gözlenen ve model tahmini puan dağılımlarının karşılaştırılması (Jang, 2005; Roussos ve diğ., 2010). .....47
- Şekil 3. Madde özelliğine sahip olanların ve madde özelliğine sahip olmayanların arasındaki performans farkını karşılaştırarak yapılan iç geçerlik kontrolü (Jang, 2005; Roussos ve diğ., 2010). .....53



## DENKLEMLER LİSTESİ

Denklem 1. Bütün Özelliklere Sahip Bireyin Maddeyi Doğru Cevaplama Olasılığı .....	19
Denklem 2. Örtük Cevaplama ve Nitelik .....	19
Denklem 3. DINA Model Parametreleri .....	20
Denklem 4. Birleşik Model için Madde-Tepki Fonksiyonu .....	25
Denklem 5. Reparametrize Birleşik için Model Madde-Tepki İşlevi .....	27
Denklem 6. Ağırlıklı Karmaşık Toplam Puanı .....	61



## KISALTMALAR

BTM: Bilişsel Tanı Modelleri

MTK: Madde Tepki Kuramı

DINA: *Deterministic inputs, noisy “and” gate*

RUM: Reparametrize Birleşik Model

MCLCM: Çok Kategorili Örtük Sınıf Modeli

MTF: Madde tepki fonksiyonu

MCMC: *Markov Chain Monte Carlo*

LI: Kısıtlı bağımsızlık

EM: Beklenti maksimizasyonu

NIDA: *Noisy inputs, deterministic “and” gate*

EAP: Sonsal beklentisi

DIF: Ayırıcı madde fonksiyonu

MAD: Ortalama kesin fark

ppm: *Mastery* sonsal olasılığı

## ÖZET

# BİLİŞSEL TANI MODELLERİNDEN DINA VE FUSION MODELLERİNİN MADDE PARAMETRELERİ, OLASILIK DAĞILIMLARI VE ÖĞRENCİ ÖRTÜK SINIF KESTİRİMLERİ ÜZERİNDEN KARŞILAŞTIRILMASI

AKBABA, Gamze

Yüksek Lisans Tezi, Eğitimde Ölçme ve Değerlendirme

Tez Danışmanı: Doç. Dr. Tahsin Oğuz Başokçu

Ağustos, 2019

Bu teze, 115K531 numaralı TÜBİTAK projesi olan “Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek İçin Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği.” isimli projenin araştırma problemleri kapsamında çalışılmıştır.

Bilişsel Tanı Modelleri ile geliştirilmiş ve analiz edilmiş bir testte, toplam puan yerine testi alan her bireyin testte bulunan becerilerden hangilerine sahip olduğu ve hangilerine sahip olmadığı konusunda bir ölçme yapılır. Bu anlamda Bilişsel Tanı Modelleri ile geliştirilmiş testler sadece değerlendirme sürecine hizmet etmektense öğrenci ihtiyaçlarının belirlenmesinde de etkilidir.

Bu alanda geliştirilen birçok model bulunmaktadır. Bu tezde, Bilişsel Tanı Modelleri içinde daha yaygın olan ve yazılım desteği bulunan modellerden DINA ve Fusion model kullanılmıştır.

DINA model ikili yetenek modellerine benzeyen bir örtük sınıf analizidir. DINA model cevaplayıcının gözlenen yeteneğinin altında yatan örtük özelliği ortaya çıkartmayı amaçlamaktadır. Modelden  $s$  ve  $g$  parametreleri şeklinde iki adet madde parametresi elde edilir. DINA modelin diğer BTM'lerden en önemli farkı, bir madde için gerekli olan özelliklerden sadece birine bile sahip olmayan bireyi yokluk sınıfında görmesidir.

Öğrenci hakkındaki kararlarda DINA kadar katı olmayan Fusion model, cevaplayıcıların bilişsel profili hakkında detaylı bir bilgi verme ve test ile maddelerin tanılama kapasitesi hakkındaki parametreler şeklinde iki şekilde bilgi sağlamaktadır. Fusion modelde yetenek parametreleri ile birlikte  $r_{ik}^*$ ,  $\pi_i^*$  ve  $c_i$  şeklinde madde parametreleri elde edilir.

Tezde, belirlenen iki Bilişsel Tanı Modelinin istatistiksel yapıları dışında, yaptıkları kestirimler üzerinden bir çalışma yürütülmüştür. Yapılan analizler sonucunda DINA ve Fusion model parametrelerinin değişmediğine ( $r_s = -0,382$ ;  $P > 0,05$ ), her iki modelin sonsal kestirimlerinin farklılaşmadığına karar verilmiştir. Bir sonraki aşamada ise model uyumu ve madde parametrelerinin benzer olmasına rağmen öğrenci örtük sınıfları kestirimlerinin değiştiği ve öğrencilerin sadece %54,94'ünün her iki model ile aynı sınıfa atandığı görülmüştür.

**EXTENDED ABSTRACT**

**COMPARISON OF DINA AND FUSION MODELS ON ITEM PARAMETERS,  
POSTERIOR PROBABILITIES AND LATENT CLASS PREDICTIONS IN  
COGNITIVE DIAGNOSTIC MODELS**

AKBABA, Gamze

MS, Department of Measurement and Evaluation

Advisor: Assoc. Prof. Dr. Tahsin Oğuz Başokçu

*Introduction*

This thesis is in 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’s research problems.

In a test, developed and analyzed with Cognitive Diagnostic Models, instead of the total score, an assessment is made to determine which skills are present and which don’t. In this sense, tests developed with Cognitive Diagnostic Models are effective in determining the needs of the students rather than serving only the evaluation process.

There are many models developed in this field. In this thesis, DINA and Fusion models, which are more common in Cognitive Diagnostic Models and which have software support, used.

DINA model is an latent class analysis that resembles dual ability models. DINA model aims to reveal the latent attribute of the responder's that is under its observed

ability. Two item parameters,  $s$  and  $g$  parameters, are obtained from the model. The most important difference of the DINA model from other CDMs is that it sees an individual in the null class who doesn't have only one of the attributes required for an item.

The Fusion model, which is not as strict as the DINA in making decisions about the student, provides information in two ways: providing detailed information about the cognitive profile of the respondents and parameters about the test and diagnostic capacity of the items. In the Fusion model, item parameters are obtained in the form of  $r_{ik}^*$ ,  $\pi_i^*$  and  $c_i$  as well as ability parameters.

### *Method*

In this study, it was aimed to compare DINA and Fusion models, one of the Cognitive Diagnosis Models, on item parameters, probability distributions and student implicit class estimations. The actual data set was examined in which cases the practical users should choose which of the two models mentioned above.

This research involves a multifaceted comparison of the DINA and Fusion model on actual data, not statistically, but based on student decisions made using BTM. Research is one of the few studies on mathematical competences, not mathematical gains.

### *Findings*

In the thesis, apart from the statistical structures of the two Cognitive Diagnostic Models chosen, a study was carried out based on their predictions. As a result of the analyzes, it was concluded that the DINA and Fusion model parameters ( $r_s = -0,382$ ;  $P > 0,05$ ), and the posterior predictions of both models didn't differ. In the next stage,

although the model fit and item parameters were similar, it was observed that the estimation of the latent classes changed through models and only 54.94% of the students were assigned to the same class with both models.

### *Discussion and Conclusion*

As a result of the research, although the DINA and Fusion model parameters were obtained close to each other, it was seen that higher class estimations of individuals were obtained as a result of the analyzes made with the Fusion model. Therefore, it was decided that the Fusion model would make more accurate predictions than the DINA model in multi-talent measurements.

Cognitive Diagnostic Models are consistent models in large samples. The large sample used in this thesis; item statistics, DINA and Fusion model item parameters, posterior probabilities and student implicit class estimations. From this point of view, it may be possible to re-examine the problems in this study in different size samples and in cases where the number of items changes.



## BÖLÜM I

### GİRİŞ

#### 1.1. Problem Durumu

Bilişsel Tanı Modellerinin (BTM) amacı, iki kategorili niteliklerin sıralanışına dayanan örtük kategorilere göre cevaplayıcıları sınıflamaktır. Burada “nitelik” kavramı vasıf, görev, alt görev, bilişsel süreç veya beceri olarak tanımlanabilir (Tatsuoka, 1995).

BTM uygulanarak geliştirilmiş bir testte, bir toplam puan ya da toplam alt ölçek puanları yerine, sınavı alan her bir bireyin spesifik olarak kapsamda hangi becerilere sahip olduğu ve hangileri konusunda eksiklikleri bulunduğu belirlendiği bir ölçüm yapılır. Bu anlamda Bilişsel Tanı Modeliyle geliştirilmiş testler sadece değerlendirme sürecine değil aynı zamanda her bir öğrencinin eğitim ihtiyacını belirleme konusuna da hizmet eder (Cheng ve Chang, 2007).

Örtük özellik kuramı ve BTM, Madde Tepki Kuramı (MTK) modelleri ile yakından ilişkilidir. BTM içindeki birçok parametre kestirimi, MTK fonksiyonlarından yararlanarak geliştirilmiştir. Fakat BTM modelleri çok boyutlu yapısı dolayısıyla MTK'dan farklı özellikler göstermektedir. MTK modelleri bireylerin maddelere verdikleri cevaplara dayanarak bireyin örtük özelliğini kestirmek konusunda oldukça başarılı araçlardır fakat madde güçlüğünün nasıl meydana geldiği konusunda bir fikir vermez. Gerçekte ise eğitim, temel olarak MTK modellerinin bize sunduklarının ötesinde öğrenme alanının yapısını, öğrencinin psikolojik yapısını ve maddede yoklanan görevi çözebilme becerisinin altındaki psikolojiyi doğru analiz edebilmeyi hedefler (Yan, Almond ve Mislevy, 2004). BTM, bu açığı gidermeye odaklanmakta ve

öğrencinin maddeyi çözerken kullandığı bilişsel süreçleri de dikkate alarak test cevaplama örüntüsünü öğrencinin çok boyutlu profilini çıkartmak amacıyla kullanmaktadır.

Q matris, bir test için nitelik ya da beceri vektörlerinin sütunlarda, maddelerin ise satırlarda temsil edildiği bir örüntüdür. Q matris 1-0 şeklinde ikili numaralar kullanarak özelliğin maddede bulunup bulunmadığını göstermektedir. Bir  $k$  niteliğinin bir  $j$  maddesinde bulunduğu durumda 1, bulunmadığı durumda ise 0 şeklinde kodlama yapılmaktadır.

BTM alanında geliştirilen birçok model bulunmaktadır. Bu modellerin çoğu pratikte kullanım alanı olmasa da süreç içinde BTM'nin gelişim basamaklarını oluşturmaktadır. Bunun yanı sıra BTM içinde daha yaygın olarak araştırılan, pratik uygulamalarda kullanılan ve yazılım desteği olan modeller de bulunmaktadır. Bu modellerden araştırmada kullanılacak olanları DINA ve Fusion Modeldir.

Haertel (1989) tarafından geliştirilen DINA model, ikili yetenek modellerine benzeyen bir örtük sınıf analizidir. Bu anlamda DINA model, Madde Tepki Kuramının bir uzantısıdır. Bununla birlikte DINA model, MTK'dan ayrı olarak öğrencilerin farklı büyüklükte sürekli biçimde dağılmış yetenekleri olduğunu varsaymamakta, daha çok öğrencileri kesin olarak belirlenmiş az sayıda farklı örtük sınıflara ayırmaktadır.

DINA model cevaplayıcıları her özellik için iki boyutta sınıflar. Birinci sınıf "yokluk sınıfı" yani belirlenen özelliğe sahip olmayanların oluşturduğu sınıf, diğeri ise "tam sınıf" yani belirlenen özelliğe sahip olanların oluşturduğu sınıftır. Görüldüğü gibi DINA model özellik parametresini sürekli değil, kategorik bir değişken olarak tanımlar.

DINA modelde  $Q$  matrisinde tanımlanan  $k$  tane özellik için  $2^k$  tane örtük sınıf mevcuttur. Modelde her özellik için cevaplayıcının o özelliğe sahip olup olmama durumu hesaplanır. Bu şekilde cevaplayıcının özelliklerin toplamının oluşturduğu  $2^k$  örtük sınıftan hangisine dâhil olduğunu belirleyen Alfa ( $\alpha$ ) parametresi belirlenir. DINA modelin diğer Bilişsel Tanı Modellerinden en önemli farkı ise bir madde için gerekli olan özelliklerden sadece birine bile sahip olmayan cevaplayıcıyı yokluk sınıfında görmesidir. Başka bir deyişle sadece bir maddeyi doğru cevaplamak için gerekli özelliklerin tamamına sahip olan cevaplayıcının maddeyi doğru cevaplama olasılığı 1'dir.

Fusion Model Reparametrize Birleşik Model'in (RUM) Hartz (2002) tarafından geliştirilmiş ve kestirimler için hiyerarşik yapıda Bayesian yöntemi kullanan bir uzantısıdır (Başokçu, 2014). Daha önce geliştirilen olasılık temelli istatistiksel modeller klasik test teorisine göre araştırmacıların veriden daha detaylı bilgi elde etmesine olanak sağlamaktadır (Mislevy, 1995). Yeni modellerden biri olan Fusion model ise iki temel noktada bilgi vermektedir (DiBello, Stout, ve Roussos, 1995; Hartz ve Roussos, 2005; Hartz, Roussos, ve Stout, 2002). Birincisi cevaplayıcıların bilişsel profili hakkında detaylı bir bilgi vermesi ikincisi ise test ve maddelerin tanılama kapasitesi hakkındaki parametrelerdir.

Beceri ve madde arasındaki ilişkinin doğası gereği Fusion model maddenin doğru cevaplanması için bireyin sahip olması gereken özelliklerle tanımlanan özellikler arasında bir bağ kurar. Bu durumda model doğru cevaplamanın maddeyle ilişkili bütün özelliklere sahip olmakla ilişkili olduğunu varsayar. Bu bir madde için yüksek doğru cevaplama olasılığının o maddeyle ilişkili özelliklere yüksek düzeyde sahip olma

gerekliliğidir. Fusion model  $\alpha_j$  ve  $\theta_j$  olmak üzere iki yetenek parametresi içerir. Bu parametrelerden  $\alpha_j$  belirlenen özellikler ile ilgili bireyin yetenek düzeyini temsil eder,  $\theta_j$  ise bireyde var olan fakat Q matris tarafından belirlenmemiş teste ölçülen özellikle ilişkili artık yeteneği gösterir. Yetenek parametrelerine ek olarak Fusion Modelde maddelere ilişkin  $r_{ik}^*$ ,  $\pi_i^*$  ve  $c_i$  parametreleri vardır.

Bu araştırmada Bilişsel Tamı Modellerinden DINA ve Fusion modellerinin madde parametreleri, olasılık dağılımları ve öğrenci örtük sınıf kestirimleri üzerinden karşılaştırılması amaçlanmıştır. Pratik kullanıcıların hangi durumlarda, yukarıda belirtilen iki modelden hangisini tercih etmeleri gerektiği gerçek veri seti üzerinden incelenmiştir.

## 1.2. Problem Cümlesi

Fusion ve DINA model kullanılarak belirlenen öğrenci yetenek profilleri, madde parametreleri ve model veri uyumları ne düzeyde farklılaşmaktadır?

### 1.2.1. Alt Problemler

1. DINA model parametreleri ve örtük sınıf sonsal kestirimleri nasıldır?
2. Fusion model parametreleri ve örtük sınıf sonsal kestirimleri nasıldır?
3. DINA ve Fusion parametreleri değişmekte midir?
4. DINA ve Fusion sonsal kestirimleri değişmekte midir?
5. DINA ve Fusion model örtük sınıfları değişmekte midir?
6. DINA ve Fusion modelde öğrencilerin özelliklere sahip olma düzeyi değişmekte midir?
  - 6.1. İletişim ve İlişkilendirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?
  - 6.2. Matematikleştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

6.3. Akıl Yürütme ve Strateji Geliştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

6.4. Sembolik ve Teknik Dil Kullanımı Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

### **1.3. Denenceler, Sınırlılık ve Sayıtlar**

Araştırmada aynı ölçümler için farklı modellerin farklı yetenek kestirimlerinde bulunması beklenmiş ve Q matris geçerliği için uzman görüşlerinin yeterli olduğu kabul edilmiştir. Ayrıca araştırma, iki modelin istatistiksel yapıları, örneklemin sadece 6. sınıf öğrencilerinden oluşması, sadece matematik dersinin ve matematiksel yeterliklerin ele alınması ile sınırlıdır.

### **1.4. Araştırmanın Önemi**

Bilişsel Tanı Modelleri yaklaşımının psikometri alanında giderek yaygınlaşması ve modeller hakkındaki çalışmaların gittikçe artması sonucu, bir araştırmacının elde ettiği veri seti, nasıl bir problemi çözmek istediği ve ne elde etmek istediğine göre hangi modeli kullanmasının daha uygun olacağına dair soru işaretleri de artmıştır. Bu bakımdan araştırma, gerçek veri üzerinde DINA ve Fusion modelin, istatistiksel olarak değil, BTM kullanılması sonucu öğrenci hakkında verilmiş kararlar üzerinden çok yönlü karşılaştırılmasını kapsamaktadır. Araştırma, matematiksel kazanımlar değil matematiksel yeterlikler üzerinden yapılan az sayıda çalışmalardan biridir.

Kunina-Habenicht, Rupp ve Wilhelm (2012) örneklemin küçük olduğu ve testteki madde sayısının oluşacak örtük sınıf sayısından az olduğu durumlarda etkileşimler için parametre kestirimlerinin iyileştirilmesinin ciddi şekilde zayıfladığını bulmuşlardır. Rule Space Yönteminin (Tatsuoka, 1983) gerektirdiği örneklem büyüklüğü hakkında yorum yapan Rupp (2007), sınıflama doğruluğunun örneklem büyüklüğü ile arttığını

belirtmiştir. Lei ve Li (2016), uygun BTM ve Q matrisi seçiminde uyum indekslerinin performansını incelemişlerdir ve örneklem boyutunun uyum indeksleri performansı üzerinde önemli bir etkisi olduğunu bulmuşlardır. Örneklem büyüklüğünün, model uyumu ve sınıflama doğruluğuna etkisi söz konusu olduğunda geniş bir örneklem oldukça yararlı olabilmektedir (Ravand ve Robitzsch, 2018), ve çalışma bu konuda yeterli büyüklükte bir örnekleme sahiptir.



## BÖLÜM II

### LİTERATÜR

#### 2.1. Bilişsel Tanı Modelleri

Bilişsel Tanı Modelleri (BTM terimiyle ifade edilen modellere literatürde; bilişsel psikometrik modeller (cognitive psychometric models) (Rupp ve Mislevy, 2007), bilişsel tanı modelleri (cognitive diagnosis models) (Tatsuoka, 1995) sınırlandırılmış örtük sınıf modelleri (restricted latent class models) (Haertel, 1989), çoklu sınıflandırma modelleri (multiple classification models) veya yapısal madde tepki kuramı modelleri (structured item response theory models) (Rupp ve Mislevy, 2007) gibi farklı isimler verilmiştir.

Bütün bu farklı tanımlamaların her biri, modellerin belirli özelliklerine yönelik çağrışımları ifade etmektedir. Bazı tanımlamalar modellerin teorik alt yapısına göre belirlenirken bazıları modellerin kendine özgü amacını ifade etmektedir. Diğerlerinde ise modellerin istatistiksel özellikleri dikkate alınmıştır. Çalışmada kullanılan Bilişsel Tanı Modeli teriminin temelinde ise modelin ayırt edici özelliğinin yerleştirme, kabul ya da sertifikasyondan çok “tanı” olduğunun düşünülmesi yatmaktadır.

##### 2.1.1. Bilişsel Tanı Modellerinin Genel Özellikleri

BTM'nin amacı, iki kategorili niteliklerin sıralanışına dayanan örtük kategorilere göre cevaplayıcıları sınıflamaktır. Bu modellerde uzmanlığı belirleyen bir vektör olan örtük değişkenler, öğrenciye yönelik tanının altında yatan nitelik setini tanımlar. Burada “nitelik” kavramı vasıf, görev, alt görev, bilişsel süreç veya beceri olarak tanımlanabilir (Tatsuoka, 1995).

BTM uygulanarak geliştirilmiş bir testte, bir toplam puan ya da toplam alt ölçek puanları yerine, sınavı alan her bir bireyin spesifik olarak kapsamda hangi becerilere sahip olduğu ve hangileri konusunda eksiklikleri bulunduğu belirlendiği bir ölçüm yapılır. Bu anlamda BTM ile geliştirilmiş testler sadece değerlendirme sürecine değil aynı zamanda her bir öğrencinin eğitim ihtiyacını belirleme konusuna da hizmet eder (Cheng ve Chang, 2007).

Örtük özellik kuramı ve BTM, Madde Tepki Kuramı (MTK) modelleri ile yakından ilişkilidir. BTM içindeki birçok parametre kestirimi, MTK fonksiyonlarından yararlanarak geliştirilmiştir. Fakat BTM modelleri çok boyutlu yapısı dolayısıyla MTK'dan farklı özellikler göstermektedir. MTK modelleri bireylerin maddelere verdikleri cevaplara dayanarak bireyin örtük özelliğini kestirmek konusunda oldukça başarılı araçlardır. MTK modelleri madde ve yetenek parametrelerini sürekli bir ölçek üzerinde kestirirken “*a*”, ayıt edicilik, “*b*”, güçlük ve “*c*” tahmin parametrelerini hesaplar. Fakat MTK modelleri madde güçlüğüne nasıl meydana geldiği konusunda bir fikir vermez. Gerçekte ise eğitim, temel olarak MTK modellerinin bize sunduklarının ötesinde öğrenme alanının yapısını, öğrencinin psikolojik yapısını ve maddede yoklanan görevi çözebilme becerisinin altındaki psikolojiyi doğru analiz edebilmeyi hedefler (Yan, Almond ve Mislevy, 2004). BTM, bu açığı gidermeye odaklanmakta ve öğrencinin maddeyi çözerken kullandığı bilişsel süreçleri de dikkate alarak test cevaplama örüntüsünü öğrencinin çok boyutlu profilini çıkartmak amacıyla kullanmaktadır.

MTK' da öğrencinin yeteneği genel yetenek parametresi olan  $\theta$  ile tahminlenir.  $\theta$  ve gözlenen puanlar belirli bir öğrencinin performansını ve grup içindeki sırasını



kestirme imkânı tanır. BTM ise bunların dışında öğrencinin zihinsel süreci ve performansın altında yatan bilginin yapısını da ortaya çıkartabilir. BTM uygulamalarında her test birden fazla özelliği ölçer ve her madde birden fazla özellikle ilişkilendirilebilir. Bu şekilde öğretmen sınav sonunda tek bir puan vermek yerine BTM ile öğrenci performansını, sahip olduğu becerileri de göstererek düzenleyebilir. Bu durumda öğretmen öğrencinin bir beceri profilini çıkartıp eksik olan becerilerini belirleyerek bu becerilerin üzerine eğilebilir. Diğer taraftan, eğitim sürecinin güçlü ve zayıf yönlerini ortaya çıkartarak daha iyi öğrenme stratejileri düzenlenebilir.

Görüldüğü gibi BTM öğrencilerin örtük ölçekteki yetenek düzeylerine odaklanmaz, her bir bilişsel öge üzerindeki öğrenci performansını ve bu ögeye sahip olma olasılığını belirler. Elde edilen olasılıklar öğrencinin uzmanlaştığı becerilerin profilini çıkartacak şekle dönüştürülür. BTM’de de tıpkı yapısal eşitlik modellerinde olduğu gibi örtük değişken ile maddeler arasındaki ilişkilerin analizden önce belirlenmesi gerekmektedir. Modellerde bu süreç Q matris olarak adlandırılan madde özellik matrisiyle tanımlanır.

### **2.1.2. Q Matris**

Q matris, bir test için nitelik ya da beceri vektörlerinin sütunlarda, maddelerin ise satırlarda temsil edildiği bir örüntüdür. Nitelikler, alan uzmanları tarafından belirlenen özellikler, prosedürler, buluş yöntemleri, stratejiler, beceriler ve başka bilgi bileşenleri olabilir. Q matris 1-0 şeklinde ikili numaralar kullanarak özelliğin maddede bulunup bulunmadığını göstermektedir. Bu numaralandırma ilk önce Fischer (1973) tarafından “ağırlıklandırılma” şeklinde tanımlanmış ve bir  $k$  niteliğinin bir  $j$  maddesinde bulunduğu durumda 1, bulunmadığı durumda ise 0 şeklinde kodlandığını belirtmiştir.

$J \times K$  matrisini  $Q = [q_{jk}]$  0 ve 1 için şu şekilde yazabiliriz;

$$q_{jk} \begin{cases} 1 \text{ eğer } k \text{ özelliği } , j \text{ maddesinde bulunuyorsa} \\ 0 \text{ eğer } k \text{ özelliği } , j \text{ maddesinde bulunmuyorsa} \end{cases}$$

Tablo 1’de 3 nitelik ve 5 madde için hazırlanmış örnek bir Q matrisi görülmektedir. Q matrisinde görüldüğü gibi 1. maddeyi doğru cevaplayabilmek için ilk niteliğe sahip olmak gerekmektedir. İkinci maddeyi doğru cevaplayabilmek için ise 1. ve 2. niteliklere öğrencilerin sahip olması gerekmektedir.

Tablo 1. Örnek Q Matrisi

Maddeler	$\alpha_1$	$\alpha_2$	$\alpha_3$
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1
5	0	0	1

Ölçülen niteliklerle birlikte düzenlenen Q matrisi, madde güçlüğü’nün psikometrik model ve bilişsel süreçle arasındaki bağlantısını kurar. Modelin seçimi ve Q matrisi birbirine eşit sınırlardan oluşan, örtük sınıf cevaplama olasılıklarını belirlemeyi sağlar.

Son dönem yapılan araştırmalarda Q matrisinin doğruluk miktarını ve etkisini ölçmeye yönelik sağlam kanıtlar sunan metotlar geliştirilmektedir. Henson ve Douglas (2005), Q matrisin kalitesinin, öğrencilerin beceri profilinin kestirimini doğrudan belirlediğini ifade etmektedir. Bu nedenle, Q matrisin yapısı ve model veri uyumu “tanısal analiz” için çok önemli bir nitelik taşımaktadır (de la Torre ve Douglas, 2004).

### 2.1.3. BTM Hata Kaynakları

BTM’de testi alan bireylerin yanlış sınıflandırılmasına neden olan farklı hata kaynakları bulunmaktadır. Maddenin yapılandırılması, Q matris ile ilişkilendirilmesi

veya cevaplanması aşamasında oluşabilecek bu hata kaynakları modellerin yapısı incelenirken ve modellere ilişkin parametreler belirlenirken öncelikli önem taşımaktadır. Roussos ve diğ. (2007), sınıflamalarda oluşabilecek bu hatalarla ilgili dört önemli faktöre işaret etmektedir.

**Strateji:** Cevaplayıcının maddeye cevap verirken Q matris tarafından ifade edilen özelliklerin dışında farklı bir yöntem kullanması BTM içinde belirlenen cevaplayıcı profilinin hatalı olmasına neden olmaktadır.

**Tamlık (*Completeness*):** Bir test maddesinin Q matris tarafından belirlenen özelliklerden farklı başka bir özelliğe daha sahip olması durumudur. Bu durumda Q matrisin tamlık özelliği göstermediği söylenebilir.

**Pozitiflik düzeyi (*Positivity*):** Bazı durumlarda özelliğe sahip olan bir birey o özellik altında tanımlanan bir maddeyi yanlış cevaplayabilir, diğer taraftan özelliğe sahip olmayan bir birey o özellik altında tanımlanan bir maddeyi doğru cevaplayabilir. Bu sonucu doğuran temel neden Q matris tarafından belirlenen özelliklerin temelde bir genellemeye dayanmasıdır. Özellikler girift yapılar değildir ve alt gruplara ayrılma potansiyeli taşımaktadırlar. Bu durumda sınıflama sonucunda belli bir özelliğe sahip olduğuna karar verilen bir bireyin, o özelliğin alt becerilerinden bir ya da birkaçına sahip olmaması mümkündür. Genel olarak bireyin özelliğe sahip olma düzeyini ikili (dikotom) bir değişken olarak tanımlayan modellerde pozitiflik düzeyi bir hata kaynağı olarak gözlenmektedir.

**Kaydırma (*Slip*):** Birey tarafından gerçekleşen rastsal hatalar kaydırma olarak tanımlanır. Kaydırma arta kalan ve sistematik olmayan bütün hataları içeren kategoridir.

## 2.2. DINA Model

BTM alanında geliştirilen birçok model bulunmaktadır. Bu modellerin çoğu pratikte kullanım alanı olmasa da süreç içinde BTM'nin gelişim basamaklarını oluşturmaktadır. Sympson (1978) tarafından geliştirilen Denkleştirici olmayan Çok Boyutlu MTK Modeli (Noncompensatory MIRT model), Whitely (1980) ve Embretson (1999) tarafından geliştirilen Çok Bileşenli Örtük Özellik Modeli (Multicomponent latent trait model), Embretson ve Reise (2000, 2013) tarafından geliştirilen Genel Bileşenli Örtük Özellik Modeli (General component latent trait model) bu modellerden birkaçıdır.

BTM içinde çok boyutlu örtük sınıf modelleri bir dönem araştırmacılar tarafından yoğun olarak çalışılmıştır. Özellikle Maris (1999), farklı özellikleri olan birkaç çok kategorili örtük sınıf modeli önermiştir. Ayırıcı Çok Kategorili Örtük Sınıf Modeli (Disjunctive MCLCM) ve Denkleştirici Çok Kategorili Örtük Sınıf Modeli (Compensatory MCLCM) bu modeller arasındadır. Bunun yanında von Davier ve Yamamoto (2004); von Davier (2005); Xu ve von Davier (2006) tarafından da yeteneğin sürekli kabul edildiği denkleştirici MTK modellerine benzer ama yetenek parametresinin kategorik olarak alındığı denkleştirici çok kategorili örtük sınıf modelleri geliştirmiştir.

BTM içinde daha yaygın olarak araştırılan, pratik uygulamalarda kullanılan ve yazılım desteği olan modeller de bulunmaktadır. Bu modellerden en önemlileri; Tatsuoka (1983) Rule Space Modeli, Hartz (2002) tarafından geliştirilen Reparametrize Birleşik Model, Gitomer ve D. Yamamoto (1991) HYBRID modeli ve Haertel (1984); Walberg ve Haertel (1990) Sınırlandırılmış Örtük Sınıf Modelidir.

Haertel (1989) tarafından geliştirilen DINA model, ikili yetenek modellerine benzeyen bir örtük sınıf analizidir. Bu anlamda DINA model, Madde Tepki Kuramının bir uzantısıdır. Bununla birlikte DINA model, MTK'dan ayrı olarak öğrencilerin farklı büyüklükte sürekli biçimde dağılmış yetenekleri olduğunu varsaymamakta, daha çok öğrencileri kesin olarak belirlenmiş az sayıda farklı örtük sınıflara ayırmaktadır.

DINA model kısaca şu şekilde tanımlanabilir:  $X_{ij}$ 'nin  $i$  cevaplayıcısının  $j$  maddesine verdiği yanıt olduğu farz edilirse,  $i= 1, \dots, I$  ve  $j= 1, \dots, J$  olur. Cevaplayıcının ikili özellik vektörü  $\alpha_i = \{\alpha_{ik}\}$  şeklinde gösterilirse,  $k= 1, \dots, K$  için cevaplayıcının  $k$ . elemanının 1 olduğu durum  $k$  özelliğine sahip olduğunu, 0 olduğu durum ise  $k$  özelliğine sahip olmadığını göstermektedir (de la Torre, 2009). Özellik olarak adlandırılan terim genel olarak nitelik, beceri, yetenek, bilginin sunumu, bilişsel süreç olarak tanımlanabilir (Tatsuoka, 1995). Çoğu BTM cevaplama özelliğinin belirlendiği  $J \times K$  şeklinde oluşturulan ve 1-0 şeklinde kodlanan bir Q matrisi üzerinden hesaplama yapar (Embretson, 1984; Tatsuoka, 1985). Q matrisinde sütunlar özellikleri ve satırlar maddeleri gösterir. Matriste  $q_{jk}$  hücresi  $j$  maddesini doğru cevaplamak için  $k$  özelliğine sahip olunması gerekip gerekmediğini işaret eder.

DINA model cevaplayıcıları her özellik için iki boyutta sınıflar. Birinci sınıf "yokluk sınıfı" yani belirlenen özelliğe sahip olmayanların oluşturduğu sınıf, diğeri ise "tam sınıf" yani belirlenen özelliğe sahip olanların oluşturduğu sınıftır. Görüldüğü gibi DINA model özellik parametresini sürekli değil, kategorik bir değişken olarak tanımlar. DINA modelde Q matrisinde tanımlanan  $k$  tane özellik için  $2^k$  tane örtük sınıf mevcuttur. Modelde her özellik için cevaplayıcının o özelliğe sahip olup olmama durumu hesaplanır. Bu şekilde cevaplayıcının özelliklerin toplamının oluşturduğu  $2^k$

örtük sınıftan hangisine dâhil olduğunu belirleyen Alfa ( $\alpha$ ) parametresi belirlenir. Örneğin  $\alpha_1$ ,  $\alpha_2$  ve  $\alpha_3$  olarak belirlen 3 özellekle temsil edilen 4 maddelik bir teste ilişkin Q matrisi aşağıdaki şekilde belirlenmiş olsun:

Tablo 2. Örnek Q Matris

Maddeler	$\alpha_1$	$\alpha_2$	$\alpha_3$
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1

Yukarıdaki Q matris, birinci maddenin doğru cevaplanması için sadece  $\alpha_1$ 'e sahip olmanın yeterli olduğunu, ikinci maddenin doğru cevaplama için ise hem  $\alpha_1$  hem de  $\alpha_2$ 'ye sahip olunması gerektiğini göstermektedir. Örnek Q matrise göre yapılan analiz sonucunda cevaplayıcılar  $2^3$  örtük sınıfa yerleştirilirler. Bu örtük sınıflar (000), (100), (010), (001), (110), (101), (011) ve (111) şeklinde belirlenir. Örtük sınıflar öğrencinin hangi özelliklere sahip olup hangilerine sahip olmadıklarını tam olarak ifade eder. Yukarıdaki örnekte belirlenen örtük sınıflardan (000)'a dâhil olan öğrenci hiçbir özelliğe sahip değildir. Benzer şekilde (100) örtük sınıfında sadece  $\alpha_1$ 'e sahip olan öğrenciler yer alırken, (011) sınıfında ise  $\alpha_1$ 'e sahip olmayıp  $\alpha_2$  ve  $\alpha_3$ 'e sahip olan öğrenciler bulunmaktadır. Bu durum BTM'nin tanı işlevini açıklamaktadır. BTM ile geliştirilmiş ve analiz edilmiş bir testte, bir toplam puan ya da toplam alt ölçek puanları yerine, sınavı alan her bir bireyin sınavın kapsamındaki becerilerden hangilerine sahip olduğu ve hangileri konusunda eksiklikleri bulunduğu belirlendiği bir ölçme yapılır. Bu anlamda Bilişsel Tanı Modeliyle geliştirilmiş testler sadece değerlendirme sürecine

değil aynı zamanda her bir öğrencinin eğitim ihtiyacını belirleme konusuna da hizmet eder (Cheng ve Chang, 2007).

DINA modelin diğer BTM'den en önemli farkı ise bir madde için gerekli olan özelliklerden sadece birine bile sahip olmayan cevaplayıcıyı yokluk sınıfında görmesidir. Başka bir deyişle sadece bir maddeyi doğru cevaplamak için gerekli özelliklerin tamamına sahip olan cevaplayıcının maddeyi doğru cevaplama olasılığı 1'dir. Aşağıda bütün özelliklere sahip olan bireyin maddeyi doğru cevaplama olasılığının fonksiyonu verilmiştir:

Denklem 1. Bütün Özelliklere Sahip Bireyin Maddeyi Doğru Cevaplama Olasılığı

$$P \left[ Y_{ij} = 1 \mid \eta_{ij}, s_j, g_j \right] = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}$$

P, aranan bütün becerilere sahip olan öğrencinin maddeyi doğru cevaplama olasılığıdır.  $\eta_{ij}$ ,  $\alpha$  tarafından belirlenen örtük cevaplama ve  $i$ . konunun niteliği ve  $q_j$ 'nin vektörüdür. Q matrisinin  $j$ . maddesine tekabül eden sıra şu şekilde gösterilebilir:

Denklem 2. Örtük Cevaplama ve Nitelik

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

DINA modelin diğer BTM'lere göre en önemli avantajı, gerek uygulama gerekse yorumlama süreçlerinin daha az karmaşık olmasıdır. Bununla birlikte DINA model basitliğine rağmen yüksek düzeyde model-data uyumu verdiği ve bazı modifikasyonlarla modelin farklı stratejilere kolaylıkla uyum sağladığı, de la Torre ve Douglas (2004, 2008) tarafından göstermiştir. Aynı zamanda de la Torre çalışmalarında daha yeni ve daha karmaşık olan BTM'lerle, DINA model sonuçları arasında yüksek

düzeyde uyum olduğunu belirtmiştir (de la Torre, 2008, 2009; de la Torre & Liu, 2008). Bu gerekçeler genel olarak BTM arasında DINA modelin öneminin bir göstergesi olduğu gibi bu çalışmada DINA modelin seçilmesinin nedenini oluşturmaktadır.

### 2.2.1. DINA Model Parametreleri

DINA model cevaplayıcıların sahip olduğu özellikleri belirlerken aynı zamanda tahmin (guess)  $g$  ve kaydırma (slip)  $s$  parametreleri olarak adlandırılan iki parametre daha hesaplamaktadır.

Denklem 3. DINA Model Parametreleri

$$\begin{aligned} s_j &= P[Y_{ij} = 0 | \eta_{ij} = 1] & s_j &= P[Y_{ij} = 0 | \eta_{ij} = 1] \\ g_j &= P[Y_{ij} = 1 | \eta_{ij} = 0] & g_j &= P[Y_{ij} = 1 | \eta_{ij} = 0] \end{aligned}$$

$s_j$  örtük özelliğe sahip bireyin  $j$  maddesine yanlış cevap verme olasılığını gösteren durumu (yanlış pozitif olasılık) ve  $g_j$  ise örtük sınıfa sahip olmayan bireyin doğru cevap verme olasılığı durumunu (doğru pozitif olasılık) ifade eder.  $s_j$  parametresi “slip=kaydırma” anlamına gelir ve bu parametre ne kadar düşük olursa, aranan özelliklere sahip bireylerin doğru cevap verme olasılığı o kadar artar.

$g_j$  parametresine “guessing” tahmin parametresi de denir. Maris (1999) alternatif olarak  $g_j$  parametresini, zihinsel tahmin yeteneğini başarıyla kullanmak olarak da açıklar. Görüldüğü gibi DINA model  $g$  parametresi MTK 3 parametrelili modelde görülen  $c$  “şans” parametresinden biraz farklıdır. DINA modelde hesaplanan  $g$  sadece bir maddeyi tahminle doğru cevaplamak değil, aynı zamanda  $Q$  matris tarafından tanımlanamamış bir özelliği kullanarak maddeyi doğru cevaplamak anlamına da gelmektedir.



DINA modelde  $s$  ve  $g$  parametreleri madde düzeyinde ortaya çıkmaktadır. Her madde, popülasyonu iki sınıfa böler ve aynı sınıfa düşen öğrencilerin o maddeye doğru cevap verme olasılıkları eşittir. Öğrencilerin sınav performansı, testte yoklanan niteliğin vektörünün tam bir göstergesi değildir. Bu nedenle olasılık temelinde bir model ancak  $s$  ve  $g$ 'nin olasılığını görmeye izin verir. “ $s$ ” durumu, öğrencinin maddede aranan niteliğe sahip olmasına rağmen alt görevi doğru cevaplayamadığı ya da maddeyi doğru cevaplayamadığında gerçekleşir. “ $g$ ” ise öğrencinin maddede aranan özelliklerden birine ya da birkaçına sahip olmamasına rağmen alt görevleri tamamlaması veya maddeyi doğru cevaplama durumudur (de la Torre ve Douglas, 2008).

DINA model öğrencilerin maddeyi doğru cevaplama için gerekli becerilere sahip olup olmama durumlarını belirlemek noktasında oldukça başarılıdır. Bununla birlikte belirlediği  $s$  ve  $g$  parametreleri yoluyla madde güçlüğü ve ayırt ediciliği şeklinde yorumlanabilecek bazı indekslerin hesaplanmasına da olanak sağlamaktadır. de la Torre (2008), maddenin kalitesini belirleyen ayırt edicilik indeksini belirlemek amacıyla hem  $s$  hem de  $g$  parametrelerinin birlikte kullanıldığı bir hesaplama yöntemi önermiştir. de la Torre'ye göre  $\delta_j$  (delta), ayırt edicilik indeksi olarak  $1 - s_j - g_j$  formülüyle hesaplanabilmektedir. Örtük özelliğe göre öğrencileri mükemmel düzeyde ayırt eden bir madde için  $\delta_j = 1$  olmalıdır ve bu değer “sıfır” a yaklaştıkça maddenin ayırt edicilik gücü düşmektedir. Modelde  $s$  ve  $g$  parametreleri 0 ile 1 arasında değer alabilmektedir. Bu nedenle madde ayırcılığını ifade eden  $\delta_j$  katsayısı  $[-1, 1]$  arasında değerler almaktadır.

DINA modelde hesaplanan  $\delta_j$  parametresi konusunda de la Torre (2008) tam bir ölçütün olmadığını belirtmiştir. Daha yüksek  $\delta_j$  değeri daha ayırcı bir maddeyi işaret

etmektedir. Bununla birlikte modelin yeni oluşu ve bu konuda yapılan çalışmaların sınırlı sayıda olması, henüz böyle bir ölçütün belirlenememesi durumunu doğurmuştur.

### 2.3. Fusion Model

Madde Tepki Kuramı (MTK), bir madde tepki fonksiyonu (MTF) modeli tabanlı değerlendirme analizi ve puanlamasına yüksek oranda başarılı ve geniş çapta uygulanan bir olasılık modellemesi yaklaşımı sağlamaktadır. Tüm MTK modelleri gibi MTK bazlı bilişsel tanılama modelleri de aday beceri parametresi ve madde parametresi açısından aday  $j$  tarafından madde  $i$ 'ye belirli bir tepki gözleminin olasılığını belirler. Sembolik olarak, aday  $j$ 'nin madde  $i$ 'ye olan yanıtının  $X_{ij} = x$  olduğu yerde bu olasılık  $P(X_{ij} = x | \theta_j, \beta_i)$  olarak yansıtılır;  $\theta_j$  aday  $j$  beceri parametrelerinin bir vektörü ve  $\beta_i$  ise madde  $i$  parametrelerinin bir vektörüdür (Roussos ve diğ., 2007).

MTK modellemesinin esas varsayımı adayın beceri parametresi koşuluna bağlıdır, adayın herhangi bir madde  $i$ 'ye olan yanıtı, adayın herhangi bir başka madde  $i$ 'ye olan yanıtından bağımsızdır. Bilişsel tanı modelini diğer MTK modellerinden ayırıcı özellik ise  $i = 1, \dots, I$  maddeler önceden belirlenmiş, test geliştirici seçimli bir dizi bilişsel becerileri, özellikle beceri tanılamayla alakalı  $k = 1, \dots, K$  ile ilişkilidir. Bu ilişki  $q_{ik}=1$ 'de  $i$  maddesinin  $k$  niteliğini gerektirdiğini ve  $q_{ik} = 0$ 'da madde  $i$  tarafından  $k$  niteliğinin gerek duyulmadığını belirten Q matrisi olarak bilinir.

Fusion Modelin bir türü olan Birleşik Model, beceri tabanlı madde parametrelerini ve yine beceri tabanlı aday parametrelerini de içermektedir. Bunun yanı sıra Birleşik Model, modelin veriye uyumunu geliştirmek için ek parametrelerde içermektedir. Samejima'nın 1995'te yeterlilik alanı (Competency Space) teorisinde tartışıldığı gibi,

aday parametresi  $\theta = \{\alpha_Q, \alpha_b\}$ 'nin tüm geçerli becerilerin bütün örtük aralıklarını işaret ettiğinde.  $\alpha_Q$  da, Q matrisinde ifade edildiği üzere önceden belirlenmiş bilişsel becerilerin vektörü olduğu kabul edildiğinde geri kalan örtük aralık  $\alpha_b$ , Q matrisi tarafından belirlenenler için geçerli beceri tamamlayıcısını içermektedir. Samejima,  $\alpha_b$ 'yi “üst seviye işlem” ile ilgili beceriler olarak tanımlakta ve bu becerilerin  $\alpha_Q$ 'dan daha da fazla önemli olabileceğini öne sürülmektedir. Birleşik Model açısından, ise,  $\alpha_b$  üst seviye işlem olarak yorumlanmak durumunda değildir, test geliştirici tanımlı Q matrisinin dışındaki aday görev performansını etkileyen örtük becerilerin parametrik olarak basit bir temsilidir (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

$\alpha_j$ 'nin  $\alpha_Q$  ile birebir örtüşen beceri uzmanlık parametrelerinin bir vektörü olduğu ve  $\eta_j$ 'nin  $\alpha_b$ 'nin öğelerinin ( $\alpha_b$ 'nin tek boyutlu ölçeklendirmede a projeksiyonu) tek boyutlu bir bileşimi olarak kavramsallaştırıldığı durumda Birleşik Model, aday  $j$  ( $\alpha_j$ ,  $\eta_j$  olarak) için beceriyi parametrelerle ifade etmektedir ve  $\eta_j$  terimi tamamlayıcı beceri olarak adlandırılmaktadır. Birleşik Modelde bu tamamlayıcı beceri olan  $\eta_j$ 'nin ilave edilmesi madde parametresinin, matrise atanan Q matrisi becerileri tarafından, bir test maddesinin iyi planlanmış olup olmadığını tanılamak adına kullanılabilen bir türüyle bağlantılıdır. Q matrisinin sınavdaki her madde için tüm beceri gerekliliklerinin bütün bir temsili olduğuna dair ortak kabul Birleşik Modeli diğer beceri tanılama modellerinden ayırmaktadır (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

Kolaylık olması açısından  $\alpha_j$  ikilisinin öğeleri ikiye bölünerek incelenmektedir, şöyle ki eğer aday  $j$ , beceri  $k$ 'ya hâkim olursa  $\alpha_{kj} = 1$  ve eğer aday, beceri  $k$ 'ya hâkim

olmazsa  $\alpha_{kj} = 0$ 'dır. Genellikle, Fusion Model beceri ediniminin iki kategorisinden daha fazlasını da ele alabilmektedir. Öte yandan,  $\eta_j$  standart bir sürekli değişken olarak modellenmiştir.  $a_j$  parametreleri için beceri dağılımı birinci ve ikinci derece önemle, yani (a) her bir beceri  $k$ 'yı öğrenen kitlenin oranı – bu oran  $p_k$  ile ifade edilmektedir ki  $k = 1, \dots, K$ , ve (b)  $\alpha_j$ 'nin  $k$  becerisi bileşenleri arasındaki ilişki, başarılı olarak modellendiği düşünülmektedir. Yorumlanabilirliği kolaylaştırmak için ilgileşim yerine dört-düzeyle ilgileşim ikiye ayrılmış yeterlik değışkenliđi arasındaki ilişkileri modellemek için kullanılacaktır (Hartz ve Roussos, 2005; Hartz ve diđ., 2002; Roussos ve diđ., 2007).

$\pi_{ik} = P(Y_{ikj} = 1 \mid \alpha_{kj} = 1)$  formülünü ve  $Y_{ikj} = 1$  formülünün aday  $j$ 'nin beceri  $k$ 'yı madde  $i$ 'ye doğru uyguladıđı gözlemlenemez durumlara işaret ettiđinde  $r_{ik} = P(Y_{ikj} = 1 \mid \alpha_{kj} = 0)$  formülü ařađıdaki řekilde tanımlanmaktadır. Örtük değışken  $Y_{ikj}$ 'nin kullanımı Maris tarafından 1999'da tartıřıldıđı üzere örtük yanıt modellerinde de kullanılır. Birleşik Model için MTF (Madde Tepki Fonksiyonu) denklem 4'te verilmektedir:

Denklem 4. Birleşik Model için Madde-Tepki Fonksiyonu

$$P(X_i = 1 \mid \underline{\alpha}_j, \eta_j) = d_i \prod_{k=1}^K \pi_{ik}^{\alpha_{jk} \cdot q_{ik}} r_{ik}^{(1-\alpha_{jk}) \cdot q_{ik}} P_{c_i}(\eta_j) + (1 - d_i) P_{b_i}(\eta_j),$$

$Ph(\eta_j) = \{1 + \exp[-1.7(\eta_j + h)]\}^{-1/2}$ ; a Rash model zorluk parametresi ile  $h$ 'nin birincide  $c_i$ 'ye ya da ikinci de  $b_i$ 'ye karşılık geldiđi negatif  $h$ 'ye eşdeđerdir.

Modeldeki çıktı terimi modelin iki önemli yönüne işaret etmektedir. İlki, kullanılan  $Q$  temelli stratejiyi sađlayan becerilerin uygulanmasındaki koşullu bađımsızlıđın istatistiksel tahminidir. İkincisi ise, verilen bir maddenin iđerisindeki çeřitli becerilerin

etkileşiminin bağlayıcı olduğu, bilişsel varsayımdır. Yani maddeyi doğru bir şekilde cevaplamanın yüksek olasılığı tüm gerekli becerilerin başarılı bir şekilde yürütümünü gerektirir. Denklem 4'te verilen herhangi bir yanıt örneği  $x$ 'in olasılığını modellemek için kullanılabilir. Eğer  $k_i$ , madde  $i$  için gerekli becerilerin sayısı ise, o zaman testteki her bir madde  $i$  için  $2k_i + 3$  madde parametreleri vardır:  $\pi_{ik}$  ve  $r_{ik}$ , iki MTK Rasch Model parametreleri  $c_i$  ve  $b_i$  ve son parametre  $d_i$ , diğer tüm stratejilerin yerine  $Q$  tabanlı strateji seçiminin olasılığıdır (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

Birleşik Model MTF yalnızca adayların  $Q$ 'dan etkilenmiş yanıtlarını biçimlendirmez aynı zamanda da  $Q$ 'suz becerilerin  $P_{ci}(\eta_j)$  terimi ile adayların yanıtlarını etkilemeye de imkân verir ve  $P_{bi}(\eta_j)$  terimi ile alternatif  $Q$ 'suz stratejiler için de olanak sağlar. Maris'in Modelleri (1999) ve Genel Bileşenli Örtük Özellik Modeli (Embretson, 1984) ile birlikte Birleşik Model kavramsal bilişsel olarak yorumlanabilir parametrelere sahiptir ancak ne yazık ki her parametre tanımlanabilir değildir ya da böylelikle istatistiksel olarak tahmin edilebilir değildir (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

Birleşik Model parametrelerinin esnekliği ya da yorumlanabilirliği bu bölümde tanımlanan Fusion Model beceri tanılama sistemi için altyapı olarak seçimine izin verir. Ancak orijinal Birleşik Model'de tanımlanamaz parametrelerin olması sebebiyle, modeli kestirilebilir hale getirmek için parametre alanında bir azalmaya gidilmesi gerekir. Bu görevi yerine getirmek için Hartz (2002), Birleşik Modeli parametrelerin yorumlanabilirliğini koruyan bir şekilde yeniden parametrize etmiştir. Parametre alanının karışıklığını daha fazla azaltmak ve parametreleri tahmin edebilme becerisinin

arttırmak için, alternatif stratejilerin olasılığının modellenmesi  $d_i = 1, i = 1, \dots, I$  ile son bulmuştur. Azaltılmış model, Reparametrize Birleşik Model (RUM) olarak adlandırılmış ve bu model, orijinal modeldeki madde başına  $2ki + 3$  parametrelerine kıyasla madde başına  $2 + ki$  parametrelerine sahiptir. Azaltılmış model, maddeden maddeye çeşitlilik sağlamak için beceri ayrımı kapasitesi ve diğer tüm modellerde eksik bulunan Birleşik Model'in ilave ve potansiyel olarak önemli bir ögesi olan artık beceri parametresi  $\eta$  gibi ek olarak en önemli öğelerini tutarak, diğer beceri tanılama modellerine istinaden tanılama test verisi setlerine uyum sağlamak adına Birleşik Modelin esnek kapasitesini taşır. Denklem 5, orijinal Birleşik Modelde kullanılmış aynı aday parametreleri,  $\alpha_j$  ve  $\eta_j$  tabanlı RUM MTF ortaya koymaktadır (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).  $P_{c_i}(\eta_j)$  terimi, bir kez daha zorluk parametresi ile Rasch modelini işaret eder ( $c_i$ 'nin değeri ne kadar düşükse  $P_{c_i}(\eta_j)$ 'nin değeri de o kadar düşük olur.).

Denklem 5. Reparametrize Birleşik için Model Madde-Tepki İşlevi

$$P(X_{ij} = 1 | \underline{\alpha}_j, \eta_j) = \pi_i^* \prod_{k=1}^K r_{ik}^{*(1-\alpha_{jk}) \times q_{ik}} P_{c_i}(\eta_j).$$

$$0 < \pi_i^* < 1, 0 \leq r_{ik}^* \leq 1, 0 \leq c_i \leq 3.$$

Bu parametrelerin yorumlanmasının açık bir şekilde anlaşılması Fusion Modeli anlamak ve uygulamak adına önemlidir.

$\pi_i^* = P$  (tüm madde  $i$  gerekli becerileri,  $\alpha_{jk} = 1$  verilmiş tüm madde  $i$  gerekli becerilerini doğru bir şekilde uygulamak)

$$\begin{aligned}
r_{ik}^* &= \frac{P(Y_{ijk} = 1 | \alpha_{jk} = 0)}{P(Y_{ijk} = 1 | \alpha_{jk} = 1)} \\
&= \frac{r_{ik}}{\pi_{ik}}
\end{aligned}$$

(bireyin beceri uygulamasının koşullu bağımsızlığı tahmini altında)

$P_{ci}(\eta_j) = P(\alpha_b)$  ile alakalı becerileri  $\eta_j$  koşullu madde  $i$ 'ye doğru bir şekilde uygulamak).

$P_{ci}(\eta_j)$  terimi zorluk parametresi  $c_i$  ile Rasch Modeline işaret etmektedir. Parametrize edilirlen  $c_i$ 'nin değeri ne kadar düşükse  $P_{ci}(\eta_j)$ 'nin değeri de o kadar düşük olacağı dikkate alınmalıdır.

Fusion Model yeniden parametrizasyonunda orijinal Birleşik Modeldeki  $2ki$  parametreleri  $\pi_{ik}$  ve  $r_{ik}$ ,  $(1 + ki)$  parametreleri  $\pi_i^*$  ve  $r_{ik}^*$  ile yer değiştirir. Yeniden parametrize edilen modelin matematiksel olarak orijinal Birleşik Modele eşdeğer olduğu ve  $(1 + ki)$  parametrelerinin artık tanımlanabilir olduğu kolayca gösterilebilir. Tanımlanabilir bir parametre setinin meydana konmasına ek olarak, yeni parametreler uygulama açısından özellikle uygun bir yoldan kavramsal olarak yorumlanabilir bir haldedir. Parametre  $\pi_i^*$  madde  $i$  için tüm  $Q$  gerekli becerilerini öğrenmiş bir adayın madde  $i$ 'yi çözerken tüm becerileri doğru uygulayacağına dair olasılıktır. Gerekli bir beceri  $k_0$  öğrenmemiş adayın doğru madde yanıtı olasılığı  $r_{ik_0}^*$ 'dir. 0'a yakın bir  $r_{ik_0}^*$  değeri ile belirtildiği üzere, madde ne kadar çok becerinin öğrenilmesine bağlı olursa, beceriyi öğrenmeyen biri için madde yanıt olasılığı o kadar düşük olur. Böylece,  $r_{ik}^*$  beceri  $k$ 'nın öğrenimi ile ilgili madde  $i$  tarafınca sağlanan kanıt gücünün ters bir göstergesi gibidir.  $r_{ik}^*$  0'a ne kadar yakın olursa, beceri  $k$  için madde  $i$ , o kadar ayırt

edici olur. Bir beceri için  $r_{ik}^*$  parametrelerinin çoğu sıfıra yakındır (örneğin ortalamasının 0,5'ten daha az alınması), o beceri adına, beceri üzerine bilgi tanılama için iyi planlanmış bir testin göstergesi olan, bilişsel yüksek bir yapı ortaya koyduğu söylenir. Açık bir şekilde  $r_{ik}^*$  parametreleri bir değerlendirme aracının tanılama kapasitesini hesaplamada önemli bir rol oynar. Diğer modellerdeki parametrelerle kıyaslandığında  $\pi^*$  ve  $r^*$  parametrelerinin ayırt ediciliği dikkate değerdir. Aslında diğer modeller  $\pi_i^*$  ve  $r_{ik}^*$  terimlerine benzer öğeler içermektedir. Maris'teki modeller (1999) Birleşik Modeldeki tanımlanamaz  $\pi_{ik}^*$  ve  $r_{ik}^*$  parametrelerine sahiptir. Öte yandan, Patz ve diğ., (2002) ile Junker'ın (2000) ayrık Markov Chain Monte Carlo'su (MCMC) tanımlanabilir beceri tabanlı madde parametrelerine sahiptir, ancak madde belirli değildir, dolayısıyla her bir bireysel madde yanıt olasılığı üzerindeki beceri etkisi kayıptır. Bu durum birinin her bir beceri için o beceriye sahip olan ile olmayan adaylar arasında en etkili ayrımı yapan maddeleri bilmek istediği beceri tabanlı test planı açısından önemlidir (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

$P_{ci}(\eta_j)$  ögesi Birleşik Modelden alınmış önemli özgün bir öğedir çünkü Q matrisinin tüm maddeler için tüm geçerli bilişsel becerileri içermesinin şart olmadığını kabul eder. İlginç bir şekilde diğer beceri tanılama modellerinde bu mevcut değildir. Bu öğede  $c_i$ , Q matrisi tarafından o maddeye atananların haricinde, beceri üzerindeki madde yanıt fonksiyonunun güvenilirliğini işaret eder. Bir benzerlik olarak, bu diğer beceriler tarafından tek boyutlu bir beceri parametresi  $\eta_j$  tarafından tüm maddeler üzerinden ortalama bir şekilde modellenmiştir.  $c_i$  3 ya da daha fazla olduğunda, madde yanıt fonksiyonu  $\eta_j$  tarafından neredeyse etkilenmez çünkü  $P_{ci}(\eta_j)$ ,  $\eta_j$ 'nin pek çok değeri için



1'e çok yakın olacaktır.  $c_i$  0'a yakın olduğunda  $\eta_j$  varyasyonu sabitlenmiş bir  $\alpha_j$  ile bile madde yanıt olasılığı üzerindeki etkisini arttırmaktadır. Böylelikle  $c_i$  kestirimi bir becerinin Q matrisinden kaybolup kaybolmadığı ya da Q matrisinde hâlihazırda bir becerinin ölçülen becerilerin belirli bir madde listesine eklemek gerekip gerekmediğine dair düzeyli tanılama bilgisini sağlar (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

Fusion Model kısıtlı bağımsızlık (LI) verildiğini ( $\alpha, \eta$ ) varsayar. Eğer çoklu madde başına beceri atamaları Q matrisinden kaybolmuşsa, o zaman kayıp çoklu becerilerin etkisi de  $\eta$  tarafından tek boyutlu olarak tutulur. Bu durumda, kısıtlı bağımsızlık ( $\alpha, \eta$ )'e ilişkin olarak ancak tahmini olacaktır.

Özetle Fusion Model, Birleşik Modelin madde parametrelerini yeniden parametrize ederken, orijinal Birleşik Modelden en kritik aday parametrelerinin kestirimini mümkün kılar, böylece yalnızca tahmin edilebilir değil aynı zamanda geleneksel tek boyutlu psikometri modellerine nazaran eğitimsel testlerin kullanıcıları için daha dikkat çekici modeller yapan bir özellik olan beceri tabanlı yorumlanabilirliğini korur (Hartz ve Roussos, 2005; Hartz ve diğ., 2002; Roussos ve diğ., 2007).

#### **2.4. Model Parametreleri İçin Kestirim Metodu**

Birleşik Modeli parametrize ettikten sonra madde parametreleri kestirimi için çeşitli metotlar keşfedildi. Veriyi ayarlamak adına RUM kapasitesini arttırmak ve kestirim sürecini basitleştirmek ve geliştirmek için RUM'un Bayes bir artış modellemesi geliştirildi. Bayes Ağlarını kullanmak, seçilecek bir Bayes yaklaşımı olsa da, bunun yerine hiyerarşik bir Bayes modelleme yaklaşımı kullanımı ile ilginin olasılık yapısının

beceriler arası ilişkinin daha doğrudan birleşebileceği bulunmuştur. Geçerli Bayes hiyerarşik yapı daha fazla parametre ve genel uygulamada olduğu gibi önsellerin ya bilgilendirmeyen, ya fazlasıyla bilgilendiren ya da veriden tahmin edilerek seçilen önsellerini ekleyerek yeniden parametrize edilmiş Denklem 5 Birleşik Modeli geliştirir. Böylece beceri ve madde parametreleri için Bayes modelleme sistemini ve sonrasında Bayes sistemindeki parametrelerin tahmini için MCMC süreci tanımlanmaktadır.

#### 2.4.1. Beceri Parametreleri İçin Bayes Yapısı

$\eta_j$  parametresi için kullanılan önsel, basitçe standart normal dağılım için hazırlanmıştır. Ancak, ikili beceri öğrenim parametrelerinin altını çizen Bayes model daha karmaşık bir yaklaşım gerektirir. İkiye bölünmüş  $\alpha_{kj}$  beceri parametreleri başarı  $p_k$  olasılığıyla, beceri  $k$ 'nın öğreniminin yoğunluk oranı ile beraber Bernoulli rasgele değişkenleri gibi modellenmiştir.  $\alpha_{kj}$  beceri parametreleri için önsel, her bir beceri için  $p_k$  parametresi ve tüm beceri öğrenimi çiftlerinde dört düzeyli korelasyondan oluşmaktadır. Bu parametreler hiyerarşik bir Bayes modelinde hiperparametreler olarak modellenir.

Özellikle ikili beceriler arasındaki açık korelasyon büyük ölçüde her bir becerinin sahiplerinin ayrı oranlarına bağlıdır. Bu problemle başa çıkabilmek için, dört düzeyli korelasyonlar adı verilen iyi bilinen bir örtük değişken aracı kullanılmaktadır. Dört düzeyli korelasyonlar, normal bir şekilde dağılan örtük değişkenlerin gözlenen her bir ikili değişkeni ortaya çıkardığı farz edildiği noktada, iki tane ölçülü ikili değişken arasındaki ilişkiyi biçimlendirir. Dört düzeyli korelasyonların kullanımı ikili becerilerdeki açık korelasyon kullanımına tercih edilir, çünkü açık korelasyonlar büyük ölçüde her bir becerinin hakiminin ayrılan oranlarına bağlıdır, böylece iki beceri

arasındaki ilişki gücünün ölçümleri olarak yorumlamayı zorlaştırır. Belirtildiği üzere, dört düzeyli korelasyonlar, sürekli normal rastgele değişkenlerin ikili  $\alpha_{kj}$  öğrenim değişkenlerini vurguladığını ve sürekli değişkenlerin kesme noktası parametrelerince ikiye ayrıldığını varsayar. Sembolik bir şekilde  $\tilde{\alpha}_{kj}$  olarak belirttiğimiz bu sürekli değişkenler “kesme noktaları” tarafından ikiye bölündüğünde bir beceride yeterliliği ya da yetersizliği saptayan, beceri yeterliliğinin sürekli düzeyleri olarak incelenir. Kesme noktası parametreleri  $\kappa_k$  olarak simgelenir ve  $\tilde{\alpha}_{kj}$ 'nin standart normal ayrımı takip ettiği düşünüldüğü  $P(\tilde{\alpha}_{kj} > \kappa_k) = p_k$  ilişkisine  $p_k$  parametreleri ile bağlantılıdır.

$\eta_j$  ve öncelikli  $\tilde{\alpha}_j$  değişkenleri arasındaki korelasyonlar negatif olmayan korelasyonlar olarak modellenir. Bu korelasyonlar hiperparametreler olarak tahmin edilir (birleşik dağılımlarının tahmini için veri sağlar) ve sırasıyla sınıf problemlerinin önüne geçmek için  $a$  ve  $b$ 'nin 0,01 ile 0,99 arasına koyulduğu, bir Uniform önselliği,  $\text{Unif}(a, b)$  verilir. MCMC algoritmasındaki korelasyonları tasarlamak için sayısal teknikler korelasyon matrisinin gereksediği üzere pozitif kesinlik sürdürmek için kullanılır.

#### 2.4.2. Madde Parametreleri İçin Bayes Yapısı

$\pi_{ik}^*$ ,  $r_{ik}^*$  ve  $c_i$  değerleri verilmiş bir veri serisi için fazlasıyla çeşitlenebildiği için üç tür madde parametreleri için önseller (dağılım işlevleri), önsellerin alabileceği şekillerin azami esnekliğini sağlayarak her biri birer Beta dağılımı olarak seçilir. Beta dağılımları Bayes analizlerinde sıkça kullanılır çünkü sınırlı aralıkta kalmış parametreleri  $\pi_{ik}^*$  ve  $r_{ik}^*$ 'de olduğu gibi doğal bir şekilde biçimlendirir. Kurallara bakılsa bile  $c_i$  gerçekten bir değer olabilir, model parametresi kestirimi için değerlerini sınırlı bir aralıkta bırakmak uygulanabilirlik hissi uyandırır. Ayrıca Beta dağılımlarını önseller

olarak tercih etmek adına matematiksel sebepler de vardır. Bu sebepler, Beta dağılımlarının iki terimli dağılım ailesinin kökeninden geliyor olması ile ilgilidir. Beta dağılımlarının değerli esnekliğinden faydalanabilmek için, Beta önsellerinin parametrelerinin kendileri, önseller olarak atanarak ve denk gelen hiperparametreleri hesaplayarak hesaplanır. Arpeggio'nun güncel sürümünde bu hiyerarşik önseller tahminen uniform dağılımları olarak endekslenir.

### **2.4.3. Markov Chain Monte Carlo (MCMC)**

Büyük oranda, Patz ve Junker (1999) ikilisinin makalelerinin etkisine bağlı olarak MCMC pek çok beceri tanılama uygulamalarında, özellikle tümüyle Bayes Fusion model ve NIDA ve DINA modelleri için popüler hale gelmiştir. EM (beklenti ve maksimizasyon) algoritmaları sayısal olarak MCMC algoritmalarından çoğu kez daha etkili olmasına rağmen, EM algoritmalarını yeni modellere ya da MCMC algoritmalarından model seçeneklerine genişletmek daha zordur. Patz ve Junker'ın da üzerinde durduğu gibi, EM algoritmaları parametrik olarak karmaşık modellere uygulamak için “açık” değildir. Burada, MCMC'nin kısaca bir kavramsal tanımlaması sağlanacaktır. İlk olarak, olasılığa dayalı bir tabanı olan hesaba bilgisayarlı metot, tüm parametreleri hesaplamak için simüle edilmiş değerlerin Markov zincirlerini oluşturmak için kullanılır. Zincirdeki her bir zaman noktası (ya da adım) simüle edilmiş bir değerler grubuna denk düşer. MCMC teorisi yeteri kadar büyük sayıdaki adımlardan sonra (zincirin ısıdırma sistemi olarak adlandırılır), kalan simüle edilmiş değerler parametrelerin istenen Bayes sonsal dağılımına giderek yaklaşacaktır. Bu sonsal dağılımı edinmenin, neredeyse her zaman, analitik olarak hesaplamının ve modeldeki parametrik karmaşa seviyesi sebebiyle modern

bilgisayarimsal yaklaşım metotlarının kullanımında bile sayısal olarak yaklaşmanın zor olacağı kayda değerdir. Bu analitik direktmeye bir çözüm ise analitik olarak karmaşık tam koşullu tasarı dağılımlarının yerine daha kolay tasarı koşulları kullanan Gibbs örneklendirmesi ile beraber Metropolis-Hastings yöntemini uygulanmasıyla sağlanmaktadır. Arpeggio'nun güncel sürümünde tasarı dağılımları (ayrıca atlama dağılımları olarak da adlandırılır) hareket eden pencerelerin şeklini alır. Bu tasarı dağılımlarını tanımlayan parametreler değerleri özelleştirmek için program tarafından çoktan sabitlenmiştir ve kullanıcılar bu değerleri öncelikle Henson, Templin, ve Porch'a (2004) danışmadan değiştirmemelidirler (Roussos ve diğ., 2007).

MCMC kestirimi; uzun zincirler işleyerek, uzun zincirin her bir zaman noktasında tüm parametreleri benzeştirerek, ısındırma adımlarını çıkarmak ve parametre değerlerini ve standart hatalarını ölçmek için her bir zincirde geri kalan adımları önsel dağılımlar olarak kullanarak gerçekleştirilir. Uygulayıcı kullanılacak zincir sayısını, zincirin total uzunluğunu ve ısındırma için kullanılacak sayıyı dikkatlice seçmelidir. Bir uygulayıcı verilen bir analiz adına az tecrübesi varsa, oldukça uzun en az iki zincir ve çok uzun ısındırmalar kullanması tavsiye edilmektedir. (örneğin 50.000 uzunluğunda bir zincir ve 40.000 uzunluğunda bir ısındırma). Bu bölümdeki sonraki kısım MCMC sonuçlarının nasıl değerlendirileceği hakkında bilgi sağlar. Isındırma sonrası zincirler parametrelerin sonsal dağılımlarının kestirimleridir. Arpeggio programı bu dağılımları EAP (bir sonsal beklentisi (expectation a posteriori)) puan kestirimleri ve değişkenlik ölçüleri olarak sonsal standart sapmaları sağlamak için kullanır.

İlginçtir ki MCMC hem kısmen Bayes değerlendirmesinden ya da bir Bayes olmayan değerlendirmeden de kullanılabilir ve her ikisi de farklı durumlarda faydalı

olabilir. Ortak bir uygulama, uygun bir deneysel Bayes niteliği katmak adına sık sık hiyerarşik yapı ile tamamıyla Bayes tarzıdır (burada Fusion Model için açıklandığı gibi). Başlıca çıkarımsal sonucu, ek dağılımdan çıkan rasgele bir çekimler grubunca yansıtıldığı gibi, ek madde ve sonra giren kişi parametre dağılımının tam sonsal dağılımıdır. Pek çok kişi bu dağılımlara ve özellikle her bir parametrenin ölçülmüş mevcut standart hatalarının olmasını kayda değer bir fayda olduğunu tartışmaktadır. Çoğu MCMC kaynakları mevcuttur ve Partz ve Junker makaleleri ve Gelman ve diğerlerinin (2013) kitabı burada incelenen modeller gibi karmaşık parametrik modelleri değerlendirmeye ilgili okuyucular için özellikle tavsiye edilen kitaplardır.

## **2.5. Model Kontrolü Yöntemi**

Bu bölümde Fusion Model sistemi ile birlikte mevcut olan model kontrol metotlarının çeşitlerini, bu metotların sonuçlarının nasıl yorumlanacağını ve bu sonuçlara karşı araştırmacının hangi eylemleri tepki olarak göstereceği tartışılacaktır.

### **2.5.1. Yakınsama Kontrolü**

Bir çözüm elde etmek için yakınsamaya kadar tekrarlama içeren tüm istatistiksel kestirim metotları, yakınsama için kontrol ihtiyacı duymaktadır. Çünkü MCMC kestiriminden elde edilen istatistiksel bilgi (tam bir sonsal dağılım), bir EM algoritmasından (bir kestirim ve onun standart hatası) elde edilenden daha fazladır, yakınsamanın gerçekleştiği ya da gerçekleşmediği değerlendirmesi MCMC’de daha zor olmaktadır.

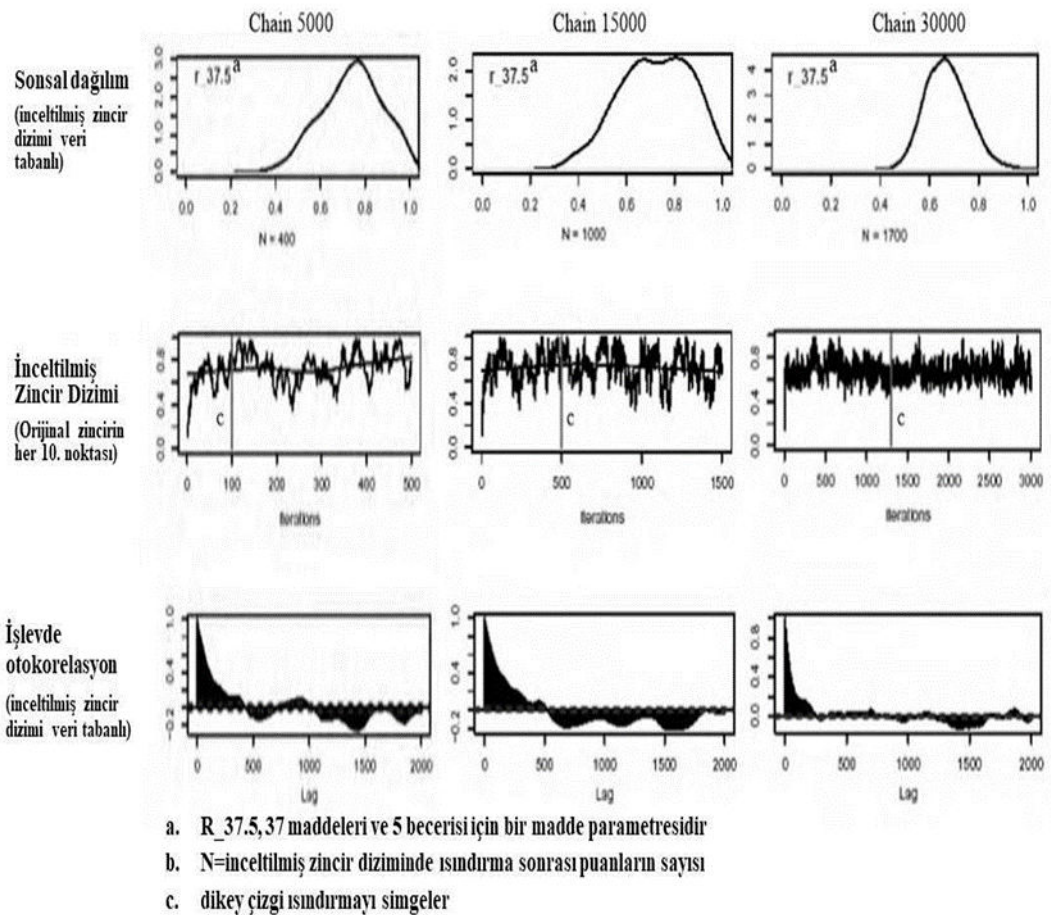
Beceri tanılama uygulamalarında MCMC yakınsaması, modelin karmaşası ve modelin tasarısı ve uygulaması ne kadar verinin gerçekliğine denk düşse de elde etmesi

zor olabilir. Bazı durumlarda karmaşık modeller istatistiksel olarak tanımlanamaz olabilir. Diğer durumlarda uymaz benzerlik fonksiyonlar yüzünden tanımlanabilir modelleri iyi bir şekilde hesaplaması zor olabilir. Tabi ki bu modelleme sorunları yalnızca MCMC için değil tüm tahmin yöntemleri için problemlere sebep olabilir. Her halükarda, Arpeggio MCMC algoritmasını işlettikten sonra yapılması gereken ilk şey zincirlerin kararlılığa yakınsanıp yakınsanmadığını saptamaktır.

MCMC tahmininde Markov Zincirlerinin yakınsamasına istinaden literatürde pek çok şey yazılmasına rağmen, her bir model parametre için Markov Zincirlerinin yakınsanıp yakınsanmadığını güvenilir bir şekilde ölçen tek bir istatistik bile yoktur. Fusion Model sistemi yakınsamayı ölçme için 4 yönde sonuçlar ortaya koymaktadır (Gelman ve diğ., 2013).

Yakınsama kontrolü için önemli bir araç her bir parametre için zincir dizimidir. Şekil 1’de Jang tarafından yapılmış madde başına (her beceri için sekiz madde kadar bir ortalamaya dönüştüren) iki beceri kadar bir ortalama ile birlikte dokuz becerilik bir Q matrisi kullanarak 1350 kişi tarafından girilen 37 maddelik bir testin analizi bağlamında bir madde parametresinin kestirimi üzerine zincir uzunluğunun etkisine dair araştırmanın bir örneği verilmiştir. Dizimler, dizimleri kolaylaştırmak ve çıktının büyüklüğünü azaltmak için zincirdeki her 10. maddeyi gösterir. Bu örnekte zincir dizimlerinden aşikâr görünen odur ki 1000 adım ısındırma olarak yeterince hizmet etmektedir. Dizimler zincirlerin sabit bir dağılıma yerleştiğini açık bir şekilde göstermektedir. Ayrıca Şekil 1’de sonsal dağılımlar da gösterilmektedir. Bu grafikler, zincirin ilk 10000 adımından sonra sonsal dağılımda küçük bir değişim olduğundan yakınsamanın gerçekleştiği zincir diziminden çıkan sonucu destekler. Şekil 1 zincirdeki

belirli bir sayıdaki adımlarla ayrılmış benzeşik veriler iki grup arasındaki bağımsızlığın derecesini elde etmek için kullanılabilir. Oto-korelasyon ne kadar düşüğe verilen bir zincir uzunluğu için o kadar çok bilgi vardır. Böylece eğer oto-korelasyon nispeten yüksekse mantıksal doğrulukla beraber sonsal dağılımı tahmin etmek için zincirler daha uzun işlenmelidir. Şekil 1, 500 ya da daha fazla adım aralıklı zincir kestirimleri arasında küçük bir korelasyon olduğunu ifade eder. Bu örnek iyi yakınsama sonuçları ortaya koymaktadır ve bu şüphesiz ki Jang tarafından yapılan dikkatli model gelişimi sayesinde.



Şekil 1. Zincir uzunluğunun madde parametreleri kestirim yakınsamasındaki etkisi (Jang, 2005; (Jang, 2005; Roussos, DiBello, Henson, Jang, & Templin, 2010).



Unutmamalıdır ki yüksek oto-korelasyonlar ille de MCMC yakınsamazlığını ifade etmez, ancak alternatif olarak düz bir benzerlik yüzeyini işaret ediyor da olabilir. Bu, en az birkaç önemsiz parametrelere sahip herhangi bir belirli veri grubu için neredeyse muhtemel olan karmaşık modellerle ilgili ortak bir problemdir. Bu olgu aşırı parametrize edilmiş modellerin tanımlanamazlığı ile aynı durum değildir. Böyle durumlarda benzerlik, tüm veri grupları için bağıl olarak düzdür.

Şekil 1’de Jang (2005) çeşitli zincir uzunluklarıyla tek bir zinciri gözlemlemiştir. Fusion Model sistemi tarafından desteklenen bir diğer önemli metot ise çoklu zincirler için Arpeggio’yu uygulamak ve zincir dizimlerini, sonsal dağılımları ve oto-korelasyonları karşılaştırmaktır. Eğer seçili zincir uzunluğu yakınsamayı gerçekleştiriyorsa farklı zincirlerden sonuçlar da hemen hemen aynı olmalıdır. Gelman ve diğ., (2013) 1,2’den az  $R^{\wedge}$  değerlerinin gerekli olduğunu ancak yakınsama koşulu için yeterli olmadığını ortaya sürer. Yakınsamasız durumlarda 1,2’den daha fazla  $R^{\wedge}$  değerleri ile nadiren karşılaşılmıştır. Karşılaşıldığı zamanlarda da genelde girdilerden birinde ya da veri dosyalarında yapılan bir kaba hatanın göstergesi olmuştur. Yakınsama Fusion Model sistemi ile birlikte gerçekleşmediğinde çıkarılan birkaç varsayım bulunmaktadır. Karşılaşılan sık bir problem,  $c_i$  parametresi (ve yanında ona eşlik eden  $Q$ ’suz beceriler için  $\eta_j$  parametresi) modele eklendiğinde,  $Q$  becerilerine ilişik model parametrelerinin ölçümünün yakınsanmamasıdır (ya da oldukça geniş olan  $p_k$  parametrelerinin çoğunu içeren bir çözüme yaklaşmıyordur). Bunun analizlerde çok sık gerçekleşme sebebi uygulamaların pek çoğunun tek bir baskın boyutu bulunan testlerle birlikte olması ve bu sürekli  $\eta_j$  parametresinin madde yanıtlarında değişkenlerin çoğunu “içine çekmesine” yol açmaktadır. Diğer yandan, pek çok maddenin  $Q$  matrisinde

belirtilen gerekli becerilerin büyük oranına sahip olduğu durumlarda  $c_i$  parametresi için bazen yakınsamasızlık meydana gelmektedir. Böyle durumlarda, modelin  $P_{c_i}(\eta_j)$  kısmı tarafınca açıklanacak bir değişken kalmamaktadır. Eğer bu durumlardan herhangi biri meydana gelirse modeldeki tüm maddelerden  $c_i$  parametresini çıkarmak iyi bir çözüm olacaktır. Fusion Modelin bu azaltılmış sürümü gerçek bir veriyi analizde sıkça kullanılan bir sürümüdür.

Eğer  $c_i$  parametresi modelde yoksa ve yakınsamazlık gerçekleşirse, kontrol edilecek ilk şey MCMC zincirinin ısınma evresinde sonsal dağılıma ulaşmak için yeterince uzun olup olmadığıdır. Bu fazlasıyla uzun bir zincir kullanımı ile kontrol edilebilir. Eğer daha uzun olan zincir yakınsama ile sonuçlanmıyorsa zincir uzunluğu problem olmaktan çıkar. Bu durumda, model geliştirme adımlarını yeniden gözden geçirilebilir ve çıkan değişikliklerin nereden geldiğini belirlemek adına Q matrisi ve seçili model yeniden incelenebilir.

### **2.5.2. Model Parametre Ölçümlerinin Yorumlanması**

Model tahmin yöntemi yakınsandığı farz edilirse beceri dağılımları için kestirimler ve madde parametreleri ölçülmelidir.

#### **2.5.2.1. Beceri Dağılımı Parametreleri**

Beceri (beceriye sahip olma/olmama) tanılama modellerinde kilit nokta her bir beceride beceriye hâkim olarak ölçülen adayların oranlarının kullanıcı beklentileri ile nispeten eşleşip eşleşmemesidir. Fusion Model sistemi  $p_k$  parametrelerinin ölçümlerini ve bu amaç için de standart hataları kestirici bulundurur. Eğer bir beceri beklenilenden çok daha zor ya da kolay hale gelirse  $p_k$  ölçümü ve bunun standart hatası tabanlı olarak

Q matrisi yeniden gözden geçirilmeli ve becerilerin atandığı maddeler için madde zorluk dereceleri incelenmelidir. Buna ek olarak, beceriyi değerlendiren görevlerin seçimi, mevcut görevlerde eklenecek ya da yer değiştirecek daha uygun görevlerin olup olmadığını görmek adına yeniden gözden geçirilebilir. Örneğin, bir beceri için özelliğe sahip olanların oranı çok düşükse o beceri için daha kolaylarını ya da her ikisinden de ekleyerek daha zor olan görevler daha kolaylar ile değiştirilebilir. En nihayetinde, becerinin tanımının, örneğin uygun Q değişikliği ile ya da yeni bir grup görevlere yol açacak daha basit bir şekilde düzenlenmesi gerekebilir. Örneğin, Jang'a (2005) göre diğer sekiz beceri zorlukları Jang'ın yaklaşık beklentileri ile tutarlı bir şekilde düzenlenmiş haldeyken becerilerden biri istatistiksel analiz tarafından beklenenden daha kolay bulunmuştur.

#### 2.5.2.2. Madde Parametreleri

Fusion Model sistemi ayrıca madde parametreleri için kestirimleri ve standart hataları da sağlar.  $\pi_i^*$  ve  $r_{ik}^*$  ve eğer verildiyse  $c_i$  parametreleri için kestirimler detaylı olarak incelenmelidir. Çünkü tanılamamanın başarısını saptamada kilit bir rol oynar.

Öncelikle daha önceki kısımlarda bahsi geçen durumlar tekrarlanmalıdır. Eğer  $c_i$  parametreleri modele dâhil edildiyse ve kestirim yakınsaması elde edildiyse modelin  $P_{ci}(\eta_j)$  kısmı uygun modelde baskın mı değil mi belirlemek için çözüm dikkatlice incelenmelidir. Eğer bu gerçekleşirse o zaman  $p_k$  parametrelerinin çoğu oldukça geniş olacaktır. Böylece bireysel farkların pek çoğu  $\eta_j$  parametresi tarafınca açıklanır. Bu analizlerin çoğunda sıkça meydana gelmektedir çünkü uygulamaların çoğu tek bir baskın boyuta sahip testlerden oluşmaktadır ve sürekli  $\eta_j$  parametresinin madde yanıtlarında değişkenlerin çoğunu yutmasına olanak sağlamıştır. Bu durumda

yakınsama gerçekleşse bile tanısal açıdan faydalı tahmini bir model elde etmek için modelden  $c_i$  parametrelerinin çıkarılması gerekir.

Yakınsamanın gerçekleştiği ve önceki  $c_i$  parametrelerinde olduğu gibi kestirimlerin herhangi önemli bir problem göstermediği farz edilirse, her bir madde parametresi türü için kestirimler, gereken becerilerle ilgili her bir maddenin performansı hakkında faydalı bilgiler sağlayacaktır.

İnceleme için en basit madde parametresi tüm gereken becerileri edinme koşulunu taşıyan adayların bir madde tarafından gereken tüm  $Q$  becerilerini doğru bir şekilde sergileme olasılığını işaret eden  $\pi_i^*$  parametresidir. Bu değer bire yakın olması istenmektedir. Uygulamalarda atanan beceriler için fazlasıyla zor olan maddeler belirlenerek  $\pi_i^*$  değerleri 0,6'dan az olarak yorumlanmıştır. Maddeye atanmış belirli becerilerin öğreniminin gösterimi olan doğru madde yanıtını elde etmek için gerek daha fazla beceri gerekse farklı beceriler olsun böylesi maddelere atanması gerekir. Geri kalan iki tür madde parametreleri  $r_{ik}^*$  ve  $c_i$ , maddelerin sırasıyla  $Q$  becerilerinde ve bir madde için  $Q$  atamalarında becerilerin kaybolduğu dereceye kadar nasıl iyi ayırım yaptığı konusunda bilgi verir.  $r_{ik}^*$  için düşük bir değer (0 – 0,5) yüksek derecede ayırım yapan bir maddenin göstergesidir ve  $c_i$  için düşük bir değer ise (0 – 0,15) hem madde için bir  $Q$  becerisinin eksik olduğunun hem de bir becerinin tüm  $Q$  matrisinde eksik olduğunun göstergesidir.  $r_{ik}^*$  ve  $c_i$  için yüksek değerler olası model sadeleştirmesinin belirtisidir. Eğer, örneğin  $r_{ik}^*$  0,9'dan büyükse, madde denk gelen beceri için çok da ayırt edici olmaz ve  $Q$  matrisi girdisi çıkartılabilir ve eğer  $c_i$  büyükse o zaman madde yaklaşık olarak “tamamlanır” ve  $c_i$  düşürülebilir.  $Q$  matrisindeki bir adet “1” girdisi anlamlı temellerde tamamı ile doğrulanabileceği ve yine de 1,0'a oldukça yakın

hesaplanmış bir  $r_{ik}^*$  değeri gösterebilir. Tanısal değerlendirme için yüksek ölçülmüş  $r_{ik}^*$  değeri, istatistiksel olarak belirli madde verilen özelliğe sahip olan ve sahip olmayan arasındaki ayırım adına çok bir bilgi katkısında bulunmuyor demektir. Bu yüzden  $Q$  matrisinden “1” girdisini çıkarmak illa ki madde kodlamasında bir “hata”nın olduğu anlamına gelmez. Buna karşılık karar, parametre kestiriminin doğruluğu üzerine bahse giren istatistiksel olarak stratejik bir karardır. Kestirilecek az bir parametre ile ve sınıflandırma doğruluğunun gelişmesi ile sonuçlanması ve yorumlanabilirlik, orijinal “1” kodlaması için sağlam gerekçelerin ışığında bile  $Q$  matrisinde “1” girdisini eksiltmeyi destekleyebilecek elle tutulur bir fayda sayılır. Tabi ki bu tür iptaller belli bir beceri gerektiren maddelerin çoğunda meydana gelir; uygulayıcı bunu, sürekli kodlama ve istatistiksel davranışı daha iyi sıralamayı sağlamak adına beceri ya da beceri kodlaması yönteminin değiştirilmesi gerektiğine dair bir işaret olarak yorumlamalıdır.

Bilgilendirici olmayan  $r_{ik}^*$  ve  $c_i$  parametrelerinin çıkarılması Fusion Model sistemin önemli bir parçası olarak kabul edilir çünkü bu modelin istatistiksel gücünü tanısal bilgilerin olduğu yere odaklamasına yardımcı olur. Bir diğer deyişle, modelin parametrelerinde bu sayıda bir azalma aynı gözlemlenen veri ile daha az parametrenin daha iyi kestirilebilmesine yardımcı olabilir. Ancak daha da önemlisi, önemsiz  $Q$  matris girdilerini azaltmak çıkarılan  $Q$  girdilerine denk gelen maddelerde neredeyse işaretli uzman adaylar kadar iyi performans gösteren işaretli uzman olmayan adaylardan doğabilecek aday karışıklığı olasılığını azaltır. Buraya kadar, istatistiksel olarak etkili olmayan  $r_{ik}^*$  ve  $c_i$  madde parametrelerini tanımlamada yardımcı olan istatistikler geliştirilmiştir.

Etkisi olmayan parametreleri saptama hedefini tamamlamak için Fusion Model sistemi, o parametreyi kullanan madde yanıt fonksiyonu üzerindeki  $r_{ik}^*$  ve  $c_i$  etkisini tahmin eder. Bu etkiyi ölçmek için ortak bir olasılık ölçeği kullanarak, aynı istatistiksel karar kuralı, modeldeki parametreyi tutma garantisi için ölçülen etkinin yeterince büyük olup olmadığını saptamada hem  $r_{ik}^*$  ve hem  $c_i$  için kullanılır. Simülasyon çalışmalarında bu kararlar yalnızca istatistiksel hipotez denemesi yapı temeline dayanır ve sıfır hipotezi inceleme altındaki madde parametresinin madde yanıt fonksiyonunda önemsiz etkiye sahip olması durumudur. Ancak, daha önce de bahsedildiği üzere uygulamada böylesi karar almanın tipik olarak istatistiksel ve temel girdilerin etkileşimine dayanması gerektiğini vurgulamak gerekir. Çünkü bir Q matrisi, matrisi desteklemek için sık sık güçlü teorik argümanlarla geliştirilir, Q matrisi girdilerini azaltmak yalnızca güçlü istatistiksel kanıt değil aynı zaman güçlü asli argümanlar gerektirir.

Belirli bir madde parametresinin etkisini kestirmek için üç ayrı madde yanıt olasılıkları, Fusion Model MTF tarafından her bir aday için hesaplanır. Tahmin edilen ortalamalarına sabitlenmiş diğer parametrelerle birlikte üç MTF olasılığı hesaplanır: (a) kendi sıfır hipotezi (etkisi yok) değerine sabitlenmiş parametre ile (bir  $r_{ik}^*$  1,0'da ve  $c_i$  10,0'da sabitlenecektir) (b) ölçülmüş standart sapmanın altında ölçülen ortalama alınan parametre ile ve (c) ortalamasına kendi standart sapmasının eklenerek belirlenen parametre ile. Bir  $r_{ik}^*$  parametresi için üç MTF olasılık hesabı beceri  $k$ 'nın uzmanı olmadığı kestirilen tüm adaylar üzerinden ortalama alınır. Bir  $c_i$  parametresi için tüm adayları ele alarak ortalama alınır. Madde yanıt olasılığının ortalaması (a), (b) ya da (c) için de yakınsa parametrenin etkisiz olduğu ve modelden çıkarıldığı söylenir.

İki ortalama olasılığının da kapalı olup olmadığının kararlılığı tamamı ile kişisel bir karardır, ancak yine de bu noktada karar almada yardımcı olacak diğer MTK ilgi alanlarından deneyimler mevcuttur. Aşına olduğumuz bu çeşit karar vermede en yaygın örnekler ayırt edici madde işleyişi (DIF) analizleri bağlamındadır. DIF analizlerinde, 0,01'den az MTF'ler arasındaki bir ortalama farkı kesin olarak önemsiz bir oran olarak görülür ve ne var ki 0,05 ile 0,10 arasındaki fark ise genelde büyük olarak düşünülür. Yapılan simülasyon çalışmalarında, böyle belirli çalışmalar için en iyi çalışan nispeten daha bağımsız bir yaklaşım bulunmuştur. Özellikle simülasyon çalışmalarında bir  $c_i$  ya da  $r_{ik}^*$  için (a) için ortalama (b) ortalamasının 0,03 içinde ya da c ortalamasının 0,01 içinde ise bir madde için etkisizdir. Uygulayıcılar tabii ki diğer mantıklı kriterleri seçebilirler ve yalnızca (a) ve (b) arasındaki farkı ya da (a) ile (c) arasındaki farkı kullanmayı tercih edebilirler. İlki, parametrenin modele ait olduğu ve onu düşürecek, sıfır hipotezinin seçildiği duruma denk gelmektedir. İkinci ise parametrenin modele ait olmadığı ve yalnızca eğer veri tam tersinde güçlü kanıtlar sağlarsa onu içerecek olan sıfır hipotezi durumuna karşılık gelir. Her iki yaklaşım da her iki kriter türünün birleşmiş kullanımı olduğundan istatistiksel olarak geçerlidir. Seçim, olması gerektiği gibi uygulamacıya kalmıştır.

Son olarak, madde parametreleri becerileri geçtiği ve maddeler birbirini etkilediği için hesaplamaları tek bir sefer de değil de tekrarlayarak gerçekleştirilmektedir. Eğer aynı anda birden fazla parametre etkisiz gözükürse o zaman bir parametrenin çıkarılması diğer bir parametrenin aynı adımda etkili hale gelmesine sebep olabilir. Eğer bir  $r_{ik}^*$  ve bir  $c_i$  parametresinin ikisi de aynı adımda etkisiz olarak belirirse,  $\eta$  kestirimini tutmanın olasılığı üzerine, maddede beceriye sahip olma kestirimini tutmanın olasılığını

kollamak için  $c_i$  parametresi tercihen çıkarılabilir çünkü beceri tanılama analizinin temel amacıdır. Ayrıca eğer belirli bir madde için iki ya da daha fazla  $r_{ik}^*$  parametreleri etkisiz bulunursa sıfır hipotezi MTF olasılığı ( $\alpha$ ) ve üst sınır MTF olasılığı arasındaki en küçük ortalama farkına sahip parametre çıkarılır.

Bir Q matrisi gerçek veri ile Fusion Modelin kestirimi için dikkatlice geliştirildiğinde bile ya da simüle edilmiş veri ortaya çıkarıldığında ve kestirim modelinde bilinen Q matrisini kullanıldığında bile verilen veri için istatistiksel olarak anlamsız madde parametreleri olabilir.

Örneğin, gerçek bir veri tabanında beceri maddeye Q matrisi tarafından atanmış olabilir, ancak adaylar maddeye doğru bir şekilde yanıt verirken beceriye ihtiyaçları olmayabilir ya da beceriyi kullanmamış olabilirler ya da bir beceri yine Q matrisi tarafından belirlenmiş diğer becerilerden çok daha düşük bir derecede olabilir. Bu durumun bir örneği de veri düşük  $r_{ik}^*$  değerleri ve yüksek  $\pi_i^*$  değerleri ve orta  $c_i$  değerleri ile simüle edildiğinde gerçekleşir. Bu durumda,  $c_i$  parametreleri doğru kestirilebilir halde olmayabilir çünkü madde yanıtları adayların becerileri edinip edinmediği ile belirlenir ve onların Q olmayan beceriler üzerindeki yeterliliklerinden az derecede etkilenirler. Ek olarak, eğer bir madde hem zor becerileri hem de kolay becerileri ölçerse ve zor beceri için  $r_{ik}^*$  çok düşükse, kolay beceri için  $r_{ik}^*$  MTF'de çok küçük bir etkiye sahip olacaktır.

Parametre çıkarımının kullanıldığı gerçek bir durum örneği ise Jang'ın (2005) analizidir. Çalışmada tahmin edilen madde parametreleri Jang'ın Q matrisi girdilerinin %9'unu çıkarmaya itmiştir çünkü kesin maddeler onlara atanan becerilerin bazılarında



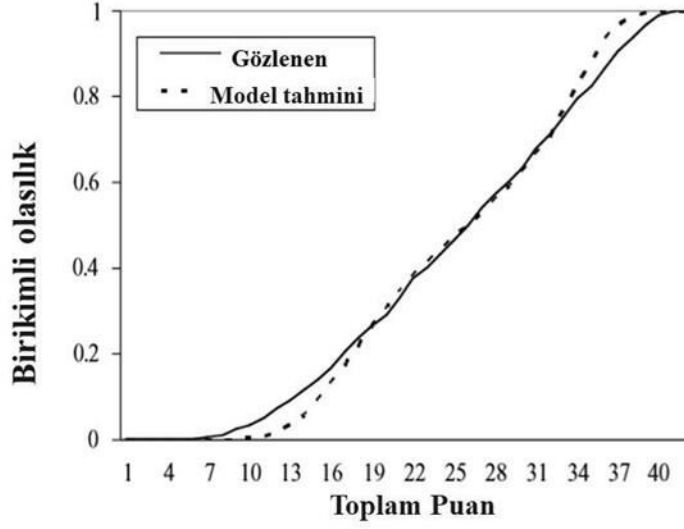
iyi ayırım yapamamaktadır. Özellikle tahmin edilen parametreler göstermektedir ki uzman olmayanlardan uzman olanlara madde-beceri kombinasyonunun yaklaşık %9'u için beceri performans oranı 0,9 ya da daha fazladır.

### 2.5.2.3. Model Uyarlık Değerleri

Fusion Model verilerden bir kere ayarlandığında, gözlenebilir istatistiklere kıyaslanan gözlenebilir özet istatistiklerini kestirmek için uygun model kullanılarak model uyarlığı hesaplanır. Beklenen istatistikler uygun modelden veri simüle edilerek ve simüle edilen verideki istatistikleri hesaplayarak elde edilir. Simülasyon modelinde kullanılan madde parametreleri asıl veri analizinden madde parametre kestirimleridir (EAPs). Simülasyon modelde kullanılan beceri parametreleri ise olasılığa dayalı olarak belirtildiği gibi elde edilmektedir. Kalibrasyon verisi grubundaki her bir aday için Fusion model sistemi, her bir beceri ve bir  $\eta$  tahmini için uzmanlık olasılığı sonsalı (ppm) sağlar. Simülasyonda, adaylar kalibrasyon örneğinden bir yerleştirme ile denenirler ve bir  $(\alpha_k, k = 1, \dots, K)$  vektörü, her bir beceri için adayın örneklenmiş uzmanlık olasılığı sonsalına eşit olan Bernoulli olasılığı ile, her bir beceri için bir tane bağımsız bir Bernoulli rastgele değişkenleri oluşturarak simüle edilir. Adayın yeniden örneklendirilmiş asıl hesaplanmış  $\eta$ 'i simüle edilmiş değer için kullanılır. Sonra madde parametreleri ve beceri parametreleri simüle edilmiş madde yanıtlarını oluşturmak için standart bir şekilde kullanılır. Bu tür bir yaklaşım sonsal kestirimci model kontrolü olarak adlandırılır. Fusion model sisteminde tipik olarak hesaplanan istatistikler, maddelerde, ikili madde korelasyonlarında ve aday ham puan dağılımında dozu ayarlanmış puanlardır. İkili madde korelasyonları ve puan dağılımı veri-model

uygunluğunu daha çok gerektiren testlerin ürünüdür, dikkatle ve değerlendirmenin tanılama amacı ışığında yorumlanmalıdır.

Belirli bir örnek olarak, Jang'ın uygulamasında tahmin edilen ve gözlenen dozu ayarlanmış puanlar arasındaki ortalama kesin fark (MAD) 0,002 ve korelasyonlar için MAD iyi uygunluğu destekleyen 0,049'dur. Gözlenen ve tahmin edilen puan dağılımları da ayrıca karşılaştırılabilir. Şekil 2, Jang'ın gerçek veri analizinden gözlenen ve tahmin edilen puan dağılımları arasındaki bir karşılaştırmayı gösterir. Dağılımın en düşük ve en yüksek kısımdaki uyumsuzluğun, adayın özelliğe sahip olma/olmama modeli en düşük puanları çıkaran adayların puanlarını fazla değerlendirmesi ve en yüksek puanları çıkaran adayların puanlarını ise eksik değerlendirmesi olması beklenir. Çünkü analizin hedefi adayları sıralamaktansa özelliğe sahip olup olmamayı değerlendirmektir, bu uyumsuzluğun aslında bilgiyi edinmiş/edinmemiş sınıflandırmasına bir etkisi yoktur. En düşük puanı çıkaranlar fazla değerlendirilse bile yine de tüm becerilerde bilgiyi edinmemiş olarak sınıflandırılırlar. Benzer bir şekilde, en yüksek puanı çıkaranlar da eksik değerlendirilse bile, tüm becerilerde bilgiyi kazanmış olarak sınıflandırılırlar.



Şekil 2. Gözlenen ve model tahmini puan dağılımlarının karşılaştırılması (Jang, 2005; Roussos ve diğ., 2010).

Bu birinin, sürekli bir parametrelili daha basit bir tek boyutlu modelin puan dağılımı verisine daha uygun olacağını keşfedebileceği bir durumun örneğidir. Ancak böylesi bir model daha fazla çalışma olmaksızın istenen bilgiyi kazanmışlık/kazanmamışlık sınıflandırma kestirimini vermediyse, tanılama değerlendirmesi de amacını yerine getirmiş olmayacaktır.

Bilişsel tanı modelleri ile ilişkili, daha geniş bir model tanısal uyarılık değerleri tartışması için okuyucu Sinharay'a (2005) sevk edilmektedir. Makalesinde Sinharay, gerçek bir veri analizi bağlamında pek çok tanı modeli yaklaşımlarını açıklar ve model parametrelerinin zayıf model uyumu ve katı tanımlanamazlığını belirler. Sinharay'ın makalesi bir Bayes ağı yaklaşımı kullanılan gerçek bir beceri tanı verisi içerirse de, açıkladığı model uyum yaklaşımları Fusion Model de dâhil olmak üzere herhangi bir parametrik beceri tanılama modeline açık bir şekilde uygulanabilir.

#### 2.5.2.4. İç Geçerlik Kontrolü

Genel olarak geçerliğin geniş kavramları model değerlendirmelerinin ve tahminlerinin gözlenen veri ile karşılaştırıldığı ölçüt geçerliğinin de ötesine gider. Örneğin, bir değerlendirme aracı için yapı geçerliği aracın yapısının ölçümlerinin, ilişkili yapıların daha geniş nomotetik yapılarda başka yapılarla ilişkisi olmasının beklenmesine kadar değinir ve sonuç, geçerliği değerlendirme bilgisinin işlenmesinden çıkan, istenen ve pozitif sonuçlara kadar değinir.

Ne var ki, gözlenen veri tabanlı ölçüt geçerliği tüm test geçerlik sürecinin kritik bir elemanı olarak kalır. Ölçüt geçerliğini iki geniş kategoriye ayırılabilir; iç geçerlik (iç tutarlılık ile eş anlamlı) ve dış geçerlik. İç geçerlik durumunda gözlenen veri, test verilerinin kendisinden gelir. Dış geçerlik istatistikleri testten elde edilmiş beceri kestirimlerini bazı diğer test dışı beceri ölçütlerine bağlar. Fusion Model sistemi iç geçerlik kontrollerini içermektedir ve kullanıcılar, güçlü bir şekilde teste dışardan giren herhangi bir ölçütün, kullanıcı tabanı için en uygun olanına göre dış geçerlik kontrolleri yapma yönünde cesaretlendirilir.

İç geçerlik kontrollerinin bir türü de farklı olarak sınıflandıran adaylar arasındaki gözlenen davranıştaki farkları ölçmektir ve bu fikir, geliştirdiğimiz iç geçerlik kontrolünün temelini sağlar. Özellikle, Fusion Model bu türde iki çeşit istatistik ortaya koyar: Madde özellik sahibi olma istatistiği için IMstats ve aday hâkimiyeti istatistiği için EMstats.

IMstats birim bazında ve tüm maddelerin ortalamaları üzerine dayalı olarak ne kadar iyi olduğunu tanımlar, adayın her bir beceri hâkimiyetinin Arpeggio MCMC

kestirimleri her bir madde üzerinde adayların asıl gözlenen performanslarına denk gelmektedir. Özellikle testteki her bir madde için adayların her birinin iyi öğrenmiş olduğu varsayılan madde tarafından ne kadar belirli becerilere ihtiyaç duyulduğuna göre, adaylar üç gruba ayrılır. Madde *i* tarafından gerekli olan tüm becerileri iyi öğrenmiş adaylar “madde *i* becerisine sahipler” olarak anılır. Ancak madde *i* için gerekli becerilerin en az birini iyi öğrenmemiş adaylar “madde *i* üst beceriye sahip olmayanlar” olarak adlandırılır. Madde *i* için gerekli becerilerin yarısından fazlasını iyi öğrenmemiş adaylar için “madde *i* alt beceriye sahip olmayanlar” adı kullanılır. Beceriye sahipler ve sahip olmayanlar terimlerinin bu kullanımı bireysel becerilerin öğrenimi hususunda önceki kullanımlarından farklıdır. Burada beceriye sahip olma bir madde için gerekli tüm becerilerle ilişkili olarak, beceriye sahip olmama ise tek bir beceriyle ilişkili olarak bahsedilmektedir. O zaman bir birim bazında, IMstats madde için bu üç gruptan birinde yer alan adaylar için bir maddede gözlenen doğru oran puanını hesaplar.

Arpeggio MCMC kestirim yönteminin  $Q$  matrisi ile iyi çalışıp çalışmadığını saptamak için, IMstats sonuçları, madde için beceriye sahip olanlarının sahip olmayanlardan kesin olarak daha iyi performans gösterip göstermedikleri ve benzer bir şekilde, madde üst beceriye sahip olmayanlar madde alt beceriye sahip olmayanlardan çok daha iyi performans gösterip göstermedikleri incelenir. Bu karar almanın gayet subjektif olduğunu düşünülebilir, şöyle ki, bu durum bir başkasının, bahsedilen oranlarda önemli bir fark olmadığını düşünmesine dayalıdır. Bu noktada resmi bir hipotez testi yaklaşımı da önerilemez çünkü bu oranlar büyük bir örneklem üzerinden

hesaplanır ve standart hatalarının çok küçük olması beklenerek bu durum önemsiz küçük farkları istatistiksel olarak anlamlı hale getirir.

EMstats “aday uzmanlık istatistiği”dir ve bu madde için gerekli tüm becerileri öğrendikleri madde üzerinde beklenmedik bir şekilde zayıf performans gösterenlerde ya da tüm gerekli becerileri öğrenmedikleri madde üzerinden beklenmedik bir şekilde iyi performans çıkarıcıları araştırmak için kullanılır. EMstats tüm adaylar üzerinden özet istatistiğin yanı sıra aday bazındaki ölçme istatistiklerini de üretir. Her bir aday her bir beceride öğrenmiş ya da öğrenmemiş olarak değerlendirildikten sonra verilen bir adayın öğrendiği belirli beceriler grubu adayın maddelerini üç gruba ayırmada temel sağlar. Özellikle seçilen kişinin iyi öğrendiği varsayılan her bir madde becerileri oranı tabanlı olarak her bir aday  $j$  için testteki maddeler üçe ayrılır:

Aday  $j$ 'nin beceriye sahip olduğu maddeler: Her bir madde için gerekli tüm  $Q$  matrisi becerilerini iyi öğrenmiş olarak varsayılan aday  $j$  için bir madde grubu.

Aday  $j$ 'nin üst beceriye sahip olmadığı maddeler: Adayın  $Q$  matrisine göre maddenin gerektirdiği becerilerin en az birini, ancak yarısından fazlasını değil, öğrenmediği tahmin edilen aday için tüm maddeler bir grubu.

Aday  $j$ 'nin alt beceriye sahip olmadığı maddeler:  $Q$  matrisine göre maddenin gerektirdiği becerilerden yarısından fazlasını öğrenmemiş aday için tüm maddeler grubu.

Son iki kategori bu aday için öğrenilmemiş tüm maddeler adına dördüncü bir genel kategori oluşturmak için birleştirilir. Her bir aday için EMstats, o adayın beceriye sahip

olduđu, üst beceriye sahip olmadığı ve alt beceriye sahip olmadığı maddelerin sayısını ve adayın her bir madde grubundaki doğru oran puanını hesaplar.

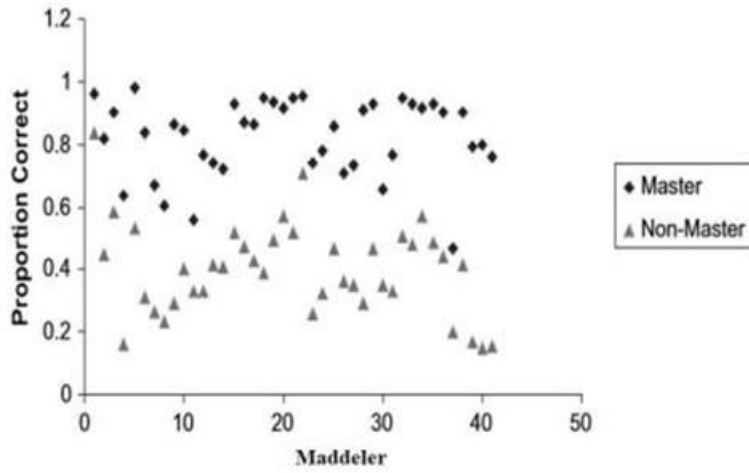
EMstats, puanın olağandışı bir şekilde düşük olup olmadığını saptamak için bir değer kriteri ile birlikte adayın uzmanlaştığı maddeleri için adayların gözlenen doğru oran puanlarını karşılaştırır. Benzer olarak, üst beceriye sahip olunmayan maddeler, alt beceriye sahip olunmayan maddeler ve tüm beceriye sahip olunmayan maddelerin hepsi üzerinden puanlar, olağandışı bir şekilde bu puanların yüksek olup olmadığını saptama kriteriyle karşılaştırılır. Çünkü bu gözlenen doğru oran puanları oldukça az sayıda maddelere bağlıdır ve karar vermede bir hipotez testi yaklaşımı, yüksek tip 1 hata oranından kaçınmak için gereklidir. Böylece her bir aday için, EMstats her madde grubu için tek uçlu bir hipotez uygular. Hipotez testleri tek uçludur çünkü beceriye sahip olunan maddeler için yalnızca adayın özel olarak bir düşük doğru oran puanı alıp almadığını önemsenir ve herhangi bir beceriye sahip olunmayan maddeler grubu içinse adayın özellikle yüksek doğru oran puanı alıp almadığına dikkat edilir. Kriter değerleri kullanıcı tarafından belirlenir ya da IMstats'dan uzmanlar, üst beceriye sahip olmayanlar ve alt beceriye sahip olunmayanlar için gözlenen madde doğru oran puanlarınca belirlenir. Kriter değerleri bu çalışmada kullanılan sonraki durumdan temel alır.

Hipotez testleri şu şekilde yürütülmektedir: İlk olarak adayın gözlenen doğru oran puanı uygun ölçüt değerinden çıkarılır. Sonra, sıfır hipotezinin sahip olduğu varsayım, standart hata hesaplama, n tane ilgili madde üzerinde,  $P_{ci}$  madde i için ölçüt değeri  $\sqrt{P_{ci}(1 - P_{ci})} \div n$  olarak hesaplanır. Daha sonra, basit bir z istatistiği standart hata ile bölünerek şekillendirilir. Bu z istatistiğinin yaklaşık olarak normal dağıldığı

varsayılarak, hipotez testinin odağı beceriye sahip olunan maddelerdeyken hesaplanan z istatistiğinin -1,645'ten az olduğu ya da odak beceriye sahip olunmayan maddelerin üzerindeyken 1,645'ten fazla olduğu saptanarak hipotez testi gerçekleştirilir.

Gerçek bir veri analizinde IMstats iç geçerlilik kontrolünün bir kullanım örneği olarak, Şekil 3, Jang'ın analiz ettiği test formlarından biri için test sonuçlarını ortaya koymaktadır. Bu sonuçlar açık bir şekilde beceri tanılama için iç geçerliğin yüksek bir derecesini işaret etmektedir çünkü madde için beceriye sahip olanlar ve beceriye sahip olmayanlar arasındaki ortalama değer farkı, maddelerin büyük çoğunluğu için oldukça geniştir. Sonuçlar ayrıca, Jang'ın incelediği kesin problematik maddeleri açıkça gösterir. Jang, bu maddelerin aşırı derecede kolay ya da aşırı derecede zor olma eğilimlerini keşfetmiştir. Jang tarafından belirtilen odur ki kullandığı test, adayları sürekli bir ölçeğe koyma amacı ve böylece çok çeşitli zorlukta maddeler elde ederek orijinal olarak norm referanslı bir test olarak kullanılmıştır. Jang oldukça haklı bir şekilde başka bir amaç için geliştirilen böylesi bir testin, beceri tanılama için iyi bir test olmadığından ve iç geçerlik analizi boyunca bu problematik maddelerin belirlenmesinin, hedefi bu noktaya getireceğinden söz etmektedir.





Şekil 3. Madde özelliğine sahip olanların ve madde özelliğine sahip olmayanların arasındaki performans farkını karşılaştırarak yapılan iç geçerlik kontrolü (Jang, 2005; Roussos ve diğ., 2010).

#### 2.5.2.5. Güvenilirlik Tahmini

Sürekli bir tek boyutlu örtük özellekle modellenen değerlendirmeler için tahmin edilmiş standart güvenilirlik katsayıları, bilişsel tanı testleri ile modellenmiş ayrık örtük uzama doğrudan dönüşmez. Ne var ki, ilk prensiplerden güvenilirlik kavramları yine de uygulanır. Tanısal nitelik sınıflandırması güvenilirliği bir tarafta, çıkarılan ve gerçek beceri durumu arasındaki uygunluk ve diğer tarafta, eğer aynı değerlendirme aynı adaya pek çok kez uygulandıysa sınıflandırmanın tutarlılığı olan iki fikirle kavramsallaştırılabilir. Fusion Model sisteminde güvenilirlik kestirimi için metot olarak faaliyete geçirileni, tutarlılık kavramıdır. Özellikle, sınıflandırma güvenilirliğini kestirmek için, ölçülü model simüle edilen paralel veri gruplarının oluşumu için ve sonra her bir gruptaki her bir simüle edilen aday için beceriye sahip olma/olmama tahmininde kullanılır. Simülasyon daha önce model uyarlık değerleri için anlatılan yolla aynı şekilde yürütülür. Her bir beceri için, her adayın testte doğru olarak

sınıflandırıldığı (doğru sınıflandırma oranı tahmini) zamanların oranı ve aynı iki paralel testte (tahmini sınama-yeniden sınama tutarlılık oranı) her adayın aynı sınıflandırıldığı zamanların oranı hesaplanır. Bu oranlar her becerinin gerçek beceriye sahip olanlar ve olmayanlar için ayrı olarak kayda geçilir ve ayrıca beceriye sahip olanlar ve olmayanlar üzerinden ortalaması alınır. Verilen bir beceri için beceriye sahip olanlar/olmayanlar tahmini, her aday için öğrenimin sonsal olasılığı (ppm - posterior probability of mastery) hesaplanarak ve eğer  $ppm > 0,5$  ise adayı uzman ilan edilerek tamamlanır. Bir farksızlık alanı kullanma seçeneği, ayrıca sağlanır; aday, eğer  $ppm > 0,6$  ise beceriye sahip,  $ppm < 0,4$  ise beceriye sahip değil olarak belirlenir ve bunun dışında bilgi verilmez. Bir farksızlık alanı kullanımı, bir grup aday için belirsiz beceriye sahip olma bildirmek pahasına güvenilirliği artırır.

## **2.6. Puan Raporlama İstatistikleri**

Fusion Model sistemindeki son öge, örneğin eğitsel bir tanı değerlendirme bağlamında öğrencilere, öğretmenler ve ebeveynlere yönlendirilecek beceri düzeyi puan raporlama için faydalı olan istatistiklerin üretimidir.

Fusion Model sistemi kullanan bilişsel tanı analizi, öğrencilere ve öğretmenlere, öğretimi ve öğrenmeyi geliştirmek adına kullanılacak detaylı beceri düzeyi öğrenci değerlendirmeleri sağlayabilir. Verilen bir testin yeterli bir kalibrasyonu tamamlandığı vakit tanı puanları, madde ve grup parametreleri ayarlanan değerlere sabitlenmiş MCMC yöntemi ya da Fast Classifier (Arpeggio) olarak adlandırılan Bayes sonsal puanlama programı kullanılarak sınıfta ya da Web aracılığı ile elde edilebilir. İkincisi tercih edilen metottur, çünkü matematiksel olarak doğru azami benzerliğin ya da bir

uzmanlığın sonsal olasılıklarının bir azami soncul kestirimlerini hesaplar ve sayısal olarak daha etkilidir.

### 2.6.1. ppm ve $\eta$

İşlemlenmiş istatistiklerin basit bir grubu, beceri kestirimleridir:  $\alpha(\alpha_k, k = 1, \dots, K)$ 'nın kestirimi ve eğer  $c_i$  parametreleri modele dahil edildiyse  $\eta$  kestirimidir.

Daha önce bahsedildiği üzere,  $\alpha_k$ 'ların kestirimi ppm'ler formundadır. Her beceri için MCMC zincirinin ısıdırma sonrası kısmından doğrudan gözlenmektedirler. MCMC kestirimi boyunca zincirdeki her bir adımda  $\alpha$ , bir aday için 1'lerin ve 0'ların bir şartı olarak öngörülür. Zincir sonsal dağılıma yakınsandığında ve pek çok adım için kullanıldığında, belirli bir beceri için zincirdeki 1'lerin oranı, ppm'dir, o aday için beceri adına uzmanlığın sonsal olasılığıdır. Eğer ppm kullanıcı şartlı üst sınırın üstündeyse (0,5 ya da daha fazla) bir aday, beceriye sahip olarak belirlenir; eğer ppm kullanıcı şartlı alt sınırın altındaysa (0,5 ya da daha az) aday, beceriye sahip değil olarak belirlenir ve eğer ppm iki aralığın arasında ise "belirsiz" olarak tanımlanır (Landis ve Koch, 1977).

Bir adayın ppm'i verilen sebeplerin biri ya da ikisinden dolayı belirsiz alana düşebilir: (a) belirli bir beceri için maddeler özelliğe sahip olmayı, olmamadan ayıracak kadar iyi değildir ve/veya (b) aday ne özelliğe sahip olan ne de olmayan gibi davranmanın ötesinde, daha çok bu iki durumun ortasında bir yerde davranarak tutarsızlık sergilemiştir. Eğer durum (a) ise, sonuçlar belirsiz alandaki aday sayısının beklenenden daha fazla olması ile sonuçlanabilir. Önceki güvenilirlik hesaplaması belirli bir beceri için (a) durumunun gerçekleşip gerçekleşmediğini saptayabilir.

$\eta_j$  ile ilgili olarak, Fusion Model sistemi MCMC tahmini sonsal dağılımı tabanlı bir Bayes EAP kestirimi sağlar. Genel olarak,  $\eta_j$  için kestirim Q matrisinde belirtilmiş becerilere doğrudan bağlı değildir ve sonuç olarak tanılama amaçları için kullanılmaz.

### **2.6.2. Puanları Beceri Profillerine İlişkilendirme: Yeterlilik Ölçekleme**

ppm'ler, beceri düzeyi tanılama bilgisini, sınananlara ve test kullanıcılarına (örneğin; öğrenciler, ebeveynler, öğretmenler, yöneticiler) iletmek için ana araçtır. Ancak, sınananlar ve test kullanıcıları için ppm'lerde yüksek güvenilirlik önemlidir, bu durum beceriye sahip olma ve testte gözlenen puanlar arasındaki ilişkiyi aydınlatan istatistiklere sahip olduğundan önemlidir.

Buraya kadar Fusion Model sistemi tüm olası test puanları ve tüm olası beceriye sahip olma modelleri arasındaki ilişkiyi kestirir ( $\alpha$ 'nın tüm olası değerleri, K becerileri için ikili özelliğe sahip olma parametrelerinin vektörüdür.). Kestirim modeli, uygun modelden 100.000 adayın madde yanıtlarının simülasyonundan yola çıkar. Bu sonuçları daha yorumlanabilir hale getirmek için grup özet istatistikleri üretilmiştir.

İstatistiklerin ilk grubu (a) verilen her bir sabit puan (ve puanın simüle edilmiş oranındaki ilgili olasılıklar) için en olası üç beceriye sahip olma modeli; (b) her bir sabit puanda olasılıkları en %50, %75 ve %90'a toplanmış, beceri özelliğine sahip olma modelinin minimum sayısının bir hesabı ve (c) her sabit puanda her beceri için tasarlanan ya da beklenen ppm'i işleyerek beceri özelliğine sahip olma modelleri verili bir test puanı dağılımını özetler.

İstatistiklerin ikinci grubu, her bir olası beceri özelliğine sahip olma modeli için test puanlarının dağılımını özetler. Fusion Model sistemi her bir beceri özelliğine sahip

olma modeli için (a) beklenen puan; (b) puanın standart sapması ve (c) en muhtemel üç puan, her birinin karşılıklı olasılığı ve bu üç olasılığın toplamını bildirir.

İyi hazırlanmış testler ve dikkatli uygulanmış beceri tanılama için bu istatistikler kullanıcılara daha çok güven verecektir çünkü beceriye sahip olma sayısı yükseldikçe beklenen bir test puanının da genellikle yükseldiğini görülecektir. Dahası, bu koşullara bağlı bilgiyi bilerek, hızlı iptidai beceriye sahip olma kestirimleri, bir adayın test puanlarından çıkarılabilir.

Böyle bilgiler ikincil değerli sınav durumlarında, öğretme ve öğrenme çevresini geliştirme yaklaşımlarında öğretmenlere ek bilgi sağlamada geniş ölçüde faydalı olabilir. Test puanları ile beceriye sahip olma modelleri arasındaki ilişkiyi bilerek öğretmenler her adayın beceriye sahip olma modelini çabucak değerlendirebilir ve adayın performansını yükseltmek için uygun bir odağın ne olabileceğini saptayabilir. Ancak, test puanı dağılımı birleşiminin bu tip özetleri ve beceriye sahip olma modeli bu kısımda bahsedilenlerle sınırlı değildir. Henüz bahsedilmeyen diğerleri de belirli amaçlar için kullanılabilir.

### **2.6.3. Beceriye Sahip Olma Kestirimleri İçin Alt Puanların Kullanımı**

Bazı test bağlamlarında, uygulayıcılar, Arpeggio MCMC puanlama motorundan ya da Fast Classier'dan gelen ppm'ler için alternatifte ihtiyaç duyabilirler. Örneğin bir durum ölçme programı, kullanıcılara aktarılacak en kolay yorumlanabilir bilgi olan testin kopma değerleri üzerinden özelliğe sahip olmayı zorlayabilir. Bu kısımda bilişsel beceri tanılaması için iyi tasarlanmış bir teste Fusion Model analizi uygulanan durumlardaki alt puanların mantıklı kullanımını tartışılacaktır. Pek çok ticari olarak

üretilen test, son zamanlarda alt puanları herhangi uygun bir psikometri kaynağı olmaksızın, “tanısal” puan bilgisi olarak ortaya koymaktadır. Bu uygulamaların aksine, burada bir Fusion Model analizi temelli basitleştirilmiş alt puanlama nasıl yapılmaktadır ve hatasızlıkta ideal Fusion Model tabanlı ppm’ler kullanımına ne kadar yakın olduğunun gösterimi ortaya koyulacaktır.

Bazı bağlamlarda her maddenin kendine kodlanan yalnızca bir becerinin bulunduğu şekilde becerilerin maddelere kodlanması vardır. Böyle Q matrisleri için “basit yapı” terimi kullanılmaktadır. Her beceri için bir alt puan ortaya koymak için bir olası metot da doğru oran ya da yalnızca o beceriye kodlanmış maddeler üzerinden toplam puanları hesaplamaktır. Eğer verilen bir testte Q matrisi çok becerili maddeler içeriyorsa tam olarak aynı yolla her beceri için hala bir alt puan çıkarılabilir. Alt puanları hesaplamak için bir yöntem olduğunda, özelliğe sahip olma/olmama sınıflandırma puanlarını sağlamak için bir ögeye daha ihtiyaç vardır. Doğru oran alt puanında bir kesme noktası alt puanı bir özelliğe sahip olma/olmama tayinine çevirebilmek için yeterlidir. İyi bir kesme noktası seçimi tanısal profillerin çıkarılması için iyi bir alt puan metodunun iyi tanısal performansı adına önemlidir ve standartlar bağlamı ile de oldukça ilişkilidir.

Bu kısımda kısaca ağırlıklı alt puan ve mantıklı model güdümlü kesme noktalarını uygun olarak hesaplamak için bir Fusion Model parametre ölçüleme sonuçlarını kullanan bir Fusion Model tabanlı alt puanlama yaklaşımının çeşitleri tartışılacaktır. Özelliğe sahip olmanın göstergesi olarak kullanıcı seçimli kesme noktaları türündekilerin alt puanları tanımlanarak uygun Fusion Model özelliğe sahip olma/olmamaya nazaran uygun performans nasıl ulaşılabileceği ileri sürülecektir. Kesme noktaları konusunda kötü bir seçim ya da ciddi olarak standart altı alt puanlar, büyük

oranda standart altı beceriler sınıflandırması doğruluğunu muhtemelen ortaya koyacaktır.

Basit olmayan yapı Q matrisi durumunda, basit olmayan yapı Q matrisini bir standart altı ama daha mantıklı bir basit yapı Q matrisini dönüştürmek için Fusion Model bilgisi kullanılarak metotlar araştırılmaktadır. Şöyle ki, çoklu beceriye sahip her madde o beceri  $k$ 'ya, maddeye Q tarafından atanan becerilerin hepsi içerisinde  $r_{ik}^*$  en küçüğü olarak hesaplandığı için atanır. Sonra o madde k. beceri için yalnızca alt puanlama yapmak adına katkıda bulunacaktır. Bu basit yapı alt puanlaması olarak adlandırılır. Bir alternatif ve mantıklı yaklaşım ise orijinal Q matrisini değiştirmeden bırakmak ve çoklu beceriler için kodlanmış maddelere çoklu alt puanlar katılmasına olanak vermektir.

Önemlidir ki eğer orijinal Q matrisi çoklu becerilere sahip maddeler içeriyorsa o zaman her iki alt puanlama metodu, Fusion Model puanlamasına göre standart altı kalacaktır. Her maddenin yalnızca bir beceriye atandığı ilk alt puan metodunda, bir madde yalnızca kendine asli olarak atanmış becerilerden birine “kredi” (doğru bir yanıt) ya da “kınama” (doğru olmayan bir yanıt) ekler ve maddenin diğer becerileri için kredi ve kınama, onların alt puanlarından eksiktir. Orijinal basit olmayan yapı Q matrisinin sürdüğü ve her alt puanın ona kodlanan maddeleri topladığı ikinci alt puan metodunda, doğru yanıt verilmeyen bir madde, bu sefer kınamayı o maddeye atanmış tüm beceriler için alt puanlara ekler. Ancak tüm gerekli becerilerin başarılı bir uygulamasının doğru bir madde performansı için gerekli olduğu Fusion model gibi birleşik bir bağlamda, doğru olmayan madde yanıtı için sebep yalnızca becerilerden birindeki özelliğe sahip olma eksikliği olabilir. İkinci alt puan metodu, bir çoklu beceri

maddesine verilen doğru olmayan bir yanıt yalnızca bir beceri için özelliğe sahip olmamadan kaynaklanması sebebinin kapmak için mekanizmaya sahip değildir.

Kestirilen madde parametrelerinde yer alan yararlı beceri ayrımı bilgisinden faydalanan ağırlıklandırmayı kullanmaktır. Şöyle ki, bazı maddeler diğerleri ile karşılaştırıldığında özelliğe sahip olmaya karşı sahip olmamanın daha iyi göstergeleridir. Bu yüzden bazen madde ağırlıklarının soruda her maddenin beceriyi nasıl ölçtüğünün miktarını belirlemesi gerektiği ağırlıklı alt puanları kullanma önerilir.

Öngörüyle mantıklı bir ağırlık seçimi  $\delta_{ik}$ 'dir. Ağırlık  $\delta_{ik}$ , 0'dan 1'e kadar uzanır ve beceri  $k$ 'nın madde yanıt fonksiyonunu ne kadar şiddetle etkilediğinin bir simgesi olarak yorumlanabilir (Roussos ve diğ., 2007). Bu ağırlıkların rapor edilmesi, alt puanlar yerine özelliğe sahip olma/olmama kararı için ppm'ler kullanılsa bile faydalıdır. Adaylar verilen bir beceride özelliğe sahip olma/olmamayı işaret eden bir rapor aldığı ve hangi maddelerin beceriyi ölçtüğü söylendiğinde, doğal olarak bu maddelerdeki performanslarını inceleyeceklerdir ve tipik olarak hem doğru hem de doğru olmayan cevapları fark edeceklerdir. İndeks, örneğin en yüksekte en düşüğe  $\delta_{ik}$  sırasıyla incelemek üzere adaylar için maddeleri listelemek gibi daha iyi kullanılabilir ve madde puanlarını daha kolay yorumlayabilir hale getirilebilir (Jang, 2005).

Bu ağırlıklandırma şemasını kullanarak, beceri  $k$  özelliğine sahip olma/olmama ikiliğine özellikle hassas olan maddelerin ağırlıklı toplamını tanımlamak mümkündür.  $k$  becerinin en etkili olduğu bu çeşit maddeler en büyük ağırlıklar sahip olacaktır. Ağırlıklı karmaşık toplam puanı  $W_{jk}$ ,  $k$  ( $q_{ik} = 1$ ), için gerekli maddelerin ağırlıklı bir toplamı olarak şu şekilde tanımlanır:



#### Denklem 6. Ağırlıklı Karmaşık Toplam Puanı

$$W_{jk} = \sum_{i=1}^I \delta_{ik}(q_{ik}x_{ij}).$$

Basit Q tabanlı alt puanların aksine, ağırlıklı Q tabanlı alt puanlar, büyük  $r_{ik}^*$  değerleriyle ilişkili küçük  $r_{ik}^*$  değerleri tarafından ilgilenir. Bu yüzden sınıflandırmanın ağırlıklı (Q tabanlı) alt puanlar metodu, sıradan (ağırlıksız) Q tabanlı alt puanlar ve bunu kanıtlayan simülasyon çalışmalarına nazaran daha yüksek doğru sınıflandırma oranları verir. Bu simülasyon sonuçlarından hareketle, pek çok bağlam adına ağırlıklı alt puanların performansının kabaca tam bir Fusion Model sistemi tarafından ortaya konan sınıflandırma sonuçları kadar doğru olması beklenir.

Herhangi bir ağırlıklı ya da ağırlıksız alt puanlama yaklaşımı kullanılırken bir sonraki kritik nokta, bir adayı özelliğe sahip olarak sınıflayan kesme noktasını belirlemektir (Roussos ve diğ., 2007). Eğer bir alt puan için kesme noktası Fusion Model parametrelerini göz ardı edecek şekilde seçilirse, doğru sınıflandırma oranı idealden büyük ölçüde düşük olacaktır. Buna ek olarak, Roussos ve diğ. (2007), Henson, Stout, ve Templin (2005) simülasyon çalışmaları aracılığıyla şu ortaya koyulmaktadır, kesme noktasını saptamak için bir Monte Carlo kestirim yaklaşımı ve bahsi geçen ağırlıklı alt puanları kullanarak Fusion Model sisteminin sınıflandırma performansını idare edecek kadar yakın tahmin edilebilir.

Alt puan tabanlı özelliğe sahip olma sınıflandırma kullanımı için bir kesme noktası belirlerken kullanılan simülasyon metodu şu şekilde gerçekleşir. Fusion Model, madde parametre kestirimlerini ve adayın beceri alanının ek dağılımının kestirimlerini elde

etmek için kullanılır. Sonra, 100.000 simüle edilmiş aday ve onların simüle edilmiş madde yanıtları oluşturulur. Basit bir araştırma kullanarak, ağırlıklı alt puan için uygun bir kesme noktası, simüle edilmiş veri grubunda olduğu gibi her beceri için maksimum doğru sınıflandırma oranları saptanır. Bu mümkün olan bir durumdur çünkü simüle edilen adayların her biri için doğru beceri örneği bilinmektedir.

Genel olarak, eğer tam bir Fusion Model analizi tanılama için etkili bir beceri düzeyi testi tasarlamaya yardım etmek için kullanılsa ve sonra test ayarlansa bile bazı bağlamlarda tam bir Fusion Model sistemi tabanlı sınıflandırma yapmak mantıksız olabilir ya da istenmez ya da uygulanamaz görülebilir. Böyle durumlarda, Fusion Model tabanlı ağırlıklı alt puanlama yaklaşımı etkili işe yarar, sezgisel olarak mantıklı ve pahalı olmayan model güdümlü bir alternatif sağlayabilir. Bu durumlarda, Fusion Model beceri yapısı kestirimi için doğrudan kullanılsa da, ayarlanmış Fusion Modeli ilgilendiren geçerli bilgi ağırlıklı alt puanlar ve “ideal” kesme noktası için kullanılan ağırlıkları saptamada kullanılır (Roussos ve diğ., 2007). Henson, Stout ve Templin (2005) simülasyon çalışmaları aracılığı ile ortaya koymuştur ki, daha önce tanımlanan ağırlıklı alt puanlama Fusion Model tabanlı metodun mantıklı bir yaklaşım ortaya koyduğu durumlarda, Fusion Model sistemi alt puanlama yaklaşımlarında daha iyi performans gösterir. Henson, Stout ve Templin, test bağlamı için belki de şart koşulan ağırlıksız alt puanlama yaklaşımları, ağırlıklı alt puanlamayla ilgili olarak açıkça yetersizdir ancak makul bir şekilde hiçbir şey etkili değildir.

## BÖLÜM III

### YÖNTEM

#### 3.1. Araştırmanın Türü

Bu araştırma, Bilişsel Tanı Modellerinden DINA ve Fusion modellerinin madde parametreleri, olasılık dağılımları ve öğrenci örtük sınıf kestirimleri üzerinden incelendiği bir temel araştırmadır.

#### 3.2. Örneklem

Araştırma evrenini İzmir İl Milli Eğitim Müdürlüğüne bağlı 30 ilçesindeki 448 devlet okulu oluşturmaktadır. Bu okullardaki 6. sınıflarda 1822 şube ve toplam 45069 öğrenci yer almaktadır. Güven düzeyi %99 ve güven aralığı  $t=2$  olarak belirlendiğinde 45069 kişilik evren için örneklem büyüklüğü  $n=3809$  olarak belirlenmiştir (Oulte, 2011; Thompson, 2012).

Daha sonraki aşamada örneklem büyüklüğü göz önüne alınarak tabakalı yöntemle tesadüfi küme örneklem yöntemi ile okul ve şubeler belirlenmiştir. Tabakalama yönteminde göz önüne alınan ölçütler; “Okulların 2015-2016 Kasım - Nisan TEOG genel başarı ortalamaları”, “Okulların 2015-2016 Kasım - Nisan TEOG matematik ortalamaları”, “Okulların 6. Sınıflarda şubeye düşen öğrenci sayıları”, “İlçelere düşen okul, şube ve öğrenci sayıları” şeklindedir. Bu ölçütler göz önüne alınarak tesadüfi yöntemle 10 ilçeden 20 okul araştırma örneklemini için belirlenmiştir. Araştırma örneklemini oluşturan ilçeler, şube ve öğrenci sayıları aşağıdaki Tablo 3’te verilmiştir.

Tablo 3. Araştırma Örneklemine Oluşturan İlçeler ve Öğrenci Sayıları

İlçe	Şube Sayısı	Öğrenci Sayısı
Aliağa	4	107
Bayraklı	10	282
Bornova	26	774
Buca	26	939
Gazimir	16	400
Çiğli	4	128
Karabağlar	12	406
Karşıyaka	14	472
Konak	32	955
Balçova	4	129
Bornova	7	249
Toplam	148	4592

Tablo 3'te görüldüğü gibi İzmir İli içerisinde okullardan araştırma örneklemini için toplam 148 şube ve 4592 öğrenci dâhil edilmiştir. Örneklem büyüklüğünün artmasının sebebi okullarda bulunan 6. sınıf şubeleri arasında ayırım yapılmamasıdır. Böylelikle hem örneklemin evreni temsil etme gücü artmış hem de okullarda bazı şubelerin mahrum kalması ve okul içi eşitliğin bozulması durumu önüne geçilmiştir.

Örneklemin dışında evren parametreleri ortalamalarına en yakın düzeyde bulunan ve örneklem büyüklüğü açısından yeterli görünen bir ortaokul, pilot uygulama okulu olarak belirlenmiştir. Belirlenen pilot okul, hazırlanan her testi önceden alarak tasarım ve uygulama hata ve eksikliklerinin ortaya çıkartılmasında yardımcı olmuştur.

Yapılan bir ön test sonuçlarına göre araştırma örneklemini oluşturan 20 okul, öğrenci ve şube sayılarına göre deney ve kontrol grupları olarak belirlenmiştir. Bu

aşamada öğrencilerin ön test puan ortalamalarına göre okul başarı sıralamaları oluşturulmuş ve buna göre Deney-1, Deney-2 ve Kontrol grupları belirlenmiştir. Grup oluşturma aşamasında okullar küme örneklem olarak alınmıştır. Bunun öncelikli sebebi aynı okulda yer alan farklı şubelere farklı uygulama yapılmamasıdır. Araştırma örnekleminde yer alan 20 okul, belirtilen ölçütler göz önüne alınarak Tablo 4'te şube ve öğrenci sayıları ile ifade edilerek gruplara ayrılmıştır.

Tablo 4. Araştırma Örneklemini Deney ve Kontrol Grupları

Grup	Okul Kodu	İlçe	Şube	N
Deney-1	28	Balçova	4	129
	3	Bornova	5	182
	18	Gazimir	6	160
	5	Buca	4	182
	6	Gazimir	10	240
	16	Bornova	6	169
	20	Konak	7	186
	22	Bornova	9	233
		Toplam	51	1481
Deney-2	7	Karşıyaka	7	272
	19	Karabağlar	7	214
	27	Buca	14	465
	23	Bornova	6	190
	17	Buca	8	292
		Toplam	42	1433
Kontrol	14	Çiğli	4	128
	10	Konak	18	596
	15	Aliğa	4	107
	21	Bayraklı	10	282

13	Karşıyaka	7	200
24	Karabağlar	5	192
25	Konak	7	173
	Toplam	55	1678

Araştırma örneklemini oluşturan toplam 2914 öğrenciden oluşan Deney-1 ve Deney-2 gruplarına izleme testlerinin uygulanmasına karar verilmiş ve Kontrol grubu bu test uygulamalarının dışında bırakılmıştır.

### 3.3. Veri Toplama Aracı

Bu tezde, proje kapsamında yürütülen, Deney-1 ve Deney-2 grubuna uygulanan 11 maddelik İzleme-1 testi verileri ile araştırmanın Q matrisi kullanılmıştır.

#### 3.3.1. Testin Geliştirilmesi

##### 3.3.1.1. Matematiksel Yeterlik Matrislerinin Belirlenmesi

115K531 numaralı TÜBİTAK projesinde matematiksel yeterliklerin belirlenmesinde PISA standartları temel alınmıştır. Genel olarak 7 temel başlıkta ifade edilen bu yeterlikler; ilişkilendirme, iletişim, matematikleştirme, çoklu temsiller, akıl yürütme ve kanıt gösterme, strateji geliştirme ile sembolik, formel ve teknik dil ve işlem kullanılmasıdır (Oecd, 2013). Bununla birlikte proje araştırmacıları, danışmalar ve alan uzmanı öğretmenler ile birlikte 6. Sınıf düzeyi göz önüne alınarak bu 7 temel yeterlik, 4 özellik altında toplanarak yeniden düzenlenmiştir. Düzenlenen bu yeterlikler ve proje analiz kodları aşağıda Tablo 5’te verilmiştir.

Tablo 5. Matematiksel Yeterlikler ve Kodları

Kod	Yeterlik
M1	İletişim ve İlişkilendirme

M2	Matematikleştirme
M3	Akıl Yürütme ve Strateji Geliştirme
M4	Sembolik ve Teknik Dil Kullanımı

---

İletişim ve İlişkilendirme (M1): İletişim becerisi, matematiksel dili günlük dil ve sembollerle ilişkilendirme, matematiksel düşüncelerin doğruluğunu ve anlamını yorumlama gibi özellikleri; ilişkilendirme becerisi ise matematik kavramlarının kendi aralarında, diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır (Basokcu, 2019).

Matematikleştirme (M2): Bir problemi kesin matematiksel formda ifade etmeye yönelik modelleme, yapısal gösterim, varsayımlarla tanımlama, formülleştirme, kurulan bir yapı ya da modele yönelik matematiksel çıktıların elde edilmesi ve yorumlanması gibi özel eylemleri ifade etmektedir (Basokcu, 2019).

Akıl Yürütme ve Strateji Geliştirme (M3): 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 (Basokcu, 2019).

Sembolik ve Teknik Dil Kullanımı (M4): 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 (Basokcu, 2019).

Yukarıda bahsi geçen matematiksel yeterliklerin kullanıldığı İzleme-1 testinin 11 sorusu için hazırlanan Q matris Tablo 6'daki gibidir.

Tablo 6. Matematiksel Yeterlikler Matrisi

Madde	M1	M2	M3	M4
9796	1	0	0	0
3690	0	1	1	0
1027	0	1	1	1
1025	1	0	0	1
1033	1	0	0	0
1021	1	0	0	0
1032	0	0	1	0
7339	1	0	1	0
6728	1	0	0	1
5158_B	1	0	1	0
5158_C	1	0	1	1

Q matristeki her bir hücre, ait olduğu soruyu doğru cevaplayabilmek için gerekli matematiksel yeterliklere sahip olup olmadığını ifade etmektedir. Eğer birey o özelliğe sahipse “1” değilse “0” şeklinde Q matriste ifade edilmektedir.

### 3.3.1.2. İzleme-1 Testi Pilot Uygulaması

İzleme-1 testi için 231 öğrenci ile gerçekleştirilen pilot uygulamada dönem kazanımlarına yönelik hazırlanan 11 madde kullanılmıştır. Pilot uygulamada kullanılan her bir maddenin kodu için ‘P’ ön eki kullanılmıştır. Uygulanan bu 11 maddelik testin her maddesine ait madde güçlük ve madde ayırt edicilik değerleri Tablo 7’de verilmiştir.



Tablo 7. İzleme-1 Testi Pilot Uygulaması Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü ( $P_j$ )	Madde Ayırt Ediciliği ( $r_{bis}$ )	No	Soru Kodu	Madde Güçlüğü ( $P_j$ )	Madde Ayırt Ediciliği ( $r_{bis}$ )
1	P9796	0,78	0,60	7	P1032	0,39	0,73
2	P3690	0,19	0,15	8	P7339	0,01	0,53
3	P1027	0,23	0,65	9	P6728	0,35	0,71
4	P1025	0,16	0,24	10	P5158_B	0,10	0,81
5	P1033	0,35	0,57	11	P5158_C	0,40	0,68
6	P1021	0,40	0,53		Ortalama	0,31	0,56

Tablo 7'ye bakıldığında İzleme-1 testi pilot uygulaması için kullanılan maddelerin madde güçlük değerlerinin 0,10 ile 0,78 arasında ve madde ayırt edicilik değerlerinin ise 0,15 ile 0,81 arasında değiştiği görülmektedir. Testin ortalama madde güçlük değeri 0,31 ve ortalama madde ayırt edicilik değeri 0,56 olarak hesaplanmıştır.

### 3.3.2. İzleme-1 Testi Gerçek Uygulaması

İzleme-1 testi pilot uygulamasında kullanılan 11 madde, düzenlenip geliştirilerek Deney-1 ve Deney-2 gruplarına uygulanmıştır. İzleme-1 testinin gerçek uygulamasına 2672 öğrenci katılmıştır. Yapılan analizler sonucunda İzleme-1 testine ait betimsel istatistikler Tablo 8'de olduğu gibidir.

Tablo 8. İzleme-1 Testi Betimsel İstatistikleri

Ortalama	3,65
Medyan	3
Standart Sapma	2,07
Varyans	4,28
Çarpıklık	0,626
Basıklık	0,016

Tablo 8’de görüldüğü gibi testin ortalama puanı 3,65 ve medyan değeri 3 olarak hesaplanmıştır. Teste ait standart sapma değerinin 2,07 ve varyans değerinin 4,28 olduğu görülmektedir. Çarpıklık ve basıklık değerleri sırasıyla 0,626 ve 0,016 olarak saptanmıştır. Testin çarpıklık değerinin 1’ den küçük oluşu, normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir.

İzleme-1 testine ait maddelerin madde güçlük ve madde ayırt edicilik değerleri Tablo 9’da gösterilmektedir.

Tablo 9. İzleme-1 Testi Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (P <sub>j</sub> )	Madde Ayırt Ediciliği (r <sub>bis</sub> )	No	Soru Kodu	Madde Güçlüğü (P <sub>j</sub> )	Madde Ayırt Ediciliği (r <sub>bis</sub> )
1	9796	0,41	0,50	7	1032	0,40	0,41
2	3690	0,21	0,22	8	7339	0,01	0,19
3	1027	0,18	0,45	9	6728	0,29	0,38
4	1025	0,27	0,55	10	5158_B	0,38	0,52
5	1033	0,51	0,48	11	5158_C	0,31	0,49
6	1021	0,68	0,45		Ortalama	0,33	0,43

Tablo 9’dan da açıkça görüldüğü gibi İzleme-1 testine ait parametrelerden madde güçlüğü değerleri 0,01 ile 0,68 arasında ve madde ayırt edicilik değerleri de 0,19 ile 0,55 değerleri arasında değişiklik göstermektedir. Bu teste ait ortalama madde güçlük değeri 0,33 ve ortalama madde ayırt edicilik değeri 0,43 olarak hesaplanmıştır. Bu parametre değerlerine bakılarak testin grubun büyük bir çoğunluğuna zor geldiği söylenebilir. Ayrıca üst grup (n=816) için en düşük puan 5 ve alt grup (n=867) için en yüksek puan ise 2’dir.

### 3.4. Verilerin Analizi

Bu arařtırmada elde edilen veriler, belirlenen alt problemlere yanıt verecek řekilde ařađıdaki istatistiksel yntemler kullanılarak analiz edilmiřtir.

1. Birinci alt problem olan ‘‘DINA model parametreleri ve rtk sınıf sonsal kestirimleri nasıldır?’’ sorusunun czmnde arařtırma iin hazırlanan Q matris ile birlikte OxEdit 8 (Jurgen A. Doornik, 2018) programında DINA model 1110 modl kullanılmıřtır.

2. İkinci alt problem olan ‘‘Fusion model parametreleri ve rtk sınıf sonsal kestirimleri nasıldır?’’ sorusunun czmnde arařtırma iin hazırlanan Q matris ile birlikte Arpeggio (Roussos ve diđ., 2007) programı kullanılmıřtır.

3. cnc alt problem olan ‘‘DINA ve Fusion parametreleri deđiřmekte midir?’’ sorunun czmnde DINA model  $g$  ve  $s$  parametreleri toplanıp Fusion model  $p_i^*$  parametresi ile aralarındaki Spearman sıra farkları korelasyon katsayısı hesaplanmıřtır.

4. Drdnc alt problem olan ‘‘DINA ve Fusion sonsal kestirimleri deđiřmekte midir?’’ sorunun czmnde her iki modelin her sınıf iin olan posterior olasılıkları birbirlerinden cıkarılıp mutlak deđerleri alınmıř ve oluřan farkın byklđne bakılmıřtır.

5. Beřinci alt problem olan ‘‘DINA ve Fusion model rtk sınıfları deđiřmekte midir?’’ sorunun czmnde đrencilerin rtk sınıfları DINA ve Fusion model ile belirlenmiř ve sonrasında sınıfların iki farklı modelde deđiřimini grmek iin elde edilen rtk sınıflar IBM SPSS Statistics 23 paket programında satırlarda DINA, stunlarda Fusion model verilerin yer aldıđı ve iki modelin rtk sınıflarının

kesişiminin her iki model ile de aynı sınıfa atanmış öğrenci sayısı anlamına geldiği bir tablo haline getirilmiştir.

6. Altıncı alt problem olan “DINA ve Fusion modelde öğrencilerin özelliklere sahip olma düzeyi değişmekte midir?” sorunun çözümünde iki modelde elde edilen örtük sınıflar her bir öğrenci için 4 ayrı değişkene ayrıştırılmış (örneğin 1101 örtük sınıfında yer alan bir öğrenci için ayrı ayrı 1, 1, 0, 1 şeklinde 4 yeni değişken oluşturulmuştur) ve ayrı ayrı her yetenek için öğrenci örtük sınıflarının iki farklı modelde değişimini görmek için elde edilen örtük sınıflar IBM SPSS Statistics 23 paket programında iki modelin örtük sınıflarının kesişiminin her iki model ile de aynı sınıfa atanmış öğrenci sayısı anlamına geldiği 2x2’lik tablolar haline getirilmiştir (İncelenen değişken için DINA’da yeteneğe sahip olup olmama ile Fusion modelde yeteneğe sahip olup olmama durumu karşılaştırılmıştır.).

## BÖLÜM IV

### BULGULAR

#### 4.1. Birinci Alt Probleme İlişkin Bulgular

*DINA model parametreleri ve örtük sınıf sonsal kestirimleri nasıldır?*

2672 öğrencinin katılımıyla gerçekleştirilen İzleme-1 testi uygulaması verileri için matematiksel yeterliklere göre hazırlanan Q matris kullanılarak BTM analizleri yapılmıştır. İzleme-1 testi için yapılan DINA analizlerine göre elde edilen madde parametreleri Tablo 10'daki gibidir.

Tablo 10. DINA Madde Parametreleri

No	Madde Parametreleri			
	g	SE (g)	s	SE (s)
9796	0,261	0,012	0,299	0,02
3690	0,189	0,009	0,694	0,035
1027	0,106	0,008	0,174	0,054
1025	0,122	0,009	0,329	0,023
1033	0,368	0,013	0,213	0,019
1021	0,561	0,013	0,103	0,014
1032	0,269	0,016	0,382	0,025
7339	0,001	0,001	0,969	0,007
6728	0,227	0,011	0,539	0,022
5158_B	0,256	0,011	0,18	0,023
5158_C	0,211	0,010	0,311	0,025
Ortalama	0,234	0,010	0,381	0,024

Tablo 10'da görüldüğü gibi DINA analizleri sonucuna göre matematiksel yeterlikler için g parametreleri değerleri 0,001 ile 0,561 arasında ve s parametre değerleri 0,103 ile

0,969 arasında deęişkenlik göstermektedir. Maddelerin  $g$  parametreleri ortalaması 0,234 deęerine sahiptir. Maddelerin  $s$  parametreleri ortalamalarının ise 0,381 olarak hesaplandıęı görölmektedir. Her iki parametrenin ortalama deęerlerine bakılarak matematiksel yeterlikler için testin model uyumluluęunun yeterli olduęunu görölmektedir.

Tablo 11. DINA Model ile Belirlenen Örtük Sınıf Posterior Olasılıkları

Sınıf	Olasılık	Sınıf	Olasılık	Sınıf	Olasılık	Sınıf	Olasılık
"0000"	0,1310	"0001"	0,0236	"0110"	0,0512	"1101"	0,0310
"1000"	0,1310	"1100"	0,1310	"0101"	0,0310	"1011"	0,0027
"0100"	0,1310	"1010"	0,0360	"0011"	0,0130	"0111"	0,1112
"0010"	0,0512	"1001"	0,0236	"1110"	0,0005	"1111"	0,1011

Tablo 11'den görölebileceęi üzere bütün örtük sınıflar birbirlerine yakın olasılık daęılımları göstermişlerdir En düşük görölme olasılıęı 1110 örtük sınıfındayken, en yüksek görölme olasılıęı 0000, 1000, 0100 ve 1100 örtük sınıflarındadır.

Tablo 12. DINA Modelde Öğrencilerin Örtük Özelliklere Sahip Olma Yaygınlıkları

	Olasılık
M1	0,4569
M2	0,5880
M3	0,3669
M4	0,3371

Tablo 12'ye göre her yeteneęin örneklem içinde yer alma olasılıęı birbirine çok yakın olmakla birlikte, öğrencilerin sahip olduęu en yaygın örtük özellik M2 yeteneęi (Matematikleştirme) iken en düşük yaygınlık gösteren örtük özellik M4 (Sembolik ve Teknik Dil Kullanımı) olmuştur.

## 4.2. İkinci Alt Probleme İlişkin Bulgular

*Fusion model parametreleri ve örtük sınıf sonsal kestirimleri nasıldır?*

İzleme-1 testi için yapılan Fusion model analizlerine göre elde edilen madde parametreleri Tablo 13'teki gibidir.

Tablo 13. Fusion Madde Parametreleri

Madde	$p_i^*$	$r_1^*$	$r_2^*$	$r_3^*$	$r_4^*$	c
9796	0,64625	0,35453	0	0	0	10
3690	0,22792	0	0,83192	0,92258	0	10
1027	0,47145	0	0,04024	0,48998	0,35811	10
1025	0,617	0,2309	0	0	0,33594	10
1033	0,7258	0,4767	0	0	0	10
1021	0,87102	0,60186	0	0	0	10
1032	0,57638	0	0	0,1779	0	10
7339	0,01893	0,09207	0	0,55817	0	10
6728	0,42267	0,68756	0	0	0,60594	10
5158_B	0,6196	0,45665	0	0,66467	0	10
5158_C	0,54546	0,55072	0	0,80401	0,54119	10

Tablo 13'e göre hiçbir madde için Q matriste tanımlı becerilerde bir eksiklik veya Q matris genelinde eksik bir beceri bulunmamaktadır. Q matriste tanımlanan yeteneği en iyi şekilde ölçen madde 1021 iken, madde 7339 Q matriste tanımlanan yeteneği çok zayıf bir şekilde ölçmektedir. M1 yeteneği 6728 maddesi için en yüksek ayırt edicilikte iken 7339 için en düşük ayırt edicilikte; M2 yeteneği sadece iki maddede yer almasına rağmen bu iki maddede ayırt ediciliği uçlarda olmuş ve 3690'da yüksek ayırt edicilikte; M3 yeteneği madde 3690'da en yüksek ayırt edicilikte ve 1032'de en düşük ayırt

edicilikte; M4 yeteneği 6728’de en yüksek ayırt edicilikte ve 1025’te en düşük ayırt edicilikte gözlenmiştir.

Tablo 14. Fusion Model ile Belirlenen Örtük Sınıf Posterior Olasılıkları

Sınıf	Olasılık	Sınıf	Olasılık	Sınıf	Olasılık	Sınıf	Olasılık
"0000"	0,0864	"0001"	0,0163	"0110"	0,0666	"1101"	0,0399
"1000"	0,0731	"1100"	0,0551	"0101"	0,0194	"1011"	0,0519
"0100"	0,0582	"1010"	0,0696	"0011"	0,0274	"0111"	0,0550
"0010"	0,0791	"1001"	0,0226	"1110"	0,0775	"1111"	0,2021

Tablo 14’ten görülebileceği üzere çoğu örtük sınıf birbirlerine yakın olasılık dağılımları göstermişlerdir En düşük görülme olasılığı 0001 örtük sınıfındayken, en yüksek görülme olasılığı 1111 örtük sınıfındadır.

Tablo 15. Fusion Modelde Öğrencilerin Örtük Özelliklere Sahip Olma Yaygınlıkları

	Olasılık
M1	0,6879
M2	0,5552
M3	0,6180
M4	0,3993

Tablo 15’e göre öğrencilerin sahip olduğu en yaygın örtük özellik M1 yeteneği (İletişim ve İlişkilendirme) iken en düşük yaygınlık gösteren örtük özellik M4 (Sembolik ve Teknik Dil Kullanımı) olmuştur.

### 4.3. Üçüncü Alt Probleme İlişkin Bulgular

*DINA ve Fusion parametreleri değişmekte midir?*

DINA modeldeki parametrelerin ( $g$  ve  $s$ ) ve Fusion modeldeki  $p_i^*$  parametresinin aynı cins değişkenler olmaması sebebi ile bu değişkenlere göre hangi maddelerin her iki



modelde de daha iyi uyum verdiğini görebilmek amacı ile Spearman sıra farkları korelasyon katsayısı elde edilmiştir.

Tablo 16. Fusion Model  $p_i^*$  Parametresi ile DINA  $g$  ve  $s$  Parametrelerinin Toplam Değerinin Karşılaştırılması için Veri Seti

Madde	$p_i^*$	$g+s$
9796	0,64625	0,560
3690	0,22792	0,883
1027	0,47145	0,280
1025	0,617	0,451
1033	0,7258	0,581
1021	0,87102	0,664
1032	0,57638	0,651
7339	0,01893	0,970
6728	0,42267	0,766
5158_B	0,6196	0,436
5158_C	0,54546	0,522

Bu verilere göre her iki modeldeki parametrelerin, maddelerin uyum verip vermediğini belirlemede birbirleriyle orta derecede ilişkileri olduğuna ve DINA ile Fusion model parametrelerinin farklılaşmadığına karar verilmiştir ( $r_s = -0,382$ ).

#### 4.4. Dördüncü Alt Probleme İlişkin Bulgular

*DINA ve Fusion sonsal kestirimleri değişmekte midir?*

DINA ve Fusion model sonsal kestirimlerinin değişip değişmediğini belirleyebilmek için her iki modelin Posterior olasılıkları her sınıf için ayrı ayrı birbirlerinden çıkarılmış ve mutlak değerleri alınarak Tablo 17’de verilmiştir.

Tablo 17. DINA ve Fusion Model Posterior Olasılıkları Mutlak Farkları

Sınıf	Mutlak Fark	Sınıf	Mutlak Fark	Sınıf	Mutlak Fark	Sınıf	Mutlak Fark
"0000"	0,045	"0001"	0,007	"0110"	0,015	"1101"	0,009
"1000"	0,058	"1100"	0,076	"0101"	0,012	"1011"	0,049
"0100"	0,073	"1010"	0,034	"0011"	0,014	"0111"	0,056
"0010"	0,028	"1001"	0,001	"1110"	0,077	"1111"	0,101

Tablo 17’den görüleceği üzere, iki modelin belirlemiş olduğu olasılıklar arasındaki farklar yeterince düşük olmakla birlikte 0,001 ile 0,101 arasında değişmektedir. Buna göre, DINA ve Fusion sonsal kestirimlerinin farklılaşmadığı kararı verilmiştir.

#### 4.5. Beşinci Alt Probleme İlişkin Bulgular

*DINA ve Fusion model örtük sınıfları değişmekte midir?*

Tablo 18. DINA ve Fusion Model ile Atanan Sınıf Üyeliklerine Göre Sınıflardaki Birey Sayıları

	Fusion										
	0000	0010	0100	0110	1000	1010	1100	1101	1110	1111	
	0000	<b>631</b>	106	0	0	0	0	0	0	0	1
	0001	55	0	0	0	0	0	0	13	0	29
	0010	31	<b>239</b>	0	6	2	0	0	0	0	0
	0011	67	0	0	0	1	0	0	0	0	131
	0100	0	0	<b>47</b>	5	0	0	1	0	0	12
Dina	0101	0	0	12	0	0	0	0	78	0	117
	0110	8	0	1	<b>0</b>	0	0	0	0	0	1
	0111	2	5	1	1	0	0	0	0	0	222
	1000	0	0	0	0	<b>165</b>	10	3	0	13	5
	1001	0	0	0	0	18	0	0	4	0	27
	1010	0	0	0	0	19	<b>73</b>	0	0	0	0

1011	0	0	0	0	7	2	0	0	0	77
1100	0	0	0	0	0	0	4	0	5	10
1101	0	0	0	0	0	0	0	9	0	73
1110	0	0	0	0	0	0	0	0	101	23
1111	0	0	0	0	0	0	0	0	0	199

Tablo 18, DINA ve Fusion model ile bir örtük sınıfa atanan bireyler için, iki analizde de aynı örtük sınıfta kalan (koyu renk) öğrenci sayısını vermektedir. Tabloya göre, örneğin, hem DINA hem de Fusion model ile 0000 sınıfına atanan birey sayısı 631, DINA ile 0000 sınıfına atanırken Fusion model ile 0010 sınıfına atanan birey sayısı 106 şeklindedir. Buna göre öğrencilerin %54,94'ü her iki analizle aynı sınıfa atanmıştır.

#### 4.6. Altıncı Alt Probleme İlişkin Bulgular

*DINA ve Fusion modelde öğrencilerin özelliklere sahip olma düzeyi değişmekte midir?*

##### 4.6.1. İletişim ve İlişkilendirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

Tablo 19. İletişim ve İlişkilendirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi

		Fusion	
		0	1
DINA	0	1217	608
	1	0	847

Tablo 19'a göre DINA ve Fusion ile yeteneğe sahip olmadığı belirlenen 1217 öğrenci, yeteneğe sahip olduğu belirlenen 847 öğrenci bulunmaktadır. Bunun yanı sıra, DINA ile 608 öğrenci yeteneğe sahip değil şeklinde sınıflandırılırken aynı öğrenciler

Fusion model ile yeteneğe sahip görülmüştür. Buna göre öğrencilerin %77,25'i her iki analizle aynı sınıfa atanmıştır.

#### 4.6.2. Matematikleştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

Tablo 20. Matematikleştirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi

		Fusion	
		0	1
DINA	0	1426	309
	1	15	922

Tablo 20'ye göre DINA ve Fusion ile yeteneğe sahip olmadığı belirlenen 1426 öğrenci, yeteneğe sahip olduğu belirlenen 922 öğrenci bulunmaktadır. Bunun yanı sıra, DINA ile 309 öğrenci yeteneğe sahip değil şeklinde sınıflandırılırken aynı öğrenciler Fusion model ile yeteneğe sahip görülmüş; DINA ile 15 öğrenci yeteneğe sahip şeklinde sınıflandırılırken aynı öğrenciler Fusion ile yeteneğe sahip değil şeklinde görülmüştür. Buna göre öğrencilerin %87,87'si her iki analizle aynı sınıfa atanmıştır.

#### 4.6.3. Akıl Yürütme ve Strateji Geliştirme Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?

Tablo 21. Akıl Yürütme ve Strateji Geliştirme Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi

		Fusion	
		0	1
DINA	0	1040	413
	1	139	1080

Tablo 21'e göre DINA ve Fusion ile yeteneğe sahip olmadığı belirlenen 1040 öğrenci, yeteneğe sahip olduğu belirlenen 1080 öğrenci bulunmaktadır. Bunun yanı sıra, DINA ile 413 öğrenci yeteneğe sahip değil şeklinde sınıflandırılırken aynı öğrenciler Fusion model ile yeteneğe sahip görülmüş; DINA ile 139 öğrenci yeteneğe sahip şeklinde sınıflandırılırken aynı öğrenciler Fusion ile yeteneğe sahip değil şeklinde görülmüştür. Buna göre öğrencilerin %79,34'ü her iki analizle aynı sınıfa atanmıştır.

#### **4.6.4. Sembolik ve Teknik Dil Kullanımı Yeteneği için DINA ve Fusion modelde öğrencilerin özelliğe sahip olma düzeyi değişmekte midir?**

Tablo 22. Sembolik ve Teknik Dil Kullanımı Yeteneği için Fusion ve DINA ile Yeteneğe Sahip Öğrenci Sayılarının Belirlenmesi

		Fusion	
		0	1
DINA	0	1470	52
	1	171	979

Tablo 22'ye göre DINA ve Fusion ile yeteneğe sahip olmadığı belirlenen 1470 öğrenci, yeteneğe sahip olduğu belirlenen 979 öğrenci bulunmaktadır. Bunun yanı sıra, DINA ile 52 öğrenci yeteneğe sahip değil şeklinde sınıflandırılırken aynı öğrenciler Fusion model ile yeteneğe sahip görülmüş; DINA ile 171 öğrenci yeteneğe sahip şeklinde sınıflandırılırken aynı öğrenciler Fusion ile yeteneğe sahip değil şeklinde görülmüştür. Buna göre öğrencilerin %91,65'i her iki analizle aynı sınıfa atanmıştır.

## BÖLÜM V

### TARTIŞMA

DINA model parametrelerinden  $g$  parametrelerine bakıldığında, her madde için istendik bir şekilde düşük olduğu, ortalamalarının 0,234 olduğu ve en yüksek  $g$  parametresinin 0,561 değeri ile 1021 maddesine ait olduğu gözlenmiştir. 1021 maddesi aynı zamanda 0,68 değeri ile testteki en güç madde olma özelliği taşımaktadır.

DINA  $s$  parametrelerine bakıldığında, çoğu madde için yeterince düşük olduğu, ortalamasının 0,381 olduğu ve en yüksek  $s$  parametrelerinin sırasıyla 7739 ( $s=0,969$ ), 3690 ( $s=0,694$ ) ve 6728 ( $s=0,539$ ) maddelerinde olduğu gözlenmiştir. Bu maddeler ayırt ediciliği en düşük maddeler olma özelliklerine sahip olmakla beraber indeksleri sırasıyla 0,19; 0,22 ve 0,38'dir.

Fusion model  $p_i^*$  parametresine bakıldığında her madde için yeterince yüksek değerlere sahip ve ortalamasının 0,522 olduğu görülmektedir. En yüksek  $p_i^*$  parametresi 0,871 değeri ile 1021 maddesine aitken, en düşük 0,018 değeri ile 7339 maddesine aittir. 1021 maddesi testteki en güç madde iken 7339 maddesi testteki güçlüğü ve ayırt ediciliği en düşük madde olma özelliğine sahiptir.

DINA  $g$  parametresi ile Klasik Test Teorisine ait madde güçlüğü indeksi arasındaki ilişki  $r_s = 0,973$ ;  $P < 0,01$  olarak;  $s$  parametresi ile ayırt edicilik indeksi arasındaki ilişki ise  $r_s = -0,56$ ;  $P > 0,05$  olarak bulunmuştur. Elde edilen korelasyon katsayısı ve maddelerin incelenmesi sonucunda, her iki modelin de kendi içlerinde tutarlı sonuçlar verdiklerine karar verilmiştir.

En yüksek  $g$  parametresine sahip madde, aynı zamanda en yüksek  $p_i^*$  parametresine sahip ve en güç madde olarak gözlenmiş; en düşük  $s$  parametresine sahip madde, aynı zamanda en düşük  $p_i^*$  parametresine ve en kolay madde ile ayırt ediciliği en düşük madde olarak gözlenmiştir. Her iki modeldeki parametrelerin, maddelerin uyum verip vermediğini belirlemede birbirleriyle orta derecede ilişkileri olduğuna ve DINA ile Fusion model parametrelerinin farklılaşmadığına karar verilmiştir,  $r_s = -0,382$ ;  $P > 0,05$ .

Her iki modelin örtük sınıf sonsal kestirimleri kendi içlerinde düzenli bir dağılım göstermekle beraber, DINA’da en yüksek olasılık 0000, 1000, 0100 ve 1100 örtük sınıflarındayken, en düşük olasılık 1110 sınıfında; Fusion’da en yüksek olasılık 1111 örtük sınıfındayken, en düşük olasılık 0001 sınıfında görülmüştür. Her iki modelin sonsal kestirimlerinin mutlak farkı alındığı zaman en büyük farkın 1111 sınıfında olduğu ve bu sınıfın olasılığının Fusion modelde daha yüksek olduğu görülmüştür. Bunun dışında Fusion modelin daha yüksek sonsal olasılık belirlediği sınıflar 0010, 1010, 0110, 0011, 1110, 1101 ve 1011 şeklindedir. DINA’da bir öğrencinin 0000 sınıfına ait olma sonsal olasılığının Fusion’dan yüksek olmasının ve Fusion’da 1111 sınıfına ait olma sonsal olasılığının DINA’dan yüksek olmasının sebebi modellerin farklılıklarından kaynaklanmaktadır. Fusion modelin aksine DINA model daha “katı” bir modeldir ve gereken yeteneklerden birine bile sahip olmayan öğrenciyi yokluk sınıfına atmaktadır.

DINA modelin Fusion’dan daha katı bir model olduğunun rahat görülebilmesi için modellerin öğrencileri atadıkları örtük sınıflardaki örtük sınıf sayısına göre bir karşılaştırma yapılmış ve aşağıda Tablo 23 ve 24’te verilmiştir.

Tablo 23. Fusion Modelde, DINA'ya Göre Daha Fazla Yeteneğe Sahip Bulunan Öğrenci Sayısı

Yetenek Sayısı		Öğrenci Sayısı
DINA	Fusion	
0	1,2,3 ve 4	107
1	2,3 ve 4	97
2	3 ve 4	373
3	4	395
Toplam		972

Tablo 23'e göre, bütün örneklemin %36,38'i Fusion ile daha fazla yeteneğe sahip olarak görülmüştür.

Tablo 24. DINA Modelde, Fusion'a Göre Daha Fazla Yeteneğe Sahip Bulunan Öğrenci Sayısı

Yetenek Sayısı		Öğrenci Sayısı
Fusion	DINA	
0	1,2,3 ve 4	163
1	2,3 ve 4	64
2	3 ve 4	3
3	4	0
Toplam		230

Tablo 24'e göre, bütün örneklemin sadece %8,61'i DINA ile daha fazla yeteneğe sahip olarak görülmüştür.

İletişim ve İlişkilendirme yeteneğinde, Fusion modelde yeteneğe sahip olmayan öğrencilerin hiçbiri DINA ile yeteneğe sahip bulunmamışken DINA'da yeteneğe sahip olmayan 608 öğrenci Fusion'da yeteneğe sahip olarak bulunmuştur. Matematikleştirme yeteneğinde, Fusion modelde yeteneğe sahip olmayan öğrencilerin sadece 15'i DINA



ile yeteneęe sahip bulunmuřken, DINA'da yeteneęe sahip olmayan 309 ğrenci Fusion'da yeteneęe sahip olarak bulunmuřtur. Akıl Yürütme ve Strateji Geliřtirme yeteneęinde, Fusion modelde yeteneęe sahip olmayan ğrencilerin 139'u DINA ile yeteneęe sahip bulunmuřken, DINA'da yeteneęe sahip olmayan 413 ğrenci Fusion'da yeteneęe sahip olarak bulunmuřtur. Sembolik ve Teknik Dil Kullanımı yeteneęinde, Fusion modelde yeteneęe sahip olmayan ğrencilerin 171'i DINA ile yeteneęe sahip bulunmuřken, DINA'da yeteneęe sahip olmayan sadece 52 ğrenci Fusion'da yeteneęe sahip olarak bulunmuřtur. Sembolik ve Teknik Dil Kullanımı yeteneęinde, Fusion modelin DINA modelden daha fazla ğrenciyi yeteneęe sahip sınıfına koymasý beklentisinin aksi durumu oluřmuřtur ve bu yetenekte DINA daha fazla ğrenci için yeteneęe sahip řeklinde kestirim yapmýřtır. Bunun sebebi bu yeteneęi ölçen maddelerde DINA modelin Fusion modele göre daha iyi kestirimler yapmasýdır ( $g_{ort} = 0,167$ ;  $s_{ort} = 0,338$ ;  $r_{ort}^* = 0,42$ ). Bu maddeler için GDINA gibi daha genelleyici modellerin kullanımı sonucu yapılacak ğrenci örtük sınıf kestirimlerinin Fusion modele yaklařacaęı düşünölmektedir.

## BÖLÜM VI

### SONUÇ VE ÖNERİLER

DINA modelin diğer BTM'lerden en önemli farkı, bir madde için gerekli olan özelliklerden sadece birine bile sahip olmayan bireyi yokluk sınıfında görmesidir. Fusion modelde, DINA modeldeki bu katılığın aksine, testteki her bir madde için adaylar üç gruba ayrılır. Madde tarafından gerekli olan tüm becerileri iyi öğrenmiş adaylar “beceriye sahipler” olarak anılır. Ancak madde için gerekli becerilerin en az birini (ve en çok yarısını) iyi öğrenmemiş adaylar “üst beceriye sahip olmayanlar ” olarak adlandırılır. Madde için gerekli becerilerin yarısından fazlasını iyi öğrenmemiş adaylar için “alt beceriye sahip olmayanlar” adı kullanılır.

Bu bakımda Fusion model, DINA gibi bir öğrenciyi hemen yokluk sınıfında görmemektedir. Yapılan araştırma sonucunda da DINA ve Fusion model parametreleri birbirine yakın elde edilmekle birlikte, Fusion model ile yapılan analizler sonucunda bireylere ait daha yüksek sınıf kestirimleri elde edildiği görülmüştür. Dolayısıyla çoklu yetenek ölçümlerinde Fusion modelin DINA modele göre daha isabetli kestirimler yapacağına karar verilmiştir.

Bilişsel Tanı Modelleri geniş örneklemelerde tutarlı sonuçlar veren modellerdir. Bu tezde kullanılan geniş örneklem; madde istatistiklerini, DINA ve Fusion model madde parametrelerini, sonsal olasılıkları ve öğrenci örtük sınıf kestirimlerini etkileyen değişkenlerden biri olmuştur. Buradan hareketle, farklı boyuttaki örneklemelerle ve madde sayısının değiştiği durumlarda bu çalışmadaki problemlerin yeniden incelenmesi söz konusu olabilir.

Bu tez kapsamında öğrencilerin sadece matematik alanındaki bilişsel profilleri üzerinde çalışılabilme imkânı bulunmuştur ve modellerin karşılaştırılmasında sadece matematik alanı ile sınırlı kalmıştır. Bu bağlamda, konuyla ilgilenen araştırmacıların, öğrencilerin farklı yaklaşımlar sergiledikleri diğer alanlarla ilgili çalışmalar yapmaları alanyazına önemli katkı sunabilir.

Çalışma kapsamında DINA ve Fusion Modellerin madde parametreleri, sonsal olasılıklar ve örtük sınıf kestirimleri üzerinden karşılaştırılması yapılmış olup diğer Bilişsel Tanı Modelleri bu karşılaştırmaya dâhil edilmemiştir. İleride yapılacak araştırmalarda farklı modellerin karşılaştırılması alanyazına katkı açısından faydalı olabilir.

Proje kapsamında elde edilen gerçek veri seti kullanılarak yapılan karşılaştırmada, modelleri istatistiksel farklılıklarından ziyade öğrenci hakkında verilen kararlar üzerinden bir çalışma yürütülmüştür. Bilişsel Tanı Modellerinin kendi içlerinde karşılaştırılmasında modellerin istatistiksel yapılarının yanında mantıksal geçerlik çalışmaları yapılan gerçek verilerin kullanılmasının daha sağlıklı olacağı, gerçek veri setlerinde karar verme aşamasında farklı geçerlik yöntemleriyle karşılaştırılmasının alanyazına katkı sağlayacağı düşünülmektedir.

## KAYNAKÇA

- 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.
- Başokçu, T. O. (2014). The Cognitive Diagnostic Models for Estimating Students' Ability and Their Applications. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 1–32.
- Cheng, Y. and Chang, H. (2007). The modified maximum global discrimination index method for cognitive diagnostic computerized adaptive testing. *CAT and Cognitive Structure Paper Session, June, 7*.
- de la Torre, J. (2008). An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications. *Journal of Educational Measurement*, 45(4), 343–362. <https://doi.org/10.1111/j.1745-3984.2008.00069.x>
- de la Torre, J. (2009). A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options. *Applied Psychological Measurement*, 33(3), 163–183. <https://doi.org/10.1177/0146621608320523>
- de la Torre, J. and Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333–353. <https://doi.org/10.1007/BF02295640>
- de la Torre, J. and Douglas, J. (2008). Model Evaluation and Multiple Strategies in Cognitive Diagnosis: An Analysis of Fraction Subtraction Data. *Psychometrika*, 73(4), 595–624. <https://doi.org/10.1007/s11336-008-9063-2>

- de la Torre, J. and Liu, Y. (2008). Cognitive diagnosticity of IRT-constructed assessment: An empirical investigation. In *National Council on Measurement in Education*. New York, NY.
- DiBello, L. V, Stout, W. F. and Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In Paul D Nichols, S. F. Chipman and R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–389). New Jersey, USA: Lawrence Erlbaum Associates.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*.
- Embretson, S. E. (1999). Cognitive psychology applied to testing. In F. T. Durso (Ed.), *Handbook of applied cognition*. New York, NY, US: John Wiley and Sons Ltd.
- Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists*. L. Erlbaum Associates.
- Embretson, S. E. and Reise, S. P. (2013). *Item Response Theory*. Taylor and Francis.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37(6), 359–374. [https://doi.org/10.1016/0001-6918\(73\)90003-6](https://doi.org/10.1016/0001-6918(73)90003-6)
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis* (Third ed.). Taylor and Francis. Retrieved from <https://books.google.com.tr/books?id=ZXL6AQAAQBAJ>
- Gitomer, D., Yamamoto, K. (1991). Performance modelling that integrates latent trait

and class theory. *Journal of Educational Measurement*.

Haertel, E. H. (1984). An application of latent class models o assessment data. *Applied Psychological Measurement*.

Haertel, E. H. (1989). Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items. *Journal of Educational Measurement*, 26(4), 301–321.  
<https://doi.org/10.1111/j.1745-3984.1989.tb00336.x>

Hartz, S. M. (2002). A Bayesian Framework for the Unified Model for Assessing Cognitive Abilities: Blending Theory With Practicality. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 63(2-B), 864.

Hartz, S. M. and Roussos, L. A. (2005). The Fusion Model for skills diagnosis: Blending theory with practice. *ETS Research Report*.

Hartz, S. M., Roussos, L. and Stout, W. (2002). Skills diagnosis: Theory and practice. *User Manual for Arpeggio Software*. ETS.

Henson, R. A., Stout, W. F., Templin, J. and Henson. (2005). *Using cognitive model-based sum-scores to do skills diagnosis*. Princeton, NJ.

Henson, R. A., Templin, J. L. and Porch, F. (2004). Description of the underlying algorithm of the improved Arpeggio. *Unpublished ETS Project Report, Princeton, NJ, USA*.

Henson, R. and Douglas, J. (2005). Test Construction for Cognitive Diagnosis. *Applied Psychological Measurement*, 29(4), 262–277.  
<https://doi.org/10.1177/0146621604272623>

- Jang, E. E. (2005). A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL. University of Illinois at Urbana-Champaign Champaign, IL.
- Junker, B. (2000). Some topics in nonparametric and parametric IRT, with some thoughts about the future.
- Kunina-Habenicht, O., Rupp, A. A., and Wilhelm, O. (2012). The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. *Journal of Educational Measurement*, 49(1), 59–81. <https://doi.org/10.1111/j.1745-3984.2011.00160.x>
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.
- Lei, P., and Li, H. (2016). Performance of Fit Indices in Choosing Correct Cognitive Diagnostic Models and Q-Matrices. *Applied Psychological Measurement*, 40(6), 405–417. <https://doi.org/10.1177/0146621616647954>
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187–212. <https://doi.org/10.1007/bf02294535>
- Mislevy, R. J. (1995). Test theory and language-learning assessment. *Language Testing*, 12(3), 341–369. <https://doi.org/10.1177/026553229501200305>
- Oecd. (2013). PISA 2015 Draft Mathematics Framework. *Oecd*, (March 2013), 52. <https://doi.org/10.1177/0022146512469014>
- Oulte, D. (2011). *Sampling Methods*. GRIN Verlag.

- Patz, R. J., and Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, 24(4), 342–366. <https://doi.org/10.2307/1165367>
- Patz, R. J., Junker, B. W., Johnson, M. S. and Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and Its Application to Large-Scale Educational Assessment Data. *Journal of Educational and Behavioral Statistics*, 27(4), 341–384.
- Ravand, H. and Robitzsch, A. (2018). Cognitive Diagnostic Model of Best Choice: A Study of Reading Comprehension. *Educational Psychology*, 38(10), 1255–1277. <https://doi.org/10.1080/01443410.2018.1489524>
- Roussos, L. A., DiBello, L. V, Henson, R. A., Jang, E. and Templin, J. L. (2010). Skills diagnosis for education and psychology with IRT-based parametric latent class models. In *Measuring psychological constructs: Advances in model-based approaches*. (pp. 35–69). Washington, DC, US: American Psychological Association. <https://doi.org/10.1037/12074-003>
- Roussos, L. A., Stout, W., DiBello, L. V, Hartz, S. M., Templin, J. L. and Henson, R. A. (2007). The Fusion Model Skills Diagnosis System. In J. P. L. M. J. Gierl (Ed.), *Cognitive Diagnostic Assessment for Education Theory and Applications* (pp. 275–318). New York: Cambridge University Press.
- Rupp, A. A. and Mislevy, R. J. (2007). Cognitive psychology as it applies to diagnostic assessment. In J. P. Leighton and M. J. Gierl (Eds.), *Cognitive diagnostic assessment in education: Theory and practice*. New York, NY, US: Cambridge



University Press. <https://doi.org/http://dx.doi.org/10.1017/CBO9780511611186>

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. <https://doi.org/10.1080/15305050701193454>

Samejima, F. (1995). A cognitive diagnosis method using latent trait models: Competency space approach and its relationship with DiBello and Stout's unified cognitive-psychometric diagnosis model. In P. D. Nichols, S. F. Chipman and R. L. Brennan (Eds.), *Cognitively diagnostic assessment*. (pp. 391–410). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Sinharay, S. (2005). Assessing Fit of Unidimensional Item Response Theory Models Using a Bayesian Approach. *Journal of Educational Measurement*, 42(4), 375–394. <https://doi.org/10.1111/j.1745-3984.2005.00021.x>

Sympson, J. B. (1978). A model for testing with multidimensional items. In D. J. Weiss (Ed.), *Computerized Adaptive Testing Conference* (pp. 82–98). Minneapolis.

Tatsuoka, K. K. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, 20(4), 345–354.

Tatsuoka, K. K. (1985). A Probabilistic Model for Diagnosing Misconceptions By The Pattern Classification Approach. *Journal of Educational Statistics*, 10(1), 55–73. <https://doi.org/10.3102/10769986010001055>

- Tatsuoka, K. K. (1995). Architecture of Knowledge Structures and Cognitive. In R. L. B. Paul D. Nichols, Susan F. Chipman (Ed.), *Cognitively Diagnostic Assessment*. L. Erlbaum.
- Thompson, S. K. (2012). *Sampling*. Wiley.
- von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report No. RR-05-16*.
- von Davier, M. and Yamamoto, K. (2004). A class of models for cognitive diagnosis. In *4th Spearman Conference*. Philadelphia, PA.
- Walberg, H. J. and Haertel, G. D. (1990). *The international encyclopedia of educational evaluation*. Pergamon Press.
- Whitely, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4), 479–494.
- Xu, X. and von Davier, M. (2006). Cognitive Diagnosis For NAEP Proficiency Data. *ETS Research Report Series*, 2006(1), i–25. <https://doi.org/10.1002/j.2333-8504.2006.tb02014.x>
- Yan, D., Almond, R. and Mislevy, R. (2004). *A Comprasion of Two Models for cognitive diagnosis*. NJ: ETS.

