



T.C.

EGE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

**BİLİŞSEL TANI MODELLERİNDE DATA/Q MATRİS
GENİŞLETME YÖNTEMİNİN ÖRTÜK SINIF DAĞILIMLARI
ÜZERİNE ETKİSİ**

**SİMGE CEYLAN
YÜKSEK LİSANS TEZİ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
ANABİLİM DALI**

**İzmir
2019**

T.C.

EGE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

**BİLİŞSEL TANI MODELLERİNDE DATA/Q MATRİS
GENİŞLETME YÖNTEMİNİN ÖRTÜK SINIF DAĞILIMLARI
ÜZERİNE ETKİSİ**

**EFFECT OF DATA/Q MATRIX AUGMENTATION ON LATENT
CLASS DISTRIBUTIONS IN COGNITIVE DIAGNOSTIC MODELS**

SİMGE CEYLAN

YÜKSEL LİSANS TEZİ

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
ANABİLİM DALI**

**Tez Danışmanı
Doç. Dr. Tahsin Oğuz BAŞOKÇU**

**2019
İZMİR**



T.C.EGE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



YÜKSEK LİSANS

TEZ SAVUNMA TUTANAĞI

ÖĞRENCİNİN

Adı Soyadı : Simge CEYLAN
Numarası : 9216000062
Anabilim Dalı : Eğitimde Ölçme ve Değerlendirme

Tez Başlığı (Türkçe) : Bilişsel Tanı Modellerinde Yapılandırılmış Yanıtlı Maddeler İçin Örtük Sınıf Değişmezliğinin İncelenmesi: Data/ Q Matris Genişletme Yöntemi

Tez Başlığı (İngilizce) : Investigating The Latent Class Invariance For The Constructed Response In Cognitive Diagnostic Models: Data/Q Matris Augmentation

Tez Savunma Tarihi : 09/09/2019

Tez Başlığı Değişildiği Varsa Yeni Başlık: Bilişsel Tanı Modellerinde Data/Q Matris Genişletme Yönteminin Örtük Sınıf Dağılımları Üzerine Etkisi

Tez Başlığı Yeni (İngilizce) : Effect of Data/Q Matrix Augmentation on Latent Class Distributions in Cognitive Diagnostic Models

JÜRİ ÜYELERİ

Jüri Başkanı

Unvan, Adı, Soyadı : Prof. Dr. Tuncay Öğretmen
Karar : Başarılı Başarısız Düzeltme
İmza :

Jüri Üyesi

Unvan, Adı, Soyadı : : Doç. Dr. T. Oğuz BAŞOKÇU (Danışman)
Karar : Başarılı Başarısız Düzeltme
İmza :

Jüri Üyesi

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

ETİK KURALLARA UYGUNLUK BEYANI

Ege Üniversitesi Sosyal Bilimler Enstitüsü Müdürlüğüne sunduğum “**Bilişsel Tanı Modellerinde Data/Q Matris Genişletme Yönteminin Örtük Sınıf Dağılımları Üzerine Etkisi**” adlı yüksek lisans tezinin tarafımdan bilimsel, ahlak ve normlara uygun bir şekilde hazırlandığını, tezimde yararlandığım kaynakları bibliyografyada ve dipnotlarda gösterdiğimi onurumla doğrularım.

Simge CEYLAN

İmza

TEŞEKKÜR

Yüksek lisans eğitimim boyunca ve bu tezin hazırlanmasında bilgi ve tecrübelerini benden esirgemeyen, bana her zaman sabırla destek olup her aşamada yol gösteren değerli danışmanım Doç.Dr. Tahsin Oğuz BAŞOKÇU'ya

Yüksek lisans eğitimim süresince benim için emek veren ve sahip oldukları bilgi birikimleriyle her zaman öğrencilerine anlam katan Ege Üniversitesi Ölçme ve Değerlendirme Anabilim Dalı değerli hocaları Sayın Prof.Dr. Tuncay ÖĞRETMEN'e ve Sayın Doç.Dr. Hakan ATILGAN'a, bu süreçte birlikte eğitim aldığım, her aşamada manevi desteklerini hissettiğim sevgili dönem arkadaşlarım Yasemin YARDIM PALAMUTLU, Pelin BAĞDU SÖYLER ve Dilay AK'a

Hem çalışıp hem yüksek lisansımı yaparken bana destek olan Özel Manisa ODTÜ Ülkem Koleji idarecilerine ve ODTÜ GVO Matematik Koordinatörü Sayın Serpil ATASOY'A, bu süreçte her zaman manevi desteklerini hissettiğim sevgili matematik zümresi arkadaşlarıma ve diğer mesai arkadaşlarıma

Tüm eğitimim hayatım boyunca eğitimim ve öğretimimde rol oynamış değerli öğretmenlerimin hepsine

Eğitimim için hiçbir zaman fedakarlıktan kaçınmayan, kendimi geliştirme gayretim arkasında duran ve bu süreçte desteklerini esirgemeyen annem Fatma CEYLAN, babam Halil CEYLAN ve biricik kardeşim ATAKAN CEYLAN'a teşekkürlerimi ve saygılarımı sunarım.

ÖZGEÇMİŞ

KİMLİK BİLGİLERİ

Adı Soyadı : Simge CEYLAN

Doğum Yeri : Muğla- Ula

Doğum Tarihi : 23.01.1993

E-posta : simge_cy@hotmail.com

EĞİTİM BİLGİLERİ

Lise : Fethiye Ömer Özyer Anadolu Öğretmen Lisesi, Muğla

Lisans : Akdeniz Üniversitesi İlköğretim Matematik Öğretmenliği

Yabancı Dil ve Düzeyi: 80 (YÖKDİL 2018)

İÇİNDEKİLER

TEŞEKKÜR.....	IV
İÇİNDEKİLER	VI
TABLolar LİSTESİ.....	X
ŞEKİLLER LİSTESİ	XII
KISALTMALAR	XIII
ÖZET	XIV
EXTENDED ABSTRACT.....	XVI
Introduction	XVI
Method	XVII
Findings.....	XVIII
Discussion and Conclusion	XX
BÖLÜM I.....	1
GİRİŞ	1
1.1. Problem Durumu.....	1
1.2. Amaç ve Önem	3
1.3. Problem Cümlesi.....	5
1.4. Alt Problemler.....	5
1.5. Sayıtlar.....	5
1.6. Sınırlılıklar	5
BÖLÜM II.....	6
İLGİLİ YAYIN VE ARAŞTIRMALAR	6
2.1. Örtük Sınıf Analizi.....	6
2.2. Madde Tepki Kuramı	8
2.3. Bilişsel Tanı Modelleri ve Kuramsal Özellikleri	9
2.3.1. İstatistiksel Temeller	12
2.3.2. Q Matris	12
2.3.3. Rule Space Metodolojisi ve Q Matris	15

2.3.3. Bilişsel Tanı Modeli Türleri.....	16
2.3.3.1. DINA.....	16
2.3.3.2. G-DINA	17
2.3.3.3. RUM.....	19
2.3.4. DINA Modelde Yetenek Kestirimleri	19
2.3.5. DINA Modelde Parametre Kestirimleri	20
2.4. Madde Formatı.....	21
2.4.1. Yapılandırılmış Yanıtlı Maddeler ve Türleri	22
2.4.1.1. Açık Uçlu Maddeler.....	24
2.4.1.2. Boşluk Doldurma Maddeleri	27
2.4.2. Seçme Yanıtlı Maddeler ve Türleri	27
2.4.2.2. Çoktan Seçmeli Maddeler.....	28
2.4.2.3. Alternatif Seçim Maddeleri	29
2.4.2.4. Doğru-Yanlış Maddeleri.....	29
2.4.2.5. Eşleştirme Madde Formatı.....	30
2.4.3. Seçme ve Yapılandırılmış Yanıtlı Maddelerin Farkları.....	30
2.4.4. Kısmi Puanlama.....	33
2.4.4.2. Yapılandırılmış Yanıtlı Maddelerde Kısmi Puanlama Yaklaşımları	34
2.4.4.3. Seçme Yanıtlı Maddelerde Kısmi Puanlama Yaklaşımları	35
2.5. İlgili Araştırmalar	36
BÖLÜM III	43
YÖNTEM.....	43
3.1. Araştırmanın Modeli.....	43
3.2. Projenin Evren ve Örneklemi.....	43
3.3. Veri Toplama Araçları	44
3.3.2. Ölçme Aracında Ölçülen Özelliklerin Belirlenmesi.....	45
3.3.3. Ölçme aracında Q Matrisin Hazırlanması	49
3.3.4. Ölçme aracına ilişkin pilot uygulamalar.....	49

3.3.5. Ölçme aracının psikometrik özellikleri	50
3.3.5.1. Klasik Test Teorisine Göre Psikometrik Özellikler.....	50
3.3.5.2. BTM'ye Göre Psikometrik Özellikler	53
3.3.6. Analize alınan maddelerin seçimi.....	56
3.3.7. Kullanılan analizler	60
BÖLÜM IV	61
BULGULAR.....	61
4.1. Maddeler Kısmi Puanlandıktan Sonra Bilişsel Tanı Modeli (BTM) Parametrelerinin ve Model Veri Uyumu Değerlerinin Değişimine İlişkin Bulgular	61
4.2. Maddeler Kısmi Puanlandıktan Sonra Örtük Sınıf Sonsal Olasılıkları ve Tahminlenen Özellik Dağılımlarına İlişkin Bulgular	65
4.3. Maddeler Kısmi Puanlandıktan Sonra Örtük Sınıflar İçin Toplam Puanlar ve Madde Tepki Kuramı (MTK) Yetenek Kestirimlerine İlişkin Bulgular	67
4.4. Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde Sahip Olunan Özellik Sayısına Göre Toplam Puanlar ve MTK Puanlarına İlişkin Bulgular	68
4.5. Maddeler Kısmi Puanlandıktan Sonra Özelliğe Sahip Olma Dağılımlarına İlişkin Bulgular	69
4.6. Maddeler Kısmi Puanlandıktan Sonra İletişim ve İlişkilendirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular	71
4.6.1. Maddeler Kısmi Puanlandıktan Sonra Matematikleştirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular	72
4.6.2. Maddeler Kısmi Puanlandıktan Sonra Akıl Yürütme, Kanıt Gösterme ve Strateji Geliştirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular	73
4.6.3. Maddeler Kısmi Puanlandıktan Sonra Sembolik ve Formel Teknik Dil Kullanma özelliğine sahip olma düzeylerine ilişkin bulgular	74
BÖLÜM V	75
TARTIŞMA VE YORUM.....	75
BÖLÜM VI.....	78
SONUÇ VE ÖNERİLER.....	78
KAYNAKÇA.....	80

EKLER.....	87
EK-1: Pilot uygulamada kullanılan maddelerin madde güçlük ve ayırt edicilik değerleri	87
EK-2: İzleme 1 testine ait Q matris	88
EK-3: İzleme 2 testine ait Q matris	89
EK-4: İzleme 3 testine ait Q matris	89
EK-5: İzleme 4 testine ait Q matris	90



TABLolar LİSTESİ

Tablo 1. Puanlama yöntemlerine göre BTM türleri	11
Tablo 2.Örnek Q Matris	13
Tablo 3.Örnek Q Matris	14
Tablo 4.Yapılandırılmış ve Seçme Yanıtlı Maddelerin Güçlü Yanları ve Sınırlılıkları.	22
Tablo 5.Yapılandırılmış Yanıtlı Madde Türleri	23
Tablo 6.Açık uçlu maddelerin avantaj ve dezavantajları	26
Tablo 7.Proje örnekleminde yer alan okullar, şube ve öğrenci sayıları	44
Tablo 8. Araştırmada belirlenen matematiksel yeterlikler	48
Tablo 9. İzleme 1 testine ait madde güçlük ve ayırt edicilik parametreleri	50
Tablo 10. İzleme 2 testine ait madde güçlük ve ayırt edicilik parametreleri	51
Tablo 11.İzleme 3 testine ait madde güçlük ve ayırt edicilik parametreleri	52
Tablo 12. İzleme 4 testine ait madde güçlük ve ayırt edicilik parametreleri	53
Tablo 13. İzleme 1 testinden BTM'ye göre elde edilen g ve s parametreleri	54
Tablo 14. İzleme 2 testinden BTM'ye göre elde edilen g ve s parametreleri	54
Tablo 15.İzleme 3 testinden BTM'ye göre elde edilen g ve s parametreleri	55
Tablo 16. İzleme 3 testinden BTM'ye göre elde edilen g ve s parametreleri	55
Tablo 17. Analiz için seçilen maddelerin formatları.....	56
Tablo 18. DPM için Q matris	57
Tablo 19. KPM için Q matris	57
Tablo 20.DPM için Matematiksel Yeterlikler.....	58
Tablo 21. KPM için Matematiksel Yeterlikler.....	59
Tablo 22.DPM için Matematiksel Yeterlikler.....	59
Tablo 23. KPM için Matematiksel Yeterlikler.....	60
Tablo 24.KPM ve DPM'nin g ve s parametreleri	62
Tablo 25.DPM'nin en düşük ve en yüksek madde parametreleri	63
Tablo 26. DPM ve KPM için test uyum istatistikleri.....	64
Tablo 27.DPM ve KPM için madde uyum istatistikleri.....	65
Tablo 28.DPM ve KPM için örtük sınıflar ve sonsal olasılıkları.....	65
Tablo 29. DPM ve KPM için tahminlenen özellik dağılımları	67
Tablo 30.DPM ve KPM Sınıflarını için yetenek kestirimleri	67
Tablo 31.DPM için Toplam Özellik Sayısına Karşılık Gelen MTK ve KTK puanları .	69
Tablo 32. KPM için Toplam Özellik Sayısına Karşılık Gelen MTK ve KTK puanları .	69
Tablo 33. KPM ve DPM için Toplam Özellik Sayısına Karşılık gelen Frekansları	70
Tablo 34.İletişim ve İlişkilendirme özelliğine sahip olma düzeyleri	71

Tablo 35. Matematikleştirme özelliğine sahip olma düzeyleri	72
Tablo 36. Akıl yürütme, kanıt gösterme ve strateji geliştirme özelliğine sahip olma düzeyleri.....	73
Tablo 37. Sembolik, Formel Teknik Dil ve İşlem Kullanımına Sahip olma düzeyleri...	74



ŞEKİLLER LİSTESİ

Şekil 1.Kısmi puanlama	34
------------------------------	----



KISALTMALAR

PISA	: Uluslararası Öğrenci Değerlendirme Programı
TIMSS	: Uluslararası Matematik ve Fen Eğilimleri Araştırması
NAEP	: National Assessment of Educational Progress
BTM	: Bilişsel Tanı Modelleri
DINA	: Deterministic Input Noisy “And”
DINO	: Deterministic Input Noisy “Or”
KTK	: Klasik Test Kuramı
MTK	: Madde Tepki Kuramı
G-DINA	: Generalized Deterministic Input Noisy “And”
RUM	: Reparametrised Unified Model/ FUSION Model
BDF	: Boole Tanımlayıcı Fonksiyonu
CPM	: Coloured Progressive Fonksiyonu
CR	: Yapılandırılmış Yanıtlı Madde
SR	: Seçme Yanıtlı Madde
IWLE	: Item-Weighted Likelihood Method
CDA	: Bilişsel Tanı Analizi

ÖZET

BİLİŞSEL TANI MODELLERİNDE DATA/Q MATRİS GENİŞLETME YÖNTEMİNİN ÖRTÜK SINIF DAĞILIMLARI ÜZERİNE ETKİSİ

CEYLAN, Simge

Yüksek Lisans Tezi, Eğitimde Ölçme ve Değerlendirme

Tez Danışmanı: Doç. Dr. Tahsin Oğuz Başokçu

Eylül, 2019

Bu araştırmada çoktan seçmeli ve yapılandırılmış yanıtli madde formatlarında kısmi puanlamanın Bilişsel Tanı Modellerine (BTM) göre örtük sınıf değışmezliđi üzerine etkisi incelenmiştir ve karma maddelerden oluşan test üzerinde data ve Q matris genişletme yöntemleri kullanılarak daha hassas yetenek kestirimi yapılması amaçlanmıştır. Ayrıca Q matrisinin genişletilip yeniden yapılandırılmasının BTM’de kullanılabilirliđi incelenmektedir.

Tezde kullanılan ölçme aracı, yapılandırılmış yanıtli ve çoktan seçmeli maddelerden oluşan BTM kullanılarak yetenek kestirimlerinin yapılabilmesi amacıyla oluşturulmuş testlerdir. Testler 115K531 no’lu TÜBİTAK Projesi kapsamında yapılan araştırma için hazırlanmıştır ve çalışmanın verileri,115K531 no’lu TÜBİTAK Projesi kapsamında 6.sınıf düzeyindeki öğrencilere uygulanan matematik alanına ait testlerden elde edilmiştir. Testte maddeler üst düzey düşünme becerilerini ölçecek nitelikte hazırlanmıştır. Maddelerin hazırlanması için öncelikle üst düzey düşünme becerileri ile ilgili madde yazma eğitimi verilmiştir. Sonra maddeler soru havuzundan, içlerinde alan uzmanlarının da bulunduğu hakemlerce puanlanarak seçilmiştir. Pilot uygulamalar sonrasında test hazırlanmış ve esas uygulamalar yapılmıştır. Oluşturulan testin PISA ve TIMMS ile aynı düzeye ve psikometrik özelliklere sahip olduđu yapılan analizler ile ortaya koyulmuştur.

Araştırmada BTM’ye uygun analizlerin gerçekleştirilebilmesi için test ve maddelerin ilişkilendirilebilecekleri özellik kümeleri belirlenmiş ve buna göre Q matrisler oluşturulmuştur. Alan uzmanları tarafından belirlenmiş Q matris ile maddeler 1-0 şeklinde puanlanıp yetenek kestirimi yapılmıştır. Daha sonra Q matris yeniden yapılandırılarak kısmi puanlanabilecek maddeler belirlenmiştir. Ardından maddeler tekrar puanlanarak yetenek kestirimi yapılmıştır. Bu şekilde madde sayısı iki katına çıkmıştır. Genişletilen data ile sınıflama düzeyinde yetenek kestirimleri

yapılmıştır ve elde edilen veriler ilk datalardan elde edilen sonuçlar ile karşılaştırılmıştır. Ayrıca çoktan seçmeli ve yapılandırılmış yanıtli maddelerin kısmi puanlanmasının parametre kestirimlerine etkisi incelenmiştir.

Araştırmanın sonuçları incelendiğinde, maddenin yapısına göre kısmi puanlanacak şekilde Q matris ve data genişletildiğinde öğrencilere ilişkin daha detaylı bilgiler elde etme, öğrencileri daha geniş bir yetenek düzleminde görebilme ve onlara daha net bir geri bildirim verme imkanına sahip olunduğu gözlenmiştir.

Anahtar kelimeler: Bilişsel Tanı modelleri, Data/Q matris genişletme yöntemi, Örtük Sınıf Değişmezliği, Karma Madde Formatı



EXTENDED ABSTRACT

EFFECT OF DATA/Q MATRIX AUGMENTATION ON LATENT CLASS DISTRIBUTIONS IN COGNITIVE DIAGNOSTIC MODELS

CEYLAN, Simgе

MS, Department of Measurement and Evaluation

Advisor: Assoc. Prof. Dr. Tahsin Oğuz Başokçu

Introduction

In this study, multiple choice and constructed response item formats were analyzed on latent class invariance according to Cognitive Diagnostic Models (CDM).

In a qualified education, formative assessment has a great place. In this way, learning deficiencies can be identified and effective learning can be provided by making necessary corrections in a timely manner. Tracking cognitive process skills is also an important part of formative assessment. Commonly used measurement instruments fail to identify these deficiencies. However, field knowledge, high-level thinking skills and cognitive process skills were measured together in the 6th grade mathematics course through the CDM used in the research. Moreover, more than one skill could be examined with one question and the level of having cognitive process skills was determined. Research on the use of CDM is important for tests that measure high level skills in mathematics. It is very important to choose the item format to be used when preparing the measurement instrument. Two factors that will be based on choosing the item format; should be to measure cognitive process and content. In this context, it is the best format to reveal the cognitive process that is highly representative of content and intended to be observed.(Haladyna, 2004). In other words, the format that provides the most accurate assessment for the purpose of the exam is the best. Because each item format and content of the measure different cognitive processes, mixed-use format is recommended to take full advantage of this situation.(Haladyna, 2004). In this research, mathematical thinking skills required for the relationship, mathematical, reasoning and evidence to show, strategy development and so on. competences have been checked. Therefore, since the use of

substance in a single format would not be sufficient for the outcomes to be achieved, constructed response and selected response item formats were used together in the study. In this sense, the study is the first in terms of approaches using different substance formats together. In addition, it will contribute to the field in terms of preparing the measurement instrument to measure high-level thinking skills.

In this study, DINA model parameters were estimated by means of a composite material test prepared on the basis of CDM. In general, the DINA model parameters are estimated by simulation data or by preparing the Q matrix of the pre-prepared tests. Conducting the study using real data contributes to the literature for this area, which remains at the theoretical level in general. Then, it is aimed to make more accurate talent estimation by using data and Q matrix expansion methods on the test consisting of mixed substances. In addition, the usability of Q matrix expansion and restructuring in CDM is examined. However, the research includes an application of the method of analyzing by reconstructing the Q matrix with partially scored items. It is also one of the first studies conducted with this method in the literature. Considering the limited number of studies based on partial scoring and based on CDM, it is expected to contribute to the literature since no study has been conducted on the determination of implicit classes for both binary and different scoring methods by Q matrix and data expansion.

While conducting research, it is essential to reach the correct information and generalize the obtained data. Considering the characteristics of the study, this is an important study in terms of access to realistic parameters due to the large sample size.

Method

The measurement instrument used in the thesis is the tests designed to make ability estimation by using CDM consisting of constructed responses and multiple choice items. The tests were prepared for the research conducted within the scope of TUBITAK Project 115K531 and the data of the study were obtained from the tests of mathematics applied to 6th grade students under TUBITAK Project 115K531. The items were designed to measure high-level thinking skills. In order to prepare the items, first of all, substance writing training related to high level thinking skills was given. The items were then selected from the question pool by the referees, including

field experts. After the pilot applications, the test was prepared and the main applications were made. The analysis revealed that the test had the same level and psychometric properties as PISA and TIMSS. In this study, feature sets that the test and items could be related to were determined and Q matrices were formed according to the cognitive diagnostic model. With the Q matrix determined by the field experts, the items were scored as 1-0 and talent estimation was performed. Then, the Q matrix was restructured and the items that could be scored partially were determined. Then, the items were re-scored and ability estimation was performed. In this way, the number of items has doubled. Ability estimations were made at the classification level with the expanded data and the obtained data were compared with the results obtained from the first data. In addition, the effect of partial scoring of multiple choice and constructed response items on parameter estimations was examined.

Findings

When examining the DINA model parameters in CDM, g and s parameters calculated at the item level for this model and the standard errors of these parameters are taken into consideration. The basis is the idea that each item divides the test group into two classes. The status of being in the same class is interpreted as equal probability of respondents to the same class. In this model which is based on probability, s parameter is; the student does not answer the item correctly even though he / she has all the necessary attributes to answer the item, g parameter; The student is interpreted as correct answer to the item even though it does not have the necessary attributes to respond to the item (De La Torre, 2008). When the g parameters of the partially scored items were examined, it was found that the lowest g parameter was 0 and the highest g parameter was 0.71. It is concluded that the values of the g parameter vary between 0 and 0.71 in the items scored dichotomically. When the s parameters related to the partially scored items were examined, it was found that the lowest s parameter was 0 and the highest s parameter was 0.77. In this case, it is concluded that s parameter takes values between 0 and 0.77. When the g parameters of dicotomic (binary) items were examined, the lowest g parameter was found to be 0.12 and the highest g parameter was 0.42. It is concluded that the values of the g parameter vary between 0.12 and 0.42 in the items

scored dichotomically. When the s parameters were examined, it was found that the lowest s parameter was 0.07 and the highest s parameter was 0.77. In this case, it is concluded that s parameter takes values between 0.07 and 0.77.

Another finding of the study is the latent classes and posterior probabilities of dichotomic scored items (DSI) and partial scoring items (PSI). In order to compare the latent class posterior probabilities observed for DSI with PSI, total probabilities for corresponding probability patterns were determined in PSI and the number of latent classes observed after analysis in PSI were also indicated. It is seen that the items for partial scores have lower probability values because they are distributed to more implicit classes. However, it is understood that the sum of probability values is close to each other with respect to dichotomic substances. There are 16 implicit classes for dicotomic substances and 256 different implicit classes for partially scored items. Any of the implicit classes (11110000), (01000000), (00010000), (01010000), (01100000), (01110000) may correspond. This is indicative of a more detailed analysis for the same students. In addition, it is observed that the probability increases especially in the classes with more attributes. When the partial scores were evaluated, it was observed that the probabilities for the 8 properties ranged between 0.5047 and 0.6342. When the communication and association skills were determined to be 0.4956 observed probability in DSI, it was observed that the probability of observation increased to 0.6097 and 0.6347. This can be considered as an indicator of increased sensitivity, in other words, increased observation.

When the items were transformed into constructed response item forms, the average difficulty values for the latent classes and the findings of the IRT ability estimations were observed, the latent class distributions of the students were close to each other. Although a high correlation was determined for both cases, it was observed that the students who fell into the same range with the dichotomic scoring method had a wider range of predictions when the partial scoring method was used. This can be considered as an indicator that more reliable measurements are made for the items scored partially for reliability. After reconstructing the Q matrix, the ability scores obtained from the partially scored items according to the IRT and the findings of Absolute Success Percentages were observed to increase as the number of implicit features increased and the predictions of the students increased in both types.

However, it was observed that both lower ability level and upper ability level can be predicted for the items scored partially. For example; While the average of the group without any characteristics in the IRT analysis was -0.92, the same situation reached up to -1.46 in the partially scored items. This shows that it is possible to evaluate students at a wider scale. After the Q matrix is structured according to partial scoring, when the data related to property distributions are examined, it is seen that the distributions of property possessions are similar in both cases, but it is seen in the students with four characteristics are distributed to higher classes in partial scoring. In addition, when the findings related to the level of possessions were examined, it was found that some features were not completely hierarchical.

Discussion and Conclusion

In this section, the findings of the tests used in the research are discussed and interpreted within the framework of the literature. When the results of the study were examined, it was found that a test prepared using the data augmentation method was more structurally reliable and could be organized in a way to give more detailed information about the students. As it was seen in the results, the results were consistent and similar between the four skill tests and the eight attributes tests. It can be said that it may be appropriate for CDM to create measurement situations that can get more detailed information for a single test by partial scoring the items. However, the findings of the study show that this method can be used not only for constructed response items but also for selected response items. As a method for the first time in the field, augmentation of the Q matrix according to the properties of the material also means increasing the number of items in the test. Furthermore, it is observed that the psychometric properties of the test are maintained after this procedure. As it is known, there are problems in statistical inferences regarding the reliability and validity of the test in structured responsive or open-ended item approaches to obtain more detailed information about the student. Therefore, although constructed response items have the potential to provide more detailed information about the student, its validity and reliability are difficult to prove. In this study, it was found that the method can be used with similar statistical robustness for partial scoring in constructed response and selected response items for CDM models. At the same time, another result shows that it is possible to set up a test structure where constructed

response and selected response items can work together by organizing the analyzes in the tests performed with mixed item format. When the studies on the mixed item format are examined, it is seen that the approach here is usable and can be examined in terms of the field, especially considering the limited empirical data in the field of CDM. When the findings obtained according to the results of the research are taken into consideration, it is seen that it is possible to expand the matrices of the investigators who want to develop tests according to CDM not only for the correct answer cases but also for the matrices of other alternative responses.

Key words: Cognitive diagnosis models, Data / Q matrix augmentation method, Latent Class Invariance, Constructed Response Items, Selected Response Items



BÖLÜM I

GİRİŞ

1.1.Problem Durumu

Ülkelerin gelişim sürecinde, eğitimin yeri ve önemi yadsınmaz. Bu sebeple eğitim süreçlerini geliştirmek, eğitimin niteliğini ve çıktılarını değerlendirmek için çeşitli sınavlar yapılmaktadır. Sınavlardan elde edilen çıktılar değerlendirilmesiyle eğitim stratejileri geliştirilmekte ve eğitim politikaları oluşturulmaktadır. Bu sınavlara sınıflarda yapılan izleme testleri, quizler, sözlüler, yazılı yoklamalar, deneme sınavları, liseye ve üniversiteye geçiş sınavları ve uluslararası yapılan sınavlar vb. örnek verilebilir. Uluslararası yapılan sınavların en bilinenleri PISA (Uluslararası Öğrenci Değerlendirme Programı) ve TIMSS (Uluslararası Matematik ve Fen Eğilimleri Araştırması)'dır.

PISA, sanayileşmiş ülkelerin hükümetler arası bir organizasyonu olan Ekonomik İş Birliği ve Kalkınma Örgütü (OECD) tarafından yürütülmektedir. Amacı, okulda öğrenilen bilgilerin günlük hayat becerilerine aktarılabilme durumlarını matematik, fen okuryazarlığı ve okuma becerileri alanlarında yoklamaktır. Bu sınav sonucunda OECD'ye bağlı olan ülkelerin eğitim politikacıları, sonuçlardan elde edilen verilerin doğrultusunda eğitim sistemleri ve eğitim politikaları adına çalışmalarına yön vermektedirler. Dünya genelinde katılımcı ülkelerin eğitim politikacılarının birinci gayesi, vatandaşlarının potansiyellerini değişen dünya şartlarına uygun olarak tam anlamıyla gerçekleştirmelerine ve onların yeteneklerini geliştirmelerine olanak sağlamaktır. Bu bağlamda PISA sonuçları, puanlardan veya sıralamalardan çok daha fazlasını ifade etmektedir. Okul sistemlerinin kalitesini, eşitliğini ve verimliliğini değerlendirmek amacıyla kullanılan önde gelen eğitim ölçütlerinden biridir. Bu şekilde hükümetlere ve eğitimcilere, yüksek performanslı bir eğitim sisteminin özelliklerini belirleyerek, kendi öz değerlerini adapte edebilecekleri etkili eğitim politikaları tanımlamalarını sağlamaktadır (PISA, 2015). Değerlendirme sonuçlarının böylesine kapsamlı bilgi vermesi amaçlanan PISA ve TIMSS gibi sınavlarda farklı madde formatları bir arada kullanılmıştır.

Yapılan birçok çalışmada öğrencilerin, çalışma yöntem ve stratejilerini girecekleri sınavın formatına göre belirledikleri ve bu sebeple de öğrenmelerinin

niteliğinin de değiştiğine dair kanıtlar bulunmaktadır. Örneğin; sınavın yapılandırılmış yanıtli maddelerden oluşması kavrama, çoktan seçmeli maddelerden oluşması ise hatırlama basamağındaki becerilerin gelişimini desteklemektedir (Michael E. Martinez, 1990). Her bir format farklı içerik ve bilişsel süreçleri ölçtüğünden dolayı, bu durumdan tam olarak yararlanmak için karma format kullanımı önerilmektedir (Haladyna, 2004). PISA ve TIMMS aynı testte farklı madde formatları kullanılarak hazırlanır. Çünkü matematiksel becerileri günlük hayat becerileri ile birleştirmek, matematiksel düşünme, iletişim kurma, ilişkilendirme gibi becerileri yoklayabilmek için tek bir formatta madde kullanmak ulaşılmak istenilen çıktılar için yetersiz kalabilir (Stankous, 2016). Ayrıca bir sınavın sadece yapılandırılmış yanıtli maddelerden oluşması ya da çoktan seçmeli maddelerden oluşması, ilgili alanın bilişsel basamaklarını yoklamada da sınırlılık yaratabilir (Frary, 1989). Bu gerekçelerden dolayı araştırmada yapılandırılmış yanıtli ve çoktan seçmeli maddeler birlikte kullanılmıştır.

PISA ve TIMMS, sınava katılanları yeteneklerine göre sınıflara yerleştirir. Bu şekilde elde edilen sonuçların eğitim politikaları açısından daha geçerli olduğuna dair kanıtlar bulunmaktadır (Fleischman, Hopstock, Pelczar, Shelley, ve Xie, 2011). Yetenek kestiriminde, öğrencilerin yetenek düzeyleri belirlenirken istatistiksel olarak Madde Tepki Kuramı modellerinden Rasch Model kullanılmaktadır (Edition, 2009). Madde tepki kuramı (MTK), klasik test kuramının zayıf yönlerini giderebilecek matematiksel modeller sunan bir yaklaşımdır. Bilişsel tanı modelleri (BTM) ise MTK modellerinin bir uzantısı olup sağladığı bilgiler sayesinde öğrencilerin güçlü ve zayıf yanlarını değerlendirmeyi sağlayan psikometrik bir modeldir. BTM'nin MTK'dan ayrıldığı noktalardan biri ise yetenek kestirimidir. MTK'da aynı puanı alan iki öğrencinin yetenek kestiriminin aynı olduğu kabul edilir. Ancak BTM'de öğrencilerin aynı puanı alması aynı yeteneğe sahip olduğunu söylemek için yeterli değildir. Farklı konulardaki maddeleri doğru yanıtlayarak aynı puanı almış olabilirler (Tatsuoka, 2009). Bu sebeple MTK'da elde edilen puanlar öğrencilerin becerilere sahip olma durumları konusunda detaylı bilgi verme açısından sınırlı olabilmektedir. Diğer bir deyişle BTM'nin ayrıntılı değerlendirme ve yetenek kestirimi için daha elverişli bir model olduğu söylenebilir.

1.2.Amaç ve Önem

Ölçme aracı hazırlarken kullanılacak madde formatının seçimi oldukça önemlidir. Madde formatını seçerken temel alınacak iki unsur; bilişsel süreci ve içeriği ölçmek olmalıdır. Bu bağlamda içeriği temsil etme gücü yüksek ve gözlemlenmesi amaçlanan bilişsel süreci ortaya çıkaracak format, en iyi formattır.(Haladyna, 2004) Başka bir deyişle sınavın amacına uygun en doğru değerlendirmeyi sağlayacak format en iyisidir. Her bir madde formatı farklı içerik ve bilişsel süreçleri ölçtüğünden dolayı, bu durumdan tam olarak yararlanmak için karma format kullanımı önerilmektedir. (Haladyna, 2004) Araştırmada matematiksel düşünme becerileri için gerekli olan ilişkilendirme, matematikleştirme, akıl yürütme ve kanıt gösterme, strateji geliştirme vb. yeterlikler yoklanmıştır. Bu sebeple tek bir formatta madde kullanmak, ulaşılmak istenen çıktılar için yetersiz kalacağı için araştırmada yapılandırılmış yanıtı ve seçme yanıtı madde formatları birlikte kullanılmış olup farklı madde formatlarının bir arada kullanıldığı yaklaşımlar açısından ilktir. Ayrıca ölçme aracının üst düzey düşünme becerilerini ölçecek nitelikte hazırlanması yönünden alana katkı sağlayacak niteliktedir. Çünkü gerek yurt dışında gerekse ülkemizde uluslararası yapılan geniş ölçekli sınavlardan elde edilen sonuçlar üst düzey düşünme becerilerine sahip olmanın önemini vurgular niteliktedir.

Çalışmada Q matris ve data genişletme ile hem ikili hem de farklı puanlama yöntemleri için örtük sınıfların belirlenebilmesi üzerinde çalışılmıştır. Bu şekilde yapılan bir çalışmanın daha önce yapılmamış olmasından dolayı literatüre katkıda bulunması beklenmektedir.

Araştırma yaparken doğru bilgilere ulaşmak ve elde edilen verileri genelleştirmek şarttır. Araştırmanın özellikleri göz önünde bulundurulduğunda örneklem büyüklüğünün geniş olmasından dolayı gerçekçi parametrelere ulaşabilme imkanı açısından bu alanda önemli bir çalışmadır.

Çalışmada BTM temel alınarak hazırlanmış karma madde formatlı bir test üzerinden DINA model parametreleri kestirilmiştir. Genel olarak DINA model parametreleri, simülasyon dataları üzerinden ya da önceden hazırlanmış testlere ait Q matris sonradan hazırlanarak kestirilmiştir. Çalışmanın, gerçek veriler kullanılarak yürütülmesi, genel olarak teorik düzeyde kalan bu alan için literatüre katkı niteliğindedir.

Nitelikli bir eğitimde, biçimlendirmeye yönelik değerlendirmenin yeri büyüktür. Bu şekilde öğrenme eksiklikleri belirlenebilir, zamanında gerekli düzeltmeler yapılarak etkili öğrenme sağlanabilir. Bilişsel süreç becerilerinin izlenmesi de biçimlendirmeye yönelik değerlendirmenin önemli bir parçasıdır. Yaygın olarak kullanılan ölçme araçları bu eksikleri belirlemede yetersiz kalmaktadır. Ancak araştırmada da kullanılan BTM ile ortaöğretim 6. Sınıf matematik dersindeki alan bilgisi, üst düzey düşünme becerisi ve bilişsel süreç becerileri bir arada ölçülmüştür. Üstelik bir soru ile birden fazla beceri yoklanabilmiş ve öğrencinin bilişsel süreç becerilerine sahip olma düzeyleri belirlenebilmiştir. Matematikte üst düzey becerileri ölçen testler için BTM kullanımı konusunda araştırma önem taşımaktadır.

Bu çalışmanın amacı; yapılandırılmış yanıtı ve seçme yanıtı maddelerin kısmi puanlanmasının BTM ile belirlenen örtük sınıflar üzerine etkisini incelemektir.

Araştırmada yapılandırılmış yanıtı ve seçme yanıtı maddelerin kısmi puanlamaya uygun olanları Bilişsel Tanı Modellerinden DINA model ile analiz edilebilmesine yönelik bir yöntem test edilmektedir.

DINA model ile yapılan çalışmalar genellikle simülasyon temellidir. Ancak araştırmada elde edilen sonuçlar gerçek veri temellidir.

Araştırma aynı zamanda örneklem büyüklüğü açısından ele alındığında geçerli ve güvenilir veriler ortaya koyacak nitelikte olup alanda yapılan diğer araştırmalar göz önünde bulundurulduğunda ayrıca önem arz etmektedir.

Araştırma, kısmi puanlamanın temel alındığı BTM' ye dayalı çalışmaların azlığı göz önüne alındığında literatüre katkıda bulunması beklenmektedir. Ayrıca, araştırma kısmi puanlanabilen maddeler ile Q matrisin yeniden yapılandırılması ile analiz edilmesi yönteminin bir uygulamasını içermektedir. Literatürde bu yöntem ile yapılan ilk çalışmalardan biri olma özelliğini de taşımaktadır.

1.3.Problem Cümlesi

Çoktan seçmeli ve yapılandırılmış yanıtli madde formatlarında kısmi puanlamanın bilişsel tanı modellerine göre örtük sınıf değişmezliğı üzerine etkisi nedir?

1.4.Alt Problemler

1. Maddeler kısmi puanlandıktan sonra Bilişsel Tanı Modeli (BTM) parametrelerinin ve model veri uyumu değerlerinin değışimi ne düzeydedir?
2. Maddeler kısmi puanlandıktan sonra örtük sınıf sonsal olasılıkları ve tahminlenen özellik dağılımları ne düzeydedir?
3. Maddeler kısmi puanlandıktan sonra örtük sınıflar için toplam puanlar ve Madde Tepki Kuramı (MTK) yetenek kestirimleri nasıldır?
4. Maddeler kısmi puanlandıktan sonra sahip olunan özellik sayısına göre toplam puanlar ve MTK puanları nasıl değışmektedir?
5. Maddeler kısmi puanlandıktan sonra özelliğıe sahip olma dağılımları nasıldır?
6. Maddeler kısmi puanlandıktan sonra A1, A2, A3, A4 özelliğıne sahip olma düzeyleri nasıl değışmektedir?

1.5.Sayıtlar

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

1.6.Sınırlılıklar

1. Bu araştırma sadece DINA model ile yapılan analizler ile sınırlıdır. Diğer modeller denenmemiş ve modeller arası karşılaştırma yapılamamıştır.
2. Yalnızca matematik alanında ölçme aracı geliştirilmiş ve bu alanda çalışılmıştır.
3. Sınıf düzeyi sadece 6. Sınıf ile kısıtlıdır.
4. Matematiksel yeterlikler sadece 4 özellik üzerinden sınıanmıştır.
5. Yöntem simülasyonlarla test edilememiştir.
6. Araştırmanın sonuçları üzerindeki madde sayısının etkisi incelenememesi bir diğer sınırlılıktır.

BÖLÜM II

İLGİLİ YAYIN VE ARAŞTIRMALAR

2.1. Örtük Sınıf Analizi

Örtük sınıf analizi; ilk olarak 1950 yılında Lazarsfeld tarafından, dikotomik olarak gözlemlenen değişkenlere dayanarak sınıflar oluşturmaya yarayan araçlar olarak tanıtılmıştır. Örtük sınıf analizinin temeli; doğru sayılan istatistiksel bir modelin bazı parametrelerinin, gözlemlenemeyen alt grupları arasında farklılık bulunduğu dayanmaktadır. Bu alt gruplar, kategorik bir örtük değişkenin kategorilerini oluşturmaktadır (Vermunt ve Magidson, 1979). McCutcheon'a göre gözlenebilen belirtiler arasındaki iç ilişkiyi anlamak ve bu ilişkinin karakterini ortaya çıkartmak, örtük olan yani gözlenemeyen özellik hakkında bilgilere ulaşmayı sağlamaktadır (Allan L. McCutcheon, 1987).

Örtük sınıf analizi, matematiksel metodolojiden ziyade olasılığın istatistiksel konseptinin temel alındığı geliştirilmiş sınıf (cluster) analizi olarak da kabul edilmektedir. Parametre tahminleri; sınıfların profillerine yani her bir sınıfın tanımına ve her bir sınıfın büyüklüğüne yöneliktir. Temel fark ise sınıfların kesin olmaması ancak her bir sınıfa ait olunma durumunun olasılığının kesin olmasıdır (Francis,1978). Esasında örtük sınıf modellerinin, araştırmacıların gözlemlenen iki ya da daha fazla değişkenler arasındaki gözlemlenen ilişkileri daha iyi anlamalarına yardımcı olacak bir araç olarak kullanılması yirminci yüzyılda çok daha kısa bir tarihe sahiptir. Ancak on dokuzuncu yüzyıl çalışmalarında daha önce kullanılmış olan bazı matematiksel modeller artık, örtük sınıf modellerinin veya diğer örtük yapı türlerinin özel durumları olarak görülmektedir. On dokuzuncu yüzyılda kullanılan modellerle ilgili olarak Peirce, bu modeli tahminlerin başarısını ölçmek için göze çarpan iki değişken arasındaki ilişkiyi daha iyi anlayabilmek için kullanılabileceği şeklinde tanıtmıştır (Goodman, 2018).

Örtük sınıf analizi zamanla geliştirilerek regresyon yöntemlerinin ve kümeleme analizlerinin de kullanıldığı çok değişkenli analiz yöntemi olmuştur. Genel olarak iki veya daha fazla gözlenen kategorik değişkenin çapraz sınıflanmasıyla çok boyutlu ayrı bir örtük özelliğin karakterize edilebilmesine olanak

tanılmaktadır. Bunların yanı sıra örtük sınıf analizi sıralama ve sınıflama ölçeğindeki veriler için de kullanılabilir (Başokçu, 2016).

Örtük sınıf analizinin matematiksel modeli şu şekilde ifade edilebilir;

Y_i 'yi j maddesine, her hangi bir gruptan gelen i . cevaplayıcının cevaplama vektörü olarak kabul edildiği durumda $Y_i = \{y_{ij}\}$, $j = 1, \dots, J$ şeklinde ifade edilmektedir. Bu durumda maddelerin cevap seçenekleri farklı bir set olarak tanımlanmaktadır. Aynı zamanda j . madde için

$r = 1, \dots, R_j$ gibi birbirini dışlayan değerler almaktadır. Buradan C farklı mutlak örtük sınıf modeli aşağıdaki gibi tanımlanır.

$$P_r(Y_i) = \sum_{c=1}^c \theta_c \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{cjr}^{\delta_{ijr}}$$

Olağan kısıtlanmasıyla θ_c 'nin karışık oranı $c=1, \dots, C$ ise;

$$\sum_{c=1}^c \theta_c = 1$$

örtük sınıfın oranıdır ve gözlenemeyen örtük sınıfların ölçüsünü verir. Maddelerle ilişkili olan koşullu olasılıklar c . örtük sınıfa üye olana ait j maddesine verilen r cevabın olasılığını verir. Bu koşullu olasılıkların toplamı örtük sınıfların hepsinde her bir madde için 1'dir. Bu durumun matematiksel ifadesi;

$$\sum_{r=1}^{R_j} a_{cjr} = 1$$

şeklindedir (Dayton ve Macready, 2006).

Örtük değişken modelleri, değişkenlerin sürekli ya da süreksiz oluşuna göre 4 temel model olarak sınıflandırılmaktadır (Bartholomew, Knott, ve Moustaki, 2011). Bu temel modeller içerisinde sosyal bilimlerde en çok bilinen ve yaygın kullanılan model faktör analizidir. Faktör analizinde örtük ve gözlenen değişkenlerin yapısı sürekli kabul edilmektedir. Faktör analizi zaman içerisinde farklı amaçlarla kullanılmış olup gözlemlenen değişkenlerin sayısını indirgeme ve kestirilen faktör yüklerine regresyon analizi uygulama amacıyla kullanılması yaygın olanlarıdır. Bir diğer örtük değişken modeli ise örtük profil analizi (latent profile analysis)'dir. Bu modelde örtük değişken sürekli yapıdayken gözlenen değişken süreksiz olarak kabul edilmektedir. Üçüncü model, örtük özellik analizi (latent trait analysis) ya da MTK

olarak bilinir. Burada örtük değişken sürekli, gözlenen değişkenin süreksiz yapıda olduğu kabul edilmektedir. Dördüncü model ise örtük sınıf analizidir (latent class analysis). Hem örtük hem de gözlenen değişkenlerin süreksiz yapıda olduğu kabul edilmektedir. Örtük özelliğe ilişkin yaklaşımlar içerisinde MTK en bilinen modellerden olup BTM de temelinde örtük sınıf analizi olan bir yaklaşımdır.

2.2. Madde Tepki Kuramı

Geleneksel psikometrik kuramların temel amacı, örtük değişkenlere sistematik olarak puan atamaktır. Gözlemlenemeyen örtük değişkenleri doğru olarak ölçmek için psikometrisler tarafından iki ana ölçme kuramı geliştirilmiştir. Bunlardan biri Klasik Test Kuramı (KTK) diğeri ise Madde Tepki Kuramıdır (MTK). MTK, 1920'lerde Louis Thurstone, 1940'larda Lawley, Mosier ve Richardson ve 1950 ve 1960'larda daha belirleyici çalışmalar yapan Alan Birnbaum, Frederic Lord ve George Rasch'ın çalışmalarına dayanmaktadır. (van der Linden, 2016)

KTK'da, ölçek ve testlerin güvenilirliğini belirlemede yaşanan sınırlılıklar ve maddeye, teste ve bireye ait parametrelerin bir gruba bağlı olması gibi nedenlerle KTK'ya alternatif olarak MTK ortaya atılmıştır. MTK, madde ve yeteneklerin sürekli bir ölçek üzerinde kestirildiği ve bunun yanı sıra "a" ayırt edicilik, "b" güçlük ve "c" tahmin parametrelerinin hesaplandığı bir modeldir. MTK'nın varsayımlarından biri testteki performansın, davranış ya da beceri olarak adlandırılan kavramlarla, sınava girenlerin özelliklerini tanımlayarak tahmin edebileceği veya açıklanabileceğidir. Kurama göre, sınava girenlerin bu özellikler ile ilgili madde ve test performanslarını tahmin ederken ürettiği puanlar, yetenek puanları olarak adlandırılmaktadır (Hambleton ve Swaminathan, 1985).

MTK'da, maddelere verilen yanıtlar ile örtük değişkenler arasındaki ilişki, bir matematiksel işlev tarafından belirlenmektedir ve ölçme işlemi sonucu olarak tek bir puan üretilmektedir. Bu ölçme işlemi sonuçlarını, öğrencinin performansını bir grup içerisinde sıralamak ve karşılaştırmak için kullanmak faydalıdır, ancak örtük değişken içindeki belirli alanlar üzerindeki performans hakkında değerli bilgiler sağlamamaktadır (Mislevy, 2014). BTM ise MTK'nın, madde ve yetenek parametrelerini sürekli bir ölçek üzerinde kestirme, öğrencinin performansını sadece tek bir puan ya da sıralama olarak verme gibi sınırlılıklarını gideren bir modeldir.

Çünkü BTM, maddelere verilen yanıtları basit bir psikolojik teori ile açıklar, test aracılığıyla ölçülen yapının altında yatan psikolojik süreçleri açık bir şekilde betimler.

Üç parametrelili MTK modelinde, testteki maddeler aracılığıyla elde edilen sonuçlara sadece üç parametrenin başarı üzerinde etkisi olduğu varsayımı yapılır. Ancak BTM’de performansı etkileyen daha gerçekçi değişkenlerin varsayımı söz konusudur. MTK gibi geleneksel yöntemlerde, kişinin (testi alanın) maddeyi doğru yanıtlaması için gerekli bilişsel süreçleri, araştırmacının varsayım ya da beklentilerine dayanmaktadır. Ancak BTM’lerde bu bilişsel süreçler ampirik kanıtlara dayandırılır. MTK’da herhangi bir ölçme işlemi sonucu elde edilen performans, örtük değişkenler yoluyla bilgi sağlarken BTM’lerde bir grup kendi içinde ilişkili, ayrı davranışlara sahip olma ya da olmama şeklinde bilgi sağlar (Ravand, 2015).

2.3. Bilişsel Tanı Modelleri ve Kuramsal Özellikleri

Bilişsel tanı modelleri, örtük sınıf analizi kullanılan ve yetenek kestirimi yapılırken MTK’nın temel alındığı istatistiksel bir modeldir. İlk olarak 2001 yılında eğitim alanındaki değişim ihtiyacına yönelik Amerika’da “No Child Left Behind” olarak bilinen hareket sonrasında ortaya çıkmıştır. Bu hareket ile öğrencilerin eğitsel anlamda zayıf ya da güçlü oldukları yönleri belirleyebilmek önem kazanmıştır. Psikometri uzmanları, bu eğitim ihtiyacını karşılayabilmek adına bilişsel psikologlar ile iş birliği içerisinde yeni bir kuramsal çerçeve olan bilişsel tanı değerlendirme sistemini geliştirmiştir. Geliştirilen modelin diğer modellerden farkı sadece öğretim çıktılarını değerlendirmeyi hedeflemekle kalmayıp aynı zamanda kalıcı ve tam öğrenme gerçekleştirebilme adına tanısallı tespit sağlayıp sonucunda geri bildirim ile eğitsel süreci kuvvetlendirmektir (Lim, 2015).

BTM’nin çıkış noktası Lineer Lojistik Test Modeli (LLTM) ile (Fischer, 1973) Tatsuoka ve Tatsuoka’nın (1995) Rule Space modelidir. LLTM, temel Rasch Modeli’nin, madde cevaplama aşamalarını kapsayan bilişsel süreçleri de içerecek şekilde genişletilmiş bir versiyonudur (DiBello, Roussos, ve Stout, 2006).

BTM’de kullanılan analiz bilişsel tanı analizidir. Bilişsel tanı analizlerinde, öğrencilerin verdiği yanıtlar, önceden belirlenmiş Q matrisler yardımı ile örtük sınıflara atanmaktadır. Daha sonra bu sınıflara ait öğrenmelerin niteliklerine göre

belirlenmiş seviyeler oluşturulmaktadır. Bu sayede sadece öğrenme çıktılarının değerlendirilmesinin aksine öğrenme ve öğretme sürecinin değerlendirilmesine de olanak sağlanmaktadır (Jang, 2008).

BTM tanısai deęerlendirmelerde disiplinler arası bir yaklaşımdır. Hatta bilişsel psikoloji ve istatistiksel analiz arası bir ara yüz olarak da düşünülebilir. Yani psikolojik süreçler ve belirli bir testin maddeleri üzerindeki öğrencilerin performansının altında yatan stratejiler ile bu maddelere verilen yanıtların karmaşık istatistiksel analizler yoluyla ilişkisini araştırır.

Teknik anlamda BTM, test katılımcılarının test maddelerine verdikleri tepkilerin benzerliğine göre bazı örtük sınıflara ayıran örtük sınıf modelleridir (Andersen, Hagenars, ve McCutcheon, 2006).

Bu türlerin puanlama yöntemleri ve model tipi göz önünde bulunduracak şekilde Rupp ve Templin (2008)' e göre gruplandırılması aşağıdaki tabloda verilmiştir.

Tablo 1. Puanlama yöntemlerine göre BTM türleri

Gözlenen Tepki Değişkenleri	Örtük Yordayıcı Değişkenler		Model Tipi
	DİKOTOMİK	ÇOKLU	
Dikotomik puanlanmış	1. RSM	1. BIN	Tamamlayıcı Olmayan Model
	2. AHM	2. MCLCM	
	3. DINA	3. F NC-RUM	
	4. HO-DINA	4. RE NC-RUM	
	5. MS-DINA		
	6. NIDA		
	7. RERUM		
	8. BIN		
	9. MCLCM		
	10. F NC-RUM		
	11. RE NC-RUM		
Çoklu Puanlanmış	1. DINO	1. BIN	Tamamlayıcı Model
	2. NIDO	2. MCLCM	
	3. BIN	3. C-RUM	
	4. MCLCM	4. GDM	
	5. C-RUM	5. H-GDM	
	6. GDM	6. LCDM	
	7. H-GDM	7. G-DINA	
	8. LCDM		
	9. G-DINA		
Çoklu Puanlanmış	1. RSM	1. BIN	Tamamlayıcı Olmayan Model
	2. AHM	2. MCLCM	
	3. BIN	3. F NC-RUM	
	4. MCLCM	4. RE NC-RUM	
	5. F NC-RUM		
	6. RE NC-RUM		
Çoklu Puanlanmış	1. BIN	1. BIN	Tamamlayıcı Model
	2. MCLCM	2. MCLCM	
	3. C-RUM	3. C-RUM	
	4. GDM	4. GDM	
	5. H-GDM	5. H-GDM	
	6. LCDM	6. LCDM	
	7. G-DINA	7. G-DINA	

*RSM, Rule-space method (Rule-space modeli). AHM, attribute hierarchy method (beceri hiyerarşi modeli). BIN, Bayesian inference network (Bayesçi çıkarım ağı modeli). DINA, Deterministic inputs, noisy 'and' gate. G- DINA, generalized DINA (genelleştirilmiş DINA), HO-DINA, higher order DINA (yüksek düzeyli DINA), MS-DINA, multistrategy DINA (çoklu stratejili DINA). DINO, Deterministic inputs, noisy 'or' gate. NIDA=Noisy inputs, deterministic 'and' gate, NIDO = Noisy inputs, deterministic 'or' gate, RUM = reparametrised unified model/fusion model (Yeniden parametrelendirilmiş birleşik model). RERUM, random effect RUM (rasgele etki RUM). C-RUM. compensatory RUM (Tamamlayıcı RUM). NC-RUM. noncompensatory RUM (Tamamlayıcı olmayan RUM). Full NC-RUM = Tam NC-RUM. Reduced NC- RUM = İndirgenmiş NC-RUM. GDM = Genel tanısıl model. H-GDM = Hierarchical GDM (hiyerarşik GDM). LCDM = Log-Doğrusal bilişsel tanı modeli (Log-Linear Cognitive Diagnosis Model). MCLCM= Çoklu sınıflama örtük sınıf modeli (multiple classification latent class model) (A. A. Rupp & Templin, 2008).

Telafi edici modellerde, bir maddeyi doğru yanıtlayabilmek için gerekli özelliklerin bir veya daha fazlasına sahip olmak gereklidir yani diğer özelliklerin olmamasını telafi edebilir. Aksine, telafi edici olmayan modellerde, bir özelliğin eksikliği, maddeyi doğru cevaplayabilme açısından diğer özellikler tarafından tam olarak telafi edilemez; Bu durum, maddeyi doğru cevaplayabilmek için tüm özelliklere sahip olma zorunluğu getirir. Genel olarak BTM'ler aynı test içindeki her iki ilişki türüne de izin verir. Örneğin DINA, DINO, NC-RUM, C-RUM ve ACDM gibi birçok spesifik bilişsel tanı modeli, GDINA'dan türetilebilir. Bu nedenle GDINA, aynı testte her bir madde için farklı bir model kullanımına izin verir. Örneğin, bir madde için, DINA modeli tercih edilebilirken, başka bir madde için de DINO modeli tercih etmek daha yerinde olabilir (Ravand, 2015).

2.3.1. İstatistiksel Temeller

Bu bölümde bilişsel tanı modellerinin istatistiksel temelleri üzerinde durulacaktır.

2.3.2. Q Matris

Bir Q matris, temel anlamda öğretmenlerin ve araştırmacıların, bir maddenin doğru yanıtlanabilmesi için hangi özelliklerin gerekli olacağına ilişkin hipotezlerinin bir özetidir (Tatsuoka, 2009). Q matriste belirlenen her bir sütun bir nitelik ya da becerinin vektörüken her satır bir maddeyi temsil etmektedir. Nitelik olarak adlandırılan alan uzmanları tarafından belirlenen özellikler; prosedürler, buluş yöntemleri, stratejiler, beceriler ve başka bilgi bileşenleri olabilir. Q matris'te ele alınan özelliklerin maddede bulunup bulunmama durumları 1-0 şeklinde ikili numaralarla gösterilmektedir. Bu numaralandırma ilk önce Fisher tarafından (1973) "ağırlıklandırma" şeklinde tanımlanmış ve bir k niteliğinin bir i maddesinde bulunduğu durumda 1, bulunmadığı durumda 0 şeklinde kodlandığı belirtilmiştir.

$J \times K$ 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 2'de 3 nitelik ve 5 madde için hazırlanmış örnek bir Q matrisi görülmektedir. Q matrisinde görüldüğü gibi 1. maddeyi doğru cevaplamak için ilk niteliğe sahip olmak gerekmektedir. İkinci madde için ise 1. ve 2. niteliklere öğrencilerin sahip olması gerekmektedir.

Tablo 2. Örnek Q Matris

Maddeler	a_1	a_2	a_3
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1
5	0	0	1

Q matriste tanımlanan k tane nitelik için 2^k tane örtük sınıf belirlenir. Tablo 2.1’de verilen Q matriste 3 nitelik bulunduğu için 2^3 tane örtük sınıf söz konusudur. Bu örtük sınıflar (1,1,1) ,(1,1,0) ,(1,0,1) , (0,1,1), (0,0,1), (1,0,0), (0,1,0), (0,0,0) şeklindedir. Örneğin; (1,0,0) anlamı bu örtük sınıfa dahil olanların 3 örtük özellikten sadece ilkinde sahip olduğu şeklinde yorumlanabilir.

Tatsuoka (1990), sütunların bilişsel görevleri temsil ettiği ve satırların maddeleri temsil ettiği bir Q matrisle maddeleri cevaplamak için gerekli bilişsel işlem becerilerini düzenlemiştir. Her sütundaki girdiler, her bir maddenin çözümü için hangi beceri ve bilginin gerekli olduğunu göstermektedir. Q matristeki 1-0 şeklindeki girdiler için;

$q_{jk} = 1$ olduğu durumda A_k özelliğinin j maddesinde bulunduğu ve $q_{jk} = 0$ olduğu durumda ise A_k özelliğinin j maddesinde bulunmadığı şeklinde yorumlanmaktadır. Tatsuoka’ya göre bu şekilde düzenlenmiş özelliklerin, maddede bulunma durumları bir Boole değişkeni olarak kabul edilmektedir. Bir Q matris aynı zamanda bir Boole matrisidir. Bununla birlikte, bir Q matristeki girdiler sadece bilişsel görevlerle sınırlı değildir. Cinsiyet gibi bazı arka plan değişkenleri kullanılabilir ancak gözlenebilir ve ölçülebilir değişkenlerin Q matriste kullanımı önerilmez. Çünkü Boole hesaplaması çok sayıda kombinasyonla ilgilenir, bu yüzden orijinal Q matrise yeni bir özellik eklenmesi büyük ölçüde yük olarak hesaplamayı arttırabilir (Tatsuoka, 2009).

Örnek:

Kesir problemleri üzerinde üç özelliğin Q matriste gösterimi;

Özellik 1 (A1) : Paydaları farklı iki kesrin paydalarını eşitleyebilme

Özellik 2 (A2) : Tam sayıyı, basit kesre dönüştürme

Özellik 3 (A3) : Paydaları aynı olan iki kesri toplama

İ1: $\frac{2}{3}$ ve $\frac{1}{4}$ 'ün paydalarını eşitleyiniz.

İ2: $\frac{2}{5}$ ile $\frac{1}{5}$ 'i toplayınız.

İ3: 2 ile $\frac{1}{2}$ 'yi toplayınız.

Maddelerin çözümleri;

İ1: $\frac{2}{3}$ ile $\frac{1}{4}$ 'ün ortak paydası 12'dir.

(Burada sadece paydaları farklı iki kesrin paydalarını eşitleyebilme özelliği var. (A1))

$$\text{İ2: } \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

(Burada sadece paydaları eşit kesirlerle toplama işlemi yapabilme özelliği var. (A3))

$$\text{İ3: } 2 + \frac{1}{2} = \frac{4}{2} + \frac{1}{2} = \frac{5}{2}$$

(Burada iki özellik var. Tam sayıyı basit kesre dönüştürme ve paydaları eşit iki kesri toplayabilme. (A1 ve A2))

Bu maddeler için Q matris aşağıdaki gibi oluşturulacaktır.

Tablo 3. Örnek Q Matris

Maddeler	a_1	a_2	a_3
İ1	1	0	0
İ2	0	0	1
İ3	0	1	1

Bu Q matrise bakarak şu yorumları yaparız;

A1 özelliği birinci maddede bulunmaktadır. A3 özelliği 2. Maddeyi doğru yanıtlamak için gereklidir ve 3. Maddeyi yanıtlamak için hem A2 hem de A3 özelliğine sahip olmak gerekmektedir.

Q matriste özellikler vektörler ile temsil edilmektedir.

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ ve } A_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Bu vektörleri kullanım kolaylığı sağlamak için transpozunu alarak yeniden yazarsak;

$$A'_1 = (1,0,0), A'_2 = (0,0,1) \text{ ve } A'_3 = (0,1,1) \text{ şeklinde olur.}$$

Teorik bir notasyonla gösterecek olursak;

$$A_1 = \{i_1\}, A_2 = \{i_2\}, A_3 = \{i_2, i_3\} \quad \text{ayrıca} \quad i_1 = \{A_1\}, i_2 = \{A_3\} \text{ ve } i_3 = \{A_2, A_3\} \text{ şeklinde de ifade edilebilir.}$$

Ölçülen niteliklerle birlikte düzenlenen Q matrisi, madde güçlüğüünün 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 (2004), Q matrisin kalitesinin, öğrencilerin beceri profilinin kestirimini doğrudan belirlediğini ifade etmektedir. Bu nedenle, bu çalışmalar “tanısal analiz” için çok önemli bir nitelik taşımaktadır (De La Torre ve Douglas, 2004).

2.3.3. Rule Space Metodolojisi ve Q Matris

Bilişsel Tanı Değerlendirmeleri; her öğrencinin bilgi durumu ve performansı ile ilgili tanısal bilgi sağlayan bir yöntemdir. Rule Space Metodolojisi (RSM) ise bu bilişsel tanı değerlendirme yöntemlerinden biridir. RSM, iki bölümden oluşmaktadır. Bunlar; Q matris teorisi ve rule space olarak anılan istatistiksel desen sınıflandırmasıdır. Bir Q matris, öğretmenlerin ve araştırmacıların, bir maddenin doğru yanıtlanabilmesi için hangi özelliklerin gerekli olacağına ilişkin hipotezlerinin bir özetidir. Değerlendirme sonrasında elde edilen sınıflandırma sonuçları, Q matrisin belirli bir testin maddeleri üzerindeki özellikleri ne kadar açıkladığı hakkında bilgi verir niteliktedir. Chipman, Nichols ve Brennan'ın (1985) belirttiği gibi RSM bilişsel modellerin bir hipotez testidir.

Q matris teorisinin yararlı uygulamalarından biri, bir Q matristeki tüm özellikler; sütun vektörlerinin bir setinin, alt kümeler (L_A) ile tanımlanan sıralama ilişkisidir. Sıralama ilişkisi; nitelikler, maddeler ve bilgi durumları arasındaki yapıları keşfetmeyi sağlamaktadır. Bir set L , “Boolean Algebra” (Boole Cebiri)

olmak için koşulu yerine getirmektedir. Boole Cebiri, alışılmış toplama ve çıkarma işlemleri yerine veya (meet, union or OR), ve (join, intersection or AND) işlemleri kullanılarak tanımlanan bir matematiksel yapıdır. Daha açık bir ifadeyle, Boole cebiri, kümelerin dahil edilmesiyle tanımlanan bir kümenin alt kümeleri üzerinde “kısmi sıralamadır.” Kısmi sıralama, bazı elemanların bir ağ, bir ağaç figürü ya da hiyerarşik ilişki figürleri gibi sıralanamayacağı anlamına gelmektedir. Bu noktada yapılan işlemler, veya (union, meet, OR), ve (intersection, join, AND) ve tümleyen (complementation, NOT)’dir. (kümelerdeki, birleşim, kesişim, tümleyen) Boole cebiri aynı zamanda bir kafes yöntemidir.

Q matris teorisindeki orijinal set (L_A) bir Q matristeki öznelikleri temsil eden sütun vektörlerinin bir kümesidir ve set, teorik kısmi-sıra ilişkisine sahip sütun vektörlerinin kümesi beceri uzayı (attribute space) olarak adlandırılmaktadır. Madde uzayı (item space), belirli bir Q matristen, benzer şekilde tanımlanmaktadır.

Q matris teorisi, örtük özellik uzayı (L_A) ve gözlenebilir madde uzayı olmak üzere bu iki matematiksel yapı ile, bir Boole tanımlayıcı fonksiyonu (BDF) kullanarak matematiksel bir bağlantı geliştirmektedir. (Tatsuoka, 2009).

2.3.3. Bilişsel Tanı Modeli Türleri

Bu bölümde bilişsel tanı modelinin, araştırmayı temellendiren türleri üzerinde durulacaktır.

2.3.3.1. DINA

DINA, bir sınıflandırma modelidir ve diğer birçok sınıflandırma modeli gibi bağlayıcı (conjunctive condensation) fonksiyonu kullanılan telafisi olmayan (noncompensatory) bir modeldir (Maris,1995,1999). Telafi edici olmayan terimi, katılımcının (respondents) bir davranışa verdiği yanıtı, başka bir davranıştaki fazlalık ile telafi edemeyeceğinden ya da tamamlayamamasından gelir. Bağlayıcı fonksiyonlar; tüm davranışların belirli bir maddede 1 olarak alınması için deterministik bir bakış açısıyla mevcut olması gerektiğine işaret eder. Model, her bir madde için, yanıtlayıcıları madde üzerinde 1 puanını almak için gerekli tüm özelliklere sahip olanlar ve en az bir gerekli özelliğe sahip olmayanlar olmak üzere iki sınıfa ayırır. Gerekli özelliğe sahip olmayanların atandığı ikinci sınıftaki yanıtlayıcılar arasında başka bir ayırım yapılmamaktadır.

DINA modelin temelleri Macready ve Dayton (1977) başta olmak üzere Tatsuoka (1983) ve Haertel (1989) tarafından atılmıştır. Haertel tarafından temellendirilen bu model önce ikili yetenek modeli olarak isimlendirilmiştir. Daha sonraları Junker ve Sijtsma (2001) tarafından geliştirilen model DINA adını almıştır. İsmi "Deterministic Input Noisy And Gate" ifadesinin baş harflerinden alır (A. Rupp ve Templin, 2008). DINA model ikili yetenek modellerini temel alan bir örtük sınıf modeli olması sebebiyle MTK modelleriyle ilişkilidir (Haertel,1989). Ancak DINA model, MTK'da olduğu gibi yeteneklerin sürekli olduğunu varsaymaz, öğrencileri kesin olarak belirlenmiş örtük sınıflara ayırır.

DINA model, madde-özellik ilişkisini temel alır. Özellikler ve maddeleri belirleyebilmek için Q matris denilen bir yapıdan faydalanılır. Q matris her bir maddenin bilişsel spesifikasyonunun yapıldığı bilişsel tasarım olarak da ifade edilebilir. Modelde kullanılan bu yapı şu şekilde oluşturulur:

X_{ij} i kişinin j maddesine verdiği cevap olsun, $i=1,2,3...I$ $j=1,2,3...J$ ve $a_i = \{a_{ik}\}$ de testi alan kişinin ikili beceri faktörü olsun $k=1,...K$ $k=1$ özelliğe sahip olma $K=0$ özelliğe sahip olmama durumunu temsil eder. Çoğu bilişsel tanı modeli uygulaması $J \times K$ 'lık 1-0 Q matrisi oluşturmayı gerektirir (Embretson,1984; K. Tatsuoka, 1985). Matriste j 'ler satırları, k 'lar sütunları, q_{jk} , j maddesini doğru yanıtlamak için gerekli k becerisini ifade eder. Örnekleme gerekirse, ortaokul seviyesinde kesirlerde karışık çıkarma işlemleri öğrenme alanını ele alalım. Bu işlemleri yapabilmek için öğrencilerde şu temel becerilerin olması gereklidir; temel kesirleri çıkartma, sadeleştirme ve genişletme, bileşik bir kesir içindeki tam sayıyı görebilme, tam kısımdan elde alma, bileşik kesri tam sayılı kesre dönüştürebilme. $7\frac{3}{5} - \frac{4}{5}$ işlemini doğru cevaplayabilmek için hazırlanacak Q matrisin satır vektörü (1,0,1,1,0) şeklinde olacaktır. DINA modele göre bu işlemi doğru yanıtlayabilmek için gerekli bu üç özellikten birine bile sahip olunmaması, maddenin yanıtlanamamasına sebep olur.

2.3.3.2. G-DINA

G-DINA (generalized deterministic inputs, noisy "and" gate) modeli, DINA modelin genelleştirilmesiyle türetilmiştir. Doymuş formunda G-DINA model, alternatif bağlantı işlevlerine dayanan bilişsel tanı için diğer genel modellere

eşdeğerdir. Aynı zamanda yine diğer modeller gibi G-DINA model $J \times K$ Q matrisine dayanmaktadır. Q matrisin j satırı ve k sütununun kesişimi q_{jk} ile gösterilmekte ve j maddesini doğru cevaplamak için gerekli k becerisini ifade etmektedir. K eğer 1'e eşitse o maddeyi doğru yanıtlamak için k becerisinin gerektiği, 0'a eşit olma durumu ise k becerisinin o maddeyi yanıtlamak için gerekli olmadığı anlamına gelmektedir.

$$K^{*j} = \sum_{k=1}^K q_{jk}$$

İfadesi j maddesini cevaplamak için gerekli becerilerin sayısını temsil etmektedir. G-DINA modelde, DINA modelde olduğu gibi ayrılmış her örtük grup bir beceri vektörüne indirgenmektedir. Aynı şekilde örtük sınıflar 2^{K^*j} sayıda örtük gruba ayrılır. Her bir örtük grup ise α_{ij}^* ile gösterilen bir beceri vektörü ile gösterilmektedir. G-DINA modelin DINA modelden farkı her örtük grubun kendine ait başarı olasılığının olmasıdır. (D. L. J. Torre ve Rutgers, 2011)

DINA modelde maddeyi doğru yanıtlama olasılığı sadece maddeye ait özelliklerin tümüne sahip olma durumuna bağlıyken G-DINA modelde her bir özellik için maddenin doğru yanıtlanma olasılığı ayrıca hesaplanır. Yani öğrencinin maddeye ilişkin özelliklerden herhangi birine ya da birkaçına sahip olması durumunda maddeyi doğru yanıtlama olasılığı, özelliğin ağırlıklandırılmasına bağlı olmaktadır (De La Torre, 2008). G-DINA için $P(\alpha_{ij}^*)$ temelli olasılık formülü aşağıda verilmiştir;

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

$\delta_{j0} = j$ maddesinin kesişimi

$\delta_{jk} = \alpha_k$ üstündeki temel etki

$\delta_{jkk'} = \alpha_k$ ve $\alpha_{k'}$ nin birbiriyle etkileşiminin etkisi

$\delta_{j12\dots K_j^*} = \alpha_1 \dots \dots \alpha_{K_j^*}$ bağlı karşılıklı etkileşimin etkisi anlamına

gelmektedir.

Formülden de görüldüğü gibi maddenin doğru yanıtlanabilmesi için maddeye ait özelliklerin tümüne sahip olma durumundan ziyade özelliklerin birbirleriyle etkileşimine yani özelliklerin ağırlandırılmasına bağlıdır.

2.3.3.3. RUM

Birleşik model; Tatsuoka'nın Rule Space ve örtük sınıf modelini temel alan ve maddelere verilen yanıtları ve bu yanıtların altındaki yetenekler arasındaki ilişkileri inceleyen bilişsel tanı modelidir (Dibello, Stout ve Roussos,1995). BTM içerisinde, Q matrisin testte yer alan bütün bilişsel gereklilikleri karşılamadığı ve Q matris dışında kalan başka örtük yeteneklerin de farklı parametrelerle temsil edilmesi ve denklemin içinde yer almasının gerektiği ilk kez bu modelle öne sürülmüştür. Birleşik Modelde sınıflama geçerliliğinin, gerçek test ve test tekrar test verilerinin uyumuyla mümkün olduğu görüşü kabul edilmiştir. Kuramsal olarak ilgi çekici olduğu halde bu modeli işlevsiz kılan neden; denklemin içinde yer alan bütün parametrelerin istatistiksel olarak hesaplanamamasıdır (DiBello , 2006). Bu sebeple Hartz (2002) tarafından daha esnek bir model olan Reparametrize Birleşik Model (RUM) geliştirilmiştir. Bu modelle daha çok sınıflamanın bilişsel anlamı üzerine yoğunlaşmıştır. (Roussos, 2007). Bu şekilde model, istatistiksel olarak basitleştirilmiş ancak bilişsel olarak daha karmaşık bir yapıya çevrilmiştir. RUM, tek bir beceri ile öğrenci performansının modellenmesi amaçlanmaktadır. Fakat bu şekilde sürekli olan beceri, kesikli bir yapıymış gibi hesaplanmaktadır. Yani madde parametreleri hesaplanırken cevaplayıcının beceri düzeyi iki kategorili olarak ele alınarak belirlenmektedir. Ayrıca RUM, Q matrisin, maddelerin gerektirdiği becerileri karşılamadığı durumlarda da kullanışlıdır. Q matris tarafından tanımlanmayan becerileri tek boyutlu MTK yöntemleri ile sürekli örtük özellikler olarak modeller (Başokçu, 2016).

2.3.4. DINA Modelde Yetenek Kestirimleri

DINA modelde, her madde için yokluk sınıfı (null class) ve tam sınıf (full class) olmak üzere iki sınıf vardır. Yokluk sınıfı, var olması gereken özelliklerden hiçbirine sahip olmayan yanıtlayanların oluşturduğu sınıftır. Tam sınıf ise tüm özelliklere sahip olanların oluşturduğu sınıftır. Bir maddeyi cevaplamak için gerekli üç özellik olsun. Eğer yanıtlayıcı iki özelliği sahip ama bir özelliğe sahip değilse bile yine yokluk sınıfına ait olur.

Maddeyi doğru cevaplamak için gerekli tüm özelliklere sahip olunma durumunda, maddenin doğru cevaplanma olasılığını gösteren fonksiyon;

$$P[Y_{ij} = 1 | n_{ij}, s_j, g_j] = (1 - s_j)^{n_{ij}} \cdot g_j^{1-n_{ij}} \text{ 'dur.}$$

Tüm becerilere sahip bir yanıtlayıcının, maddeyi doğru cevaplama olasılığı P ile ifade edilmiştir. n_{ij} , α tarafından belirlenen örtük cevap olup i . konunun niteliğini ve

q_j 'nin vektörünü ifade eder.

Q matrisin j . Maddesine karşılık gelen sıranın matematiksel ifadesi şu şekildedir; $n_{ij} = \prod_{k=1}^k \alpha_{ik}^{q_{jk}}$

Tatsuoka'nın terminolojisinde $a_i = (a_{i1}, \dots, a_{ik})$ "knowledge states" yani bilgi durumları ve $n_{ij} = (n_{i1} \dots n_{ij})$ 'yi "ideal response patterns" ideal yanıt desenleri olarak adlandırılır. Çünkü her bir yanıtlayıcının bilgi durumundan performansa dair belirleyici bir tahmini temsil ederler. (Junker, 2000) Burada $a_{ik} = 0$ veya 1 olması i öğrencisinin k niteliğine sahip olma durumuna bağlıdır. j , toplam madde sayısını göstermektedir. Bu maddeler i öğrencisinin istenilen özelliklere sahip olup olmadığını ve Y_{ij} gözlenen puanının belirlenmesini sağlar (J. De Torre ve Lee, 2010).

Belirli bir k niteliği için 2^k kadar olası örtük sınıf vardır. Örneğin 1 nitelik için 2^1 'den 2 tane olası örtük sınıf vardır ve bu da 1 ya da 0'dır. 2 nitelik için 2^2 'den 4 tane örtük sınıf oluşur ve şu şekilde ifade edilir; (1,1) iki niteliğe de sahip olma durumu. (1,0) birinci niteliğe sahip olup ikincisine sahip olmama durumu. (0,1) ilkinde sahip olmayıp ikincisine sahip olma durumu. Son olarak da (0,0) iki niteliğe de sahip olmama durumu. Bu şekilde nitelik sayısına bağlı olarak olası örtük sınıf sayısı arttırılabilir.

Örtük yanıtların oluşmasında öğrenci yeteneği ve maddenin gereklilikleri belirleyici rol oynar.

$$n_{ij} = 0 \text{ olduğunda } g_j$$

$$n_{ij} = 1 \text{ olduğunda } 1 - s_j \text{ olur}$$

Matematiksel olarak örtük yanıt, öğrenci yeteneği ve maddenin gerekliliklerinin bir fonksiyonudur diye ifade edilir (J. De Torre, 2009).

2.3.5. DINA Modelde Parametre Kestirimleri

DINA model isminde de içerdiği noisy and gate unsuru ile maddeye verilen yanıt davranışının deterministik değil olasılıksal olduğuna vurgu yapar. Çünkü maddeye yanıt verenler, madde düzeyinde yanlış pozitif ve yanlış negatif hatalar

yaparlar. Bir maddeye doğru bir şekilde yanıt verme olasılığı, maddeyi yanıtlayan kişiler arasında ayırt ettiği iki gruba da bağlı olarak iki farklı hata olasılığı ile belirlenir. Bu hata olasılıkları tahmin ve kaydırmadır. Bu olasılıklar hesaplandıktan sonra tahmin (guess) parametresi ve kaydırma (slip) parametresi diye isimlendirilir. Kısaca g ve s parametresi diye de anılırlar (A. Rupp ve Templin, 2008).

Tahmin parametresi (g parametresi): g_j ifadesi bireyin j maddesini doğru yanıtlayabilmek için gerekli özelliklere sahip olmadığı halde maddeyi doğru yanıtlamasını ifade eder. Başka bir deyişle o maddeyi, gerekli olduğu düşünülen özellikleri kullanmadan onun yerine başka özellikleri kullanarak doğru yanıtlamak anlamına gelir. Marris (1999) g parametresini zihinsel tahmin yeteneği olarak ifade etmiştir. Ayrıca, doğru pozitif olasılık diye de isimlendirilir. Maddeler için g parametresinin düşük olması sadece gerekli özelliğe sahip yanıtlayıcılar tarafından doğru yanıtlanma olasılığının yüksek olduğu anlamını taşır.

Kaydırma parametresi (s parametresi): s_j ifadesi bireyin j maddesini doğru yanıtlamak için gerekli özelliklere sahip olduğu halde o maddeyi yanlış yanıtlamasını ifade eder. Bir diğer ifade şekli yanlış pozitif olasılıktır. Maddeler için s parametresinin düşük olması, gerekli özelliğe sahip bireylerin maddeyi doğru yanıtlama olasılığının yüksek olacağı anlamına gelir.

s ve g parametreleri aşağıdaki gibi gösterilmektedir.

$$g_j = P[Y_{ij} = 1 | n_{ij} = 0]$$

$$s_j = P[Y_{ij} = 0 | n_{ij} = 1]$$

Y_{ij} , j yanıtlayıcısının i maddesine verdiği yanıtı temsil eder. Yukarıda verilen denklemlerden yola çıkarsak, yanıtlayıcının maddeyi doğru yanıtlama olasılığı $P[Y_{ij} = 1 | n_{ij}] = (1 - s_j)^{n_{ij}} \cdot g_j^{1-n_{ij}}$ olur. $(1 - s_j)$ ifadesi kaydırmama olasılığıdır. Eğer yanıtlayıcı, madde üzerindeki gerekli tüm özelliklere sahip ise, cevap olasılığı $(1 - s_j)^1 \cdot g_j^0 = (1 - s_j)$ olur. Yanıtlayıcının özelliklere sahip olmama durumunda ise $P = (1 - s_j)^0 \cdot g_j^1 = g_j$ olur (Yeongyu Lim, 2015).

2.4. Madde Formatı

Madde formatlarının çeşitli sınıflandırılmaları bulunmaktadır. Çalışmada madde formatlarının sınıflandırılması (Osterlind, 2012) ve (Downing & Yudkowsky, 2009)'nin üzerinde çalışmış olduğu sınıflandırmaya dayalı olarak yapılmıştır. Bu

sınıflandırmada madde formatları; yapılandırılmış yanıtli ve seçme yanıtli maddeler olarak ikiye ayrılmıştır. Aşağıdaki tabloda bu iki türün bazı karakteristik özellikleri verilmiştir.

Tablo 4. *Yapılandırılmış ve Seçme Yanıtli Maddelerin Güçlü Yanları ve Sınırlılıkları*

	Yapılandırılmış Yanıtli	Seçme Yanıtli
Güçlü Yanları	<ul style="list-style-type: none"> • İpucu içermeme • Oluşturması kolay • Mantık, muhakeme, adımlı problem çözme • Kısmi puanlama kolaylığı • Ayrıntılı değerlendirme 	<ul style="list-style-type: none"> • İçeriği geniş temsil etme • Doğru, nesnel ve tekrarlanabilir puanlar • Savunabilirlik • Doğru ve zamanında geri bildirim • Zaman ve maliyet açısından ekonomiklik
Sınırlılıkları	<ul style="list-style-type: none"> • Puanlamanın öznel olması • İçeriği sınırlı temsil • Tekrarlanması güç • Uygulaması, puanlaması zaman gerektirir • Sınırlı psikometrik ve kalite kontrolü 	<ul style="list-style-type: none"> • Maddenin yazımı zor • Doğru yanıt tahmin yoluyla bulma • Akılda kalıcı***

*Tablo Assessment in Health Profession Education kitabından alınmıştır (Downing & Yudkowsky, 2009).

2.4.1. *Yapılandırılmış Yanıtli Maddeler ve Türleri*

Yapılandırılmış yanıtli maddeler; öğrencilerin yanıtları, seçenekler arasından seçmesinden ziyade sahip olduğu bilgi birikimi ile kendinin oluşturduğu madde formatıdır. Bu madde türünde beklenen, öğrencinin kapsamlı ve doğru bir yanıt oluşturmasıdır. Böyle bir yanıtın değerlendirilmesi için konu alanında uzman bir kişinin olması gerekmektedir (Hickson, Reed, ve Sander, 2010). Ülkemizde yapılan geniş çaplı sınavlarda yapılandırılmış yanıtli madde kullanımı hala deneme aşamasındadır. Bunların dışında PISA ve TIMMS’de yapılandırılmış yanıtli madde kullanıldığı görülmektedir.

Yapılandırılmış yanıtli maddeler genellikle uyarıcı olarak adlandırılır çünkü bu uyarılar performans testinde birçok biçimde yer alabilir: yazılı sorular, fotoğraflar, veri tabloları, grafikler, çeşitli türdeki etkileşimli bilgisayar uyarıları vb.

Bu genel uyarılar, yapılandırılmış bir yanıt istemek için kullanılır ve daha sonra puanlanabilir (Downing ve Yudkowsky, 2009).

Yapılandırılmış yanıtı madde formatının kendi içerisinde türleri vardır. Aşağıdaki Tablo 5’te Osterlind’in derlemiş olduğu yapılandırılmış yanıtı madde türleri verilmiştir.

Tablo 5. *Yapılandırılmış Yanıtı Madde Türleri*

Cloze Prosedürü	Gözlemler
Tartışmalar	Sözlü Raporlar
Açık Uçlu Sorular	Performanslar
Sergiler	Portfolyolar
Deneyler	Projeler
Boşluk doldurma	Araştırma kağıtları
Grid in response	Kişisel / Akran Değerlendirmeleri
Etkileşimli Video	Kısa cevap
Röportajlar	Örnek Yazma

*Tablo Constructing Test Items (Multiple-Choice, Constructed Response, Performance and Other Formats) kitabının 6. Bölümünden alınmıştır. (Osterlind, 2012)

Tabloda verilen her bir madde türünün seçimi yoklanmak istenen niteliğe göre değişiklik gösterir. Örneğin açık uçlu sorular (essays) herhangi bir konu alanında ayrıntılı ve derinlemesine analiz yapmaya olanak sağlamaktadır. Bu maddeler öğrencilerin yaratıcı yazma becerilerinin de değerlendirilmesine izin vermektedir. Böylece, öğrencilerin analiz, sentez ve değerlendirme becerileri yoklanabilmektedir. Kısa cevap ve cümle tamamlama madde türlerinin doğru şekilde kullanılması ile hatırlama gerektiren alanların değerlendirilmesi yapılmaktadır (Osterlind, 2012). Portfolyo, öğrencilerin herhangi bir dersteki başarılarını ve yeteneklerini gösteren o dersle ilgili birikimlerini bir dosyada tutmasıdır. Portfolyo bir madde özelliği göstermekten daha çok, üst düzey düşünme süreçlerinin ölçülmesi amacıyla oluşturulmuş bir kavramdır (Haladyna, 1997).

Yazmanın öğrenme sürecinde kullanımı temel ve etkili bir yöntemdir. Çünkü bu süreçte sorgulama ve açıklama yapma gibi becerileri aktif hale getirir. Ancak bu etkili yöntem, yazma eyleminin ön planda olduğu yapılandırılmış yanıtı maddeler aracılığıyla olması gerektiği gibi doğru ve yerinde kullanılmamaktadır (Fisher ve Frey, n.d.). Çünkü yapılandırılmış yanıtı madde formatı, tüm sınav formatları arasında en yanlış anlaşılan, bu sebeple kötüye kullanılmış olan madde türlerinden

biridir. Badgett ve Christmann(2009)'a göre bu madde türünün mutlak güçlü yanları olmasının yanı sıra sınırlılıkları ve yanlış kullanıldıkları durumlarda olumsuz yanları da bulunmaktadır. Yapılandırılmış yanıtli madde türünün yanlış kullanılma sebeplerine öncelikle madde oluşturma ve uygulama sonrasında maddeyi puanlamanın zorluğu atfedilebilir (Badgett & Christmann, 2009). Bu madde türünün sınırlılıklarına ise tek bir seferde kısıtlı miktarda soru sorulabilmesi örnek verilebilir. Çünkü bu madde türünde öğrencilerin yanıtları oluşturmaları zaman alıcıdır. Bir diğer sınırlılık ise cevapların değerlendirilmesinin maliyetli oluşudur. Yapılandırılmış yanıtli maddelerin en büyük dezavantajlarından biri puanlamanın uzun süre gerektirmesidir. Her öğrencinin vereceği yanıt birbirinden farklı olabileceği için kısmi puanlama daha fazla zaman gerektirir.

Yapılandırılmış yanıtli maddelerin güçlü yanlarına, hatırlama gerektiren becerilerin yanı sıra geri çağırma becerilerin de sınanmasını sağlaması verilebilir. Geri çağırma, maddeyi yanıtlamak için gerekli birkaç ipucunun karşılığında belli zihinsel süreçlerin sonrasında oluşturulan tepkilerdir. Hatırlama ise alternatifler arasından seçme durumunda etkin olan zihinsel durumdur. Bu sebeple geri çağırma becerisi; yapılandırılmış yanıtli maddelerin yanıtlanabilmesi için gerekli zihinsel süreçleri kapsamaktadır (Sternberg ve Sternberg, 2011). Chaoui (2011), "Etkili eğitimin bir parçası, öğrencilere düşüncelerini yazılı, açıklayıcı, çok adımlı, düzenleyici ve yazılı cümleler kullanarak açıklamak için fırsatlar sunmaktır" demiştir. Bu sebeple değerlendirme yaparken yapılandırılmış yanıtli maddeleri doğru ve yerinde kullanabilmek etkili bir eğitim sürecinin önemli bir işlevini yerine getirecektir.

2.4.1.1. Açık Uçlu Maddeler

Klasik bir yapılandırılmış yanıtli madde türüdür. Bir açık uçlu soru, belirli bir konuya odaklanmış, doğrudan ilgili sorudan oluşur ve soruları yanıtlayacak kişiye yeterli bilgileri sağlar. Sorunun cevaplanmasıyla ilgili tüm talimatlar, örneğin beklenen cevap süresi, zaman sınırlamaları ve beklenen spesifik cevap açıkça belirtilmelidir (Race, Brown, ve Smith, 2005). Açık uçlu maddelerin iki türü vardır. Kısa cevaplı açık uçlu madde ve uzun cevaplı açık uçlu madde. Uzun cevaplı madde, yanıt olarak 1-2 sayfa ya da daha fazla yazmayı gerektirirken, kısa cevaplı açık uçlu

maddeler ise 1-2 paragraftan oluşan yanıtlar gerektirmektedir (Downing, 2009). Downing ve Yudkowsky'e göre yapılandırılmış yanıtli maddenin uzunluđu ile ilgili karar vermede; deđerlendirmenin amacı, örneklem büyüklüđu ve kapsama bađlı geçerlik dikkate alınır. Örneđin başarı deđerlendirmesi yapılırken, örneklemin büyüklüđu önemlidir. Çünkü testin amacı, sınava giren kişinin sınırlı örnekleminde geniş içerik bilgisi vermek üzere genelleştirmektedir. Kısa cevaplı açık uçlu sorular, uzun yanıtli sorulardan daha geniş içerik örneklemesine izin vermektedir. Çünkü sınırlı sürede daha fazla soru sorulabilmekte ve cevaplanmaktadır. Testin amacı spesifik bir bilgi alanını derinlemesine örnekleme ise, uzun cevaplı maddeler tercih edilmelidir. Uzun cevaplı açık uçlu maddeler, sınava giren kişiden, tek bir konu veya içerik alanı hakkındaki bilgilerinin sınırlarını ve derinliklerini sorgulayan büyük detaylı cevaplar üretmesini istemektedir (Downing ve Yudkowsky, 2009).

Aşađıda Race'in (2005) açık uçlu maddelerin avantaj ve dezavantajlarını açıklamak için oluşturduđu tablo verilmiştir.

Tablo 6. Açık uçlu maddelerin avantaj ve dezavantajları

AVANTAJ	DEZAVANTAJ
<p>Açık uçlular, öğrencilerin cevaplarının öznel olmasına ve kendilerini ifade etmesine izin verir: Bu sorular, öğrencilerin kendilerinin ayırt edebilmesini sağlayan “en iyi” araçtır. Bununla birlikte, açık uçluların cevaplanma ölçütlerinin, öğrencinin yanıtlarının özneliğini adil bir şekilde puanlayacak kadar esnek olması gerektiği anlamına gelir.</p>	<p>Geçerlik: Açık uçlu madde yanıtlanırken güçlü bir giriş, tutarlı ve mantıklı bir gelişme ve kararlı bir sonuç iyi bir birleşimdir. Ancak bazı yanıtlayıcılar bu madde türünde yeterince deneyimli olmadıkları için diğer öğrencilere göre dezavantajlıdır. Yoklanmak istenen özelliği çok fazla açık uçlu madde ile değerlendirmek aslında hep aynı işaretçileri ölçmek anlamına gelecektir. Bu durumda sahip olunan diğer güçlü ve zayıf yönlerden bağımsız olarak yazma biçimlerini geliştirilmiş öğrenciler daha fazla puan almış olur bu da geçerliliği olumsuz anlamda etkiler.</p>
<p>Açık uçlular, öğrencilerin öğrenme derinliğini yansıtabilir: Bir konu hakkında, öznel bir şekilde yazı yazabilme ilgili konunun anlaşıldığını ve kavrandığını gösteren bir süreçtir.</p>	<p>Açık uçlu soruları, objektif olarak değerlendirmek zaman alır: Öncesinde iyi düşünülmüş değerlendirme ölçütleri bile, gerektiği gibi çalışmayabilir.</p>
<p>Açık uçlu sorular öğrencilerin cevaplarında tutarlılık sergilemelerini sağlar: Öğrencilerin akıcı ve mantıklı argümanlar oluşturma kabiliyetine dair kanıt niteliğindedirler.</p>	<p>Halo etkileri : Yanıtlanan son açık uçlu sorunun cevabı çok iyiye, puanlayıcının bir sonraki soruya daha büyük beklentilerle yaklaşması mümkündür. Bu sebeple de puanlamada gereğinden fazla puan verme eğiliminde olabilir.</p>
<p>Özet: Öğrendiklerini doğru şekilde özetleyebilmek önemlidir. Açık uçlular, öğrencilerin düşüncelerinin farklı alanlarını tek bir argüman haline getirme olanağı sağlar.</p>	<p>Açık uçlu soruların yanıtlarını yazmak zaman alır: Bu sebeple az soru sorulur. Bu durum da değerlendirmede, direk olarak kapsanan müfredatın çok azı test edilir. Test edilmemiş konu kalabilir.</p>

*Tablo 500 Tips on Assessment kitabının 3.bölümünden alınmıştır. (Race et al., 2005)

2.4.1.2.Boşluk Doldurma Maddeleri

Boşluk doldurma maddeleri cümle tamamlama maddeleri ve kısa cevap maddelerinden oluşur. Boşluk doldurma maddeleri adı, alan uzmanlarınca önerilmemesine karşın bu isimle anılması yaygın olarak görülmektedir. Kısa cevap maddeleri ve cümle tamamlama maddeleri, hem kavramsal olarak hem de birçok kullanım açısından, benzer bir formatı paylaşmaktadır. Yapıları gereği, kısa cevaplı maddeler ve cümle tamamlama maddeleri hatırlama gerektiren bilişsel süreci gerektirmektedir. Çünkü bu formatta sadece tek bir kelimeyle veya birkaç cümle ile yanıtın ifade edilmesi beklenir. Kavramsal bilgi ve becerilerin yoklanması için kullanmak uygundur. Cevapların puanlanmasının kolaylığı ve birçok bilginin değerlendirilmesine imkan sağlaması yönünden tercih edilen bir madde formatıdır. Bununla birlikte, kısa cevaplı veya cümle tamamlama formatını doğru-yanlış veya çoktan seçmeli formatlara dönüştürmek mümkündür. Bu şekilde dönüştürülmüş maddeler düşük seviyede geri çağırma bilişsel süreci gerektirir hale gelmiş olur (Osterlind, 2012).

2.4.2. Seçme Yanıtlı Maddeler ve Türleri

Seçme yanıtlı maddeler; bir maddeye verilebilecek olası yanıtların içerisinde doğru ya da en iyi yanıtı seçmeyi gerektiren madde türüdür (Downing & Yudkowsky, 2009). En çok bilinen ve yaygın olarak kullanılan seçme yanıtlı madde türü çoktan seçmeli maddelerdir. Çoktan seçmeli maddelerden, ilk olarak Frederick J.Kelly 1914'te bahsetmiştir. Geliştirilme amacı, 1.Dünya Savaşında askeri personeli işe almak içindir. İlk çoktan seçmeli test; Army Alfa Test olarak bilinmektedir (Madarus ve O'Dwyer,1999). Daha sonrasında, çoktan seçmeli maddelerin kullanımı 1950'lerin ortalarında yüksek hızlı optik okuyucuların icadı ile daha da artmıştır (Baker,1971). Günümüzde ise gerek okullarda gerekse geniş çaplı sınavlarda yaygın olarak kullanılmaktadır. Bu maddelerin yaygın olarak kullanılmasının nedenleri; temel içerik alanlarının her biri için bilişsel düzeylere kolayca uyarlanabilirliği, seçme yanıtlı maddelerin yapılandırılmış yanıtlı maddelere göre daha nesnel olması, seçme yanıtlı maddeleri hazırlamanın ve değerlendirmenin ekonomikliği verilmektedir. Seçme yanıtlı maddeler yüksek güvenilirliği olan objektif bir değerlendirme biçimidir (Wilson ve Wang,1995). Ayrıca, bilimsel kavram ve gerçekleri anlama, anladıklarını uygulama ve aynı zamanda birleştirilmiş bütünleri

bağlantılı ilişkilere ayırma yeteneğini ölçmek için de kullanılmaktadır. Bunun yanı sıra taksonominin hiyerarşik seviyelerinin her birinde nispeten kısa bir süre içinde geniş bir madde örnekleme sağlamaktadır. Fakat sentez ve değerlendirme basamakları bu seviyelerin dışındadır (Badgett ve Christmann, 2009).

Seçme yanıtı maddelerin sınırlılıkları da vardır. Nesnelliklerinden ötürü, öğrencilere cevaplarını açıklamak için fırsat sunmaz, bu durumda da potansiyel olarak öğrencilerin topladığı bilginin derinliği ve kapsamını sınırlandırmaktadır. Bilimsel değerlendirmede, seçme yanıtı maddeler farklı gerçeklere odaklanma eğilimindedir ve karmaşık argümanlar veya tutarlı anlayış gibi araştırma biliminin belirli yönlerini ölçen farklılıkları bulunmaktadır. Seçme yanıtı maddeler de öğrencilerin seçimlerini açıklamak ya da doğrulamak için yapılan akıl yürütme işlemi yetersiz kalmaktadır. Öğrencilerin akıl yürütmeleri konusundaki yetersizlik, özetleyici (summative) değerlendirmeler için sorun olmayabilir. Ancak çoğu sınıf temelli değerlendirmeler için, öğretmenlerin öğretim amaçlarına yönelik öğrencilerin akıl yürütme şekillerini görme ve anlamaları önemlidir. Sınıflardaki öğrenmeleri değerlendirmek için çoktan seçmeli maddeleri tercih etmek bilimsel gerçekleri derinlemesine özümsemekten ziyade, bilimin gerçeklerini ezberlemelerine teşvik edebilir (Linn ve Hsi,2000;Nichols ve Sugrue,1999; Resnick ve Zurawsky,2007).

2.4.2.2.Çoktan Seçmeli Maddeler

Çoktan seçmeli maddeler yaygın kullanılan bir seçme yanıtı madde türüdür. Diğer tüm seçme yanıtı madde formatları çoktan seçmeli madde formatının bir versiyonu olarak kabul edilmektedir. Bunlara doğru-yanlış, alternatif seçenek, eşleştirme maddeleri örnek verilebilir (Downing, 2009). Çoktan seçmeli madde, problem durumu veya eksik ifade içeren bir kök ve alternatif çözümü bulunduran seçeneklerden oluşur. Seçenekler bir doğru yanıt ve çeldiricileri içermektedir. Çeldiricilerin amacı doğru cevaptan emin olmayan ya da doğru cevaba ulaşacak bilgiye sahip olmayanları şaşırtmaktır (Eignor ve Hambleton, 2007).

Standartlaştırılmış başarı ve yetenek testlerinde yaygın olarak kullanımının yanı sıra okullarda da eğitim çıktılarını değerlendirmek adına yaygın olarak kullanılmaktadır. Bu duruma uygulamasının kolay olması, değerlendirmenin zaman almaması öncelikli olarak örnek verilebilir. Ancak tüm bunlar en iyi format olduğunu göstermez. En iyi madde formatı; sınavın amacına uygun en doğru değerlendirmeyi

sağlayacak olmalıdır. Yine de çoktan seçmeli maddelerin en güçlü yanı ana içerik alanların her birindeki bilişsel düzeylerinin akademik başarı ölçümüne kolaylıkla uygulanabilirliğidir (Badgett ve Christmann, 2009).

Çoktan seçmeli maddeler, üzerinde bazı düzenlemeler yaparak üst düzey düşünme becerilerini ölçmek için de kullanılabilir. Bu düzenlemelerden kasıt seçeneklerin homojenleştirilmesi bir diğer deyişle çeldiricilerin kuvvetlendirilmesidir. Böylece maddelerin ayırt edici olması sağlanmaktadır. Bu şekilde bir maddenin tasarlanmasının elbette zorlukları bulunmaktadır. Çünkü oluşturulan seçeneklerin hem homojen olması hem de birinin en iyi seçenek (doğru seçenek) olması gerekmektedir. Aynı zamanda da çeldiricilerin benzer olması, (biri diğerine göre belli başlı yönlerden ön plana çıkmamalı) gerekmektedir. Bu şartlara uygun çeldiricileri oluşturmak zordur. Bu yüzden dört seçenekli bir maddede iki çeldiricinin açıkça yanlış olduğu görülüyorsa, maddenin formatının çoktan seçmeli olması yerine doğru-yanlış formatında ya da alternatif seçim madde formatında olması daha doğru olur (Badgett ve Christmann, 2009).

2.4.2.3. Alternatif Seçim Maddeleri

Alternatif seçim maddeleri aslında iki seçenekli çoktan seçmeli madde türüdür. Ebel(1982), alternatif seçim madde formatının yazılması ve değerlendirilmesinin kolaylığından dolayı kullanımı elverişli bir format olduğunu öne sürmüştür. Ayrıca bu madde türü çeşitli içerik ve bilişsel becerileri ölçmek için kullanılmaktadır. Haladyna ve Downing(1993) tarafından yapılan bir araştırma ile işlevsel olmayan çeldiriciler kaldırıldığında aslında çoğu çoktan seçmeli maddenin alternatif seçim maddesi olduğu görülmüştür. Alternatif seçim maddelerinde şans ile doğru yanıt bulma çoktan seçmeli maddelere göre daha yüksek olasılıklı olarak gözükse de testin uzunluğunu doğru bir şekilde belirledikten sonra bu durum dengelenebilmektedir (Haladyna, Downing ve Rodriguez, 2004).

2.4.2.4. Doğru-Yanlış Maddeleri

Doğru-yanlış maddeleri, verilen önermeleri doğru veya yanlış olarak cevaplamayı gerektiren seçme madde formatı türüdür (Frisbie, 1992). Basit ama doğrudan yoklama işlevi vardır. Ebel (1979), D-Y maddelerinin daha yüksek bilişsel seviyeleri değerlendirebileceğini ileri sürmüştür. Yeterli sayıda D-Y maddeleri ile tahmin yoluyla doğru cevaba ulaşmaya bağlı düşük güvenilirlik ve kapsam

geçerliliğinin etkisinin en aza indirileceğini belirtmiştir. Ayrıca Ebel ve Frisbie (1986) D-Y maddelerinin eğitimin temel sonuçlarının basit ve doğrudan bir ölçümünü sunduğunu öne sürmüştür. Genellikle D-Y maddelerinin yetkin bir format olarak görülmemesinin sebebinin maddenin formatıyla ilgili bir kusur olmadığı ancak vasıfsız soru yazarları kaynaklı olduğu görüşünü belirtmiştir (Downing, 1992). Ebel (1972) bu formatın konu içeriğini uygun şekilde kapsayabileceği on bir alan tanımlamıştır. Bu alanlara; bir konudaki genellemeler, bir konu alanındaki kavramlar arasındaki karşılaştırmalar, bir kavram veya ilkenin örnekleri, bir süreç veya prosedürdeki adımlar örnek verilmektedir (Osterlind, 2012). Doğru-yanlış ve evet-hayır temel olarak iki tür doğru yanlış madde formatı vardır. Doğru-yanlış formatındaki maddeler basit bildirim içeren cümlelerdir. Evet-hayır formatı ise sorgulayıcı yapıdadır.

2.4.2.5. Eşleştirme Madde Formatı

Eşleştirme formatı, bir dizi ifade ya da sorudan oluşan eşleşme gövdesini izleyen bir dizi seçenek gerektirir. Sınıf içerisinde değerlendirme yapmak için kullanım için idealdir. Ancak standartlaştırılmış başarı testlerinde kullanımı önerilmemektedir (Haladyna, 2004). Eşleştirme maddeleri kavramsal olarak zor değildir. Genel yargı, bu formatın olumsuz yanının sadece düşük seviyeli, gerçek içerikle sınırlandırılmış olması yönündedir. Ancak yapılan araştırmalar bu yargının tamamen doğru olmadığı yönündedir. Bazı durumlarda eşleştirme maddelerini doğru yanıtlamak için karmaşık düşünme becerileri gerekmektedir. Sınava giren, her bir kategori ile ilgili bilgileri sentezlemek için karmaşık bir düşünme becerisine ihtiyaç duymaktadır ve belirli bir özelliğin ifade ile ilişkili olup olmadığını muhakeme etmek durumundadır. Eşleştirme maddeleri üst düzey düşünme becerilerini değerlendirebilmesine rağmen düşük seviyeli becerileri değerlendirmek için karakteristik özelliklere sahip olması nedeniyle az miktarda ve diğer madde formatlarıyla birlikte kullanımı önerilmektedir (Osterlind, 2012).

2.4.3. Seçme ve Yapılandırılmış Yanıtlı Maddelerin Farkları

Seçme yanıtlı ve yapılandırılmış yanıtlı madde formatlarının farklı etkilerini ölçmek için birçok çalışma yapılmaktadır. Bu çalışmaların çoğu bir grup öğrenciye seçme yanıtlı ya da yapılandırılmış yanıtlı madde testlerinden oluşan bir sınav yapma ve sonuçlarını karşılaştırma şeklinde olmaktadır. Ancak bu konu ile ilgili kesin bir

bilimsel dayanak yoktur. Tartışmanın iki tarafını da destekleyen bir çok araştırma bulunmaktadır. Bu nedenle eğitim sistemi içerisinde bu konuda çok fazla tartışma süregelmektedir.

Test oluştururken madde formatı seçimi oldukça önemlidir. Madde formatı seçiminde temel ilke, içeriği ve bilişsel süreci ölçmek olmalıdır. İçeriği temsil edebilecek ve gözlemlenmesi amaçlanan bilişsel süreci ortaya çıkaracak format en iyi formattır (Haladyna, 2004).

Bu iki test formatını karşılaştırırken ortaya atılan konulardan biri, problemlerin çözümünde yer alan bilişsel süreçtir. Çünkü, “Çoktan seçmeli testler, simple factual recognition; yapılandırılmış yanıtli maddeler, higher order thinking becerisi gerektirir (Chaoui, 2011).

Matematik eğitiminde, öğrencilerin sadece problem çözme yetenekleriyle ilgilenmek ve onların açıklamalarına yeterince odaklanmamak yaygın olarak görülen bir durumdur. Ancak amaç öğrencilerin bilgileri kalıcı bir şekilde öğrenmesi ve onları gerçek hayat durumlarına ilişkilendirebilmesi ise o zaman bu süreçte kazandırılması gereken beceri, bildiklerini açıklayabilmeleridir. Chaoui’ye göre matematiksel problemleri ve bilimsel düşünmeye teşvik eden performans görevlerini açık uçlu maddeler aracılığıyla yoklamak, öğrencileri daha özgüvenli ve etkili problem çözümler olarak yetiştirmektedir. Bu sebeple öğretmenler, matematik okuryazarlığını ve matematik becerilerini değerlendirmek için yaptıkları sınavlara çoktan seçmeli maddelerin yanı sıra açık uçlu maddeleri de dahil etmelidirler. (Chaoui,2011) O’Neil ve Brown (1997) çalışmasında, metacognition (üst biliş) konusundaki matematik değerlendirmesinde çoktan seçmeli ve yapılandırılmış yanıtli madde formatlarının hepsinin farklı etkilerini incelediği yaklaşımları ele almıştır. Çalışmada yapılandırılmış yanıtli maddelerin, çoktan seçmeli maddelerden daha çok bilişsel strateji kullanımını tetiklediği ortaya koyulmuştur. Bunun bir anlamı da öğrencilerin, yapılandırılmış yanıtli maddelerin çoktan seçmeli maddelere göre maddeyi yanıtlayabilmek için kendi düşünme becerilerini geliştirmelerine yardımcı olduğu anlamına gelmektedir (Badgett ve Christmann, 2009). Bununla birlikte, yapılandırılmış yanıtli maddeler seçme yanıtli maddelere kıyasla öğrencilerin kendi bildiklerini değerlendirme konusunda daha avantajlıdır. Ayrıca yapılandırılmış yanıtli maddeler öğrencilerin tüm yeteneklerini sergilemeleri için fırsatlar

sunmasından dolayı daha özgündürler. Üstelik öğrencilerin karmaşık bilim fikirlerini tanımlamaları, bilimsel konularda açıklama yapmaları ve öğrencilerin bilimsel kanıtlar kullanarak çıkarımlarını ayrıntılandırmaları için bağlam yaratmaktadırlar.

Yapılandırılmış yanıtli maddeler ile öğrenci yanıtları analiz edilip, öğrencilerin yanlış öğrenmeleri ve tutarsız anlayışları öğretmenler tarafından belirlenebilmektedir (Black ve William, 1998). Bu anlayışla aslında yanlış cevaplar, öğrenme durumları hakkında daha fazla bilgi vermektedir. Öğrenci, çoktan seçmeli testte, yanlış bir cevabı seçtiğinde bu cevaba nasıl ulaştığı görülememektedir. Ancak yapılandırılmış yanıtli maddelerde öğrencinin tam olarak nerede hata yaptığı ya da yanlış öğrenmeye sahip olduğu görülmektedir. Bu yüzden yapılandırılmış yanıtli maddeler, öğrencilerin yanlış öğrenmeleri konusunda uygulayıcıya bilgi sağlamaktadır ve öğretmenin kullandığı teknikleri değerlendirmeye izin vermektedir (Chaoui, 2011). Bunu görmek sadece öğretmenin değil öğrencinin de yararına olmaktadır. Çünkü bu sayede öğrenci tam olarak nerede yanlış öğrenmesi olduğunu görmekte ve aynı durumda tekrar benzer hatayı yapma olasılığı düşmektedir.

Seçme yanıtli ve yapılandırılmış yanıtli maddelerin en temel farklarından biri de çeldiricilerdir. Seçme yanıtli madde formatında bazen maddeyi zorlaştırmak için doğru seçeneğe çok yakın bir yanlış seçenek eklenir. Böyle bir durum, öğrenci maddeyi doğru yanıtlamak için gerekli becerilere sahip olduğu halde, öğrencinin yanlış yanıtlamasına neden olabilir. Böyle bir durumda eğer madde, yapılandırılmış yanıtli madde türünde sorulsaydı doğru cevap alınabilirdi (Frery,1985) .Başka bir çalışmada çoktan seçmeli maddelerin öğrenciler tarafından yanlış bilgi üretmelerine yol açabileceği şeklinde bir çıkarımda bulunmuştur. (yanlış öğrenme) (Roediger, 2005)

Yapılandırılmış yanıtli maddelerin puanlanma yöntemlerinin çoktan seçmeli maddelere göre farklılaşması yapılandırılmış yanıtli maddelerin daha zor olduğu yönünde bir algı yaratmaktadır. Çünkü; çoktan seçmeli maddelerde tahmin ile doğru yanıtı bulabilme olasılığı yapılandırılmış yanıtli maddelere göre fazladır. Öğrenci, soru hakkında bilgi sahibi olmadığı halde yüksek puan almaktadır ve bu durum öğrencinin konu hakkındaki gerçek bilgisi ile ilgi doğru bir bilgi yansıtmaz. Bu yüzden puanların, yapılandırılmış yanıtli maddelere göre yüksek olabileceği açıktır. Ancak yapılandırılmış yanıtli maddeler de yanlış maddelere kısmi puanlama

sağlayabilme açısından geleneksel çoktan seçmeli maddelerin puanlanmasından ayrılmaktadır. Araştırmacılar, öğrenme ve puan arasındaki ilişkiyi belirlemeye çalışmışlardır ve bunun zor olduğu sonucuna ulaşmışlar. Genel olarak elde edilen sonuçlar, çoktan seçmeli maddelerden elde edilen puanların, yapılandırılmış yanıtli maddeler ile hazırlanmış sınavlara göre daha yüksek çıkma eğilimde olduğu yönündedir.

Çoktan seçmeli maddeler; genellikle puanlama kolaylığı, puanlama tutarlılığı, çok sayıda konuyu kapsayabilmesi ve genel olarak yüksek puanlı sonuçlar sağladığı için tercih edilmektedir. Yapılandırılmış yanıtli maddeler ise çoktan seçmeli maddelerden puanlama nesneliliği bakımından farklılaşmaktadır (Rodriguez, 2003).

Her ne kadar çoktan seçmeli ve yapılandırılmış yanıtli maddeler için öğrencilerin verdiği yanıtların biçimleri farklı da olsa, bu iki madde türünün yapısal eşitliğine dair kanıtlar vardır. (Construct equivalence) . Çoktan seçmeli ve yapılandırılmış yanıtli maddeler arasındaki yapı eşdeğerliğini araştıran çalışmaların geniş çaplı bir gözden geçirilmesi ile Rodriguez (2003), çoktan seçmeli ve yapılandırılmış yanıtli maddelerin aynı item stem i paylaştığı zaman mean corrected correlation .95 gibi yüksek olabileceğini bulmuştur. Thissen, Wainer ve Wang'ın (1994) belirttiği gibi tanıma (recognition), üretim (generation) ile aynı değildir fakat yüksek düzeyde ilişkilidirler. Her iki formatın da standart bilimsel testlerde genel bilişsel kurguların ölçütü olarak karşılaştırılması ikisi arasındaki benzerliği de doğrulamıştır (Liu, Lee, Linn, ve Liu, 2011).

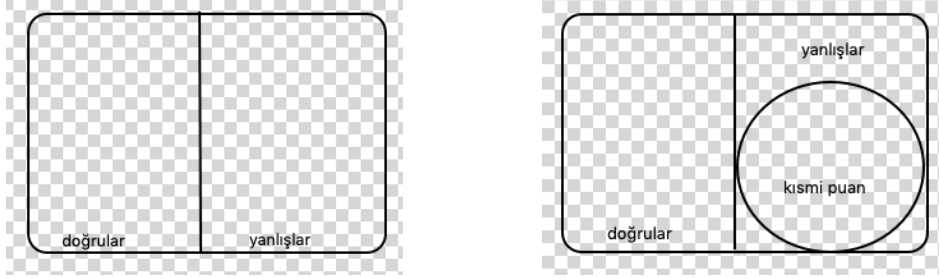
2.4.4. Kısmi Puanlama

Ölçme ve değerlendirme yaklaşımları incelendiğinde temel amaçlardan biri; madde ile ölçülmesi hedeflenen özelliğe sahip oluş derecesi hakkında mümkün olduğunca fazla bilgiye ulaşmaktır. Bu bağlamda yanıtların incelenmesi ve puanlanması ölçülmesi hedeflenen özelliğe sahip oluş ile ilgili bilgi konusunda kritik bir noktadır.

Geleneksel puanlama yöntemlerinden en yaygın olarak kullanılanı doğru ve yanlışın 1-0 olarak puanlanmasıdır. Bu yöntemle doğru 1, yanlış, boş bırakılan, ya da geçersiz sayılan yanıt 0 olarak puanlanır. Geleneksel yöntemlerin dışında maddeleri kısmi puanlamaya dayalı çeşitli alternatif yöntemler bulunmaktadır. Bu yöntemler, geleneksel yöntemlere göre yanlış kabul edilecek maddeye kısmi puanlanma ile

ölçülmesi hedeflenen özelliğe ilişkin daha kapsamlı bilgi alma amacına hizmet etmektedir. Bu şekilde yanlış cevaplar da yetenek kestiriminde rol almakta ve böylece güvenilirlik artmaktadır.

Şekil 1. *Kısmi puanlama*



Birinci şekilde, geleneksel yöntemlerle puanlanmış bir teste verilmiş doğru ve yanlış yanıtlar verilmiştir.

İkinci şekilde ise doğru yanıtların yanı sıra yanlış yanıtlarında belli bir bölümünün kısmi puanlama aracılığıyla anlamlandırıldığı görülmektedir.

Çalışmada yapılan kısmi puanlamalarla yeniden örtük sınıf belirlenmeye çalışılmaktadır. Bu araştırmada kısmi puanlamaya yönelik kullanılan yaklaşım Q matrisin yeniden tanımlanmasına ve sınıf sayısını arttırarak öğrencinin yetenek kestirimine ve örtük özelliğin kısmileştirilmesine dayanmaktadır. Genel olarak kısmi puanlama modellerinde amaç, yetenek düzeylerini arttırarak daha doğru bir sınıflama ve yetenek kestirimi yapmaktır. Bu çalışmada ayrıca kısmi puanlanan maddeler ile birlikte soru sayısı arttırılarak analiz yapılmıştır.

2.4.4.2. *Yapılandırılmış Yanıtlı Maddelerde Kısmi Puanlama Yaklaşımları*

Puanlama, yapılandırılmış yanıtlı maddeler için büyük bir geçerlilik sorunudur. Yapılandırılmış yanıtlı maddelerin puanlaması öznel ve bu nedenle öznelliğin puanlama geçerliliği üzerindeki olumsuz etkisini azaltmak için bazı konulara dikkat edilmesi gerekmektedir. Bu bağlamda, puanlama yöntemleri ile ilgili analitik ve bütünsel (küresel, global, holistic) olmak üzere iki farklı yaklaşım vardır. Analitik puanlama yaklaşımlarında açık uçlu maddeler birkaç farklı kategoride veya birkaç farklı özellik için puanlanmıştır. Örneğin; soruya verilen yanıtın doğruluğunu değerlendirmenin yanı sıra cevabın özgünlüğü, yazılı cevabın organizasyonu, yazma becerisi gibi birkaç farklı yöne yoğunlaşmaktadır. Bütünsel veya küresel puanlamalar, açık uçlu maddeye verilen yanıtın genel değerlendirmesini yaparak

sadece bir puan oluşturmaya gerektirir. Analitik veya bütünsel puanlama yöntemlerinden hangisinin seçilmesi gerektiği yapılandırılmış yanıtli maddenin amacına göre deęişir. Analitik puanlama yöntemleri, becerinin farklı yönleri ile ilgili deęerlendirme saęladığı için belirli açılardan geribildirime izin vermektedir. Ayrıca analitik puanlama yöntemi bütünsel puanlama yöntemine göre daha çok süre gerektirir. Bu nedenle deęerlendirme açısından hızlılık önemli olduęu durumlarda pratiklik açısından bütünsel puanlamanın tercihi yapılmaktadır. Analitik yöntemler, kısmi puanların aęırlıklandırılması açısından, küresel yöntemlerden daha elverişlidir (Downing ve Yudkowsky, 2009).

2.4.4.3. Seçme Yanıtli Maddelerde Kısmi Puanlama Yaklaşımları

Çoktan seçmeli sınavlarda puanların yorumlanması ile ilgili bazı sorunlar vardır. Bu sorunlar, yanlış ya da boş bırakılan cevaplarla bağlantılı olarak ortaya çıkmaktadır. Oysa ki çoktan seçmeli sınavların genel kullanımında olduęu gibi sadece doęru sayısını saymak ve doęruların sayısından yola çıkarak sadece neyin doęru bilindięi konusundaki elde edilen bilgi oldukça eksik ve yüzeyseldir. Öğretimin niteliğini, öğrenmenin derecesini, sahip olunan becerilerin deęerlendirilmesi aşamasında, yapılan hatalar, eksik öğrenmeler, kısmi bilgiler, doęru cevapların sayısından daha çok veri saęlayıcı niteliktedir.

Geleneksel olarak çoktan seçmeli maddeler, geleneksel doęru sayısını sayma yöntemiyle deęerlendirilir. Doęru yanıtli pozitif bir deęer, yanlış ya da boş yanıtli sıfır deęeri vererek. Doęru yanıtli puanlarının toplamı tüm testin puanını verir. Bu yöntem en büyük eleştirisi sınava girenlerin tahmin ile maddeyi doęru yanıtlayabilme faktörüdür.

Dressel ve Schmid (1953) ve Coombs, Milholland ve Womer (1956) tarafından yapılan araştırmalar kısmi bilginin yani doęru cevaba ulaşmak için yetersiz olan bilginin ölçülebilir bir özellik olduęunu ortaya koymaktadır. Bunun için geçerlik ve güvenilirlik adına farklı yaklaşımlar vardır. Bu yaklaşımlardan biri diferansiyel aęırlıklandırma prosedürleridir. Diferansiyel aęırlıklandırma prosedürleri, çoktan seçmeli testlerin geleneksel puanlama yönteminden farklılaşmaktadır. Doęru cevaplar için 1, yanlış ve boş cevaplar için 0 puan vermek yerine, bir maddeye alternatif olacak her cevap için farklı puanlama aęırlıkları atanır. Bir test puanı, seçilen doęru ve yanlış yanıtli puanların aęırlıklarının toplamından

oluşur. Ağırlıklı puanlama yöntemlerinin tarihçesine bakıldığında, 1917'de ordu alfa grubu zihin testinin geliştirilmesi ve sonraki yıllarda yapılan çalışmalara dayanmaktadır (Wang ve Stanley, 2008).

Jacops ve Vandeventer (1968) facet design (Guttman & Schlesinger, 1967) gösterimini Raven'in (1963) Coloured Progressive Matrices (CPM) cevap alternatiflerinin doğruluk derecesini sistematik olarak sıralama prosedürü olarak kullanmıştır. En doğru ya da doğruya en yakın çeldirici olarak sınıflandırılacak yanlış cevapların oranından oluşan puanların orta derecede test tekrar test güvenilirliğine sahip olduğu ve çoğunlukla, aynı ve farklı bir CPM formundaki doğru cevap sayısının ölçütüne karşı orta derecede yüksek pozitif eş zamanlı ve yordama geçerliği bulunmuştur.

2.5. İlgili Araştırmalar

1. Seçme Yanıtlı Maddeler ve Yapılandırılmış Yanıtlı Maddeler için Üç Puanlama Yönteminin Karşılaştırılması

Bu çalışmada Schaeffer, Henderson-Montero, Julian, ve Bene (2002) çoktan seçmeli ve yapılandırılmış yanıtlı maddeleri puanlamak için üç puanlama yöntemini karşılaştırmıştır. Ağırlıklandırılmamış ham puanlar, MTK desen puanları ve ağırlıklandırılmış ham puanlar (unweighted raw scores, IRT pattern scores, and weighted raw scores). Öğrencilerin puanlarına ait datalar, geniş ölçekli olarak yapılan lisedeki Biyoloji ve İngilizce dersinin sınavlarından elde edilmiştir. Bu çalışmada değerlendirilen ağırlıklı ham puan yönteminde, CR maddeleri, SR ve CR maddelerinin toplam puana aynı sayıda puan vermesi için ağırlıklandırılmıştır. Puanlama yöntemleri, sonuçlanan ölçeklendirilmiş puan dağılımları, standart ölçüm hataları ve yeterlilik düzeyi sınıflandırmaları açısından toplam grup ve öğrenci alt grupları için karşılaştırılmıştır. Öğrenci yetenek dağılımının çoğu için, üç puanlama yöntemi benzer sonuçlar vermiştir.

Öğrencilerin yapılandırılmış yanıtlı ve çoktan seçmeli maddelerden aldığı toplam puanlar ve alfa test güvenilirlik katsayısı hesaplanmıştır. Önce klasik madde analizleri yapılmıştır. Bu analizler sonucu testlerin zor olma eğilimi gösterdiği ancak maddelerin büyük çoğunluğunun 0.15'in üzerinde madde toplam korelasyonu olduğu bulunmuştur. Her yapılandırılmış yanıt maddesi için en az birkaç öğrenci her puan seviyesinde puan almıştır. Tüm maddeler kullanılmıştır. Her bir içerik alanındaki

madde kalibrasyonu amacıyla, çoktan seçmeli maddeler için 3PL modeli (Lord, 1980), yapılandırılmış yanıtli maddeler için genelleştirilmiş (Muraki, 1990) veya iki parametrelili (Yen, 1993) kısmi puanlama modeli kullanılmıştır bu şekilde parametreler aynı ölçek üzerine benzer şekilde yerleştirilmiştir (same scale simultaneously). Parametreler, EM algoritması ile uygulanan maksimum olabilirlik prosedürlerini kullanan PARDUX (Burket, 1998) bilgisayar programı ile hesaplanmıştır (Bock ve Aitken, 1981). Yakınsamak problemi ile karşılaşılmanış ve tüm maddeler için parametreler üretilmiştir. Madde parametrelerinin puan ölçeğine yerleştirilmesi, WINFLUX bilgisayar programı kullanılarak gerçekleştirilmiştir (Burket, 1999) (Schaeffer, Henderson-Montero, Julian, & Bene, 2002).

2. İkili ve Çoklu Puanlanmış Maddelerden Oluşan Testlerde Yetenek Tahmini İçin Madde Ağırlıklandırma Olasılık Yöntemi

Araştırmada Tao, Shi, ve Chang (2012) İkili ve çoklu puanlanmış maddelerden oluşan testlerde yetenek tahmini üzerine çalışmıştır. İkili ve çoklu puanlanmış maddelerden oluşan karışık testler için çoklu maddeler ikili maddelerden daha fazla bilgi verir. İki madde arasındaki farkı yansıtmak için çoklu puanlanmış maddeler genellikle daha çok ağırlıklandırılarak ön tanımla yapılır. Çalışmada sınava girenlerin yetenek düzeylerini daha iyi değerlendirmek için ağırlıklandırılmış madde olabilirlik yönteminin kullanılması önerilmiştir. Önerilen prosedürü değerlendirmek için simülasyon çalışması yapılmıştır. Yapay olarak hem ikili ve çoklu puanlanmış maddelerden oluşan karma madde türlerinde 30 maddelik bir test oluşturulmuştur. Hem teorik hem de ampirik kanıtlar, genel olarak çoklu puanlanmış maddelerin, özellik düzeyinde kestirimin kesinliği ile ilgili iki maddeden daha fazla bilgi sağladığını göstermiştir. (Donoghue, 1994; Embretson ve Reise, 2000, s. 95; Jodoin, 2003; Penfield 2000; Bergeron,2005).

Simülasyon çalışması sonuçları, yeni prosedüre dayanarak tahmin edilen yetenek düzeyinin, sınava girenlerin gerçek yetenek düzeyleri ile maksimum olabilirlik yönteminden daha tutarlı olduğu sonucuna ulaşılmıştır. Ayrıca ağırlıklandırılmış madde olabilirlik (IWLE)'nin özellikle örtük özelliklerin aşırı seviyelerinde yanlılığın azalması konusunda maksimum olabilirlik (MLE)'den daha iyi performans gösterdiği gözlenmiştir.

IWLE yönteminin geniş ölçekli değerlendirmelerde uygulanabilirliğini incelemek için 2000, kişiden oluşan bir örneklem üzerinde pilot çalışma yapılmıştır. Bu çalışmada da 50 ikili puanlanmış madde ve 5 kategoriden oluşan 1 çoklu puanlanmış madde kullanılmıştır. Elde edilen sonuçlara göre yetenek kestiriminde IWLE'nin kullanılma durumunda açık bir fark görülmüştür. Çoklu puanlanmış maddenin sadece bir tane olması bile farklılık yaratırken daha fazla kullanımının daha da fark yaratacağı belirtilmiştir (Tao, Shi, ve Chang, 2012).

3. *Uzman Tanımlı Çoklu Özellikler için Genel Bir Bilişsel Tanı Modeli*

Bu çalışmada Jimmy de la Torre ve Jinsong Chen, pG-DINA modeli tanımlanmış ve gerçek data üzerinden modelin uygulanabilirliği değerlendirilmiş ve sınıflandırma doğruluğu G-DINA modeli ile karşılaştırılmıştır. PG-DINA modelinin uygulanabilirliğini sistematik olarak değerlendirmek için iki bölümlü bir simülasyon çalışması tasarlanmıştır. İlk bölümde, madde parametreleri incelenmiştir. İkinci bölümde ise pG-DINA modelin sınıflandırma doğruluğu üzerinedir. Bunun için modifiye edilmiş çoklu özellik (mG-DINA) ile G-DINA modelin sonuçları karşılaştırılmıştır. Ox programı kullanılmıştır. Gerçek data üzerinde ise 393 öğrenci, çoktan seçmeli maddeler kullanılmıştır. Araştırma sonucunda pG-DINA modelin daha yeterli fit verdiği bulunmuştur (J. Chen ve de la Torre, 2013).

4. *Matematikte Seçme Yanıtlı ve Yapılandırılmış Yanıtlı Maddelerde: Amerikan Deneyimi ve Tartışması (Derleme)*

Bu çalışmada Stankous, (2016) matematik dersini değerlendirmek için kullanılan çoktan seçmeli ve yapılandırılmış yanıtlı madde formatları karşılaştırılmıştır. Aynı konuda araştırmacıların yapmış olduğu bu ikinci çalışmadır. İlkinde iki madde formatının da avantaj ve dezavantajlarına değinilmiştir. Bu çalışmada ise matematik dersinde çoktan seçmeli maddelerin öğrenci ve öğretmen başarısını değerlendirmede daha sık kullanıldığı, yapılandırılmış yanıtlı maddelerin ise öğrencilerin öğrenme durumlarını göstermede daha iyi olduklarına vurgu yapılmıştır. Çoktan seçmeli maddelerde tahmin problemi, madde formatlarında kısmi puanlama, madde formatlarına göre puanlamanın etkisi ve yetenek gibi faktörler konusuna değinilmiştir.

Elde edilen sonuçlar; yapılandırılmış yanıtlı maddelerin çoktan seçmeli maddelere göre öğrenmeye teşvik konusunda daha etkili olduğu bulunmuştur.

Değınilen bir başka sonu ise matematik eđitimde bařarıyı problem özme yeteneđinden daha başka yollarla da tanımlayabilmektedir. oktan semeli maddelerde puanlama ařamalarının ayrıntılı belirlenmesi öđrenci bařarısını betimlemede daha kesin sonular verecektir(Stankous, 2016).

5. *Beceri Hiyerarřileri İeren BTM: Kořullu Q Matris Tasarımı ile Model Tahmini*

Bu alıřmada Tu, Wang, Cai, Douglas, & Chang, (2019) kısıtlanmıř Q matrix dzenlenmesi ile biliřsel tanı modellerinde sınıflama farklılıkları ile ilgili farklı tahmin yöntemleri verilmiřtir. Kısıtlanmıř bir Q matris tasarımıyla modelin tanımlanabilirliđini arařtırmak iin yapılmıř ampirik bir alıřma olarak ele alınabilir. alıřmada kısıtlanmamıř Q matris ile yapılan sınıflama tahminlerin neden yanlıř olduđunu aıklamak iin simlasyon alıřması yapılmıřtır. Sınıflama dođruluđunu arttırmak iin zellik hiyerarřileri kullanan kısıtlanmıř tahmin prosedrlerinin avantajları, aynı zamanda kıyasal akıl yrtme tanısal deđerlendirmesine iliřkin gerek bir veri analizi ile de gsterilmiřtir. alıřmada zellik hiyerarřisi yapısının Q matris ile yansıtıldıđı varsayılmıřtır ve sınıflandırma dođruluđunu etkileyebilecek faktrler arařtırılmıřtır. Hem simlasyon alıřması hem de gerek data alıřması ile, hangi tr biliřsel tanı modellerinin kullanıldıđına bakılmaksızın rtk sınıf yapısının (permissible attribute profile space) izin verilen zellik profili bořluklarında kısıtlanmasının nemli olduđu sonucuna ulařılmıřtır. zellikle kısıtlanmıř tahmin prosedr kullanırken, DINA gibi temel zelliklere sahip model, gerek veri analizinden elde edilen sonulara gre zellik hiyerarřisi altında da makul ve iyi sınıflandırma sonuları retebileceđi elde edilmiřtir(Tu, 2019).

6. *Q Matris Tabanlı DCM'nin İstatistiksel Analizi*

Bu alıřmada Y. Chen, Liu, Xu, & Ying, (2015), DINA ve DINO modelleri altındaki Q matrisin tanımlanabilmesi zerine teoriler geliřtirilmiřtir. Q matris iin dzenlenmiř maksimum olabilirlik olasılıđı kullanılarak bir tahmin prosedr nerilmiřtir. Bu prosedrn uygulanabilirliđinin DINA ve DINO model ile sınırlı olmadıđı ve diđer tm Q matris temelli tanı temelli sınıflandırma modellerine uygulanabilir olduđuna ulařılmıřtır. Yapılan simlasyon alıřmasında, nerilen yöntemin gerek Q matrisin iyileřtirilmesiyle yksek olasılıklar verdiđini gstermiřtir. Ayrıca alıřmada iki rnek olay alıřması sunulmuřtur. İlk alıřma

eđitim uygulamasından alınan kesirlerle çıkarma işlemleri konulu bir çalışmadan elde edilen verilerle ikinci çalışma ise psikiyatrik bir uygulama olan sosyal anksiyete bozukluğu konulu çalışmadan elde edilen verilerle yapılmıştır. (Y. Chen, 2015)

7. *Mevcut Büyük Ölçekli Deđerlendirme Verisinin Teşhis Modellemesi için Bir Prosedür: Okumadaki Uluslararası Öğrenci Deđerlendirme Programı Örneđi*

Araştırmada J. Chen & Torre, (2014) yapılan geniş ölçekli sınavlardan elde edilen verileri tanısal modellerle deđerlendirmek için sistematik bir prosedür önerilmiştir. Önerilen prosedür dört aşamadan oluşmaktadır. İlk özelliklerin belirlenmesi ve Q matrisin oluşturulması, son özelliklerin belirlenmesi ve Q matrisin oluşturulması, indirgenmiş bilişsel tanı modellerinin deđerlendirilmesi ve seçilen modellerin çapraz geçerlilik çalışması. Prosedürde, genel ve indirgenmiş bilişsel tanı modellerinin geliştirilmesi, çeşitli mutlak ve bađımlı fit indeksleri ve genel bir Q matris doğrulama prosedürü dahil olmak üzere son metodolojik gelişmeler benimsenmiş ve entegre edilmiştir. Analizler, dil uzmanlarıyla çalışarak, PISA 2000 okuma deđerlendirmesinden elde edilen verilerle yapılmıştır. Çalışmada PISA okuma deđerlendirmesinin seçilme amacı, pratikte modelleme prosedürünü göstermek için olduğu belirtilmiştir. Sonuç olarak anlamlı tanımlar ve uygun bir Q matrise sahip niteliklere ulaşılmıştır. Modelleme prosedürünün NAEP ve TIMSS gibi diđer büyük ölçekli sınavlara uygulanabileceđi elde edilen bir diđer sonuçtur(J. Chen ve Torre, 2014).

8. *Bilişsel Tanı Deđerlendirme: Sorunlar ve Düşünceler*

Çalışmada, Javidanmehr, Reza, & Sarab, (2017) bilişsel tanı deđerlendirmelerinin teorik temeli ve test analizi prosedürleri açıklanmıştır. Eğitimde bilişsel tanı modellerinin ne olduğu, neden kullanılması gerektiđi ve ikinci dil eğitimi deđerlendirmesinde kullanımına odaklanmıştır. Öncelikle tanısal deđerlendirme ve tanısal olmayan deđerlendirme yapıları sonuçlar karşılaştırılmıştır. Bilişsel tanı deđerlendirmesi için G-DINA model kullanılmıştır. Örneđim, İran'da çeşitli lisans programlarında eğitim alıp yüksek lisans programına devam etmek için yabancı dil sınavına giren rasgele seçilmiş erkek ve kadın 2500 adaydan oluşmaktadır. Analiz için 20 çoktan seçmeli maddeden oluşan İngilizce okuduđunu anlamayı içeren maddeler seçilmiştir. Alan uzmanları önce bireysel daha sonra grup olarak çalışarak

Q matrisi oluşturmuşlardır. Q matrisin geçerliliği için R programı üzerinde G-DINA paketi ile çalışılmıştır(Javidanmehr , 2017).

9. *Bilişsel Tanı Değerlendirme ve Bilişsel Tanı Modeli Çerçevesi*

Bu çalışmada, De La Torre, (2009) bilişsel tanı modelleri ve bilişsel tanı değerlendirmelerinin yararı ve gereksinimi üzerinde durulmuş ve bunların geliştirilmesinde ve kullanımında yer alan çeşitli hususlar konu alınmıştır. Bilişsel tanı değerlendirmenin değerlendirme üçgeni çerçevesine ve kanıt merkezli tasarım çerçevesine nasıl uyduğu ayrıntılarıyla anlatılmıştır. Bilişsel tanı modelleri bağlamında öznitelik tanımlaması ve madde geliştirme yöntemleri tartışılmış ve ilgili araştırmalardan örnekler verilmiştir. Aynı zamanda bilişsel tanı analizlerinin pratik uygulamalarını destekleyen istatistiksel modeller olan Bilişsel tanı modellerinin de detayları tartışılmıştır. Bilişsel tanı değerlendirmelerinin sınıf içi öğretimi ve öğrenmeyi geliştirebilecek tanısallık bilgileri sağlayabileceği sonucuna ulaşılmıştır (De La Torre, 2009).

10. *DINA Model Parametrelerinin Değişmezliği Üzerine Bir Not*

Bu çalışmada J. De Torre & Lee, (2010) bilişsel tanı modellerinden olan Deterministic Inputs, Noisy “and” Gate (DINA) modeline ve parametrelerinin değişmezlik özelliklerine odaklanmıştır. DINA model parametrelerinde, model ve veri yüksek uyum gösterdiğinde, kesinlikle değişmezlik olduğu sonucuna ulaşılmıştır. Ayrıca farklı yetenek gruplarını içerecek ek bir örnek çalışma sunulmuş ve gerçek verilerle parametre değişmezliğinin eksikliğine nasıl katkıda bulunabileceğine değinilmiştir. Örneklem 536 ortaokul öğrencisinden seçilmiştir. Maddeler 15 tane kesirlerle çıkarma işlemi konusundan seçilmiştir. Makalede elde edilen sonuçlar, DINA model parametrelerinin kesinlikle değişmez olduğunu göstermektedir, ancak bu değişmezlik özelliğinin yalnızca model ve veri uyumu sağlandığı zaman olduğu sonucuna ulaşılmıştır. Model ve verinin yüksek uyum gösterdiği simülasyon verileriyle aynı madde parametresi tahminleri, belirgin şekilde kalibre edilmiş örnekleri için de elde edilmiştir. Bir diğer analiz sonucu, gerçek veriler söz konusu olduğunda, DINA model parametrelerinin değişmezlik özelliklerinin korunamayabileceğini göstermiştir. Ancak, bu durumun DINA modelin pratik kullanılabilirliğini küçümsemek için bir neden olmadığı ifade edilmiştir. Yaygın IRM'lerde paradoksal bir sonuç gözlemlenmiştir. Bu sonuçlar; üç parametrelilik

model gibi daha karmaşık modeller daha iyi model-veri uyumu sunmasına rağmen, tek parametrelili lojistik model gibi daha basit modellerin daha değişmez olduğudur. (J. De Torre ve Lee, 2010)

11. Sürekli Yanıtlı Maddeler İçin Bilişsel Tanı Modeli

Araştırmada Minchen, de la Torre, & Liu, (2017) ikili puanlamaya alternatif olarak sürekli yanıt formatı (continuous response formats) kullanılmıştır. Uygun bir şekilde bölünmüş (finer-grained) beceriler ya da davranışları değerlendirmek ve sürekli yanıt verilerinden tanısal değeri olan bilgileri çıkarmak için, sürekli yanıt için çok boyutlu beceri tanı modeli önerilmiştir. Parametrelerin tahmini için marjinal maksimum olabilirlik tahmininin bir beklenti maksimizasyon uygulaması geliştirilmiştir. Önerilen modelin uygulanabilirliği bir simülasyon çalışması ve gerçek bir veri örneği ile gösterilmektedir. Önerilen modelin, ikili hale getirilmiş sürekli yanıtlara dayanan modele kıyasla nitelik sınıflandırmasında da önemli bir gelişme sağladığı sonucuna ulaşılmıştır (Minchen, 2017).

12. Bilişsel Tanı Değerlendirmede Dereceli Veriler için Çoklu Değerlendirme Modeli

MTK'daki farklılık modellerinin çizgisini izleyen bu çalışmada BTM'deki DINA model bazında ikili-çoklu veriler için derecelendirilmiş veriler için DINA model adında yeni bir model önerilmiştir. Simülasyon üzerinden tam bir bayes tahmininin performansı araştırılmıştır. Ayrıca, DINA-GD modeli için sınıflandırma doğruluğu ve madde iyileştirmeleri incelenmiştir. Sonuçlar, önerilen modelin kabul edilebilir olduğu ve özellikleri doğru sınıflandırma oranına ve madde parametrelerini iyileştirir nitelikte olduğunu göstermektedir. Ayrıca yapılan simülasyon çalışmasının yanı sıra bu yeni modelin dereceli verilerle veya düzeyli ya da çoklu (polytomous) puanlanan maddelerle uygulaması için gerçek veri örneği kullanılmıştır (Tu, Zheng, Cai, Gao, ve Wang, 2018).

BÖLÜM III

YÖNTEM

Bu bölümde, araştırmanın türü, çalışma grubu, verilerin elde edilmesi, verilerin analiz yöntemleri ve bu analizlerde kullanılacak bilgisayar programları açıklanmıştır.

3.1.Araştırmanın Modeli

Araştırma, farklı madde formatlarında kısmi puanlama yapmanın bilişsel tanı modellerindeki örtük sınıf değişmezliği üzerine etkisini belirlemek amacı taşımaktadır. Araştırma, bilişsel tanı modelleri için yapılandırılmış yanıtli maddeleri puanlamaya yönelik yöntem önermesi açısından temel, aynı zamanda gerçek data üzerinden puanları karşılaştırmaya olanak sağlaması açısından uygulama araştırması türündedir (Kite ve Whitley, 2018).

3.2.Projenin Evren ve Örneklemi

Araştırmada 115K531 no'lu TÜBİTAK Projesinin verilerinden yararlanılmıştır. Projenin evrenini İzmir İli Milli Eğitim Müdürlüğüne bağlı 448 Devlet Okulu oluşturmaktadır. Bu okullarda 6.Sınıf düzeyinde toplam 1822 şube ve 45069 öğrenci yer almaktadır. Örneklem büyüklüğü belirlenirken güven düzeyi ve güven aralığı istatistiklerinden yararlanılmıştır (Oulte, 2011). Güven düzeyi %99 ve güven aralığı $t=1,77$ olarak belirlendiğinde 45069 kişilik evren için örneklem büyüklüğü $n=4592$ belirlenmiştir. Daha sonra tabakalı yöntem ile tesadüfi küme örneklem yöntemi kullanılarak okul ve sınıflar belirlenmiştir (Lodico, Spaulding, & Voegtle, 2010). Tabakalama yöntemi kullanılırken dikkate alınan kriterler; öğrencilerin 2015-2016 TEOG genel başarı ortalamaları, 2015-2016 TEOG matematik ortalamaları, okulların 6.sınıf düzeyindeki öğrencilerinin şubeye düşen öğrenci sayıları, ilçe bazındaki okul, şube ve öğrenci sayılarıdır. Bu kriterler göz önüne alınarak tesadüfi olarak 10 ilçeden 20 okul proje örneklemini oluşturmuştur. Proje örnekleminde yer alan okullar, şube ve öğrenci sayıları Tablo 7'de verilmiştir.

Tablo 7. Proje örnekleminde yer alan okullar, şube ve öğrenci sayıları

İLÇE	6. SINIF ŞUBE SAYISI	ÖĞRENCİ SAYISI
ALIAĞA	4	107
BAYRAKLI	10	282
BORNOVA	5	182
	6	169
	9	233
	6	190
BUCA	4	182
	8	292
	14	465
ÇİĞLİ	4	128
GAZİEMİR	10	240
	6	160
KARABAĞLAR	7	214
	5	192
KARŞIYAKA	7	272
	7	200
KONAK	18	596
	7	186
	7	173
BALÇOVA	4	129
BORNOVA	7	249
TOPLAM	148	4592

Tablo 7’de görüldüğü gibi İzmir ilinde yer alan okullar ilçelere göre belirlenmiş olup proje örnekleminde 20 okuldan toplam 148 şube ve 4592 öğrenci araştırmaya dahil edilmiştir.

3.3. Veri Toplama Araçları

Tezde kullanılan ölçme aracı, yapılandırılmış yanıtı ve çoktan seçmeli maddelerden oluşan ve aynı zamanda BTM kullanılarak yetenek kestirimlerinin yapılabilmesi

amacıyla oluşturulmuş testlerdir. Testler 115K531 no'lu TÜBİTAK Projesi kapsamında yapılan araştırma için hazırlanmıştır. Çalışmanın verileri, proje kapsamında 6.sınıf düzeyindeki öğrencilere uygulanan matematik alanına ait izleme testlerinden seçilen 18 adet karma formatta maddeden oluşmaktadır. Bunlardan 10 tanesi çoktan seçmeli, 8 tanesi ise yapılandırılmış yanıtli maddelerden oluşmaktadır. Testlerde maddeler üst düzey düşünme becerilerini ölçecek nitelikte hazırlanmıştır. Maddelerin hazırlanması süreci öncelikle, gönüllü öğretmenlere üst düzey düşünme becerileri ile ilgili madde yazma eğitimi verilmesiyle başlamıştır. Sonra maddeler soru havuzundan, içlerinde alan uzmanlarının da bulunduğu hakemlerce puanlanarak seçilmiştir. Pilot uygulamalar sonrasında test hazırlanmış ve esas uygulamalar yapılmıştır.

3.3.1. Veri Çözümleme Teknikleri

Bu bölümde test geliştirme aşamaları üzerinde durulmuştur.

3.3.2. Ölçme Aracında Ölçülen Özelliklerin Belirlenmesi

Projenin amacı doğrultusundan Bilişsel Tanı Modeline uygun analizlerin gerçekleştirilebilmesi için test maddelerinin ilişkilendirilebilecekleri özellik kümelerinin belirlenmesi ve buna uygun Q matrislerin oluşturulması gerekmektedir. BTM analizleri birden fazla özellik ile ilişkilendirilmiş maddeleri içeren testler için uygun olmakla birlikte özellik sayısının çok olduğu ya da sadece tek madde ile ölçülen özelliklerin bulunduğu testlerde model uyumu konusunda sıkıntılar yaşanmaktadır. Bu nedenle kazanımlar ve matematiksel yeterlikler Q matris için yeniden yapılandırılmıştır. Bunu yaparken uluslararası geniş ölçekli sınavlardan olan PISA'nın belirlediği matematiksel yeterliklerden faydalanılmıştır. PISA'nın belirlediği matematiksel yeterliklerin matematik okuryazarlığı için ön koşul niteliğinde olduğu belirtilmiştir. Belirlenen yeterlikler her maddede dönüşümlü olarak sınanmıştır. Genel olarak 7 temel başlıkta ifade edilen bu yeterlikler; İlişkilendirme, İletişim, Matematikleştirme, Çoklu Temsiller, Akıl Yürütme ve Kanıt Gösterme, Strateji Geliştirme, Sembolik, Formel ve Teknik Dil ve İşlem Kullanılmasıdır. Bu yeterlikler kısaca aşağıdaki gibi tanımlanmıştır (Oecd, 2013).

İlişkilendirme: Matematik, sadece kurallar, semboller, şekiller ve işlemlerden ibaret değildir. İçinde bir anlam bütünlüğü olan düzenler ve ilişkiler ağıdır. Matematikle diğer disiplinler ve yaşam arasında da ilişkiler bulunmaktadır. Buna

bağlı olarak ilişkilendirme becerisi, matematik kavramlarının kendi aralarında da, bir matematiksel kavramın diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

İlişkilendirme becerilerinin geliştirilmesi için dikkate alınması gereken bazı göstergeler şunlardır: Kavramlar ve işlemler arasında ilişki kurma, matematiksel kavram ve kuralları farklı temsil biçimleriyle gösterme, matematiksel kavram ve kuralların farklı temsil biçimlerini birbiriyle ilişkilendirme ve birbirine dönüştürme, farklı matematik kavramlarını birbiriyle ilişkilendirme, matematiği diğer derslerde ve günlük yaşamda karşılaşılan konu ve durumlarla ilişkilendirmedi.

İletişim: Temel matematik yeteneği olarak iletişim, problem çözme sürecinin aşamaları içerisindeki ya da arasındaki iletişimin sağlanabilmesi yeteneğini ifade etmektedir.

Bir problemin çözümü; okuma, yeniden kodlama, ifade, soru, görev ve verilenleri yorumlama gibi problemi anlama ve tanımaya yönelik birçok bireysel eylemin gerçekleştirilmesini, problemin netleştirilmesi ve formüle edilmesi aşamalarını, çözüm için gerekli işlemler gerçekleştirilmekte ve çözümler üretilmesi aşamalarını içerir. Bir çözüme ulaşıldığında, bu çözümün sunulması, açıklanması ve doğruluğunun kanıtlanması, problemi çözen kişi ya da kişilerin eylemleri arasındadır.

İletişim yeterliği ile ilgili dikkate alınması gereken bazı göstergeler şunlardır: matematiğin sembol ve terimlerini etkili ve doğru kullanma, matematiksel dili matematiğin kendi içinde, farklı disiplinlerde ve yaşantısında uygun ve etkili bir biçimde kullanma, somut model, şekil, resim, grafik, tablo, sembol vb. farklı temsil biçimlerini kullanarak matematiksel düşünceleri ifade etme, matematiksel düşünceleri sözlü ve yazılı ifade etme, günlük dili, matematiksel dil ve sembollerle; matematiksel dili, günlük dil ve sembollerle ilişkilendirme, matematiksel düşüncelerin doğruluğunu ve anlamını yorumlamadır.

Matematikleştirme: Gerçek dünyada karşılaşılabilecek bir problemi kesin bir matematiksel forma dönüştürebilme becerilerinin bir bütünü ya da yeteneğini içermektedir. Gerçek yaşamda karşılaşılan problemler, doğası gereği açık ve kesin matematiksel biçimde değildir. Bu tür problemlerin çözümünde, öncelikle matematiksel biçime dönüştürerek tanımlama ve açıklama gerekebilmektedir.

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

Çoklu temsiller: Çoklu temsil ya da gösterim, matematiksel obje ya da durumların betimlenmesinde sıklıkla kullanılmaktadır. Matematik eğitiminde çoklu temsillerin önemi hem öğrenmeye araç olmaları hem de öğrenmenin bir göstergesi olmaları gibi iki önemli noktada kendini göstermektedir. Öğrencilerin matematiksel kavramları anlamaları, bunları kullanarak etkili bir şekilde iletişim kurabilmeleri bunların değişik temsillerini kullanabilmeleri ve bu temsiller arasında geçiş yapabilmeleriyle ilgilidir.

Uygulamada çoklu temsiller, grafik, tablo, diyagram, resim, denklem, formül ve diğer görsel araçları ifade etmektedir.

Akıl yürütme ve kanıt gösterme: Akıl yürütme (muhakeme), eldeki bilgilerden hareketle matematiğin kendine özgü araç (semboller, tanımlar, ilişkiler, vb.) ve düşünme tekniklerini (tümevarım, tümdengelim, karşılaştırma, genelleme, vb.) kullanarak yeni bilgiler elde etme süreci olarak tanımlanabilir. Akıl yürütme ve kanıt gösterme, matematik okuryazarlığının farklı aşamalarını ve bu aşamalardaki ilişkili eylemleri ifade eden bir beceriler bütünü ya da yeterlidir. Çıkarımların doğruluğunu ve geçerliliğini savunma, mantıklı genellemelerde ve çıkarımlarda bulunma, bir matematiksel durumu analiz ederken matematiksel örüntü ve ilişkileri açıklama ve kullanma akıl yürütme ve kanıt göstermenin göstergeleridir.

Strateji geliştirme: Strateji belirleme ya da tasarlama, etkili bir tanıma, formülleştirme ve problem çözmede bireylere kılavuzluk eden bir dizi kritik kontrol süreci içermektedir. Bu yeterlik, problem çözmede matematiksel bilgi ve becerilerin kullanılmasına yönelik bir plan ya da stratejinin seçilmesi ya da tasarlanmasını ifade etmektedir. Aynı zamanda problem çözme sürecinin herhangi bir aşamasında kullanılabilecek bir yeterlik olarak görülmektedir.

Sembolik, formel ve teknik dil ve işlemlerin kullanılması: Matematik okuryazarlığı, sembolik, formel ve teknik dil ve işlemlerin kullanılmasını gerektirmektedir. Matematik okuryazarlığı açısından sembolik, formel ve teknik dilin kullanılması yeteneği, matematiksel kurullarla ifade edilen bir matematiksel içeriğe

yönelik olarak sembol gösterimlerini anlama, yorumlama, manipüle etme ve kullanma davranışlarının bir bütünü, tanımlama ve kurallara dayalı formel yapıların ve formel sistemlerin anlaşılmasını ve kullanılmasını, ayrıca bu anlama ve kullanma süreçlerinde algoritmaların da kullanılmasını içermektedir. (Oecd, 2013)

Projede temel alınan matematiksel yeterlikler, proje araştırmacıları, danışmanlar, ve alan uzmanı öğretmenler ile birlikte 6.sınıf düzeyi göz önüne alınarak tekrar düzenlenmiştir. İlişkilendirme, İletişim, Matematikleştirme, Çoklu Temsiller, Akıl Yürütme ve Kanıt Gösterme Strateji Geliştirme, Sembolik, Formel ve Teknik Dil ve İşlem Kullanılması yeterlikleri 4 özellik altında toplanmıştır. Bu yeterlikler ve yeterliklere ait analiz kodları aşağıda Tablo 8’de verilmiştir.

Tablo 8. *Araştırmada belirlenen matematiksel yeterlikler*

KOD	YETERLİK
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 kapsar. İlişkilendirme becerisi; matematik kavramlarının kendi aralarında, diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

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

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.

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, yorumlama davranışlarının bir bütünüdür (Basokcu, 2019).

3.3.3. Ölçme aracı Q Matrisin Hazırlanması

Bilişsel Tanı Modellerinde Q matris, testin geliştirme aşamasından başlanarak, titizlikle üzerinde çalışılması gereken önemli bir adımdır. Çünkü BTM'nin hangi modeli olursa olsun Q matrisin hatalı belirlenmesi, ona bağlı yapılan tüm analizlerin ve dolayısıyla kestirimlerin de hatalı olmasına sebep olacaktır (De La Torre, 2008; Templin ve Henson, 2006).

Proje için Q matrisin belirlenmesi:

Projede kullanılan 4 izleme testine ait Q matrisler tabloda verilmiştir. M1, M2, M3 ve M4 olarak isimlendirilen matematiksel yeterlikler sırasıyla “İletişim ve İlişkilendirme”, “Matematikleştirme”, “Akıl Yürütme ve Strateji Geliştirme” ve “Sembolik ve Teknik Dil Kullanımı” şeklindedir.

3.3.4. Ölçme aracına ilişkin pilot uygulamalar

Pilot uygulamada, yazılan sorulardan dönem kazanımlarına uygun olarak seçilen maddeler kullanılmıştır. Pilot uygulama Suphi Koyuncu İlköğretim okulunda 231 öğrencinin katılımı ile gerçekleştirilmiştir. Projede geliştirilen her bir soru için özel kod bulunmaktadır. Bu kodlar pilot uygulamalarda “P” ön ekiyle kullanılmaktadır.

İzleme 1 testinde madde güçlük değerleri 0,10 ile 0,78 arasında ve madde ayırt edicilik değerleri 0,15 ile 0,81 arasında değişmektedir. Teste ait ortalama madde güçlüğü 0,31 olup ortalama madde ayırt ediciliği 0,56 olarak hesaplanmıştır. Yapılan analizlere göre test yeterli ayırt edicilik değerindedir. İzleme 2 testinin ise madde güçlük değerleri 0,02 ile 0,37 arasında ve madde ayırt edicilik değerleri 0,40 ile 0,82 arasında değişmektedir. İzleme 2 testinin ortalama madde güçlüğü 0,20'dir, ortalama madde ayırt ediciliği ise 0,57 düzeyinde hesaplanmıştır. Yapılan analizler sonucunda testin yeterli ayırt edicilik değerinde olduğu görülmüştür. İzleme 3 testinin tabloda görüldüğü gibi madde güçlük değerleri 0,00 ile 0,45 arasında ve madde ayırt edicilik değerleri 0,00 ile 0,62 arasında değişmektedir. Testin ortalama madde güçlüğü 0,25'tir ortalama madde ayırt ediciliği 0,37

düzeyinde hesaplanmıştır. Yapılan analizler sonucunda testin yeterli ayırt edicilik değerinde olduğu görülmüştür.

3.3.5. Ölçme aracının psikometrik özellikleri

Ölçme aracının psikometrik özellikleri klasik test teorisine göre ve bilişsel tanı modelleri olmak üzere iki başlık altında incelenmiştir.

3.3.5.1. Klasik Test Teorisine Göre Psikometrik Özellikler

Klasik test teorisine göre psikometrik özellikler madde güçlük ve ayırt edicilik parametreleri bazında incelenmiştir. Aşağıda verilen tablolarda, 4 izleme testine ait istatistikler verilmiştir.

Tablo 9. İzleme 1 testine ait madde güçlük ve ayırt edicilik parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ediciliği (rbis)	Ayırt
1	9796	0,41	0,5	
2	3690	0,21	0,22	
3	1027	0,18	0,45	
4	1025	0,27	0,55	
5	1033	0,51	0,48	
6	1021	0,68	0,45	
7	1032	0,4	0,41	
8	7339	0,01	0,19	
9	6728	0,29	0,38	
10	5158_B	0,38	0,52	
11	5158_C	0,31	0,49	
	Ortalama	0,33	0,42	

Tablo 9’da görüldüğü gibi 1. İzlemede bulunan her bir maddeye ait madde güçlükleri 0,18 ile 0,68 arasında ve madde ayırt edicilik değerleri ise 0,22 ile 0,55 arasındadır. Ortalama madde güçlüğü 0,33 ve ortalama ayırt edicilik değeri 0,42 olarak hesaplanmıştır. Buradan İzleme testi 1 ait maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir.

Tablo 10. *İzleme 2 testine ait madde güçlük ve ayırt edicilik parametreleri*

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ediciliği (rbis)	Ayırt
1	1011_A	0,29	0,45	
2	1011_B	0,27	0,46	
3	1012_A	0,29	0,40	
4	1011_C	0,35	0,40	
5	1012_B	0,30	0,38	
6	1014_A	0,43	0,45	
7	1015_A	0,08	0,32	
8	1014_B	0,05	0,32	
9	1015_B	0,38	0,43	
10	1014_C	0,24	0,27	
	Ortalama	0,27	0,39	

Tablo 10’da görüldüğü gibi 2. İzlemede bulunan her bir maddeye dair madde güçlükleri 0,05 ile 0,43 arasında, madde ayırt edicilik değerleri ise 0,27 ile 0,46 arasında değerler almıştır. Ortalama madde güçlüğü 0,27 ve ortalama ayırt edicilik değeri 0,39 olarak hesaplanmıştır. Buradan İzleme 2 testindeki maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir.

Tablo 11.İzleme 3 testine ait madde güçlük ve ayırt edicilik parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ediciliği (rbis)	Ayırt
1	1023_A	0,42	0,39	
2	1023_B	0,52	0,46	
3	1024	0,55	0,49	
4	1028	0,39	0,45	
5	1022_A	0,07	0,06	
6	1022_B	0,33	0,37	
7	1026	0,18	0,25	
8	1020_A	0,29	0,41	
9	1020_B	0,05	0,29	
10	1020_C	0,19	0,22	
11	1021_A	0,13	0,4	
12	1021_B	0,43	0,44	
13	1021_C	0,16	0,22	
	Ortalama	0,29	0,34	

Tablo 11’de görüldüğü gibi 3. İzlemede bulunan her bir maddeye dair madde güçlükleri 0,05 ile 0,55 arasında, madde ayırt edicilik değerleri ise 0,22 ile 0,49 arasında değerler almıştır. Ortalama madde güçlüğü 0,29 ve ortalama ayırt edicilik değeri 0,34 olarak hesaplanmıştır. Buradan maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir.

Tablo 12. *İzleme 4 testine ait madde güçlük ve ayırt edicilik parametreleri*

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	5017_A	0,3	0,42
2	5017_B	0,02	0,23
3	5036_A	0,42	0,54
4	5036_B	0,25	0,37
5	5036_C	0,14	0,26
6	5058_A	0,74	0,5
7	5058_B	0,53	0,52
8	5059	0,22	0,42
9	5057	0,74	0,44
	Ortalama	0,37	0,41

Tablo 12’de görüldüğü gibi 4. İzlemede bulunan her bir maddeye dair madde güçlükleri 0,02 ile 0,74 arasında, madde ayırt edicilik değerleri ise 0,23 ile 0,54 arasında değerler almıştır. Ortalama madde güçlüğü 0,37 ve ortalama ayırt edicilik değeri 0,41 olarak hesaplanmıştır. Buradan 4. izlemedeki maddelerin öğrencilerin büyük çoğunluğuna zor geldiği çıkarımı yapılabilir.

3.3.5.2.BTM’ye Göre Psikometrik Özellikler

Yapılan analizler sonucu BTM’ye göre elde edilen g ve s parametreleri ve ortalamaları tabloda verilmiştir.

Tablo 13. *İzleme 1 testinden BTM'ye göre elde edilen g ve s parametreleri*

soru kodu	g	s
9796	0,261	0,299
3690	0,189	0,694
1027	0,106	0,174
1025	0,122	0,329
1033	0,368	0,213
1021	0,561	0,103
1032	0,269	0,382
7339	0,001	0,969
6728	0,227	0,539
5158_B	0,256	0,18
5158_C	0,211	0,311
Ort.	0,234	0,381

Tablo 13'te görüldüğü gibi İzleme testi 1'in BTM analizi sonucunda maddelerin matematiksel yeterlikler için g parametreleri ortalamaları 0,234 ve s parametrelerinin ortalamaları ise 0,381 olarak hesaplanmıştır. Her iki parametrenin matematiksel yeterlik için g ve s parametre değerleri testin model uyumunun yeterli olduğunu göstermektedir.

Tablo 14. *İzleme 2 testinden BTM'ye göre elde edilen g ve s parametreleri*

soru kodu	g	s
1011_A	0,113	0,451
1011_B	0,17	0,425
1012_A	0,25	0,048
1011_C	0,294	0,485
1012_B	0,26	0,176
1014_A	0,325	0,33
1015_A	0,043	0,699
1014_B	0,001	0,83
1015_B	0,312	0,211
1014_C	0,228	0,743
Ort.	0,2	0,44

İzleme Testi 2'de maddelerin matematiksel yeterlikler için g parametreleri ortalamaları 0,200 ve s parametrelerinin ortalamaları ise 0,440 olduğu elde

edilmiştir. Her iki parametrenin de matematiksel yeterlikler değerlerinin genel olarak düşük olduğu görülmektedir.

Tablo 15. *İzleme 3 testinden BTM'ye göre elde edilen g ve s parametreleri*

soru kodu	g	s
1023_A	0,346	0,336
1023_B	0,356	0,186
1024	0,396	0,024
1028	0,254	0,345
1022_A	0,1	0,989
1022_B	0,172	0,517
1026	0,169	0,776
1020_A	0,222	0,46
1020_B	0,023	0,843
1020_C	0,19	0,801
1021_A	0,067	0,601
1021_B	0,264	0,242
1021_C	0,151	0,833
Ortalama	0,208	0,535

İzleme Testi 3'te matematiksel yeterlikler için maddelerin g parametreleri ortalamaları 0,208 ve s parametrelerinin ortalamaları ise 0,535 olarak hesaplanmıştır. Her iki parametrenin matematiksel yeterlikler için genel olarak düşük olduğu görülmektedir.

Tablo 16. *İzleme 3 testinden BTM'ye göre elde edilen g ve s parametreleri*

soru kodu	g	s
5017_A	0,171	0,329
5017_B	0,03	0,641
5036_A	0,274	0,167
5036_B	0,196	0,611
5036_C	0,121	0,811
5058_A	0,544	0,067
5058_B	0,298	0,235
5059	0,033	0,595
5057	0,672	0,061
Ort.	0,26	0,391

İzleme Testi 4’te yapılan BTM analizleri sonucunda matematiksel yeterlikler için maddelerin g parametreleri ortalamaları 0,26 ve s parametrelerinin ortalamaları ise 0,391 olarak bulunmuştur. Her iki parametrenin matematiksel yeterlikler için g ve s parametre değerleri testin model uyumunun yeterli olduğunu göstermektedir.

3.3.6. Analize alınan maddelerin seçimi

Projede kullanılan 4 izleme testine aynı öğrenciler katılmıştır. Dolayısıyla analize alınan maddeler seçilirken tüm izleme testindeki toplam maddeler, madde havuzu olarak kabul edilmiştir. İzleme testlerinin hepsinde toplam 43 madde kullanılmıştır bunların içerisinde araştırmada kullanılmak üzere 17 madde seçilmiştir. Bu seçim yapılırken araştırmanın amaçları doğrultusunda kısmi puan verilebilme özelliği taşıyan maddeler ve cevaplara ilişkin örüntü olması gibi kriterler göz önünde bulundurulmuştur. Aşağıda seçilen maddelerin listesi soru kodları ve madde türleri ile birlikte verilmiştir.

Tablo 17. Analiz için seçilen maddelerin formatları

	Madde Kodu	Madde Türü
İzleme 1	3690	Çoktan Seçmeli
	1027	D-Y
	6728	Çoktan Seçmeli
	5158_B	Çoktan Seçmeli
	9796	D-Y
İzleme 2	1015_B	Çoktan Seçmeli
	1014_A	D-Y
	1014_B	Yapılandırılmış Yanıtlı
	1014_C	Çoktan Seçmeli
İzleme 3	1023_B	Çoktan Seçmeli
	1024	Çoktan Seçmeli
	1022_A	D-Y
	1022_B	Çoktan Seçmeli
İzleme 4	5017_A	Çoktan Seçmeli
	5036_A	Çoktan Seçmeli
	5036_B	Çoktan Seçmeli
	5036_C	D-Y

Araştırmada kullanılan dikotomik puanlanmış maddeler için Q matris aşağıda verilmiştir.

Tablo 18. *DPM için Q matris*

	A1	A2	A3	A4
1027	0	1	1	1
6728	1	0	0	1
1025	1	0	0	1
1032	0	0	1	0
5158_B	1	0	1	0
9796	1	0	0	0
1016	0	1	0	0
1014_A	1	0	0	0
1023_B	1	1	0	0
1024	0	0	1	0
1028	0	0	1	0
1022_B	0	1	0	0
1026	1	0	0	1
1020_A	1	1	0	0
5017_A	0	0	1	1
5036_B	1	1	0	0
5058_B	0	1	0	0

Araştırmada kullanılan kısmi puanlanmış maddeler için Q matris aşağıda verilmiştir.

Tablo 19. *KPM için Q matris*

	A1	A2	A3	A4	A5	A6	A7	A8
1027v1	0	0	1	1	1	1	1	1
1027v2	0	0	1	0	0	0	1	0
6728v1	1	1	0	0	0	0	1	1
6728v2	0	0	0	0	0	0	1	1
1025v1	1	1	0	0	0	0	1	1
1032v1	0	0	0	0	1	1	0	0
5158_Bv1	1	1	0	0	1	1	0	0
5158_Bv2	1	0	0	0	0	0	0	0
9796v1	1	1	0	0	0	0	0	0
9796v2	1	0	0	0	0	0	0	0
1016v1	1	1	0	0	0	0	0	0
1014_Av1	1	1	0	0	0	0	0	0
1014_Av2	1	0	0	0	0	0	0	0
1023_Bv1	1	1	1	1	0	0	0	0
1023_Bv2	1	0	0	0	0	0	0	0
1024v1	0	0	0	0	1	1	0	0
1024v2	0	0	0	0	1	0	0	0

1028v1	0	0	0	0	1	1	0	0
1022_Bv1	1	1	1	1	0	0	0	0
1022_Bv2	1	0	0	0	0	0	0	0
1026v1	1	1	0	0	0	0	1	1
1020_Av1	1	1	1	1	0	0	0	0
5017_Av1	0	0	0	0	1	1	1	1
5017_Av2	0	0	0	0	1	0	1	0
5036_Bv1	1	1	1	1	0	0	0	0
5036_Bv2	1	1	1	0	0	0	0	0
5058_Bv1	0	0	1	1	0	0	0	0

Analize seçilen maddelere ait örnekler aşağıda verilmiştir.

SORU 3.

BU SORUNUN CEVABINI KİTAPÇIK ÜZERİNE YAZINIZ VE OPTİK FORMDA "Y" KUTUSUNU DOLDURUNUZ

1027

Arzu ve Simge hem fotoğraf hem video çekilmesini ve bunların paylaşılmasını sağlayan Snapchat uygulamasına üye olmuşlar ve takipçi sayılarını arttırmak istemişlerdir. Arzu ve Simge'nin haftalara göre artan takipçi sayıları aşağıdaki tabloda gösterilmiştir.



Hafta	Arzu	Simge
1. Hafta	2	3
2. Hafta	4	9
3. Hafta	8	27
4. Hafta	16	81

Yukarıdaki bilgilere göre aşağıda verilen her bir ifade için uygun olan **Doğru** ya da **Yanlış** seçeneğini yuvarlak içine alınız.

Simge'nin ve Arzu'nun 5. haftadaki takipçi sayısı arasındaki fark 211'dir.	DOĞRU	YANLIŞ
Arzu'nun 3. Haftadaki takipçi sayısını 2^3 ile ifade edebiliriz.	DOĞRU	YANLIŞ
Simge'nin 2. Haftadaki takipçi sayısını 3^3 ile ifade edebiliriz.	DOĞRU	YANLIŞ
Arzu'nun 8. Haftadaki takipçi sayısı 9. Haftadaki takipçi sayısının yarısına eşittir.	DOĞRU	YANLIŞ

Maddenin dikotomik puanlama için hazırlanan Q matrisin oluşum aşamasındaki matematiksel yeterlikler dağılımı aşağıdaki tabloda verilmiştir.

Tablo 20. DPM için Matematiksel Yeterlikler

Matematiksel Yeterlikler	
İletişim/İlişkilendirme (A1)	0
Matematikleştirme (A2)	1
Akıl Yürütme ve Kanıt Gösterme/Strateji Oluşturma (A3)	1
Sembolik, Formel ve Teknik Dil ve İşlemlerin Kullanılması (A4)	1

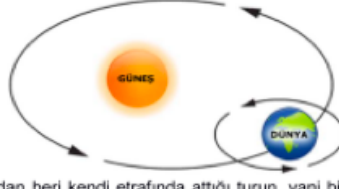
Maddenin kısmi puanlama için hazırlanan Q matrisin oluşum aşamasındaki matematiksel yeterlikler dağılımı aşağıdaki tabloda verilmiştir.

Tablo 21. KPM için Matematiksel Yeterlikler

	A11	A12	A21	A22	A31	A32	A41	A42
Tam	0	0	1	1	1	1	1	1
Kısmi	0	0	1	0	0	0	1	0

DÜNYA YAVAŞLIYOR!

Dünya'nın kendi etrafında attığı her tura bir gün denir. Dünya'nın Güneş etrafında attığı tura 1 yıl denir. Biliminsanları Dünya'nın Güneş etrafında attığı turun süresinin değişmediğini görmekle birlikte Dünya'nın kendi etrafında attığı tur sayısının değişim gösterdiğini keşfetmişlerdir. Bu keşif tıpkı ağaç kütüklerinin halkalarının yılları temsil etmesi gibi fosillerdeki bantların da günleri gösterdiğinin bulunmasıyla gerçekleştirilmiştir. Keşifle Dünya'nın oluşumundan beri kendi etrafında attığı turun, yani bir günün süresinin değiştiğini kanıtlamıştır. Dünya'nın dönme hızını belirleyen temel etkenin Dünya'nın ağırlığı ve enerjisi olduğunu fark eden biliminsanları Dünya'nın günümüzde kendi etrafında geçmişe göre daha yavaş döndüğünü bulmuşlardır. Örneğin Buzul Çağlarında Dünya'nın ağırlığı arttığı için enerjisi azalmış ve günlerin süresi uzamıştır. Biliminsanlar aşağıdaki tabloyu bu bilgilere dayanarak elde etmişlerdir.



Dönemin Adı	Zaman Dilimi	Yırdaki Gün Sayısı	1 Günün Süresi
Günümüz	Şuan	365	24 saat
Üst Kretase	70 milyon yıl önce	370	
Pensilvaniyen	290 milyon yıl önce	383	
Missisipiyen	340 milyon yıl önce	398	
Orta Devoniyen	395 milyon yıl önce	405	
Alt Siluriyen	440 milyon yıl önce	421	
Üst Ordovisyen	450 milyon yıl önce	414	
Orta Kambriyen	510 milyon yıl önce	424	
Edikara	600 milyon yıl önce	417	

SORU 4. BU SORUNUN DOĞRU SEÇENEĞİNİ OPTİK FORMDA İŞARETLEYİNİZ

P5036_B

Aşağıdaki metnin altı çizili kısmında belirtilen sayısal değerleri elde etmek için gerekli eşitlik hangisidir?

"Güneş etrafında bir tam tur atma süresinin hep sabit olduğu dikkate alındığında Missisipiyen Döneminde bir günün yaklaşık 22 saat olduğu elde edilebilir."

Bu soruyu "çözebileceğimi düşünüyorum" "çözebileceğimi düşünmüyorum" Bu seçeneği işaretlediyseniz bile, yine de soruyu size en doğru gelen yanıtla cevaplayınız!

A) $a = \frac{(365 \times 22)}{398}$ B) $a = \frac{(365 \times 24)}{398}$ C) $a = \frac{(398 \times 24)}{365}$ D) $a = \frac{(398 \times 22)}{365}$

Maddenin dikotomik puanlama için hazırlanan Q matrisin oluşum aşamasındaki matematiksel yeterlikler dağılımı aşağıdaki tabloda verilmiştir.

Tablo 22.DPM için Matematiksel Yeterlikler

Matematiksel Yeterlikler	
İletişim ve İlişkilendirme	1
Matematikleştirme	1
Akıl Yürütme ve Strateji Geliştirme	0
Sembolik ve Teknik Dil Kullanımı	0

Maddenin kısmi puanlama için hazırlanan Q matrisin oluşum aşamasındaki matematiksel yeterlikler dağılımı aşağıdaki tabloda verilmiştir.

Tablo 23. *KPM için Matematiksel Yeterlikler*

	A11	A12	A21	A22	A31	A32	A41	A42
Tam	1	1	1	1	0	0	0	0
Kısmi	1	1	1	0	0	0	0	0

3.3.7. *Kullanılan analizler*

Cdm analizleri ve klasik test teorisi için karşılaştırma amaçlı istatistiksel testler uygulanmıştır. Cdm için R , klasik test için iteman 4.2 diğer analizler için SPSS 23 kullanılmıştır.



BÖLÜM IV

BULGULAR

Bu bölümde, araştırma sonucunda elde edilen bulgulara ve bulgular hakkındaki yorumlara yer verilmiştir.

4.1. Maddeler Kısmi Puanlandıktan Sonra Bilişsel Tanı Modeli (BTM) Parametrelerinin ve Model Veri Uyumu Değerlerinin Değişimine İlişkin Bulgular

BTM’de DINA model parametreleri incelenirken bu modele ilişkin madde düzeyinde hesaplanan g ve s parametreleri ve bu parametrelere ait standart hatalar göz önünde bulundurulur. Temelinde her maddenin, testi alan grubu iki sınıfa ayırdığı düşüncesi yatmaktadır. Aynı sınıfa düşme durumu, aynı sınıfa düşenlerin maddeyi doğru yanıtlama olasılıklarının eşit olduğu şeklinde yorumlanır. Olasılık temelli olan bu modelde s parametresi; öğrencinin maddeyi yanıtlamak için gerekli tüm özelliklere sahip olduğu halde maddeyi doğru yanıtlamaması, g parametresi; öğrencinin maddeyi yanıtlamak için gerekli özelliklere sahip olmadığı halde maddeyi doğru yanıtlaması şeklinde yorumlanır (De La Torre, 2008).

Aşağıdaki Tablo 24’de kodları verilmiş kısmi puanlanmış maddelerin (KPM) ve dikotomik puanlanmış maddelerin (DPM) g ve s parametre değerleri ve bu değerlere ait hatalar verilmiştir. Dikotomik puanlanmış 1027 kodlu maddenin kısmi puanlanmış hali 1027v1 ve 1027v2’dir. 1027v1 ve 1027v2 aynı maddeyi temsil etmektedir. Bir öğrencinin 1027v1’i 0 iken 1027v2’sinin 1 olması, 1027 kodlu maddesinin kısmi puanlandığı anlamına gelir. 1027v1 ve 1027v2’nin ikisinin de 1 olması, 1027 kodlu maddenin doğru yanıtladığı anlamına gelir. 1027v1 ve 1027v2’nin 0 olması, 1027 kodlu maddenin yanlış yanıtladığı anlamına gelir.

Tablo 24. *KPM ve DPM'nin g ve s parametreleri*

KOD	KPM				DPM				
	g	SE(g)	s	SE(s)	kodu	g	SE(g)	s	SE(s)
1027v1	0,16	0,01	0,02	0,06	1027	0,12	0,01	0,36	0,04
1027v2	0,38	0,02	0,00	0,06					
6728v1	0,25	0,01	0,38	0,04	6728	0,23	0,01	0,50	0,03
6728v2	0,6	0,02	0,16	0,03					
1025v1	0,21	0,01	0,2	0,03	1025	0,13	0,01	0,26	0,03
1032v1	0,37	0,01	0,53	0,02	1032	0,28	0,02	0,45	0,02
5158_Bv1	0,34	0,01	0,42	0,02	5158_B	0,23	0,01	0,20	0,03
5158_Bv2	0,58	0,02	0,32	0,01					
9796v1	0,01	0,02	0	0,01	9796	0,20	0,02	0,31	0,02
9796v2	0,02	0,02	0	0,03					
1016v1	0,38	0,01	0,41	0,02	1016	0,29	0,02	0,34	0,02
1014_Av1	0,38	0,01	0,49	0,02	1014_A	0,28	0,02	0,41	0,02
1014_Av2	0,71	0,02	0,16	0,01					
1023_Bv1	0,19	0,01	0,77	0,03	1023_B	0,18	0,01	0,77	0,02
1023_Bv2	0,56	0,02	0,47	0,01					
1024v1	0	0,02	0	0,03	1024	0,22	0,02	0,42	0,02
1024v2	0,19	0,03	0	0,03					
1028v1	0,47	0,01	0,39	0,02	1028	0,36	0,02	0,32	0,02
1022_Bv1	0,39	0,01	0,27	0,04	1022_B	0,32	0,02	0,44	0,02
1022_Bv2	0,63	0,02	0,32	0,01					
1026v1	0,49	0,01	0,03	0,02	1026	0,42	0,01	0,07	0,02
1020_Av1	0,4	0,01	0,42	0,04	1020_A	0,36	0,01	0,43	0,03
5017_Av1	0,26	0,01	0,45	0,04	5017_A	0,24	0,01	0,53	0,03
5017_Av2	0,58	0,02	0,25	0,03					
5036_Bv1	0,23	0,01	0,64	0,04	5036_B	0,23	0,01	0,69	0,02
5036_Bv2	0,52	0,01	0,44	0,03					
5058_Bv1	0,43	0,02	0,13	0,04	5058_B	0,32	0,02	0,23	0,02
ORT.	0,36		0,28			0,26		0,40	

Tablodaki ilk bölüm olan kısmi puanlanmış maddelere ilişkin g parametreleri incelendiğinde en düşük g parametresine sahip maddenin 0 ile 1024v1 kodlu madde olduğu bulunmuştur. En yüksek g parametresine sahip maddenin 0,71 ile 1014_Av2 kodlu madde olduğu belirlenmiştir. Buradan dikotomik olarak puanlanan maddelerde g parametresinin değerlerinin 0 ve 0,71 arasında değiştiği sonucuna varılır. Kısmi puanlanmış maddelere ilişkin s parametreleri incelendiğinde en düşük s parametresine sahip maddenin 0 değeri ile 1027v2, 9796v2, 1024v1, 1024v2 kodlu

maddeler olduğu görülmüştür. En yüksek s parametresine sahip maddenin ise 0,77 değeri ile 1023_Bv1 kodlu madde olduğu bulunmuştur. Bu durumda s parametresinin 0 ile 0,77 arasında değerler aldığı sonucuna ulaşılır.

Tabloda ikinci bölüm olan dikotomik (ikili) olarak puanlanmış maddelere ilişkin g parametreleri incelendiğinde en düşük g parametresine sahip maddenin 0,12 ile 1027 kodlu madde olduğu bulunmuştur. En yüksek g parametresine sahip maddenin 0,42 ile 1026 kodlu madde olduğu belirlenmiştir. Buradan dikotomik olarak puanlanan maddelerde g parametresinin değerlerinin 0,12 ve 0,42 arasında değiştiği sonucuna varılır. Maddelere ilişkin s parametreleri incelendiğinde en düşük s parametresine sahip maddenin 0,07 değeri ile 1026 kodlu madde olduğu görülmüştür. En yüksek s parametresine sahip maddenin ise 0,77 değeri ile 1023_B kodlu madde olduğu bulunmuştur. Bu durumda s parametresinin 0,07 ile 0,77 arasında değerler aldığı sonucuna ulaşılır.

Tablo 25.DPM'nin en düşük ve en yüksek madde parametreleri

g parametresi		s parametresi		
	madde kodu	Q matris	madde kodu	Q matris
en düşük	1026	"1001"	1023_B	"1100"
en yüksek	1027	"0111"	1026	"1001"

Q matris incelendiğinde en düşük g parametresine sahip 1026 kodlu maddenin (1001) şeklinde Q matrise sahip olduğu ve A1 (iletişim/ilişkilendirme) ve A4 (sembolik, formel ve teknik dil ve işlem kullanma) özellikleri ile ilişkili olduğu görülmüştür. Bu maddenin g parametresinin düşük olması A2,A3,A4 özelliğine sahip olmayan öğrenciler tarafından maddenin doğru yanıtlanma olasılığının düşük olduğu, başka bir deyişle de bu maddeyi doğru yanıtlamak için gerekli özellikler dışında başka bir ihtiyaç duyulmadığı anlamına gelmektedir. Maddenin g parametresinin düşük çıkması bu maddenin q matriste doğru bir şekilde tanımlandığının da bir göstergesi olarak kabul edilebilir. En yüksek g parametresine sahip 1027 kodlu maddenin ise (0111) şeklinde Q matrise sahip olduğu ve A2 (matematikleştirme), A3 (akıl yürütme ve kanıt gösterme/strateji oluşturma), A4 (sembolik, formel ve teknik dil ve işlem kullanma) özellikleri ile ilişkili olduğu görülmüştür. Maddenin yüksek g parametresine sahip olması, maddeyi cevaplamak için gerekli olan özelliklere sahip

olmayan öğrencilerin de maddeyi doğru yanıtlama olasılığının yüksek olduğu anlamına gelir. En düşük s parametre değerine sahip olan 1023_B kodlu maddenin Q matrisi (1100)'dır. Bu madde A1 ve A2 özellikleri ile ilgilidir. Eğer öğrenci A1 ve A2 özelliklerine sahipse bu maddeyi doğru yanıtlama olasılığı yüksektir. En yüksek s parametresine sahip olan 1026 kodlu madde (1001) şeklinde Q matrise sahiptir. A1 ve A4 özellikleri ile ilişkilidir. Bu madde için s parametresinin yüksek olmasının anlamı; A1 ve A4 özelliklerine sahip olan öğrencilerin de bu maddeyi yanlış yanıtlama olasılıklarının yüksek olmasıdır.

DINA model veri uyumu değerlendirilirken "Test-Level Fit Statistics" analiz sonuçlarına bakılmıştır. Akaike Bilgi Kriteri (AIC) ve Bayesian Bilgi Kriteri (BIC) istatistiklerinden elde edilen sonuçlara göre veriler yorumlanmıştır. AIC temel olarak farklı boyutlu modellerin karşılaştırılmasında kullanılan güçlü bir model seçme kriteridir. ((Bandalos, 1993); (Akıncı, 2007)) BIC ise daha çok regresyonda seçilmiş model problemleri için üretilmiş bir kriterdir. (Ucal, 2006) Her iki kriter için de, uyum katsayıları daha düşük olan modelin daha uygun olduğuna karar verilir. (Cavanaugh,2009)

Aşağıda yapılan analizler sonucunda elde edilen AIC ve BIC istatistik değerleri verilmiştir.

Tablo 26. DPM ve KPM için test uyum istatistikleri

	DPM		KPM
-2LL	475.166.851	-2LL	742.412.476
AIC	476.146.851	AIC	748.592.476
BIC	478.947.043	BIC	766.250.829

Tablolar incelendiğinde kısmi puanlama yöntemiyle madde sayısı arttığı için parametre sayısı da artmıştır. Dolayısıyla sonuçlar kısmi puanlama yöntemi için daha yüksek çıkmıştır.

Aşağıda yapılan analizler sonucu elde edilen dikotomik puanlanmış ve kısmi puanlanmış maddeler için madde uyum istatistik değerleri olan ortalamaların mutlak dağılımı (ORT. MD), maksimum mutlak dağılım, (MAKS. MD) ve standart hatası verilmiştir.

Tablo 27.DPM ve KPM için madde uyum istatistikleri

	DPM			KPM		
	prop	Z	Log(OR)	Prop	Z(Corr)	Log(OR)
ORT. MD	0.0053	0.0452	0.1964	0.0103	0.0610	+.Inf
MAKS. MD	0.0136	0.1753	0.7340	0.1035	0.7473	+.Inf
Standart hata (MAKS. MD)	0.0105	0.0211	0.0966	0.0101	0.0211	+.Inf

Araştırma sonuçları incelendiğinde kısmi puanlanmış maddeler için analizler yapılırken madde sayısı arttığı için parametre sayısı da artmıştır. Bu sebeple de kısmi puanlanmış maddelere göre dikotomik maddeler sonuçlar daha yüksek çıkmıştır.

4.2.Maddeler Kısmi Puanlandıktan Sonra Örtük Sınıf Sonsal Olasılıkları ve Tahminlenen Özellik Dağılımlarına İlişkin Bulgular

Bu bölümde önce dikotomik puanlanan maddeler (DPM) için elde edilen örtük sınıflar ve onların sonsal olasılıklarına ait sonuçlar verilmiştir. Daha sonra kısmi puanlama yapılan maddeler (KPM) için örtük sınıflar ve sonsal olasılıkları incelenmiştir. DPM için gözlenen örtük sınıf sonsal olasılıkları, KPM ile karşılaştırılabilmesi için KPM’de karşılık gelen olasılık desenleri için toplam olasılıklar belirlenip yazılmıştır. KPM’de analiz sonrası gözlenen örtük sınıf sayıları(KPM Ö.S.S) ayrıca belirtilmiştir.

Tablo 28.DPM ve KPM için örtük sınıflar ve sonsal olasılıkları

Örtük sınıf	DPM	KPM	KPM Ö.S.S
"0000"	0,0811	0,0029	1
"1000"	0,0665	0,0106	1
"0100"	0,0583	0,0084	3
"0010"	0,0652	0,0104	3
"0001"	0,0811	0,0086	3
"1100"	0,0554	0,0303	9
"1010"	0,0525	0,0389	9
"1001"	0,0432	0,0307	9
"0110"	0,0530	0,0301	9
"0101"	0,0583	0,0226	9
"0011"	0,0632	0,0302	9

"1110"	0,0596	0,115	27
"1101"	0,0531	0,089	27
"1011"	0,0507	0,1162	27
"0111"	0,0442	0,0825	27
"1111"	0,1146	0,3763	81

Tablo incelendiğinde kısmi puanlar için maddeler daha fazla örtük sınıflara dağıldığından dolayı daha düşük olasılık değerleri aldığı görülmektedir. Bununla birlikte olasılık değerlerinin toplamının dikotomik maddelere göre birbirine yakın olduğu anlaşılmaktadır. Dikotomik maddeler için 16 tane örtük sınıf varken kısmi puanlanmış maddeler için 256 tane farklı örtük sınıf ile karşılaşmıştır. Örneğin (1100) dörtlü örtük sınıfa karşılık, 8'li örtük sınıfta (10000000), (11000000), (10100000), (11100000), (11110000), (01000000), (00010000), (01010000), (01100000), (01110000) şeklindeki örtük sınıflardan herhangi biri karşılık gelebilir. Bu durum aynı öğrenciler için daha detaylı bir analiz yapıldığının göstergesidir. Ayrıca özellikle daha fazla özelliğe sahip olan sınıflarda olasılıkların giderek arttığı gözlenmiştir.

DPM ve KPM için özelliklerin dağılımlarına ilişkin tahminlere ait veriler Tablo 29'da verilmiştir. A1 ile ifade edilen özellik iletişim ve ilişkilendirme olup A11 iletişim ve ilişkilendirme birinci düzey, A12 iletişim ve ilişkilendirme ikinci düzeyi ifade etmektedir. A2 matematikleştirme, A21 ve A22 sırasıyla matematikleştirme özelliğinin birinci ve ikinci özelliğini ifade etmektedir.

A3 akıl yürütme, kanıt gösterme ve strateji geliştirme özelliği yine A31 ve A32 akıl yürütme, kanıt gösterme ve strateji geliştirme birinci ve ikinci düzeyini ifade etmektedir. A4 sembolik ve formel teknik dil ve işlem kullanma özelliği ile ilişkili olup A41 ve A42 yine bu özelliğin birinci ve ikinci düzeyini temsil etmektedir. A1, A2, A3 ve A4 dikotomik puanlanan özellikler olup A11, A12, A21, A22, A31, A32, A41, A42 kısmi puanlanan özelliklerdir.

Tablo 29. DPM ve KPM için tahminlenen özellik dağılımları

A1	0,49	A11	0,61
		A12	0,63
A2	0,49	A21	0,50
		A22	0,50
A3	0,50	A31	0,60
		A32	0,60
A4	0,50	A41	0,51
		A42	0,51

Tablo incelendiğinde bir dağılım görüldüğü söylenebilir. Kısmi puanlanan özelliklere bakıldığında 8 özellik için olasılıkların 0,5047 ile 0,6342 arasında değiştiği söylenebilir. A11 ve A12 özelliği DPM’de ait özelliklerde A1’e karşılık gelmektedir. Tablo 29’da A1 için 0,4956 gözlenme olasılığı varken bu özellik ikiye bölündüğünde gözlenme olasılığı A11 için 0,6097 ve A12 için 0,6347 düzeyine yükseldiği görülmektedir. Bu durum hassasiyetin arttığının diğer bir deyişle de gözlenme durumunun arttığının bir göstergesi olarak kabul edilebilir.

4.3.Maddeler Kısmi Puanlandıktan Sonra Örtük Sınıflar İçin Toplam Puanlar ve Madde Tepki Kuramı (MTK) Yetenek Kestirimlerine İlişkin Bulgular

Bu bölümde dikotomik ve kısmi puanlanan maddeler için yetenek kestirimleri sınıflar bazında incelenerek toplam puanların ortalaması ve MTK puanlarının ortalaması verilmiştir. Toplam puan değerlerini karşılaştırabilmek için ortalama güçlük değerleri hesaplanmıştır. Tabloda ortalama güçlük ve MTK değerleri verilmiştir.

Tablo 30.DPM ve KPM Sınıflarını için yetenek kestirimleri

DPM SINIF	DPM ORT.	DPM MTK	P KPM	P KPM MTK(min/max)
0 0 0 0	0,19	-0,05	0,22	-0,05
0 0 0 1	0,25	-0,04	0,30	-0,04
0 0 1 0	0,28	-0,03	0,33	-0,95 (-1,16/-0,75)
0 0 1 1	0,33	-0,02	0,40	-0,82 (-0,97/-0,67)
0 1 0 0	0,30	-0,02	0,28	-0,04
0 1 0 1	0,36	0,00	0,39	-0,52 (-0,86/-0,15)

0 1 1 0	0,41	0,00	0,37	-0,71 (-1,01/-0,47)
0 1 1 1	0,45	0,02	0,47	-0,26 (-0,72/0,61)
1 0 0 0	0,29	-0,02	0,36	-0,40 (0,53/-0,28)
1 0 0 1	0,37	0,01	0,44	0,10 (-0,67/0,51)
1 0 1 0	0,40	0,00	0,41	-0,17 (-0,50/0,14)
1 0 1 1	0,51	0,04	0,49	0,18 (-0,72/0,81)
1 1 0 0	0,42	0,01	0,45	0,01 (-0,36/0,30)
1 1 0 1	0,51	0,04	0,51	0,34 (-0,27/1,06)
1 1 1 0	0,51	0,03	0,51	0,19 (-0,31/0,75)
1 1 1 1	0,67	0,08	0,60	0,64 (0,02/1,74)

Tablo 30 incelendiğinde dikotomik ve kısmi puanlanmış maddelerin yetenek kestirimlerinde ortalama güçlük değerlerine ilişkin en fazla fark (0000) örtük sınıfı ile (1111) arasındadır. Özellikle düşük sınıflar ve yüksek sınıflar arasında yüksek düzeyde fark görülmüştür.

Dikotomik puanlanmış maddelerde yokluk (null, (0000)) sınıfında ortalama güçlük 0,19'dur. Kısmi puanlanmış maddelerin yokluk sınıfı (00000000) ortalama güçlük 0,22'dir.

Maddeler yapılandırılmış yanıtı madde formuna dönüştürüldüğünde örtük sınıflar için ortalama güçlük değerleri ve MTK yetenek kestirimlerine ilişkin bulgular incelendiğinde öğrencilerin örtük sınıf dağılımlarının birbirine yakın olduğu gözlenmektedir. İki durum için de yüksek bir korelasyon belirlenmekle birlikte dikotomik puanlama yöntemiyle aynı aralığa düşen öğrencilerin kısmi puanlama yöntemiyle olduğunda daha geniş bir ranjda kestirimler yapıldığı gözlenmiştir. Bu durum güvenilirlik açısından kısmi puanlanan maddeler için daha güvenilir ölçümler yapıldığının bir göstergesi olarak kabul edilebilir.

4.4.Maddeler Yapılandırılmış Yanıtı Madde Formuna Dönüştürüldüğünde Sahip Olunan Özellik Sayısına Göre Toplam Puanlar ve MTK Puanlarına İlişkin Bulgular

Q matris yeniden yapılandırıldıktan sonra dikotomik puanlanan maddelerden MTK'ya göre elde edilen yetenek puanları ve KTK için Mutlak Başarı Yüzdeleri(MBY) Tablo 31'de verilmiştir.

Tablo 31. *DPM için Toplam Özellik Sayısına Karşılık Gelen MTK ve KTK puanları*

DKM _{top}	MTK	MBY
0	-0,92	19,24
1	-0,43	28,94
2	0,01	38,82
3	0,55	50,41
4	1,32	67,41

Q matris yeniden yapılandırıldıktan sonra kısmi puanlanan maddelerden MTK'ya göre elde edilen yetenek puanları ve KTK için Mutlak Başarı Yüzdeleri(MBY) Tablo 32'de verilmiştir.

Tablo 32. *KPM için Toplam Özellik Sayısına Karşılık Gelen MTK ve KTK puanları*

KPM _{top}	MTK	MBY
0	-1,46	22,30
1	-1,08	29,37
2	-0,66	35,41
3	-0,34	41,41
4	-0,11	45,11
5	0,23	52,30
6	0,68	60,15
7	1,09	67,56
8	1,74	76,67

Tablodaki veriler incelendiğinde örtük özellik sayısı arttıkça iki puan türünde de öğrencilere ait kestirimlerin arttığı görülmektedir. Bununla birlikte kısmi puanlanan maddeler için hem daha alt yetenek düzeyi hem üst yetenek düzeyi kestirilebildiği gözlenmiştir. Örneğin; MTK analizlerinde hiçbir özelliğe sahip olmayan grubun ortalaması -0,92'ken aynı durum kısmi puanlanmış maddelerde -1,46'ya kadar gelmiştir. Bu durum öğrencileri daha geniş bir ölçek düzeyinde değerlendirmenin mümkün olduğunu göstermektedir.

4.5.Maddeler Kısmi Puanlandıktan Sonra Özelliğe Sahip Olma Dağılımlarına İlişkin Bulgular

Dikotomik ve kısmi puanlanan maddelerin özelliğe sahip olmalarına ilişkin veriler Tablo 33'te verilmiştir. Tabloda sahip olunan özellik sayısına ait öğrencilerin

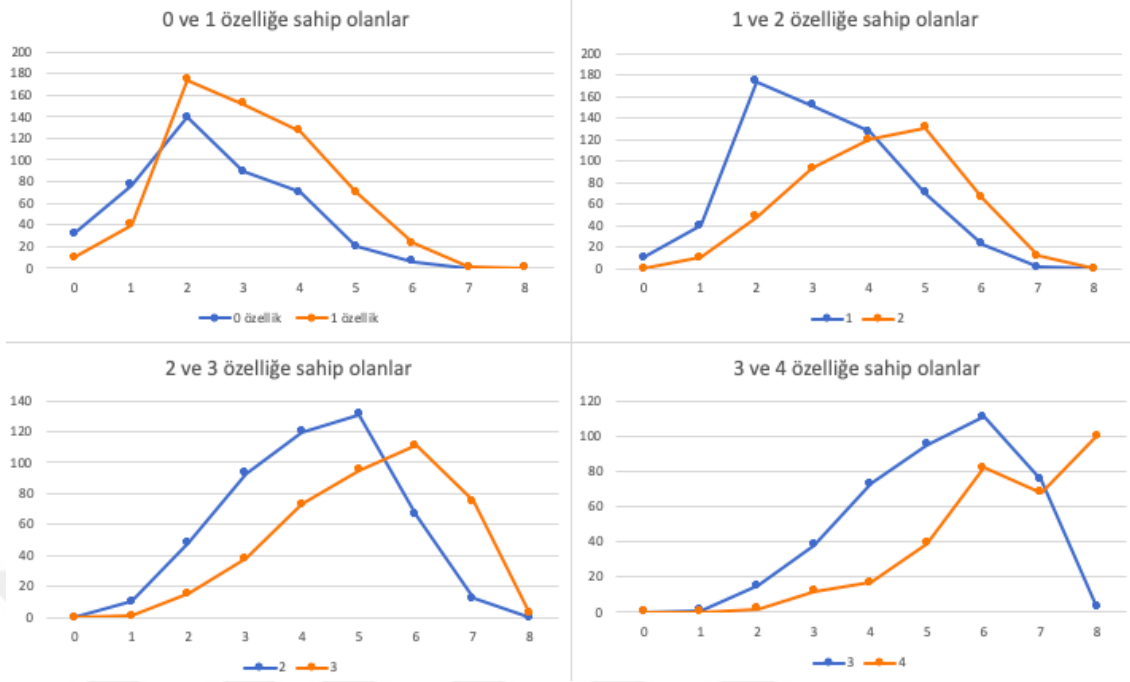
frekansları verilmiştir. DPM'ye göre toplam 4 özellik , KPM'ye göre 8 özellik üzerinden inceleme yapılmıştır.

Tablo 33. KPM ve DPM için Toplam Özellik Sayısına Karşılık gelen Frekansları

DİKOTOMİK					
	0	1	2	3	4
0	32	10	0	0	0
1	77	40	10	1	0
2	139	174	48	15	2
3	89	152	93	38	12
4	70	127	120	73	17
5	20	70	131	95	39
6	6	23	66	111	82
7	0	1	12	75	68
8	0	0	0	3	100

Tabloda Q matrisin kısmi puanlamaya göre yapılandırılmasından sonra özelliğe sahip olma dağılımlarına ilişkin veriler incelendiğinde iki durumda da özelliğe sahip olma dağılımlarının benzer olduğu gözlenmekle birlikte 1,2,3 ve 4 özelliğe sahip olan öğrencilerin kısmi puanlamalarda daha yüksek sınıflara doğru dağıldıkları grafikte görülmektedir.

Aşağıda sırasıyla 0 ve 1, 1 ve 2, 2 ve 3 ve 3 ve 4 özelliğe sahip olan kişilerin dağılımlarına ilişkin grafikler verilmiştir. Grafiklerde dikey eksen kişilerin frekanslarını temsil ederken, yatay eksen kısmi puanlanmış maddeler için sahip olunan özellik sayılarını ifade etmektedir



Grafikler incelendiğinde, 4 grafikte de kırmızı ile ifade edilen kısımlar yani maviye göre bir fazla özellikli gruplar, genelde yatay eksenle ifade edilen 8’li özellik grubunda daha yüksek özellik göstermektedir.

4.6.Maddeler Kısmi Puanlandıktan Sonra İletişim ve İlişkilendirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular

Tablo öncelikle A1 özelliğine sahip olanlar ve olmayanlar olarak ikiye ayrılmıştır. A1 özelliği Q matriste iletişim ve ilişkilendirme becerisini temsil etmektedir.

Tablo 34.İletişim ve İlişkilendirme özelliğine sahip olma düzeyleri

		İletişim ve ilişkilendirme birinci düzey		
		0	1	
		İletişim ve ilişkilendirme ikinci düzey		
		0	0	1
İletişim	ve 0	733	306	166
İlişkilendirme	1	126	86	824

A1 özelliğine sahip olmayan 1025 kişi, A1 özelliğine sahip olanlar 1036 kişidir. Daha sonra iki düzeye ayrılmış olan özellikler A11 ve A12 olarak ifade edilmiştir. A11, A1 özelliğine 1. Düzeyde sahip olup olmama durumunu temsil eder.

A12 ise A1 özelliğine 2.düzye sahip olup olmama özelliğini temsil eder. A1 özelliğine 1.düzye sahip olmayan ve A1 özelliği gerektiren maddeleri doğru yanıtlayan 733 kişidir. A1 özelliğine sahip 1.düzye sahip olmadığı halde A1 özelliği gerektiren maddeleri doğru yanıtlayanlar ise 126 kişidir. A1 özelliğine hem 1 hem de 2.düzye sahip olan kişi 824 kişi olup bu kişiler A1 özelliği gerektiren maddeleri yanıtlamıştır.

4.6.1. Maddeler Kısmi Puanlandıktan Sonra Matematikleştirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular

Tablo öncelikle A2 özelliğine sahip olanlar ve olmayanlar olarak ikiye ayrılmıştır. A2 özelliği Q matriste matematikleştirme becerisini temsil etmektedir. Daha sonra matematikleştirme özelliği iki düzye ayrılmıştır.

Tablo 35. Matematikleştirme özelliğine sahip olma düzeyleri

		Matematikleştirme birinci düzey			
		0		1	
		Matematikleştirme ikinci düzey			
		0	1	0	1
Matematikleştirme	0	644	128	306	87
	1	86	227	177	586

A2 özelliğine sahip olmayan 1165 kişi, A2 özelliğine sahip olanlar 1076 kişidir. Daha sonra iki düzye ayrılmış olan özellikler A21 ve A22 olarak ifade edilmiştir. A21, A2 özelliğine 1. Düzye sahip olup olmama durumunu temsil eder. A22 ise A2 özelliğine 2.düzye sahip olup olmama özelliğini temsil eder. A2 özelliğine 1.düzye sahip olmayan ve A2 özelliği gerektiren maddeleri doğru yanıtlayan 644 kişidir. A2 özelliğine sahip 1.düzye sahip olmadığı halde A2 özelliği gerektiren maddeleri doğru yanıtlayanlar ise 86 kişidir. Hem A1 özelliğine hem 1 hem de 2.düzye sahip olan kişi 586 kişi olup bu kişiler A1 özelliği gerektiren maddeleri yanıtlamıştır. Tablo 35'te 306 kişiyi temsil eden durum ise özelliğe birinci düzye sahipken ikinci düzye sahip olmamadır.

128 kişi ise özelliğe birinci düzeyde sahip olmadığı halde ikinci düzeyde sahip olma ile ilişkilidir. Bu durum özelliğin aslında tam olarak hiyerarşik olmaması ile açıklanabilir.

4.6.2. Maddeler Kısmi Puanlandıktan Sonra Akıl Yürütme, Kanıt Gösterme ve Strateji Geliştirme Özelliğine Sahip Olma Düzeylerine İlişkin Bulgular

Tablo öncelikle A3 özelliğine sahip olanlar ve olmayanlar olarak ikiye ayrılmıştır. A3 özelliği Q matriste akıl yürütme, kanıt gösterme ve strateji geliştirme becerisini temsil etmektedir. Daha sonra bu özellik iki düzeye ayrılmıştır.

Tablo 36. Akıl yürütme, kanıt gösterme ve strateji geliştirme özelliğine sahip olma düzeyleri

		Akıl yürütme, kanıt gösterme birinci düzey		
		0	1	
		Akıl yürütme, kanıt gösterme ikinci düzey		
		0	0	1
Akıl yürütme, kanıt gösterme ve strateji geliştirme	0	537	456	153
	1	198	156	741

A3 özelliğine sahip olmayan 1146 kişi, A3 özelliğine sahip olanlar 1095 kişidir. Daha sonra iki düzeye ayrılmış olan özellikler A31 ve A32 olarak ifade edilmiştir. A31, A3 özelliğine 1. Düzeyde sahip olup olmama durumunu temsil eder. A32 ise A3 özelliğine 2.düzeyde sahip olup olmama özelliğini temsil eder. A3 özelliğine 1.düzeyde sahip olmayan ve A3 özelliği gerektiren maddeleri doğru yanıtlamayan 537 kişidir. A3 özelliğine sahip 1.düzeyde sahip olmadığı halde A3 özelliği gerektiren maddeleri doğru yanıtlayanlar ise 198 kişidir. A3 özelliğine hem 1 hem de 2.düzeyde sahip olan kişi 741 kişi olup bu kişiler A3 özelliği gerektiren maddeleri yanıtlamıştır.

4.6.3. Maddeler Kısmi Puanlandıktan Sonra Sembolik ve Formel Teknik Dil Kullanma özelliğine sahip olma düzeylerine ilişkin bulgular

Tablo öncelikle A4 özelliğine sahip olanlar ve olmayanlar olarak ikiye ayrılmıştır. A4 özelliği Q matriste sembolik, formel teknik dil ve işlem kullanma becerisini temsil etmektedir.

Tablo 37. Sembolik, Formel Teknik Dil ve İşlem Kullanımına Sahip olma düzeyleri

		Sembolik, Formel Dil ve İşlem birinci düzey			
		0		1	
		Sembolik, Formel Dil ve İşlem ikinci düzey			
		0	1	0	1
Sembolik ve	0	737	224	257	160
Formel Teknik Dil ve İşlem	1	88	98	126	551

A4 özelliğine sahip olmayan 1378 kişi, A4 özelliğine sahip olanlar 863 kişidir. Daha sonra iki düzeye ayrılmış olan özellikler A41 ve A42 olarak ifade edilmiştir. A41, A4 özelliğine 1. Düzeyde sahip olup olmama durumunu temsil eder. A42 ise A4 özelliğine 2.düzeyde sahip olup olmama özelliğini temsil eder. A4 özelliğine 1.düzeyde sahip olmayan ve A4 özelliği gerektiren maddeleri doğru yanıtlamayan 737 kişidir. A4 özelliğine sahip 1.düzeyde sahip olmadığı halde A4 özelliği gerektiren maddeleri doğru yanıtlayanlar ise 88 kişidir. Hem A4 özelliğine hem 1 hem de 2.düzeyde sahip olan kişi 551 kişi olup bu kişiler A4 özelliği gerektiren maddeleri yanıtlamıştır.

BÖLÜM V

TARTIŞMA VE YORUM

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

5.1.Yapılandırılmış Yanıtlı Maddelerde Bilişsel Tanı Modeli (BTM) Parametrelerinin ve Model Veri Uyumu Değerlerinin Değişimine İlişkin Bulguların Tartışılması ve Yorumu

Birinci alt probleme ilişkin bulgular incelendiğinde g ve s parametrelerinde değişimler gözlenmektedir. Maddeler dönüştürüldükten sonra parametre değerleri farklılaşmıştır. Kısmi puanlanmış maddelere ait sonuçlar incelendiğinde dikotomik maddelere ait elde edilen sonuçlara göre daha yüksek parametre değerleri aldığı gözlenmiştir. Fit istatistik değerleri incelendiğinde ise kısmi puanlanmış maddelere ait sonuçlar daha yüksek olduğu saptanmıştır. Bunun temel sebebi kısmi puanlama sonrasında parametre sayısının artması olarak görülebilir.

5.2.Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde Örtük Sınıf Sonsal Olasılıkları ve Tahminlenen Özellik Dağılımlarına İlişkin Bulguların Tartışılması ve Yorumu

İkinci alt probleme ilişkin bulgular incelendiğinde kısmi puanlar için maddeler daha fazla örtük sınıflara dağıldığından dolayı daha düşük olasılık değerleri aldığı görülmektedir. Bununla birlikte olasılık değerlerinin toplamının dikotomik maddelere göre birbirine yakın olduğu anlaşılmaktadır. Dikotomik maddeler için 16 tane örtük sınıf varken kısmi puanlanmış maddeler için 256 tane farklı örtük sınıf ile karşılaşmıştır. Bu durum aynı öğrenciler için daha detaylı bir analiz yapıldığının göstergesidir. Örtük sınıf dağılımlarına bakıldığında maddelerin özellikleri gözlenme olasılıklarının kısmi puanlanmış maddeler için daha yüksek olduğu gözlenmektedir. Bu da öğrencileri aynı testle daha fazla olasılıkla tahmin edebilme durumunun bir göstergesi olarak kabul edilebilir.

5.3.Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde Örtük Sınıflar İçin Toplam Puanlar ve Madde Tepki Kuramı (MTK) Yetenek Kestirimlerine İlişkin Bulguların Tartışılması ve Yorumu

Maddeler yapılandırılmış yanıtlı madde formuna dönüştürüldüğünde örtük sınıflar için toplam puanlar ve madde tepki kuramı (mtk) yetenek kestirimlerine ilişkin bulgular incelendiğinde öğrencilerin toplam puanlar ve madde tepki kuramı yetenek kestirimlerine göre örtük sınıf dağılımlarının birbirine yakın olduğu gözlenmektedir. İki durum için de yüksek bir korelasyon belirlenmekle birlikte dikotomik puanlama yöntemiyle aynı aralığa düşen öğrencilerin kısmi puanlama yöntemiyle olduğunda daha geniş bir ranjda kestirimler yapıldığı gözlenmiştir. Bu durum güvenilirlik açısından kısmi puanlanan maddeler için daha güvenilir ölçümler yapıldığının bir göstergesi olarak kabul edilebilir.

5.4.Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde Sahip Olunan Özellik Sayısına Göre Toplam Puanlar ve MTK Puanlarına İlişkin Bulguların Tartışılması ve Yorumu

Maddeler yapılandırılmış yanıtlı madde formuna dönüştürüldüğünde sahip olunan özellik sayısına göre toplam puanlar ve mtk puanlarına ilişkin bulgular incelendiğinde örtük özellik sayısı arttıkça iki puan türünde de öğrencilere ait kestirimlerin arttığı gözlenmektedir. Bununla birlikte kısmi puanlanan maddeler için hem daha alt yetenek düzeyi hem üst yetenek düzeyi kestirilebildiği gözlenmiştir. Örneğin; MTK analizlerinde hiçbir özelliğe sahip olmayan grubun ortalaması -0,92'ken aynı durum kısmi puanlanmış maddelerde -1,45'e kadar gelmiştir. Bu durum öğrencileri daha geniş bir ölçek düzeyinde değerlendirmenin mümkün olduğunu göstermektedir.

5.5.Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde Özelliğe Sahip Olma Dağılımlarına İlişkin Bulguların Tartışılması ve Yorumu

Maddeler yapılandırılmış yanıtlı madde formuna dönüştürüldüğünde özelliğe sahip olma dağılımlarına ilişkin bulgular incelendiğinde iki durumda da özelliğe sahip olma dağılımlarının benzer olduğu gözlenmekle birlikte 1,2,3 ve 4 özelliğe sahip olan öğrencilerin kısmi puanlamalarda daha yüksek sınıflara doğru dağıldıkları grafikte görülmektedir.

5.6.Maddeler Yapılandırılmış Yanıtlı Madde Formuna Dönüştürüldüğünde İletişim ve İlişkilendirme, Matematikleştirme, Akıl Yürütme, Kanıt Gösterme ve Strateji Geliştirme, Sembolik ve Formel Teknik Dil Kullanma Özelliğine Sahip Olma Düzeylerine İlişkin Bulguların Tartışılması ve Yorumu

Maddeler yapılandırılmış yanıtlı madde formuna dönüştürüldüğünde iletişim ve ilişkilendirme, matematikleştirme, akıl yürütme, kanıt gösterme ve strateji geliştirme, sembolik ve formel teknik dil kullanma özelliğine sahip olma düzeylerine ilişkin bulgular incelendiğinde testte ölçülen dört beceri için öğrencilerin özelliklere sahip olma durumundaki değişimler gözlenmiştir. Her dört özellik için de özellikle kısmi puanlanan maddelerle yapılan analizlerde öğrencilerin belirtilen özelliğe ait daha detaylı bilgiler elde edildiği görülmüştür. Aynı zamanda bu bilgilerin hepsinin yapılan ilk analizlerle tutarlı olduğu görülmüştür.

BÖLÜM VI

SONUÇ VE ÖNERİLER

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

6.1.Sonuç

Araştırma sonuçları incelendiğinde data genişletme yöntemi kullanılarak hazırlanan bir testin yapısal olarak daha güvenilir olduğu ve öğrenciler hakkında daha detaylı bilgi verecek şekilde organize edilebileceğinin mümkün olduğu görülmüştür. Bulgularda da görüldüğü gibi sonuçlar dört beceri ölçen test ile sekiz özelliği ölçen test arasında uyumlu ve benzer çıkmıştır. Bu durum maddeleri kısmi puanlayarak tek bir test için daha detaylı bilgi alabilecek ölçü durumları yaratmanın ve BTM için uygun olabileceği söylenebilir. Ancak araştırmanın içinde elde edilen bulgular sadece yapılandırılmış yanıtı maddeler için değil seçme yanıtı maddeler için de bu yöntemin kullanılabileceğini göstermektedir. Alanda ilk kez denenen bir yöntem olarak Q matrisi madde özelliklerine göre genişletme aynı zamanda testin madde sayısını arttırmak anlamına gelmektedir. Ayrıca bu işlem sonrasında testin psikometrik özelliklerinin korunduğu gözlenmektedir. Bilindiği gibi yapılandırılmış yanıtı ya da açık uçlu maddelerde öğrenci hakkında daha detaylı bilgi elde etmeye yönelik yaklaşımlarda testin güvenilirlik ve geçerliğine ilişkin istatistiksel çıkarımlarda sorunlar yaşanmaktadır. Bu nedenle yapılandırılmış yanıtı maddeler her ne kadar öğrenci hakkında daha detaylı bilgi verme potansiyeline sahip olsa da psikometrik olarak geçerliği ve güvenilirliği kanıtlanması zordur.

Bu araştırmada BTM modelleri için de yapılandırılmış yanıtı ve seçme yanıtı maddelerde kısmi puanlama yapmak için yöntemin benzer istatistiksel sağlamlıkla kullanılabileceğine dair bulgulara ulaşılmıştır. Aynı zamanda elde edilen bir diğer sonuç karma madde formatıyla yapılan testlerde analizleri düzenleyerek yapılandırılmış yanıtı ve seçme yanıtı maddelerin birlikte çalışabileceği bir test yapısını kurmanın mümkün olduğunu göstermektedir. Karma madde formatına ilişkin yapılan çalışmalar incelendiğinde özellikle BTM alanında bu konuda sınırlı amprik veriye sahip olduğu göz önüne alındığında buradaki yaklaşımın kullanılabilir ve alan açısından incelenebilir olduğu gözükmektedir.

Araştırma sonuçlarına göre elde edilen bulgular göz önüne alındığında bundan sonra yapılacak çalışmalar BTM'ye göre test geliştirmek isteyen araştırmacıların maddelerin sadece doğru cevaplama durumlarına ilişkin matrislerini değil aynı zamanda diğer alternatif cevaplara ilişkin matrisleri genişletebilmenin mümkün olduğu görülmüştür.

6.2.Öneriler

Test geliştiriciler için tek bir madde formatının kullanılması nasıl bir etki yaratmaktadır.

Özellik sayısının değişkenliği ile model çalışma yöntemi ne kadar uyumludur. Araştırmanın verileri projeden elde edilen verilerle sınırlı olduğu için bu çalışma göz önünde bulundurulmadığı için sadece bu yöntemin kullanılacağı şekilde tasarlanan ve buna göre geliştirilen testlerle yöntem denenebilir. Örneğin bu çalışma sadece yapılandırılmış yanıtli maddelerle ya da seçme yanıtli maddelerle denenebilir. Madde sayısının etkisini incelemek için simülasyon çalışmaları ya da buna benzer çalışmalar denenebilir. Testin Q matrisi sadece seçme yanıtli maddeler için ya da sadece yapılandırılmış yanıtli maddeler için ayrı ayrı yapılandırılmasıyla bunların farkları üzerinde çalışılabilir. Uygulamada sadece DINA model kullanılmıştır. Farklı modeller için de aynı çalışmalar denenebilir ve bu modellerin bu tür yaklaşımlara nasıl tepki verdiği incelenebilir. Soru sayısı ile yöntem ne kadar uyumludur.

KAYNAKÇA

- Akıncı, E. D. (2007). *Information criteria in structural equation models/ Yapısal eşitlik modellerinde bilgi kriterleri*. Doktora tezi, Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul.
- Allan L. Mccutcheon. (1987). *Latent class analysis (Quantitative Applications in the Social Sciences)*, (1st ed) London: United Kingdom).
- Andersen, R., Hagenars, J. A., ve McCutcheon, A. L. (2006). Applied Latent Class Analysis. *Canadian Journal of Sociology / Cahiers Canadiens de Sociologie*, 25(1) 143-145. <https://doi.org/10.2307/3341848>
- Badgett, J. L., ve Christmann, E. P. (2009). *Designing elementary instruction and assessment: Using the cognitive domain*, (1st ed) London: United Kingdom.
- Bandalos, D. L. (1993). Factors Influencing Cross-Validation of Confirmatory Factor Analysis Models. *Multivariate Behavioral Research*, 28(3), 351-374. https://doi.org/10.1207/s15327906mbr2803_3
- Bartholomew, D., Knott, M., ve Moustaki, I. (2011). *Latent variable models and factor analysis: a unified approach*: (3rd ed) London: United Kingdom . <https://doi.org/10.1002/9781119970583>
- Başokçu, O. T. (2016). Comparison of Validity of Classification Made According to DINA Model with Criterion-Referenced and Normative Assessment/Bağıl ve Mutlak Değerlendirme ile DINA Modele Göre Yapılan Sınıflamaların Geçerliğinin Karşılaştırılması. Doktora Tezi, Hacettepe Üniversitesi, Ankara. *E-International Journal of Educational Research*. <https://doi.org/10.19160/e-ijer.82227>.
- 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.

- Chen, J., ve de la Torre, J. (2013). A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes. *Applied Psychological Measurement* 37(6),420-437 . <https://doi.org/10.1177/0146621613479818>
- Chen, J., ve Torre, J. de la. (2014). A Procedure for Diagnostically Modeling Extant Large-Scale Assessment Data: The Case of the Programme for International Student Assessment in Reading. *Psychology*,5(3),1967-1978. <https://doi.org/10.4236/psych.2014.518200>
- Chen, Y., Liu, J., Xu, G., ve Ying, Z. (2015). Statistical Analysis of Q-Matrix Based Diagnostic Classification Models. *Journal of the American Statistical Association* 110(510),850-866. <https://doi.org/10.1080/01621459.2014.934827>
- Dayton, C. M., ve Macready, G. B. (2006). Latent Class Analysis in Psychometrics. *Handbook of Statistics.Elsevier*,26, 421-446 . [https://doi.org/10.1016/S0169-7161\(06\)26013-9](https://doi.org/10.1016/S0169-7161(06)26013-9)
- 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., ve Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3) 333-353. <https://doi.org/10.1007/BF02295640>
- DiBello, L. V., Roussos, L. A., ve Stout, W. (2006). 31A Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models. *Handbook of Statistics*. [https://doi.org/10.1016/S0169-7161\(06\)26031-0](https://doi.org/10.1016/S0169-7161(06)26031-0)
- Downing, S. M. (1992). True- False, Alternate- Choice, and Multiple- Choice Items. *Educational Measurement: Issues and Practice*(979-1027).

<https://doi.org/10.1111/j.1745-3992.1992.tb00248.x>

Downing, S. M. (2009). *Written tests: constructed-response and selected-response formats in assessment in health professions education*. (1st ed)

<https://doi.org/10.4324/9780203880135>

Downing, S. M., ve Yudkowsky, R. (2009). *Assessment in health profession education* (1st ed) New York: Routledge.

Edition, S. (2009). *PISA Data Analysis Manual: SPSS, Second Edition. Analysis*.

<https://doi.org/10.1787/9789264056275>.

Eignor, D. R., ve Hambleton, R. K. (2007). Book Reviews : Norman E. Gronlund, Constructing Achievement Tests (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall(1977) *Educational and Psychological Measurement*.

<https://doi.org/10.1177/001316447903900140>

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)

Fisher, D. ve Frey, N. (2007) *Checking for understanding: formative assessment techniques for your classroom*, (1st ed), Virginia:USA.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., Shelley, B. E., ve Xie, H. (2011). Highlights From PISA 2009 : Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context (NCES 2011-004) 1–56.

Frary, R. B. (1989). Partial-Credit Scoring Methods for Multiple-Choice Tests. *Applied Measurement in Education*, 2(1), 79–96.

https://doi.org/10.1207/s15324818ame0201_5

Frisbie, D. A. (1992). The Multiple True- False Item Format: A Status Review. *Educational Measurement: Issues and Practice*, 11(4), 21-26.

<https://doi.org/10.1111/j.1745-3992.1992.tb00259.x>

- Guttman, L., ve Schlesinger, I. M. (1967). Systematic Construction of Distractors for Ability and Achievement Test Items. *Educational and Psychological Measurement* 27(3) 569-580. <https://doi.org/10.1177/001316446702700301>
- Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking. Writing Test Items to Evaluate Higher Order Thinking* 13(55) 264-278.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items. Lawrence Erlbaum Associates Publisher* (3rd ed). [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Haladyna, T. M., Downing, S. M., ve Rodriguez, M. C. (2004). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education* 15(3),309-334 . https://doi.org/10.1207/s15324818ame1503_5
- Hambleton, R. K., ve Swaminathan, H. (1985). *Item Response Theory: Principles and applications*, (1st ed) Boston: Kluwer- Nijhoff Pub.
- Hickson, S., Reed, W. R., ve Sander, N. (2010). To Use Constructed-Response Questions, Or Not To Use Constructed-Response Questions? That Is The Question. *IDEAS Working Paper Series from Repec* 11(6) 1-29.
- Jang, E. E. E. (2008). A framework for cognitive diagnostic assessment. In *Towards adaptive CALL: Natural language processing for diagnostic language assessment* 117-131 Iowa State University, Toronto.
- Javidanmehr, Z., Reza, M., ve Sarab, A. (2017). Cognitive Diagnostic Assessment : Issues and Considerations, *International Journal of Language Testing* 7(2), 73–98.
- Junker, B. W. (2000). *Monotonicity and conditional independence in models for student assessment and attitude measurement* (1st ed) .
- Lim, Y. (2015). *Cognitive Diagnostic Model Comparisons Introduction Of Cognitive Diagnostic Assessment (CDA.)*. Unpublished doctoral dissertation, Georgia

Institute of Technology, Atlanta .

Liu, O. L., Lee, H., Linn, M. C., ve Liu, O. L. (2011). An Investigation of Explanation Multiple- Choice Items in Science Assessment(September), *Educational Assessment*, 16(3) 37–41. <https://doi.org/10.1080/10627197.2011.611702>

Lodico, M. G., Spaulding, D. T., ve Voegtler, K. H. (2010). *Methods in educational research: From theory to practice* (2nd ed.) America. https://doi.org/10.1111/j.1467-8527.2007.00388_4.x

Michael E. Martinez. (1990). A Comparison of Multiple-Choice and Constructed. *Journal of Educational Measurement* 28(2) 131-145.

Minchen, N. D., de la Torre, J., ve Liu, Y. (2017). A Cognitive Diagnosis Model for Continuous Response. *Journal of Educational and Behavioral Statistics* 42(6), 651-677 . <https://doi.org/10.3102/1076998617703060>

Mislevy, R. J. (2014). Probability-Based Inference in Cognitive Diagnosis. *ETS Research Report Series* 5(3)253. <https://doi.org/10.1002/j.2333-8504.1994.tb01576.x>

Oecd. (2013). PISA 2015 Draft Mathematics Framework. *Oecd*, (March 2013), 52. <https://doi.org/10.1177/0022146512469014>

Osterlind, S. J. (2012). *Constructing test items: multiple-choice, constructed-response, performance, and other formats*, (2nd ed.)London: Boston. <https://doi.org/10.1007/s13398-014-0173-7.2>

PISA. (2015). *PISA 2015 Results in Focus. Oecd* (Vol. I). <https://doi.org/10.1787/9789264266490-en>

Race, P., Brown, S., ve Smith, B. (2005). *500 tips on assessment*, (2nd ed) London:United Kingdom . <https://doi.org/10.4324/9780203307359>

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., ve Templin,

- J. L. (2007). The fusion model skills diagnosis system. In *Cognitive Diagnostic Assessment for Education: Theory and Applications* 2(3) 12-19. <https://doi.org/10.1017/CBO9780511611186.010>
- Rupp, A. A., ve Templin, J. L. (2008). Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State of the Art. *Measurement: Interdisciplinary Research & Perspective* 275-318, Cambridge: Cambridge University Press . <https://doi.org/10.1080/15366360802490866>
- Rupp, A., ve Templin, J. (2008). Educational and Psychological Measurement The Effects of Q-Matrix Estimates and Classification, *Educational and Psychological Measurement*, 68(1), 78-96. <https://doi.org/10.1177/0013164407301545>
- Schaeffer, G. A., Henderson-Montero, D., Julian, M., ve Bene, N. H. (2002). A Comparison of Three Scoring Methods for Tests with Selected-Response and Constructed-Response Items. *Educational Assessment*, 8(4),317-340. https://doi.org/10.1207/S15326977EA0804_2
- Stankous, N. V. (2016). Constructive Response Vs. Multiple-Choice Tests In Math: American Experience And Discussion (Review). *European Scientific Journal, ESJ*. <https://doi.org/10.19044/ESJ.2016.V12N10P%P>
- Sternberg, J. R., ve Sternberg, K. (2011). Cognitive psychology. *Science*, 609. <https://doi.org/10.1126/science.198.4319.816>
- Tao, J., Shi, N.-Z., ve Chang, H.-H. (2012). Item-Weighted Likelihood Method for Ability Estimation in Tests Composed of Both Dichotomous and Polytomous Items. *Journal of Educational and Behavioral Statistics* 37(2): 298-315. <https://doi.org/10.3102/1076998610393969>
- Tatsuoka, K. K. (2009). *COGNITIVE ASSESSMENT An Introduction to the Rule Space Method(1st ed.)* Newyork: London .
- Templin, J. L., ve Henson, R. A. (2006). Measurement of psychological disorders

- using cognitive diagnosis models. *Psychological Methods* 11(3), 287-305.
<https://doi.org/10.1037/1082-989X.11.3.287>
- Torre, D. L. J., ve Rutgers, E. (2011). The Generalized DINA Model Framework, *Psychometrika*, 76(2) 179–199.
- Torre, J. De. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal Of Educational And Behavioral Statistics* 34(1): 115-130.
<https://doi.org/10.3102/1076998607309474>
- Torre, J. De, ve Lee, Y. (2010). A Note on the Invariance of the DINA Model Parameters, *Psychometrika* ,47(1), 115–127.
- Tu, D., Wang, S., Cai, Y., Douglas, J., ve Chang, H. H. (2019). Cognitive Diagnostic Models With Attribute Hierarchies: Model Estimation With a Restricted Q-Matrix Design. *Applied Psychological Measurement* 43(4) 255-271.
<https://doi.org/10.1177/0146621618765721>
- Tu, D., Zheng, C., Cai, Y., Gao, X., ve Wang, D. (2018). A Polytomous Model of Cognitive Diagnostic Assessment for Graded Data. *International Journal of Testing* 18(3):1-21. <https://doi.org/10.1080/15305058.2017.1396465>
- van der Linden, W. J. (2016). *Handbook of item response theory. Handbook of Item Response Theory Volume one Models (1st ed.)* Monterey, California.
<https://doi.org/10.1201/9781315374512>
- Vermunt, J. K., ve Magidson, J. (1979). *Latent class analysis*, 175-198, Thousands Oakes: Sage.
- Wang, M. W., ve Stanley, J. C. (2008). Differential Weighting: A Review of Methods and Empirical Studies. *Review of Educational Research*. 40(5) 663-705 <https://doi.org/10.3102/00346543040005663>

EKLER

EK-1: Pilot uygulamada kullanılan maddelerin madde güçlük ve ayırt edicilik

değerleri

No	Soru Kodu	Madde Güçlük (Pj)	Madde Ayırt Edicilik (rbis)
1	P9796	0,78	0,6
2	P3690	0,19	0,15
3	P1027	0,23	0,65
4	P1025	0,16	0,24
5	P1033	0,35	0,57
6	P1021	0,4	0,53
7	P1032	0,39	0,73
8	P7339	0,01	0,53
9	P6728	0,35	0,71
10	P5158_B	0,1	0,81
11	P5158_C	0,4	0,68
	Ortalama	0,31	0,56
1	P1011_A	0,28	0,54
2	P1011_B	0,12	0,75
3	P1011_C	0,02	0,43
4	P1012_A	0,37	0,69
5	P1012_B	0,33	0,44
6	P1013_A	0,11	0,82
7	P1013_B	0,25	0,6
8	P1014_A	0,28	0,55
9	P1014_B	0,04	0,4
10	P1014_C	0,18	0,4
11	P1015_A	0,07	0,61
12	P1015_B	0,35	0,55
	Ortalama	0,2	0,57

izleme 3	1	P1023_A	0	0
	2	P1023_B	0,23	0,31
	3	P1024	0,28	0,45
	4	P1028	0,38	0,58
	5	P1022_A	0,05	0,17
	6	P1022_B	0,39	0,52
	7	P1026	0,43	0,62
	8	P1020_A	0,41	0,43
	9	P1020_B	0,22	0,39
	10	P1020_C	0,21	0,4
	11	P1021_A	0,05	0,28
	12	P1021_B	0,45	0,4
	13	P1021_C	0,18	0,24
	Ortalama	0,25	0,37	

izleme 4	1	P5017_A	0,32	0,45
	2	P5017_B	0,04	0,12
	3	P5036_A	0,39	0,65
	4	P5036_B	0,23	0,27
	5	P5036_C	0,17	0,32
	6	P5058_A	0,78	0,53
	7	P5058_B	0,62	0,64
	8	P5059	0,34	0,43
	9	P5057	0,78	0,54
	Ortalama	0,41	0,44	

EK-2: İzleme 1 testine ait Q matris

	M1	M2	M3	M4
5017_A	1	0	0	1
5017_B	0	0	1	1
5036_A	0	0	1	1
5036_B	0	1	1	0
5036_C	0	1	1	0

5058_A	0	0	1	0
5058_B	0	0	1	0
5059	0	1	0	0
5057	0	0	1	1

EK-3: İzleme 2 testine ait Q matris

	M1	M2	M3	M4
1011_A	0	0	1	0
1011_B	1	0	1	0
1012_A	1	0	0	1
1011_C	"	0	1	0
1012_B	1	0	0	1
1014_A	1	0	0	0
1015_A	1	1	0	0
1014_B	1	0	0	0
1015_B	0	1	1	0
1014_C	1	0	0	0

EK-4: İzleme 3 testine ait Q matris

	M1	M2	M3	M4
1023_A	1	0	1	0
1023_B	0	1	0	0
1024	1	0	0	1
1028	0	1	0	0
1022_A	0	0	1	0
1022_B	1	0	0	0
1026	1	0	1	0
1020_A	1	0	1	0
1020_B	1	1	1	0
1020_C	1	0	1	0
1021_A	1	1	1	0
1021_B	0	0	1	0

1021_C	1	0	0	1
--------	---	---	---	---

EK-5: İzleme 4 testine ait Q matris

	M1	M2	M3	M4
5017_A	1	0	0	1
5017_B	0	0	1	1
5036_A	0	0	1	1
5036_B	0	1	1	0
5036_C	0	1	1	0
5058_A	0	0	1	0
5058_B	0	0	1	0
5059	0	1	0	0
5057	0	0	1	1

