

**İKİ KATEGORİLİ MADDE TEPKİ KURAMI MODELLERİ İLE
SINIFLAMALI TEPKİ MODELİNİN KARŞILAŞTIRILMASI**

SERPİL ÇELİKTEN

YÜKSEK LİSANS TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EYLÜL, 2015

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU

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

YAZARIN

Adı : Serpil

Soyadı : ÇELİKTEN

Bölümü : Eğitim Bilimleri

İmza :

Teslim tarihi :

TEZİN

Türkçe Adı :İki Kategorili Madde Tepki Kuramı Modelleri ile Sınıflamalı Tepki Modelinin Karşılaştırılması

İngilizce Adı : Comparison of Dichotomous Item Response Theory Models and Nominal Response Model

ETİK İLKELERE UYGUNLUK BEYANI

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

Yazarın Adı Soyadı: Serpil ÇELİKTEN

İmza:

Jüri onay sayfası

Serpil ÇELİKTEN tarafından hazırlanan “İki Kategorili Madde Tepki Kuramı Modelleri İle Sınıflamalı Tepki Modelinin Karşılaştırılması” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Yrd. Doç. Emine ÖNEN

Eğitimde Ölçme ve Değerlendirme Gazi Üniversitesi

Başkan:

Üye:

Tez Savunma Tarihi: 07/09/2015

Bu tezin Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Servet KARABAĞ

Eğitim Bilimleri Enstitüsü Müdürü



Aileme...

TEŞEKKÜR

Çalışmalarım boyunca bana destek olan, beni yönlendiren, zamanını ayırıp en iyisini yapabilmem için uğraşan, başarabileceğime inanarak şevkle çalışmamı sağlayan ve akademik gelişimime önemli katkıları olan çok değerli danışman hocam sayın Yrd. Doç. Dr. Emine ÖNEN'e

Yüksek lisans eğitimim boyunca akademik gelişimime katkı ve teşvikte bulunan değerli hocalarım Prof. Dr. Şener BÜYÜKÖZTÜRK, Prof. Dr. Mehtap ÇAKAN, Doç. Dr. Şeref TAN ve Doç. Dr. İsmail KARAKAYA'ya

Dersine katılma fırsatı bulduğum, tezimle ilgili çıkmazda kaldığım noktalardaengin bilgisiyle beni aydınlatan, sorduğum soruları içtenlikle yanıtlayarak yardımlarını esirgemeyen Sayın Prof. Dr. Hülya KELECİOĞLU'na,

Tez yazma sürecini eğlenceli hale getiren, birlikte öğrenmenin tadına vardığım, desteklerini hiçbir zaman esirgemeyen, tanıştığımдан bu yana hayatıma neşe ve enerji katan güler yüzlü dostlarım Arş. Gör. Ayşenur ERDEMİR ve Arş. Gör. Derya ÇAVDAR'a,

Hayatımın her anında bana destek olan, varlıklarını hissederek kendimi hep güvende hissettiğim, sahip olduğum için şanslı olduğumu düşündüğüm sonsuz güç kaynağım değerli annem Yaşar ÇELİKTEN'e, babam Yusuf ÇELİKTEN'e, ağabeyim Seçkin ÇELİKTEN'e ve kardeşim Kadir ÇELİKTEN'e,

Bütün eğitimim boyunca bana destek veren adını saymadığım bütün arkadaşlarıma ve kıymetli hocalarıma sonsuz teşekkürlerimi sunarım.

Serpil ÇELİKTEN

İKİ KATEGORİLİ MADDE TEPKİ KURAMI MODELLERİ İLE SINIFLAMALI TEPKİ MODELİNİN KARŞILAŞTIRILMASI

(Yüksek Lisans Tezi)

Serpil ÇELİKTEN

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

Eylül, 2015

ÖZ

Bu araştırmanın amacı, ikili kodlama yoluyla puanlanan maddeler için madde tepki kuramı modelleri ile çoklu puanlanan sınıflamalı tepki modelini model-veri uyumu, madde parametrelerinin kestirimlerinin doğruluğu, yetenek kestiriminin doğruluğu, madde bilgi fonksiyonları ve test bilgi fonksiyonları açısından karşılaştırmalı olarak incelemektir. Çalışma grubu 2012 Seviye Belirleme Sınavına katılan öğrencilerden seçilen 1500 öğrenciden oluşmaktadır. Araştırma amacı doğrultusunda istatistiksel analizler R studio yazılımı, Multilog, Lisrel 8.80 ve SPSS paket programları kullanılarak yapılmıştır. Model uyumunu test etmeye yönelik yapılan analizler sonucuna göre genel uyum ve madde uyumu bazında STM, 2PLM ve 3PLM'nin araştırma verisine yeterli düzeyde uyum sergilediği, bu modeller arasında en yüksek uyum sergileyen modelin ise 3PLM olduğu gözlenmiştir. Yetenek parametrelerinin kestirimi açısından ise STM'nin ikili kodlama yoluyla puanlanan MTK modellerine göre daha doğru yetenek kestirimleri ile sonuçlandığı gözlenmiştir. Madde bilgi fonksiyonları karşılaştırıldığında STM ile 2PLM altında maddelerin çalışma grubundaki öğrencilerin yetenek düzeyleri hakkında sağladıkları bilgi miktarı açısından benzer sırada yer aldığı sonucuna ulaşılmıştır. Test bilgi fonksiyonları dikkate alındığında ise, STM'nin düşük yetenek ranjında ele alınan bütün ikili kodlama yoluyla puanlanan MTK modellerinden daha fazla bilgi sağladığı gözlenmiştir. Ancak STM'ye dayalı olarak yetenek

kestirimleri daha doğru yapılmış olmasına rağmen söz konusu testin bu model altında, 3PLM'ye göre yetenek kestirimi konusunda daha az bilgi sağladığı gözlenmiştir. Bu duruma ilişkin olası nedenler tartışma bölümünde detaylı bir şekilde açıklanmıştır.



Bilim kodu: 10100589

Anahtar Kelime: İki Kategorili MTK Modelleri, Sınıflamalı Tepki Modeli, Kestirimin Doğruluğu, Bilgi Fonksiyonları

Sayfa Adedi: 121

Danışman: Yrd. Doç. Emine ÖNEN

**COMPARISON OF DICHOTOMOUS ITEM RESPONSE THEORY
MODELS AND NOMINAL RESPONSE MODEL**

(Master Thesis)

SERPİL ÇELİKTEN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

MEASUREMENT AND ASSESSMENT IN EDUCATION

September, 2015

ABSTRACT

The purpose of this study is to examine the dichotomous item response theory models and nominal response model in terms of model-data fit, accuracy of item parameters, accuracy of ability estimation, item information functions and test information functions in a comparative way. The study group was composed of 1500 students taken 2012-SBS. For the research purposes, the analysis were conducted through the R Studio software, Multilog, Lisrel 8.80 and SPSS. In terms of model data fit, it was found that nominal response model, 2 parameter logistic model and 3 parameter logistic model fit the data both in terms of overall fit and item fit, and 3PLM is the best fitting model among them. The abilities were most precisely estimated with NRM . When item information functions were compared, it is seen that NRM and 2PLM show the same order of items in terms of amount of provided information about the students'ability level in the study group. Lastly, in terms of test information functions, NRM provided the most information than the dichotomus IRT models at the low ability range. However, NRM provided less information in terms of total test information function than 3PLM, although the ability parameters were estimated with better

accuracy with NRM than 3PLM. These results about this matter were addressed with the reasons at the discussion section in detail.



Science code: 10100589

Key Words: Dichotomous Item Response Models, Nominal Response Model, Accuracy of Estimation, Information Functions

Page Number: 121

Supervisor: Asst. Prof. Emine ÖNEN

İÇİNDEKİLER

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
Jüri onay sayfası.....	iii
TEŞEKKÜR	v
ÖZ.....	vi
ABSTRACT.....	viii
İÇİNDEKİLER	x
TABLolar LİSTESİ.....	xiii
ŞEKİLLER LİSTESİ	xv
SİMGELER VE KISALTMALAR LİSTESİ.....	xvi
BÖLÜM I.....	1
GİRİŞ	1
Problem Durumu.....	1
Madde Tepki Kuramı.....	5
MTK Modelleri	8
İkili (Dichotomous) Puanlanan Maddeler için MTK Modelleri.....	9
Çoklu Puanlanan (Polytomous) Maddeler İçin MTK Modelleri	11
İlgili Araştırmalar	15

Araştırmanın Amacı	21
Araştırmanın Önemi	22
Sınırlılıklar	24
BÖLÜM II	25
YÖNTEM.....	25
Araştırmanın Modeli.....	25
Çalışma Grubu	25
Veri Toplama Araçları.....	26
Verilerin Toplanması	26
Ölçüm Güvenirliği.....	27
Ölçüm Yorumlarının ve Kullanımlarının Geçerliliği	28
Verilerin Analizi	28
Varsayımların Test Edilmesi	29
Tekboyutluluk	29
Yerel Bağımsızlık	32
MKE'nin Monotonik Artış Göstermesi	33
Parametrelerin Değişmezliğinin İncelenmesi.....	33
Madde Parametrelerinin Değişmezliği	33
Birey Parametrelerinin Değişmezliği	35
Model-veri Uyumu.....	35
Model Parametrelerini Kestirimi ve Bilgi Fonksiyonlarının Hesaplanması	36
BÖLÜM III.....	39
BULGULAR, SONUÇ VE TARTIŞMA.....	39
Bulgular ve Yorum.....	39
Sonuç ve Tartışma	59

Öneriler	63
KAYNAKLAR	65
EKLER.....	74
EK 1. İzin Yazısı.....	75
EK 2. Uzman Değerlendirme Formu.....	76
EK 3. 2012-SBS Matematik Alt Testi.....	77
EK 4. Araştırma Kapsamında Yapılan AFA ve DFA Sonuçları – (N=50.000)	83
EK 5. 1PLM, 2PLM, 3PLM için Yerel Bağımsızlık Varsayımı Testi Sonuçları	84
EK 6. 1PLM, 2PLM, 3PLM'ye ilişkin madde karakteristik eğrileri ve STM'ye ilişkin Kategori Karakteristik Eğrileri	91
EK 7. 1PLM, 2PLM, 3PLM Altında İki Farklı Gruptan Elde Edilen Verilere Dayalı Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Saçılma Grafikleri.....	95
EK 8. STM Altında İki Farklı Gruptan Elde Edilen Verilere Dayalı Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Saçılma Grafikleri	97
EK 9. İki Farklı Gruptan Elde Edilen Verilere Dayalı Olarak Her Bir Model Altında Kestirilen Yetenek Parametreleri Arasındaki İlişkiler.....	99
EK 10. 1PLM, 2PLM, 3PLM ve STM'ye Dayalı Hesaplanan Madde Bilgi Fonksiyonları.....	101
EK 11. Uzman Görüşleri	105

TABLolar LİSTESİ

Tablo 1. Marjinal Güvenirlik Katsayıları.....	27
Tablo 2. Matematik Alt Testinden Elde Edilen Verilere Dayalı Yapılan Açımlyıcı Faktör Analizi.....	30
Tablo 3. İkili Kategorili Modeller Altında İki Farklı Gruba İlişkin Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Hesaplanan Korelasyon Katsayıları.....	33
Tablo 4. STM Altında İki Farklı Gruba İlişkin Madde Parametreleri Arasındaki İlişkilere Yönelik Korelasyon Katsayıları.....	34
Tablo 5. Bütün Modeller Altında İki Farklı Gruba İlişkin Yetenek Parametreleri Arasındaki İlişkilere Yönelik Hesaplanan Korelasyon Katsayıları.....	35
Tablo 6. Model Veri Uyumuna İlişkin Hesaplanan Uyum İstatistikleri.....	39
Tablo 7. İki Kategorili Modellerin Uyum Düzeylerinin Karşılaştırılmasına Yönelik Varyans Analizi Sonuçları.....	40
Tablo 8. Madde Uyumuna İlişkin Hesaplanan S-X ² İstatistikleri.....	41
Tablo 9. İki Kategorili Puanlanan MTK Modelleri Altında Kestirilen Madde Parametreleri ve Madde Parametre Kestirimlerine İlişkin Standart Hata (SH) Değerleri.....	43
Tablo 10. STM altında kestirilen Zeta Parametreleri ve Bu parametre kestirimlerine İlişkin Standart Hata Değerleri.....	45
Tablo 11. Lamda Parametreleri ve Standart Hata Değerleri.....	48
Tablo 12. STM'ye İlişkin a ve b Parametreleri.....	49
Tablo 13. Yetenek Parametre Kestirimine İlişkin Standart Hatalara Yönelik Betimsel İstatistikler.....	52

Tablo 14. Farklı Modeller Altındaki Yetenek Kestirimlerinin Standart Hatalarına Ait ANOVA Sonuçları.....	53
Tablo 15. İkili Karşılaştırmalardaki Anlamlı Farklara İlişkin Etki Büyüklüğü η^2	53
Tablo 16. 1- PLM, 2PLM, 3PLM ve STM'ye Dayalı Olarak Hesaplanan Madde Bilgi Fonksiyonları	55
Tablo 17. Test Bilgi Fonksiyonları	56



ŞEKİLLER LİSTESİ

Şekil 1. Matematik alt testine ilişkin yapılan AFA sonucu oluşturulan özdeğerler grafiği.	29
Şekil 2. SBS-2012 Matematik alt testinin faktör yapısına ilişkin tanımlanan tek faktörlü ölçme modeli.....	31
Şekil 3. 2012-SBS Matematik alt testi A kitapçığı-20.madde	51
Şekil 4. 1PLM'ye dayalı test bilgi fonksiyonu grafiği.....	57
Şekil 5. 2PLM'ye dayalı test bilgi fonksiyonu grafiği.....	58
Şekil 6. 3PLM'ye dayalı test bilgi fonksiyonu grafiği.....	58
Şekil 7. STM'ye dayalı test bilgi fonksiyonu grafiği.....	58

SİMGELER VE KISALTMALAR LİSTESİ

MTK	Madde Tepki Kuramı
KTK	Klasik Test Kuramı
MBF	Madde Bilgi Fonksiyonu
TBF	Test Bilgi Fonksiyonu
MKE	Madde Karakteristik Eğrisi
1PLM	1 Parametrelı Lojistik Model
2PLM	2 Parametrelı Lojistik Model
3PLM	3 Parametrelı Lojistik Model
STM	Sınıflamalı Tepki Modeli
MEB	Milli Eğitim Bakanlıđı
SBS	Seviye Belirleme Sınavı
SH	Standart Hata
NRM	Nominal Response Model
AFA	Açımlayıcı Faktör Analizi

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, problem cümlesi, araştırmanın amacı, önemi ve sınırlılıklar yer almaktadır.

Problem Durumu

Eğitim süreci boyunca öğrencilere bir takım bilgi ve becerilerin kazandırılması amaçlanmaktadır. Ancak öğrenciler her ne kadar aynı eğitimden geçseler de bireysel farklılıklardan dolayı her öğrenci için eğitimin çıktıları farklı olmaktadır. Cox ve Godfrey (1997), bir ders ünitesinin öğretiminden sonra öğrenciler arasında önceki duruma göre hem niteliksel hem de niceliksel farklılaşmaların meydana geldiğini belirtmektedirler. Dolayısıyla, bu farklılıkları ortaya çıkarmada ölçme ve değerlendirme işlemleri önem kazanmaktadır.

Ölçme, hem fiziksel hem de sosyal bilimlerde önemli bir yere sahiptir. Bir bilim dalındaki gelişmeler o daldaki ölçme araç ve yöntemlerinin gelişmişliğine paralel bir şekilde gerçekleşmektedir. Eğitim bilimleri ve psikoloji alanlarında da ölçme araçlarının gelişmesiyle önemli ilerlemeler kaydedilmiştir. Ölçme ve değerlendirme işlemleri, eğitimde eğitim programlarının işlevselliğini anlama, öğretimde başvurulan yöntemlerin etkililik derecesini belirleme, öğrencileri başarılı olabilecekleri düşünülen alanlara yönlendirme, öğrenci tanılama ve öğrenci başarısını belirlemeye yönelik çalışmalara hizmet etmek amacıyla gerçekleştirilmektedir (Baykul, 2010). Crocker ve Algina (1986, s.67) ise ölçme ve değerlendirmenin bireyleri seçme, yerleştirme ya da bireyler hakkında tanılayıcı kararlar verme amaçları için yapılabileceğini belirtmektedir.

Geçmişten bu yana öğrencilerin öğrenme düzeyleri çoğunlukla öğretmenler tarafından geliştirilen testler aracılığıyla ölçülmektedir. Topluma hesap verilebilirlik açısından, öğretmen yapımı testlerin alan yazında yeterli görülmediği belirtilmiştir. Çünkü söz konusu sorumluluk öğrenciler için bir takım karşılaştırmaların yapılmasını gerektirmektedir, dolayısıyla, bu doğrultuda önceden belirlenmiş bir içeriğe sahip ve standart bir şekilde uygulanan testlerin kullanılması daha uygundur (Baker, 1988).

Öğretmen yapımı testlerde bir standartlaşmanın olmaması ve öğrencilerin karşılaştırılmasına izin vermemesi nedeniyle, ulusal ve uluslararası düzeyde geniş ölçekli sınavlar uygulanmaktadır. Bu sınav sonuçlarına dayalı olarak öğrenciler hakkında önemli kararlar verildiğinden, bu sınavlar önem taşımaktadır. Öğrenciler hakkında verilen kararların doğruluğu, bu sınavlarda uygulanan testlerden elde edilen puanların geçerlik ve güvenilirlik düzeylerine bağlıdır. Bu puanlara her zaman hata karışmaktadır çünkü bu sınavlarda ölçülen özellikler başarı, beceri ve yetenek gibi doğrudan gözlenemeyen soyut yapılardır. Soyut yapılar doğrudan gözlenemedikleri için davranışlar aracılığıyla dolaylı bir şekilde ölçülmektedir. Bu dolaylı ölçme işlemi ise psikolojik ölçme araçlarından yararlanılmaktadır. Bireylerin ölçme aracında yer alan maddelere verdikleri tepkilere dayalı olarak bireylerin söz konusu psikolojik özelliğine ilişkin bilgi edinilmeye çalışılmaktadır. Farklı bir ifadeyle psikolojik yapılar, ilgili maddelere verilen tepkilerin yani davranışların gözlenmesi yoluyla ölçülerek, bireylerin özellikleri davranışa dayalı olarak ölçülebilmektedir (Embretson ve Reise, 2000, s.43-45; Hambleton & Swaminathan, 1985, s.9; Lord ve Novick, 1968, s.358).

Ölçülmesi amaçlanan psikolojik yapıların doğrudan gözlenememesi bir takım sorunları da beraberinde getirmektedir. Crocker ve Algina (1986, s.5-7) bu problemleri şu şekilde belirtmiştir:

1. Herhangi bir yapıyı ölçen evrensel olarak kabul görmüş tek bir yaklaşım yoktur.
2. Genellikle psikolojik ölçmeler sınırlı sayıda davranış örneğine dayanır.
3. Elde edilen ölçümler daima hata içerir.
4. Ölçekler iyi tanımlanmış birimlerden yoksundur.
5. Psikolojik yapılar sadece onların işe vuruk tanımlarıyla açıklanamaz. Diğer yapılarla ve gözlenebilir olaylarla da ilişkisi gösterilmelidir.

Test kuramı, bu problemlerin ve çözümlerinin çalışılmasına dayanan bir disiplindir (Crocker ve Algina, 1986, s.7). Geçmişten bu yana yapılan ölçme çalışmaları incelendiğinde, ölçmelerin kalitesinin incelenmesi ve artırılması amacıyla geliştirilen iki temel test kuramı bulunduğu gözlenmektedir ve bunlar Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK)'dır (Adedoyin, 2010; Crocker ve Algina, 1986; Embretson ve Reise, 2000).

KTK ilk geliştirilen test kuramıdır. KTK'ya ilişkin birçok çalışma, 1907-1913 yılları arasında İngiliz Psikolog Charles Spearman tarafından gerçekleştirilmiştir. Spearman 1930'lu yıllarda KTK'ya dayalı olarak test geliştirme sürecinin aşamalarını tanımlamıştır (Crocker ve Algina, 1986, s.106-107; Embretson ve Reise, 2000, s.13-14; Lord ve Novick 1968). Özellikle 20.yy'ın büyük bir kısmında ölçme çalışmalarına temel olan KTK'nın temel avantajı kuramsal olarak oldukça kolay karşılanabilir varsayımlarının olması ve birçok test etme durumlarına kolaylıkla uygulanabilir olmasıdır (Hambleton ve Jones, 1993).

KTK ölçme hatasına dayalı bir kuramdır. Birçok yazar, özellikle, Guilford (1936), Gulliksen (1950), Magnusson (1967), ve Lord & Novick (1968) bu modeli yeniden şekillendirip detaylandırarak, aşağıdaki gibi tanımlamışlardır:

$$X = T + E \quad (\text{Eşitlik 1})$$

Bu modelde "T" ile gösterilen değer sabittir ve bireyin gerçek puanı olarak nitelendirilmektedir. "E" seçkisiz hata ve "X" gözlenen puan olarak nitelendirilmekte ve bu terimler seçkisiz değişkenler olarak belirtilmektedir (Crocker ve Algina, 1986, s.106-107; Lord ve Novick, 1968, s.27).

Temel odak noktası test bazında bilgi sağlamak olan KTK'da, madde istatistikleri de modelin önemli bir kısmını oluşturmaktadır. KTK madde bazında oldukça basittir. Çünkü KTK bir katılımcı havuzunu düşünmekte ve karmaşık teorik modeller yerine görgül olarak madde üzerinde bireylerin başarı oranını doğrudan incelemektedir (Fan, 1998).

KTK'nın temel işlevlerinden biri de, bireyin teste dayalı elde edilen puanının güvenilirliğini kestirmek olduğu için Klasik Güvenirlik Teorisi olarak da nitelendirilmektedir. KTK, psikolojik ölçme tarihinin başlangıcından itibaren test geliştirmede, analizde ve psikolojik ölçeklerin puanlanmasında yaygın olarak kullanılmıştır. Ancak, KTK'nın bir takım sınırlılıkları yeni bir kuramın geliştirilmesi ihtiyacını ortaya çıkarmıştır. Bu sınırlılıklar şöyle sıralanmaktadır: (a) Madde nitelikleri gruba bağımlıdır. (b) Bireyin yeterliğini gösteren puanlar teste bağımlıdır. (c) Madde bazında değil test bazında ifade edilen bir modeldir. (d)

Güvenirliğin kestirimi için tam anlamıyla paralel testler gerektirir. (e) Kayıp verilerin üstesinden gelmek zordur.

KTK'nın bu sınırlılıklarından dolayı modern test kuramı olarak da anılan MTK, KTK'ya alternatif olarak geliştirilmiştir (Bock, 1997b; Crocker ve Algina, 1986, s.339-340; Hambleton, Swaminathan ve Rogers, 1991, 1-5). MTK'nın matematiksel modellemelerine ilişkin ilk açıklamalar, lojistik test modelleri ve model parametre kestirimleri ile birlikte Lord ve Novick'in "Statistical Theories of Mental Test Scores" (1968) kitabında Örtük Özellik Kuramı (Latent Trait Theory) olarak yer almaktadır. Bu açıklamalar Allen Birnbaum'un, 1957 ve 1958'de yazdığı teknik raporlara dayanmaktadır (aktaran, Hambleton ve Jones, 1993). Daha sonra Rasch (1960), madde tepki modellerini açıkladığı kitabını yayımlamıştır. Test etme ve ölçme sürecinde karşılaşılan problemlere çözüm sunan MTK, zamanla daha popüler olmaya ve yaygın bir şekilde kullanılmaya başlamıştır (Hambleton, 1994; Penfield, 2014; Reise, Ainsworth ve Haviland, 2005).

KTK'da madde istatistikleri ve yetenek ölçümleri ilgili maddenin ya da testin uygulandığı gruba bağlı kalmaktadır, test ve madde nitelikleri gruba bağlı olarak değişmektedir. Dolayısıyla eğer seçilen gruptaki bireylerin yetenek düzeyleri yüksek ise testin bütün maddeleri kolay görünecektir, gruptaki bireylerin yetenek düzeyleri düşük ise aksine aynı maddeler zor görünecektir. Bu sınırlılık, bireyin yeteneğinin farklı test formlarıyla ölçülmesini zorlaştırmaktadır. Ancak MTK'da, KTK'nın aksine, madde nitelikleri ve bireyin yeteneği gruptan ve testten bağımsızdır (Hambleton vd., 1991, s. 2).

KTK'da güvenilirliğin tanımı, pratikte gerçekleşmesi zor olan paralel testlerden elde edilen ölçümlere dayanmaktadır ve KTK'da uzun testler her zaman kısa testlere göre daha güvenilir ölçümler sağlamaktadır. Ölçmenin hatadan arınlığı olarak da nitelendirilebilen ölçmenin doğruluğu ise, belli bir örneklemden bütün gözlenen puanlar için aynıdır. Ancak, MTK'da güvenilirlik, ölçülen psikolojik yapının kestiriminde koşullu bir fonksiyon (conditional function of psychological construct) olarak tanımlanmaktadır. Ölçmenin doğruluğu (accuracy) örtük yapının düzeyinde değişir. Yani yetenek ölçeği boyunca kestirime ilişkin elde edilen standart hata değerleri değişmekte bu doğrultuda farklı yetenek düzeylerinde elde edilen kestirimin doğruluğu da değişmektedir. Ayrıca, MTK'nın önemli bir özelliği ve KTK'ya göre üstünlüğü olan değişmezlik özelliği sayesinde örneklemden elde edilen sonuçlar bütün hedef yanıtlayıcı evrenine genellenebilmektedir.

KTK detaylı olarak ele alındığında bu kuramın bir takım sınırlılıklarının olduğu gözlenmektedir. KTK'nın bu sınırlılıkları ölçme sürecinde ve ölçümlerin analizinde bir takım sorunlara yol açtığı için bu kurama alternatif olarak MTK geliştirilmiştir.

Madde Tepki Kuramı

MTK, kişinin ölçülen özellik boyutundaki yerini, test maddelerine verdiği tepkileri kullanarak matematiksel bir model ile açıklayan bir madde fonksiyonunu tanımlayan ve bunlara dayalı olarak test geliştirme ve testlerden elde edilen puanların niteliklerinin incelenmesine temel oluşturan bir kuramdır. MTK'ya göre bireyin yanıtı, bireyin ölçülen özelliğine bağlı olarak belirlenmektedir. Genellikle örtük özellik olarak ifade edilen bu özelliklerin her birinin, tek bir ölçek (theta (θ) continuum) üzerinde sürekli olarak değiştiği varsayılmaktadır. MTK, örtük özellik olarak adlandırılan insana ait gözlenemeyen nitelikleri tanımlayarak, bunları nicel olarak ifade ettiği için Örtük Özellik Kuramı olarak da bilinmektedir (Crocker ve Algina, 1986, s.339; Embretson ve Reise, 2000, s.43-44; Hambleton vd., 1991, s.7; Lord ve Novick, 1968, s.158-159).

MTK, bir test maddesini ve bunun yanıtlayıcının yeteneğiyle olan etkileşimini niteleyen bir takım parametrelerin kullanıldığı matematiksel modeller seti tarafından nitelenmektedir. Bireyin yeteneği (θ), madde parametrelerine ve ilgili modele dayalı olarak kestirilmektedir. Modeller belli bir yetenek düzeyindeki bireyin maddeye doğru yanıt verme olasılığını sağlamaktadır (Crocker ve Algina, 1986, s.340; Hambleton vd., 1991, s.7).

MTK'nın en önemli özelliği parametre değişmezliğidir. Bu özellik bir maddeyi niteleyen parametrelerin katılımcıların yetenek dağılımından bağımsız olduğunu, bir yanıtlayıcıyı niteleyen parametrelerin ise test maddelerinden bağımsız olduğunu belirtmektedir (Adodeyin, 2010; Hambleton vd., 1991, s.8; Han Ho, 2010).

MTK, doğrudan gözlenemeyen özellikleri (örtük özellikleri) ele aldığı için, eğitim bilimleri alanında ve psikolojide sıklıkla kullanılmaktadır. Hambleton ve Linden (1982) MTK modellerinin kullanım amaçlarını şu şekilde belirtmiştir: (a) Test geliştirmek (b) Paralel olmayan test puanlarını eşitlemek (c) Madde yanlılığını incelemek (d) Puanları raporlamak (e) Diğer ölçme sorunlarını ele almak.

MTK farklı amaçlarla kullanılabilmesine rağmen, MTK modellerinden faydalanabilmek için MTK'nın varsayımlarının karşılanması gerekmektedir. Bu varsayımlar tek boyutlu MTK

modelleri için tek boyutluluk, madde karakteristik eğrisinin monotonik olarak artması ve yerel bağımsızlık olarak sıralanmaktadır. Tek boyutluluk, bireyin ölçme aracının maddelerine tepki vermesinin altında yatan tek bir baskın yetenek (özellik: θ) bulunması gerektiğine işaret etmektedir. MTK'nın diğer bir varsayımı, madde karakteristik fonksiyonunun örtük değişkenler ile madde yanıtları arasındaki doğru ilişkiyi (true relationship) yansıtmasıdır. Başka bir ifadeyle, madde karakteristik eğrisi (MKE) – monotonik olarak artan – yetenek düzeyi arttıkça doğru yanıt verme olasılığının arttığını ifade eden bir eğridir. Yerel bağımsızlık ise belli bir maddeye verilen yanıtın olasılığının, yeteneğe koşullu olmak şartıyla, herhangi bir maddeye yanıt verme olasılığından istatistiksel olarak bağımsız olması gerektiğine işaret etmektedir (Hambleton vd., 1991, s.10-11).

MTK'nın varsayımları karşılandığında MTK, KTK'ya göre bir takım avantajlar sağlamaktadır. Ancak MTK modellerinin kullanılabilirliği (usefulness) modelin veri ile uyumuna bağlıdır. Model-veri uyumunu değerlendirmede uyum iyiliği indeksleri kullanılmaktadır. Uyum iyiliği indeksi, veri setine ilişkin gözlenen değerler ile beklenen değerler arasındaki farkı (discrepancy) istatistiksel bir model altında ele alır ve test edilen modelin uyum derecesi hakkında bilgi verir. Model-veri uyumu, model bazında ve madde bazında olmak üzere iki farklı şekilde ele alınmaktadır. Modelin genel uyum düzeyi yaygın olarak -2loglikelihood değerleri ile karşılaştırılmalı olarak ele alınmaktadır. Bu değerlerin tek yönlü varyans analizi ile karşılaştırılması sonucunda elde edilen olasılık değerinin anlamlı olması, daha kompleks modelin veriye daha iyi uyum sergilediğine işaret etmektedir. Bunun yanı sıra SRMRS, CFI, TLI uyum istatistikleri ve RMSEA hata değerleri de genel uyum düzeyini değerlendirmede kullanılmaktadır (Chalmers, 2012; Kang ve Chen, 2007; Maydeo-Olivares, 2013; Maydeo-Olivares ve Joe, 2014; Orlando ve Thissen, 2000; Thissen 1991). SRMRS uyum indeksi sadece sıralı veri için elde edilmektedir. Hem sınıflama düzeyindeki veri hem de sıralı veri için eş zamanlı olarak genel model uyumunu değerlendirmek amacıyla ise hata değerlerinden RMSEA ve uyum istatistiklerinden CFI, TLI yararlanılmaktadır. Burada yer alan hata değeri (RMSEA) basitçe test edilen model ile evren olasılıkları arasındaki farkın dönüşümü olarak ele alınmaktadır ve model uyumunu değerlendirmek için bu fark istatistiksel ölçütler ile ifade edilmektedir (Maydeo-Olivares ve Joe, 2014).

Madde bazındaki uyumu değerlendirmek amacıyla ise Yen (1981)'in Q1 istatistiği, McKinley ve Mill's (1985)'in G2 istatistiği gibi bir takım geleneksel yöntemler mevcuttur. Bunun yanı sıra Orlando ve Thissen (2000, 2003) tarafından geliştirilen (proposed) S-X²

madde uyum istatistiği de son yıllarda kullanılan ve geleneksel yöntemlere (Q1 istatistiği, G2 istatistiği) göre etkililiği kanıtlanmış olan bir yöntem olarak ortaya çıkmaktadır (Orlando ve Thissen, 2000; Orlando ve Thissen, 2003; Stone ve Zhang, 2003). Bu yöntemde, geleneksel yöntemlerde de olduğu gibi, istatistiksel hesaplamalar için gruplama işlemleri yapılmakta ve kesme puanları kullanılmaktadır. Geleneksel yöntemlerde kesme puanları örneklem ve modele bağımlıyken, S-X² istatistiğine dayalı yöntemde test puanına temellendirilmiştir. Bu durum, S-X² 'nin diğer yöntemlere göre daha iyi ve etkili sonuçlar vermesini sağlamaktadır. Bunun yanı sıra bu yöntem son yıllarda çoklu puanlanan MTK modellerine de genellenmiş ve çoklu puanlanan MTK modellerindeki etkililiği de ortaya çıkarılmıştır (Kang ve Chen, 2008; Kang ve Chen, 2011).

MTK'da önemli olan bazı kavramlar bulunmaktadır. Bilgi fonksiyonları da bu kavramlardan biridir. Bilgi fonksiyonu bir testin ya da bir maddenin, bireylerin yetenek (θ) düzeylerini kestirmeye ne kadar katkı sağladığına işaret eder. Bilgi fonksiyonlarının MTK'da göze çarpan bir rolü vardır. Maddeleri ve testleri nitелеmek, test maddelerini seçmek ve testleri karşılaştırmak amacıyla kullanılan yöntemlerden biri de, MTK tarafından sağlanan madde ve test bilgi fonksiyonlarıdır (Hambleton vd., 1991, s. 91). Madde bilgi fonksiyonu (MBF) isminden de anlaşılacağı üzere madde bazında ele alınır ve bir maddenin tek başına sağladığı bilgi miktarını belirtir. Maddeye ilişkin en kesin (precise) ölçme, yetenek ölçeği üzerinde, madde güçlük parametresine karşılık gelen noktada yapılmakta ve bu noktadan uzaklaştıkça sağlanan bilgi miktarı azalmaktadır (Baker, 2001, s.108). MBF genel olarak aşağıdaki formül ile hesaplanmaktadır:

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta) Q_i(\theta)} \quad i = 1, 2, \dots, n \quad (\text{Eşitlik 2})$$

$I_i(\theta)$, i maddesinin madde bilgi fonksiyonunu, $P_i(\theta)$, θ yetenek düzeyindeki bir bireyin i . maddeye tepki verme olasılığını, $Q_i(\theta)$ maddeye yanlış tepki verme olasılığını yani formül ile gösterilecek olursa $(1 - P_i(\theta))$ 'yı ve $P_i'(\theta)$ ise $P_i(\theta)$ 'nın birinci türevini göstermektedir. Birnbaum (1968) tarafından geliştirilen test bilgi fonksiyonu (TBF), yetenek parametresinin kestiriminin doğruluğuna yöneliktir ve aşağıdaki formül kullanılarak hesaplanır:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) \quad (\text{Eşitlik 3})$$

Bir test tarafından sağlanan bilgi, temel olarak ilgili yetenek düzeyinde bütün MBF'lerinin toplamı olarak ifade edilir. Formülden, her bir maddenin TBF'na bağımsız olarak katkıda bulunduğu anlaşılmaktadır. Bu nedenle, bir test maddesinin katkısı, testteki diğer maddelere ilişkin herhangi bir bilgi olmaksızın belirlenebilir ki bu özellik KTK yöntemlerinde bulunmamaktadır (Hambleton vd., 1991, s. 94).

Yetenek parametrelerinin kestiriminin varyansı ($\text{Var}(\theta)$), TBF'unun sağladığı bilgi ile ters orantılıdır ve aşağıdaki gibi ifade edilir:

$$\text{Var}(\theta) = \frac{1}{I(\theta)} \quad (\text{Eşitlik 4})$$

Benzer şekilde belirli bir yetenek (θ) düzeyinde testin sağladığı bilginin miktarı, yetenek ölçeği üzerinde o noktadaki kestirimin doğruluğu ile ters orantılıdır ve aşağıdaki gibi ifade edilir:

$$\text{SEM}(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (\text{Eşitlik 5})$$

(Partchev, 2004).

MTK'da madde ve yetenek parametrelerinin kestirimleri farklı MTK modelleri altında yapılmaktadır. Aşağıda MTK modellerinin çeşitlerine yer verilmiş, çalışmada ele alınacak modeller detaylı bir şekilde açıklanmıştır.

MTK Modelleri

MTK modellerinden yararlanmak için, model veri arasındaki uyumun tatmin edici bir düzeyde olması gerekmektedir. Çünkü MTK'nın temel özelliklerinden biri olan değişmezlik özelliğinin sağlanması için, model-veri uyumu bir gereklilik olarak düşünülmektedir. Bu nedenle zayıf uyum gösteren MTK modelleri, madde ve yetenek kestirimlerinde değişmezliği sağlamayacaktır (DeAyala, 2006, s.110; Hambleton vd., 1991. s.53-73). Bir MTK modeli ile veri arasındaki uyumsuzluk, model varsayımlarının karşılanamaması, uygun MTK modelinin seçilememesi ya da örneklem büyüklüğünün çok küçük olması gibi nedenlerden kaynaklanabilmektedir (Hambleton vd., 1993).

MTK'da madde ve yetenek parametrelerinin kestiriminde; ölçülen özelliğin tek boyutlu veya çok boyutlu olmasına, doğrusal ilişkilerin bulunup bulunmamasına, kestirilen parametre sayısına ve maddelerin puanlanma şekline göre farklı modeller kullanılmaktadır. MTK

modellerinin sınıflandırılması, kullanım amaçları dikkate alınarak çeşitli ölçütlere göre yapılmaktadır. MTK modelleri, boyutsallık açısından; (1) tek boyutlu modeller (unidimensional models) (2) çok boyutlu modeller (multidimensional), doğrusal ilişkiler açısından; (1) doğrusal modeller (linear models) (2) doğrusal olmayan modeller (nonlinear models), puanlama şekline göre; (1) 1-0 şeklinde puanlanan maddeler için tepki modelleri (dichotomous response models) (2) çoklu puanlanan maddeler için tepki modelleri (polytomous response models) şeklinde sınıflandırılmaktadır (McDonald, 1982). Bu çalışmanın amacı doğrultusunda puanlama şekline göre sınıflandırılan MTK modelleri dikkate alınarak, ikili kodlama yoluyla puanlanan maddeler için ve çoklu puanlanan maddeler için tek boyutlu MTK modelleri kullanılacaktır. Aşağıda bu modeller açıklanmıştır.

İkili (Dichotomous) Puanlanan Maddeler için MTK Modelleri

Bu modeller, modelde kullanılan parametre sayısına göre nitelendirilmektedir. En yaygın olarak kullanılan parametreler madde güçlüğü (b), madde ayırt ediciliği (a) ve şans parametresi (c)'dir.

Bir Parametrelili Lojistik Modelde madde ayırt edicilik parametresi bütün maddeler için sabittir ve modelde sadece madde güçlük parametresi kestirilmektedir. Modele ilişkin matematiksel gösterim aşağıda yer almaktadır:

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \quad (\text{Eşitlik 6})$$

$P_i(\theta)$, θ düzeyinde yer alan ve seçkisiz olarak seçilen bir yanıtlayıcının i maddesine doğru yanıt verme olasılığını göstermektedir. b_i , i maddesine ilişkin madde güçlük parametresidir. b_i parametresi, maddeyi doğru yanıtlama olasılığının yetenek ölçeğinde 0.5 olduğu noktaya, yani maddeyi doğru yanıtlama olasılığının %50 olduğu değere karşılık gelmektedir. Bu parametrenin alabileceği değerler teorik olarak $(-\infty, +\infty)$ aralığında belirtilmesine karşın, pratikte genelde aldığı değerlerin -3 ile +3 aralığında değiştiği belirtilmektedir (Baker, 2001, s.22). Bu parametrenin sayısal değeri ne kadar büyükse maddeyi doğru yanıtlamak için gerekli olan yetenek düzeyinin de o kadar yüksek olacağı ifade edilmektedir. Dolayısıyla, zor maddeler yetenek ölçeğinin sağında ya da üst yetenek düzeyinde gösterilirken; kolay maddeler ise ölçeğin solunda ya da alt yetenek düzeyinde gösterilmektedir (Hambleton vd.,

1991, s.13). Modelde yer alan “e” terimi ise sabit bir değerdir ve $e=2.718$ olarak ele alınır. 1PLM’e ilişkin madde bilgi fonksiyonu ise aşağıdaki gibi gösterilmektedir:

$$I_i(\theta) = P_i(\theta) Q_i(\theta) \quad (\text{Eşitlik 7})$$

$I_i(\theta)$, θ yetenek düzeyinde i maddesinin sağladığı bilgi miktarına, $P_i(\theta)$, θ düzeyindeki bireyin i maddesini doğru yanıtlama olasılığına ve $Q_i(\theta)$ maddeyi doğru yanıtlamama olasılığına yani $1 - P_i(\theta)$ ’ya işaret etmektedir.

İki Parametrelili Lojistik Model (2PLM) ilk olarak 1952 yılında Lord tarafından yığılmalı (cumulative) normal dağılım yaklaşımına göre geliştirilmiştir. Daha sonra, 1968 yılında Birnbaum yığılmalı dağılım yaklaşımı yerine madde karakteristik fonksiyonunu kullanarak modeli yeniden tanımlamıştır. Modelin matematiksel gösterimi aşağıda yer almaktadır:

$$P_i(\theta) = \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}} \quad (\text{Eşitlik 8})$$

1PLM’de olduğu gibi, bu modelde de b_i madde güçlük parametresini nitelendirmektedir. 1PLM’den farklı olarak burada modele a_i ile i maddesine ilişkin madde ayırt edicilik parametresi eklenmiştir. Bu parametre $\theta=b$ olduğu noktada madde karakteristik eğrisinin eğimiyle orantılıdır. Maddelerin eğim değerinin yüksek olması maddenin eğimi düşük olan maddelere göre yetenek düzeyleri arasında daha iyi ayırım yaptığını göstermektedir (Hambleton vd., 1991, s. 15-16). Farklı bir ifadeyle, yüksek bir a değeri, maddenin yetenek (θ) ile güçlü bir ilişkisi olduğunu gösterirken; düşük a değeri zayıf bir ilişkiye işaret etmektedir. “ a ” parametresi kestirilen yetenek değerinin gerçek yetenek (true theta) değeri hakkında ne kadar bilgi verdiğini ortaya çıkardığı için, maddenin geçerliliğiyle ilgili kanıtlar elde edilmesine olanak sağlamaktadır (Lord ve Novick, 1986). $D=1.7$, lojistik fonksiyonları, normal ogive fonksiyonuna yaklaştırmak amacıyla kullanılan sabit bir değerdir ve ölçek sabiti olarak da ifade edilmektedir (Hambleton vd., 1991, s.15). 2PLM’e ilişkin madde bilgi fonksiyonu ise aşağıdaki gibi gösterilmektedir:

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta) \quad (\text{Eşitlik 9})$$

$I_i(\theta)$, θ yetenek düzeyinde i maddesinin sağladığı bilgi miktarı, $P_i(\theta)$, θ düzeyindeki bireyin i maddesini doğru yanıtlama olasılığı, $Q_i(\theta)=1 - P_i(\theta)$, a_i , madde ayırt edicilik parametresi, olarak ifade edilmektedir. Formülden anlaşılacağı üzere 1PLM’in madde bilgi fonksiyonunun formülünden farklı olarak madde ayırt ediciliği de dikkate alınarak maddenin sağladığı bilgi miktarı hesaplanmaktadır.

Üç Parametrelili Lojistik Model (3PLM), maddeye şansa dayalı olarak yanıt verme durumunun, doğru yanıt verme olasılığına katkısını ele almak amacıyla, Birnbaum (1968) tarafından 2PLM'in, şans parametresinin de modele dahil edilmesi için yeniden düzenlenmesiyle oluşturulmuştur (Baker, 2001, s.28). Modele ilişkin matematiksel gösterim aşağıda yer almaktadır:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D_{ai}(\theta - b_i)}}{1 + e^{D_{ai}(\theta - b_i)}} \quad (\text{Eşitlik 10})$$

1PLM ve 2PLM'den farklı olarak bu modelde, c_i (pseudo-chance-level parameter) ile gösterilen şans parametresi ele alınmaktadır. Bu parametre düşük yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığını temsil etmekte ve düşük asimptotu sıfırdan farklı olan madde karakteristik eğrileri sağlamaktadır (Hambleton vd., 1991, s.17). Şans parametresinin yüksek olması maddenin ayırt edicilik parametresinin düşme eğiliminde olduğu belirtilmektedir. Bu nedenle bir maddeyi şansa doğru yanıtlama olasılığının yüksek olması maddenin sağladığı bilgi miktarının da düşmesine yol açmaktadır (Harvey ve Hammer, 1999). Şans parametresi (c_i) teorik olarak ($c_i = 0$, $c_i = 1.0$) aralığında yer almaktadır, ancak 0.35 den yüksek “c” değerleri kabul edilmediğinden pratikte bu aralığın $c_i=0$ ile $c_i=0.35$ arasında olduğu belirtilmektedir (Baker, 2001, s.28-29). 3PLM'e ilişkin madde bilgi fonksiyonu ise aşağıdaki gibi gösterilmektedir:

$$I_i(\theta) = a_i \frac{2Q_i(\theta)}{P_i(\theta)} \frac{P_i(\theta) - c^2}{1 - c^2} \quad (\text{Eşitlik 11})$$

$I_i(\theta)$, θ yetenek düzeyinde i maddesinin sağladığı bilgi, $P_i(\theta)$, θ düzeyindeki bireyin i maddesini doğru yanıtlama olasılığı, $Q_i(\theta) = 1 - P_i(\theta)$, a_i , madde ayırt edicilik parametresi, c_i şans parametresi olarak ifade edilmektedir.

Çoklu Puanlanan (Polytomous) Maddeler İçin MTK Modelleri

Çoklu puanlanan maddeler için MTK modelleri, iki ya da daha fazla tepki kategorisi bulunan maddelere yanıt verme davranışlarını açıklamaya yönelik modellerdir. Bunlar ikili kodlama yoluyla puanlanan maddeler için kullanılan modellerin bir uzantısıdır. Bu modeller, maddelerin ikili puanlandığı MTK modellerinden farklı olarak, her bir yanıt kategorisinin basamak zorluklarını (step difficulties) ya da kategorilerin sınırlarını (category boundaries) kullanan modellerdir. Bu modellerde, madde karakteristik eğrisinin (Item Characteristic Curve) yerini, her bir kategoriye tepki verme olasılığı ile yetenek düzeyi arasındaki ilişkiyi

temsil eden kategori karakteristik eğrisi (Category Characteristic Curve) almaktadır (Ostini ve Nering, 2006, s.16-20).

Çoklu puanlanan maddeler için kullanılan MTK modelleri, tepki kategorileri arasındaki ilişkilerin sıralı olup olmamasına göre ya da kategorilerin basamak zorluklarının ve kategori sınırlarının ele alınması durumlarına göre farklılaşmaktadır. Thissen ve Steinberg (1986) tarafından oluşturulan MTK modelleri sınıflamasında, çoklu puanlanan maddeler için olan MTK modelleri iki kategoride ele alınmıştır:

- Fark modelleri (Difference Models)
 - ✓ Aşamalı Tepki Modeli (Samejima's (1969) Graded Response Model)
- Toplama bölünen modeller (Divide-by-total Models)
 - ✓ Kısmi Puan Modeli (Partial Credit model (Masters, 1982)
 - ✓ Dereceli Ölçek Modeli (Rating Scale Model)
 - ✓ Genelleştirilmiş Kısmi Puan Modeli (Generalized Partial Credit Model)
 - ✓ Sınıflamalı Tepki Modeli (Nominal Response Model)

Bu çalışmada, çoklu puanlanan maddeler için MTK modellerinden biri olan Sınıflamalı Tepki Modeli (Nominal Response Model) ile 1PLM, 2PLM ve 3PLM; madde-test bilgi fonksiyonları, model parametrelerinin kestiriminin doğruluğu ve model-veri uyumu bağlamında karşılaştırılmıştır.

Sınıflamalı Tepki Modeli (STM), bir maddeye ilişkin bütün kategorileri dikkate alan çoklu puanlanan maddeler için kullanılan bir MTK modelidir. STM, iki veya daha fazla tepki kategorisinden oluşan maddelerin her bir kategorisine tepki verme olasılığını, bireyin yeteneğiyle olan ilişkisi bağlamında incelemektedir. Bu model diğer çoklu puanlama modellerinden farklı olarak sınıflama ölçeğinde elde edilen veri ile çalışma imkanı sağlamaktadır.

STM, Bock (1972) tarafından iki veya daha fazla sınıflamanın söz konusu olduğu çok kategorili maddelere verilen tepkileri modelleyen bir yol olarak tanıtılmıştır. Bu tür maddelerin standart bir örneği çoktan seçmeli test maddesidir (Baker, 1992). Farklı bir ifadeyle, bu kategoriler karşılıklı olarak birbirini dışarda bırakan (mutually exclusive) ve farklı kapsamlara sahip olan kategorilerdir (mutually exhaustive), ve her biri kendi içerisinde bağımsız olarak ele alınmaktadır (DeAyala, 2009, s.238; Ostini ve Nering, 2006, s.16-18).

STM, sınıflama ölçeği kullanılarak elde edilen verinin kullanıldığı bir modeldir ve çoklu yanıt seçenekleri olan maddelerde kullanımı uygundur. Bu doğrultuda STM çoktan seçmeli test maddelerine ve kişilik envanterlerine dayalı olarak elde edilen verinin modellenmesinde kullanılabilir (De Ayala, 2009, s.238; Embretson ve Reise, 2000, s.120-122). Modelin matematiksel gösterimi şu şekildedir:

$$P_{ih} = \frac{e^{Z_{ih}}}{\sum_h^m e^{Z_{ih}}} \quad h= 1, 2, \dots, g, \dots, m. \quad (\text{Eşitlik 12})$$

$$Z_{ih} = \zeta + \lambda\theta \quad (\text{Eşitlik 13})$$

Matematiksel gösterimden de anlaşılacağı üzere, sınıflamalı tepki modelinde, temelde ζ (zeta) ve λ (lambda) olmak üzere iki parametre kestirilmektedir. Bu modelde parametreler madde bazında değil, maddelerin her bir kategorisi bazında ele alınmaktadır. Bu nedenle bir maddeye ilişkin kestirilen parametre sayısı, ilgili maddenin kategori sayısının iki katıdır. Denklemdaki “ ζ ” ile gösterilen parametre, sabit parametresidir (intercept parameter) ve i maddesinin h . kategorisinin genel popülaritesini (overall popularity) göstermektedir. Zeta parametresine bir takım doğrusal dönüşümler uygulanarak bu parametre aşına olunan b parametresine dönüştürülmektedir. Baker (1992) bu dönüşüm için şu formülün kullanılmasını önermektedir:

$$b = -\frac{\lambda}{\zeta} \quad (\text{Eşitlik 14})$$

Baker (1992) bu parametrenin doğrudan güçlük parametresi olarak yorumlanmadığını, bunun yerine yer parametresi (location parameter) olarak ele alındığını belirtmektedir. DeAyala (1993) da b parametresini yer parametresi olarak ele almaktadır ancak DeAyala bu yer parametresini ardışık kategoriler için elde edilen madde kategori yanıt fonksiyonunun yetenek ölçeği üzerindeki yerine göre tanımlamaktadır. Bu tanımlamaya göre zeta parametresine aşağıdaki dönüşüm uygulanarak b parametresi elde edilmektedir:

$$b = \frac{\zeta(h-1) - \zeta(h)}{\lambda(h) - \lambda(h-1)} \quad (\text{Eşitlik 15})$$

Bu eşitlikler ile doğrusal dönüşümler elde edildiği için, ζ (zeta) ve b parametrelerine ilişkin değerler paralellik göstermekte ve ancak belirtildiği gibi, b parametresi doğrudan güçlük parametresi olarak yorumlanmayıp, yer parametresi (location parameter) olarak ele alınmaktadır. “ λ ” parametresi ise, yeteneği çoklu logit ile ilişkilendiren eğim olarak

nitelendirmektedir ve madde ayırt ediciliğini temsil etmektedir Bu parametreye uygulanan dönüşüm aşağıdaki eşitlikte yer almaktadır:

$$\lambda = a \quad (\text{Eşitlik 16})$$

Eşitlikten de görüldüğü gibi λ parametresi doğrudan ayırtedicilik olarak ele alınmaktadır (Bock, 1997a, s.33-37; Ostini ve Nering, 2006, s.17-18).

Sınıflamalı tepki modeli için türetilen istatistiksel yöntemler, ikili kodlama yoluyla puanlanan maddeler için MTK modellerinden farklı olarak, yanıtlayıcının kısmi bilgisini de dikkate almaktadır (Bock, 1972). Farklı bir ifadeyle, bu modelde doğru yanıtların yanı sıra, ikili kodlama yoluyla puanlanan maddeler için MTK modellerinde göz ardı edilen çeldiriciler de analize alınarak, farklı yetenek düzeylerindeki bireylerin farklı çeldiricilere yönelmesinin altında yatan örtük özelliğe ilişkin bilgi de göz önünde bulundurulmaktadır.

İkili kodlama yoluyla puanlanan maddeler için tek boyutlu MTK modelleri, bireyin bilgisinin yalnızca varlığını ve yokluğunu dikkate alan iki kategorili modeller olarak bilinirler. Ancak, yanıtlayıcıların maddenin her bir kategorisiyle olan ilişkilerinin, iki kategorili modellerle incelenmesi mümkün olmamaktadır. Örneğin kısmi olarak doğru yanıtlanmış bir maddeye puan vermek için ikiden fazla kategori içeren bir modele ihtiyaç duyulmaktadır. Bu noktada STM, ikili puanlanan maddeler için MTK modelleriyle elde edilemeyen kısmi bilginin, çoklu puanlama yaparak elde edilmesini sağlamaktadır (DeAyala, 1993; Tomkowicz, 2000).

STM'nin tercih edilmesinin temelinde, çoklu puanlama ile yanıtlayıcının kısmi bilgisinin de dikkate alınması, dolayısıyla daha hassas ölçümler yapmaya elverişli olması yatmaktadır. Bu modeli kullanmak için bir takım gerekçeler bulunmaktadır. Bunlardan en önemlisi STM'nin çeldiricileri de dikkate almasıdır. Çoktan seçmeli test maddeleri ikili (dichotomous) puanlandığı zaman çeldiricilerin içerdiği bilgi kaybolur, çünkü doğru olmayan yanıtlar ayrı ayrı değil de hepsi tek bir kategoride ele alınır ve sıfır puan verilerek değerlendirilir. Ancak doğru olmayan yanıtlar (çeldiriciler) da doğru yanıtlar kadar öğrencinin yeteneği hakkında önemli bilgi içermektedir. STM ile yanıt örüntüsünün içerdiği bütün bilgiler kullanılır ve bu durum yetenek kestiriminin doğruluğunu artırır (Tomkowicz, 2000). STM, maddenin yetenek (θ) çevresindeki davranışlarının daha net resmini ortaya koyar. Bir maddeye ilişkin çeldiricilerin yetenek dağılımı boyunca, her bir noktada yani her bir yetenek düzeyinde seçilebilme olasılığına ilişkin bilgi sağlar. Diğer bir ifadeyle, hangi

çeldiricinin hangi yetenek düzeyindeki bireyler tarafından yüksek olasılıkla seçilebileceği bilgisini sağlar. Ayrıca, STM'nin kavram yanlışlarının ortaya çıkarılmasında etkili bir şekilde kullanılabilmesi belirtilmektedir. STM bu kavram yanlışlarının özüne inerek bunların ne olduğuna ışık tutabilir. Öğrencilerin öğretim sürecinde yaptıkları hata ve kavram yanlışlarının tam yeri belirlenerek, bunların üstesinden gelmek için yeni öğretim tekniklerinin geliştirilmesine yardımcı olabilir (Stone, 2006). Benzer şekilde, DeAyala (2009, s.238) da STM'nin kavram yanlışlarını ortaya çıkarmada etkili olduğunu belirtmekte ve bu doğrultuda çeldiricilerin rastgele değil de kavram yanlışını ortaya çıkarabilecek şekilde hazırlanmasını önermektedir.

MTK modellerinin eğitimsel ve psikolojik testlerden elde edilen verilere uygulanmasına yönelik birçok çalışma yapılmıştır. Bu çalışmalarda genel olarak MTK kapsamında model- veri uyumu, model parametrelerinin kestiriminin doğruluğu, madde ve test bilgi fonksiyonları, madde ayırimsal işlevliği ve madde yanlılığı konularına odaklanıldığı görülmektedir (Çepni, 2011; DeAyala, 1988; DeMars, 2008; Eser, 2015; Koğar, 2014; Şahin, 2012; Tay, Huang ve Vermunt, 2015; Tomkowicz, 2000).

İlgili Araştırmalar

Türkiye'de ve yurtdışında MTK çerçevesinde iki kategorili MTK modelleri ile ilgili çok sayıda çalışma (Çalışkan, 2000; Çelik, 2001; Can, 2003; Brown, Templin ve Cohen, 2015; Chen ve Thissen, 2010; Kılıç, 1999; Özkurt, 2002; Tay, Huang ve Vermunt, 2015) bulunmaktadır. Türkiye'deki çalışmalara (Çalışkan, 2000; Çelik, 2001; Can, 2003; Kılıç, 1999; Özkurt, 2002) bakıldığında da 1PLM, 2PLM, 3PLM ile ilgili araştırmalara sıklıkla rastlanmıştır. Fakat çoklu puanlanan maddeler için MTK modellerine ilişkin yapılan çalışmaların (Çıtak, 2007; Demirtaşlı, Yalçın ve Ayan, 2014; Dolma, 2009; Korkmaz, 2005) sınırlı sayıda olduğu, STM'nin ise sadece bir çalışmada (Demirtaşlı, Yalçın ve Ayan, 2014) ölçek verilerine uygulandığı gözlenmiştir.

MTK ile ilgili Türkiye'de yapılan ilk çalışmalarda (Baykul, 1979; Berberoğlu, 1988) KTK ve MTK'nın karşılaştırması yoluyla MTK'nın fayda ve üstünlüklerine yönelik analizler yapıldığı görülmektedir. Bunu takip eden çalışmalarda ise MTK'ya ilişkin 1PLM, 2PLM ve 3PLM'nin ilgili test verilerine uyum düzeyleri incelenerek modeller arası karşılaştırmalar yapılmıştır.

Baykul (1979) KTK'ya dayalı olarak geliştirdiği bir matematik yetenek testinin geçerlik ve güvenilirlikleri ile 3PLM'ye dayalı olarak geliştirdiği bir matematik yetenek testinin geçerlik ve güvenilirliklerini karşılaştırmıştır. Araştırma bulgularına göre, 3PLM'nin daha güvenilir sonuçlar verdiği belirtilmiştir. Berberoğlu (1988) ise üniversite seçme sınavlarında Rasch modelinin objektif ölçmeyi sağlama ve artırmadaki katkısını araştırmış ve KTK ile karşılaştırma yapmıştır. Bu çalışmada, 1986 Öğrenci Seçme Sınavına (ÖSS) katılan öğrencilerden seçilen 3390 kişilik bir örneklem üzerinde çalışılmıştır. Araştırmada MTK (Rasch Modeli) ve KTK bağlamında ilgili test puanlarına ilişkin kestirilen geçerlik ve güvenilirlik katsayıları karşılaştırılmış ve KTK'ya göre geçerlik ve güvenilirlik katsayılarının matematiksel değerlerinin daha yüksek olduğu gözlenmiştir.

Çelik (2001) çalışmasında 1PLM, 2PLM ve 3PLM'nin Milli Eğitim Bakanlığı ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavında uygulanan matematik ve fen bilgisi alt testlerinden elde edilen verilere uyum düzeylerini incelemiştir. Bu doğrultuda bu sınava katılan 300.624 öğrenciden 22292 öğrenci seçkisiz olarak seçilmiş, analizler bu örneklemde elde edilen veriler üzerinde gerçekleştirilmiştir. Model uyumunu değerlendirmek amacıyla elde edilen χ^2 istatistik değerlerine göre matematik alt testi verisine en iyi uyum sergileyen modelin 3PLM olduğu, fen bilgisi alt testi verisine en iyi uyum sergilenen modelin ise 2PLM olduğu sonucuna ulaşılmıştır.

Özkurt (2002) çalışmasında, MTK'nın 1PLM, 2PLM ve 3PLM'in bir devlet üniversitesinde uygulanan İngilizce Yeterlik Testi verilerine (N=361) uyum düzeyini incelemiştir. 1PLM, 2PLM ve 3PLM için model-veri uyumu madde bazında test edilmiştir. Araştırmada 2PLM altında sadece 2 maddenin uyum sergilemediği, diğer modeller altında ise uyum sergilemeyen madde sayısının daha fazla olduğu gözlenmiştir. Bu bulgulara dayalı olarak, 2PLM'in İngilizce Yeterlik Başarı Testi verilerine daha fazla uyum sergilediği olduğu sonucuna varılmıştır.

Önder (2007) çalışmasında, ÖZDEBİR ÖSS 2004 D-II sınavının Fen Testinden elde edilen veriye MTK'ya dayalı modellerden hangisinin en iyi uyum sergilediğini incelemiştir. Elde edilen sonuçlara göre genel model uyumu açısından en iyi uyum sergileyen modelin 3PLM olduğu gözlenmiştir. Madde bazındaki uyum ele alındığında uyumsuz madde yüzdesinin 1PLM'den 3PLM'ye doğru azaldığı ve sırasıyla, %40 (1PLM), % 8.9 (2PLM) ve %6.7 (3PLM) şeklinde olduğu gözlenmiş, bu sonuçlara göre 2PLM ve 3PLM'nin madde bazında iyi uyum sergilediği ancak 1PLM'nin iyi uyum sergilemediği sonucuna ulaşılmıştır. Bunun

yanı sıra her bir model altında madde ve yetenek parametrelerinin değişmezliği incelenmiştir. Bu incelemelere göre değişmezlik özelliğinin 2PLM için daha iyi sağlandığı gözlenmiştir. Bütün bu bulgular birlikte değerlendirildiğinde en iyi uyum sergileyen modelin 2PLM olduğu sonucuna ulaşılmıştır.

Gültekin (2011) yaptığı çalışmada, farklı madde türlerinin ve bunların test içerisinde yer alma yüzdelerinin, MTK kapsamında model parametre kestirimlerini, madde ve test bilgi fonksiyonlarını ve ölçüt dayanaklı geçerlik kanıtlarını etkileyip etkilemediğini incelemiştir. Parametre kestirimleri, BILOG-MG programı kullanılarak yapılmıştır. Araştırma bulguları uzun karma testlerin, sadece çoktan seçmeli ve sadece yanıtı sınırlı açık uçlu maddelerden oluşan testlere göre orta güçlükte ve daha ayırt edici testler olduğunu göstermektedir. Ayrıca, orta yetenek düzeyinde uzun karma testler ile daha az standart hatayla yetenek kestiriminin yapıldığı gözlenmiştir. Buna ek olarak testlerde, yanıtı sınırlı açık uçlu maddelerin çoktan seçmeli maddelerden daha fazla bilgi verdiği sonucuna ulaşılmıştır.

Teker, Kelecioğlu ve Eroğlu (2013) tarafından yapılan çalışmada iki kategorili MTK modellerinin, 2009 SBS Fen Bilgisi alt testinden elde edilen veriye uyum düzeyleri incelenmiştir. Bu doğrultuda öncelikle modelin varsayımları kontrol edilmiş, MTK'nin beklenen özellikleri incelenmiş ve modellere ilişkin -2loglikelihood değerlerine göre modeller karşılaştırılmıştır. Bu karşılaştırma için elde edilen χ^2 istatistiklerine göre veriye en iyi uyum sergileyen modelin 3PLM olduğu sonucuna ulaşılmıştır.

İlgili alan yazında iki kategorili MTK modellerine ilişkin Türkiye'de çok sayıda çalışma (Kılıç, 1999; Çalışkan, 2000; Çelik, 2001; Özkurt, 2002; Can, 2003) bulunmasına rağmen STM'ye ilişkin tek bir çalışmaya (Demirtaşlı, Yalçın ve Ayan, 2014) rastlanmıştır.

Demirtaşlı, Yalçın ve Ayan (2014) tarafından yapılan çalışmada, ölçme ve değerlendirme dersine yönelik tutumu ölçmek üzere geliştirilmiş tutum ölçeği, n=393 üniversite öğrencisine uygulanmıştır. Veriler, MTK'nın temel varsayımları olan tek boyutluluk ve yerel bağımsızlık bakımından incelendikten sonra bir boyut altında toplanan maddeler çok kategorili puanlanan MTK modellerinden Samejima (S) Dereceli Tepki Modeli, Kısmi Puan Modeli ve Sınıflamalı Tepki Modeline göre ölçeklendirilmiştir. Ardından ölçeğin, S-Dereceli Tepki Modeline göre geliştirilen yeni formunun daha yüksek test bilgi fonksiyonu verdiği sonucuna ulaşılmıştır.

Yurtdışındaki çalışmalar (Ching-Fung, 2002; DeAyala ve Koch, 1987; DeMars, 2008; Borgatto, Azevedo, Pinherio ve Andrade, 2015) incelendiğinde STM'nin genel olarak simülasyon verilerine dayalı olarak incelendiği, gerçek veriye dayalı çalışmaların (Bock, 1972; DeAyala ve Koch, 1988; Drasgow, Levine, Tsien, Williams ve Mead, 1995; Tomkowicz, 2000) az sayıda olduğu gözlenmektedir.

Bock (1972) çalışmasında iki kategorili ve çok kategorili modelleri karşılaştırmıştır. Çalışma 20 maddelik bir testten elde edilen verilere dayalı olarak gerçekleştirilmiştir. Farklı yetenek düzeylerindeki bireylerden elde edilen verilere dayalı olarak oluşturulan madde ve test bilgi fonksiyonlarının grafikleri karşılaştırmalı olarak incelenmiştir. Elde edilen bulgulara dayalı olarak, orta yetenek düzeyinin altındaki bireyler için, çok kategorili model altında, iki kategorili modele göre, yetenek kestirimlerinin daha doğru bir şekilde yapıldığı; orta yetenek düzeyinin üzerindeki bireyler için ise yetenek kestirimlerinin doğruluğunun modeller arası – yaklaşık olarak – aynı olduğu sonucuna varılmıştır.

DeAyala ve Koch (1987) çalışmasında STM ve 3PLM'ye dayalı olarak CAT uygulamalarına dayalı olarak yetenek parametrelerini simülasyon veri kullanarak kestirmişlerdir. Elde edilen sonuçlara göre iki modele ilişkin yetenek kestirimleri yüksek düzeyde ilişkilidir. Ancak STM'ye dayalı yetenek kestirimlerinin, 3PLM'ye dayalı yetenek kestirme göre daha düşük standart hata ile sonuçlandığı gözlenmiştir.

DeAyala ve Koch (1988) tarafından yapılan bir diğer çalışmada ise STM ve dereceli tepki modeline dayalı CAT uygulamaları ile yetenek kestirimleri elde edilmiş ve modeller karşılaştırılmıştır. Bu çalışmada matematik başarı testi verileri (n=275) kullanılmıştır. Çalışma sonucunda yetenek ölçeği boyunca STM'ye göre elde edilen yetenek kestiriminin yanlılık içermediğini, ancak üst yetenek ranjında dereceli tepki modelinin bireylerin yeteneğini olduğundan düşük kestirmeye eğilimli olduğu gözlenmiştir.

DeAyala (1989) çalışmasında deneysel dayalı bir veri seti kullanarak STM'ye dayalı CAT simülasyonu yapmış ve bunu 3PLM'ye dayalı CAT simülasyonu ile yetenek kestirimi açısından karşılaştırmıştır. Araştırma bulgularına göre, STM altında yetenek ölçeği üzerinde bilgi miktarının düşük olduğu bölgelerde oldukça düşük yanlılıkla (slightly bias) yetenek kestirimleri gözlenmiştir. Ayrıca, STM'ye dayalı oluşturulan bilgi fonksiyonunun grafiği pozitif kayışlı olmasına rağmen, θ yetenek ölçeğinin üst bölgesinde oldukça net yetenek kestirimlerinin elde edildiği gözlenmiştir.

DeAyala (1992) çalışmasında STM ile 3PLM'e dayalı CAT uygulamalarını karşılaştırmalı olarak incelemiştir. Bu incelemeleri farklı test uzunluklarını (15, 20, 25 ve 30) da dikkate alarak gerçekleştirmiştir. Bu çalışmadan elde edilen sonuçlar, bilgi miktarı en az 0.16 olan maddelerin, 15 ve 20 maddelik test uzunluklarında makul düzeyde (reasonably) doğru yetenek kestirimlerinde bulunduğunu göstermektedir. STM ve 3PLM karşılaştırmalı olarak ele alındığında ise, STM'ye göre elde edilen yetenek kestirimlerinin 3PLM ile karşılaştırılabilir nitelikte olduğu, ancak STM'nin $\theta = -1.0$ yetenek düzeyinin altında sağladığı bilgi miktarının 3PLM'ye göre daha fazla olduğu gözlenmiştir.

Drasgow, Levine, Tsien, Williams ve Mead (1995) çalışmalarında Bock (1972)'un sınıflamalı tepki modeli, Samejima (1993)'nin çoktan seçmeli C modeli, Thissen ve Steinberg (1984)'in çoktan seçmeli modeli ve Levine (1993)'nin maksimum olabilirlik formül puanlama (maximum-likelihood formula scoring model) modellerini karşılaştırmıştır. Çalışma üç tane başarı testinden elde edilen veriler (N=3000) üzerinde gerçekleştirilmiştir. İlk üç model için madde parametreleri Multilog programı ile, son model için madde parametreleri ise Levine (1993)'nin FORSCORE programı ile kestirilmiştir. Model uyumunu değerlendirmede grafikler ve χ^2 istatistiklerinden yararlanılmıştır. Araştırma bulgularına göre en iyi model-veri uyumun STM için sağlanmıştır.

Tomkowicz (2000) tarafından yapılan çalışmada düşük, orta ve yüksek olmak üzere üç farklı yetenek düzeyi için yetenek kestirimlerinin doğruluğu açısından karşılaştırmalar yapılmıştır. Araştırmada, sosyal bilimler ve kimya olmak üzere iki farklı test 4000 lise öğrencisine uygulanmış ve bu uygulamadan elde edilen verinin ise 1PLM, 2PLM, 3PLM ve STM'ye uyum düzeyleri test edilmiştir. Çalışmada iki kategorili modeller ile STM, yetenek kestirimlerinin doğruluğu bazında her bir yetenek düzeyinde karşılaştırılmıştır. Araştırma bulguları, lojistik modeller ile STM'nin düşük yetenek düzeyinde yetenek kestirimlerinin doğruluğu açısından farklılaştığını, STM'nin daha doğru kestirimlerde bulunduğunu belirtmiştir. Orta ve yüksek yetenek düzeyindeki bireylerin yetenek kestirimlerinin doğruluğu açısından ise modeller birbirine yakın sonuçlar vermektedir.

Ching-Fung (2002) çalışmasında farklı puanlama yöntemlerinin ve farklılaşan madde parametrelerinin (item parametrization), yetenek kestirimine olan etkisini Monte Carlo simülasyon çalışması ile incelemiştir. Bu doğrultuda ikili kategorili modellerden 1PLM, 2PLM, 3PLM; çoklu puanlanan modellerden kısmi puan modeli, genelleştirilmiş kısmi puan modeli, sınıflamalı tepki modeli ve çoktan seçmeli modeller kullanılmıştır. Model

karşılaştırmalarından elde edilen sonuçlar, çoklu puanlanan MTK modelleri altında iki kategorili MTK modellerine göre daha doğru yetenek kestirimleri yapıldığını göstermektedir. Ayrıca çoklu puanlanan MTK modelleri kendi aralarında karşılaştırıldığında; en doğru yetenek kestiriminin kısmi puan modeli altında yapıldığı, STM'nin ise genelleştirilmiş kısmi puan modeli ve çoktan seçmeli modele göre daha doğru yetenek kestirimleri sağladığı sonucuna ulaşılmıştır.

DeMars (2008) tarafından gerçekleştirilen çalışmada, çoklu puanlanan ve ikili kodlama yoluyla puanlanan maddeler için MTK ve KTK'ya dayalı karşılaştırmalar yapılmıştır. Araştırmada kullanılmak üzere modellere uygun ampirik veri üretilmiş, üretilen veri MTK'ya dayalı ikili ve çoklu puanlanan modeller altında test edilmiştir. Bu çalışmada, STM ve 3PLM karşılaştırılarak incelenmiştir. Elde edilen bulgulara göre, yetenek ölçeği boyunca, düşük yetenek düzeyindeki bireylerin yetenek düzeyleri 3PLM altında STM'ye göre daha yanlış bir şekilde kestirilmektedir. Yanlılık güvenirlilik ile ters orantılı olduğu için 3PLM'ye dayalı yetenek kestirimleri düşük yetenek düzeyindeki katılımcılar için diğer bireylere göre daha az güvenilirdir. Ayrıca 3PLM altında model parametrelerinin kestirimine ilişkin standart hata, STM'dekinden fazladır. Bu durum STM'nin daha çok bilgi sağladığını göstermektedir. Sonuç olarak, çoklu puanlanan modellerin daha güvenilir yetenek kestirimleriyle sonuçlandığı belirtilmiştir.

Borgatto, Azevedo, Pinherio ve Andrade (2015) tarafından yapılan çalışmada MTK'ya dayalı yetenek kestiriminin, testin güçlük derecesine bağımlılığı incelenmiş ve farklı kestirim yöntemlerinin (Expected A Posteriori, Maximum A Posteriori, Weighted Likelihood Estimation) yetenek kestirimine ilişkin standart hatayı nasıl etkilediği değerlendirilmiştir. Çalışma simülasyon verisi üzerinde gerçekleştirilmiştir. Elde edilen bulgulara dayalı olarak test bilgi fonksiyonlarının sağladığı bilgi miktarının en düşük olduğu yetenek ölçeğinin üst kısmında ağırlıklandırılmış olabilirlik yetenek kestirim yöntemi (Weighted Likelihood Estimation) ile daha etkili ve doğru yetenek kestirimlerinin yapıldığı sonucuna ulaşılmıştır.

Bu araştırmalara bakıldığında genel olarak STM kullanılarak başarı testi verileri üzerinden yetenek kestiriminin doğruluğu, simülasyon veri üzerinden yetenek kestirimlerinin yanlılığı ve bilgi fonksiyonlarının incelendiği anlaşılmaktadır. Ancak yurtdışında ve Türkiye'de yapılan çalışmalar incelendiğinde, bilgi fonksiyonlarının sınırlı sayıda çalışmada dikkate alındığı görülmüştür. MTK'nın kuramsal temelleri dikkate alındığında, bu kurama dayalı

olarak elde edilen kestirimlerin doğruluğunun yetenek ölçeği boyunca farklılaştığı gözlenmektedir. DeAyala (1992) ve Tomkowicz (2000)'in çalışmalarından elde edilen sonuçlar incelendiğinde de kestirimlerin doğruluğunun farklı yetenek düzeylerine göre farklılaştığı ve MTK'ya ait bu özelliğin etkililiği gözlenmektedir. Bu çalışmada test bilgi fonksiyonları üç farklı yetenek ranjında ele alınarak detaylı olarak incelendiğinden, modellerin farklı yetenek düzeylerindeki etkililiğine ilişkin daha net bir resim ortaya koyduğu düşünülmektedir. Türkiye'de STM'nin başarı testi verilerine uygulanmış olduğu herhangi bir çalışmaya ise rastlanılmamıştır. Sınıflama düzeyindeki veri yapısı için geliştirilen STM çoktan seçmeli test maddelerinin incelenmesi için oldukça uygun bir modeldir. Çünkü çoktan seçmeli test maddelerinin yapısı ele alındığında bu maddelerin çeldiricilerden oluştuğu, maddenin doğru yanıtını bilmeyen bireylerin bu çeldiricilere yöneldiği görülmektedir. Ancak bu çeldiricilere yönelen öğrencilerin yetenek düzeyleri farklılaşmaktadır ve farklı yetenek düzeylerindeki bireyler sahip oldukları kısmi bilgiye göre farklı çeldiricilere yönelmektedir. Dolayısıyla, maddeye doğru yanıt vermeyen bireylerin tek bir kategoride ele alınması yerine, bu bireylerin sahip olduğu kısmi bilginin dikkate alınması ile söz konusu başarı testi verilerine dayalı olarak daha doğru kestirimlerin elde edilmesine ve bu doğrultuda bireyler hakkında daha sağlıklı tanılama, seçme ve yerleştirme kararlarının verilmesine katkıda bulunulacağı düşünülmektedir. Bu doğrultuda bu araştırmanın problemini 1PLM, 2PLM, 3PLM ve STM'nin model- veri uyumu, madde-test bilgi fonksiyonları ve model parametrelerinin kestirimini doğruluğu açısından karşılaştırmalı olarak incelenmesi oluşturmaktadır.

Araştırmanın Amacı

Bu araştırmanın amacı SBS-2012 Matematik alt testinden elde edilen verilere dayalı olarak 1-, 2-, 3- parametrelili lojistik modeller ile Sınıflamalı Tepki Modelinin karşılaştırılmasıdır. Bu modellerin model veri uyumu, madde ve test bilgi fonksiyonlarının sağladığı bilgi miktarı, model parametrelerinin kestirimlerinin doğruluğu açısından karşılaştırılmıştır. Bu genel amaç doğrultusunda bu çalışmada aşağıdaki sorulara yanıt aranmıştır:

- 1) 1PLM, 2PLM, 3PLM ve STM'den hangisi ilgili test verisine daha fazla uyum sağlamaktadır?

- 2) 1PLM, 2PLM, 3PLM ve STM altında yetenek parametresinin kestiriminin doğruluğu modeller arasında manidar bir şekilde farklılaşmakta mıdır?
- 3) 1PLM, 2PLM, 3PLM ve STM'ye göre madde parametrelerinin kestiriminin doğruluğu farklılaşmakta mıdır?
- 4) 1- PLM, 2PLM, 3PLM ve STM'ye dayalı olarak hesaplanan madde bilgi fonksiyonlarının sağladığı bilgi miktarı farklılaşmakta mıdır?
- 5) 1PLM, 2PLM, 3PLM ve STM'ye dayalı elde edilen test bilgi fonksiyonlarının sağladığı bilgi miktarı düşük, orta ve yüksek yetenek düzeylerindeki bireyler için farklılaşmakta mıdır?
 - a) 1PLM, 2PLM, 3PLM ve STM'ye dayalı elde edilen test bilgi fonksiyonlarının sağladığı bilgi miktarı düşük yetenek düzeyindeki bireyler için farklılaşmakta mıdır?
 - b) 1PLM, 2PLM, 3PLM ve STM'ye dayalı elde edilen test bilgi fonksiyonlarının sağladığı bilgi miktarı orta yetenek düzeyindeki bireyler için farklılaşmakta mıdır?
 - c) 1PLM, 2PLM, 3PLM ve STM'ye dayalı elde edilen test bilgi fonksiyonlarının sağladığı bilgi miktarı yüksek yetenek düzeyindeki bireyler için farklılaşmakta mıdır?

Araştırmanın Önemi

Bu araştırmada 2012 yılında uygulanan SBS Matematik alt testinden elde edilen veriye dayalı olarak, 1PLM, 2PLM, 3PLM ve STM karşılaştırılmıştır. Doğru olmayan yanıtları (çeldiricileri) da dikkate alan STM, ikili puanlamanın aksine öğrencinin kısmi bilgisini dikkate alarak öğrencinin yetenek düzeyi hakkında daha fazla bilgi elde edilmesini sağlamaktadır.

Çoktan seçmeli test maddelerinin ikili puanlanması sonucu öğrencinin doğru yanıtı işaretlememesi, öğrencinin o maddeye ilişkin bilgisini yok sayarak, ilgili maddeden sıfır puan almasına neden olmaktadır. Ancak eğitimde ve psikolojide yapılan ölçmeler doğrudan gözlenemeyen örtük özelliği dikkate almaktadır. Kuramsal olarak, öğrencinin yanlış yanıtının o maddeye ilişkin bilgiye sahip olmadığı şeklinde yorumlanması ise dolaylı ölçme

ve değerlendirme işlemlerinde uygun değildir. Dolayısıyla, çeldiriciler de doğru yanıtlar kadar öğrenci hakkında bilgi içeriyor olabilir.

De Mars (2008) tarafından gerçekleştirilen çalışmada, 3PLM ve STM karşılaştırılmıştır. Elde edilen bulgulara göre, 3PLM'in uygulandığı durumda, düşük yetenek düzeyindeki bireyler için yetenek parametrelerinin (θ) daha yanlı bir şekilde kestirildiği ve 3PLM altında bu bireyler için test puanlarının güvenilirliğinin düşük olduğu gözlenmiştir. Çoklu puanlama durumunda ise daha az yanlı yetenek kestirimlerinin yapıldığı belirtilmiştir. Bu açıklamalar ve araştırma bulguları doğrultusunda, çoktan seçmeli test maddelerinin, ikili puanlama yapılırken öğrencinin kısmi bilgisini dikkate almadığı için bilgi kaybına neden olduğu sonucuna varılmaktadır. Yapılacak bu araştırmanın, ülkemizde de sıklıkla kullanılan çoktan seçmeli test maddelerine yönelik yapılan bu eleştiriye ışık tutarak, öğrencilerin yetenek düzeyine ilişkin daha doğru kestirimler yapılmasına katkıda bulunacağı düşünülmektedir.

Ülkemizde öğrenciler ilköğretimden itibaren bir takım çoktan seçmeli testlere tabi tutulmaktadır. Bu testlerden elde edilen sonuçlara dayalı olarak birey hakkında teşhis edici ve yerleştirmeye yönelik kararlar verilmekte, çoktan seçmeli testlerden elde edilen puanlara göre bireyler çeşitli eğitim kurumlarına seçilmekte ve yerleştirilmektedir. Bu kararlar bireylerin hayatlarına yön verdiğinden birey hakkında doğru kararların verilmesi önem taşımaktadır. Bu durum ölçme ve değerlendirme sürecinin hatalardan olabildiğince arındırılarak yürütülmesi ile gerçekleştirilebilir. Dolayısıyla bu çalışma ölçme ve değerlendirmenin en az hatayla yapılmasına katkıda bulunmak amacıyla farklı modelleri karşılaştırmalı olarak değerlendirip en uygun modele karar vermeyi amaçladığından önem taşımaktadır. Ayrıca bu çalışmada model-veri uyumu, bilgi fonksiyonları, parametrelerin kestirimlerinin netliği gibi MTK'ya ilişkin bir takım özellikler, farklı modeller altında eş zamanlı olarak karşılaştırılmıştır. Dolayısıyla bu çalışma söz konusu modellere ilişkin daha net bir resmin ortaya çıkarması açısından önem taşımaktadır. Bu doğrultuda bu çalışmanın eğitimsel ve psikolojik ölçmelerde farklı durumlarda uygun modele karar verilmesi açısından araştırmacılara ve bir takım sınavlar doğrultusunda öğrenci hakkında karar vermeye çalışan kurumlara katkı sağlayacağı düşünülmektedir.

STM altında maddeler tepki kategorileri bazında ele alınarak çeldiricilerin işlevselliği incelenmekte, aynı zamanda çeldiricinin yanlılık gösterip göstermediği de belirlenebilmektedir. Maddeye ilişkin bu tür niteliklerin ele alınmasıyla daha nitelikli

maddeler geliştirilebilir. Bu durumun ülkemizde gerçekleştirilmesi amaçlanan madde havuzu oluşturma çalışmalarına önemli katkılarda bulunması beklenmektedir.

Demirtaşlı (2014) tarafından yapılan çalışmadakinden farklı olarak bu çalışmada tutum ölçeğinden elde edilmiş veri yerine, başarı testinden elde edilmiş veri kullanılmıştır. Ayrıca, STM çoklu puanlanan maddeler için MTK modelleri yerine lojistik modeller (1PLM, 2PLM, 3PLM) ile karşılaştırılmıştır. STM'nin ikili kodlama yoluyla puanlanan maddeler için MTK modellerine göre ne derece farklılaştığına ve ne derece etkili olduğuna vurgu yapılacağından alan yazına önemli katkıda bulunması beklenmektedir.

Alan yazındaki çalışmalar incelendiğinde, iki kategorili MTK modellerine ilişkin çalışmaların çok sayıda olduğu ancak çoklu puanlanan MTK modellerine dayalı çalışmaların sınırlı olduğu, ayrıca STM'nin de az çalışılan bir model olduğu gözlenmiştir. Alan yazındaki STM'ye ilişkin çalışmaların çoğunluğunun ise simülasyon veri üzerinde gerçekleştirildiği gözlenmektedir. Bu çalışmada alan yazında daha az incelenmiş olan STM ele alınmış ve aynı zamanda gerçek veri üzerinde çalışılmıştır. Dolayısıyla bu çalışmanın uygulamaya da bilgi sağlaması beklenildiğinden ilgili alan yazına katkısı olacağı düşünülmektedir.

Bu çalışmada birey hakkında daha fazla bilgiyi ortaya çıkararak daha sağlıklı kararlar alınmasını sağlaması beklenen, çoklu puanlanan maddeler için MTK modellerinden STM ile lojistik modeller karşılaştırılmıştır. Karşılaştırmada ayrıca farklı yetenek düzeylerindeki bireyler için testin sağladığı bilgi miktarları da incelenmiştir. Bu karşılaştırmanın da testin kullanım amacına hizmet etme derecesini değerlendirme bağlamında geçerliğe ilişkin kanıt sağlaması açısından da önemli olduğu düşünülmektedir.

Sınırlılıklar

Bu çalışma, 2012 SBS Matematik alt testinden elde edilen veriler üzerinden gerçekleştirilen analizlerden elde edilen bilgiler ile sınırlıdır.

Çalışmada ayrıca STM altında yerel bağımsızlık varsayımını test edebilecek bir paket programa ulaşılamamış, bu varsayım sadece iki kategorili MTK modelleri için test edilmiştir. Bu durum da bir sınırlılık teşkil etmektedir.

BÖLÜM II

YÖNTEM

Araştırmanın Modeli

Bu çalışmada, 2012 SBS Matematik alt testine ait veri üzerinde 1PLM, 2PLM, 3PLM ve Sınıflamalı Tepki Modeli'ne dayalı olarak model-veri uyumu, madde ve birey parametrelerinin doğruluğu, madde ve test bilgi fonksiyonları karşılaştırmalı olarak incelenmiştir. Bu çalışmada mevcut kuramsal bilgiye yenilerinin eklenmesi ve söz konusu bilginin gelişmesine katkıda bulunulması amaçlandığından temel araştırma niteliğindedir. Temel araştırma genellikle bir kuram ile ifade edilen hipotezin altındaki süreçlere açıklık getirmek ve yeni bilgi elde etmek amacıyla yapılan çalışmadır (Fraenkal, Wallen ve Hyun, 2012, s.7).

Çalışma Grubu

Çalışma grubunu 2012 yılında uygulanan SBS'ye katılmış öğrencilerden; MEB'den elde edilen 50.000 kişilik gruptan seçkisiz olarak seçilen, 1500 öğrenci oluşturmaktadır. Çalışmada ilgili MTK modellerinin model-veri uyumunu madde düzeyinde değerlendirmek için uyum istatistikleri hesaplanmıştır. Uyum istatistikleri örneklemden etkilendiği için (Khalid, 2009) evren üzerinde çalışılmamış, kullanılan MTK modelleri için gerekli örneklem büyüklüğü göz önünde bulundurulmuştur. Bu çalışmada incelenen MTK modellerinin gerektirdiği örneklem büyüklüğü için ilgili çalışmalara (de la Torre ve Hong, 2010; Foley, 2010) bakıldığında, 1000 kişilik örneklem büyüklüğünün yeterli olduğu görülmektedir. Bu doğrultuda çalışma, MEB'den elde edilen 50000 kişilik veri setinden seçkisiz olarak seçilen 1500 öğrenciden elde edilen veriler üzerinde gerçekleştirilmiştir.

Veri Toplama Araçları

SBS-2012, 8. Sınıf öğrencilerine Matematik, Fen ve Teknoloji, Sosyal Bilgiler ve İngilizce ders alanlarında uygulanmıştır. Bu çalışma matematik alt testinde yer alan 20 maddeye ilişkin öğrenci verileri üzerinde gerçekleştirilmiştir. Matematik alt testinde sayılar, geometri, olasılık ve istatistik, ölçme ve cebir alt öğrenme alanlarına ilişkin maddeler yer almıştır (MEB, 2009). Çalışma grubundaki bireylerin SBS 2012 Matematik alt testine verdikleri yanıtlara dayalı olarak bu çalışma gerçekleştirilmiştir.

Çalışma için gerekli veriler, Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri tarafından sağlanmıştır. MEB YEĞİTEK'ten sağlanan ham veriler düzenleme ile analizler için hazır hale getirilmiştir. Çalışmada iki kategorili ve çoklu puanlanan olmak üzere modeller iki farklı şekilde ele alındığından veriler iki farklı veri seti olacak şekilde düzenlenmiştir. İkili kategorili MTK modelleri için doğru yanıt 1, diğer bütün yanıtlar 0 olacak şekilde veriler 0-1 formatında düzenlenmiştir. Çoklu puanlanan model için ise uzman görüşleri doğrultusunda çeldiricilerin doğru yanıtla olan yakınlık derecesi belirlenmiştir. Beş uzmandan alınan görüşler değerlendirilmiş, ortak görüşler doğrultusunda her bir maddenin doğru yanıtla en yakın ve en uzak olan çeldiricisi belirlenmiştir. Bu doğrultuda doğru yanıt 4, doğru yanıtla en yakın çeldirici 3, daha az yakın çeldirici 2 ve en uzak çeldirici 1 olacak şekilde kodlamalar yapılarak veri seti düzenlenmiştir (Bakınız EK 11).

Verilerin Toplanması

Araştırma amacı doğrultusunda ihtiyaç duyulan ham veriler, Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü Ölçme Değerlendirme ve Yerleştirme Grup Başkanlığı tarafından yazılı izin doğrultusunda elde edilmiştir. Yazılı izin ekte sunulmuştur (Bakınız EK 1). Çoklu puanlanan STM'nin çeldiricilerinin puanlaması için ilköğretim matematik öğretmenliği anabilim dalındaki biri doçent, ikisi yardımcı doçent olmak üzere üç uzmandan; ilköğretim matematik öğretmenliğinde yüksek lisans programını tamamlamış ve aynı zamanda eğitimde ölçme ve değerlendirme alanında doktora programına devam etmekte olan bir araştırma görevlisinden; MEB'de çalışan bir ilköğretim matematik öğretmeninden olmak üzere beş uzmandan görüş alınmıştır. Uzman değerlendirme formunda 2012-SBS Matematik alt testinde yer alan maddelere ait çeldiricilerin doğru yanıtla yakınlığı ve uzaklığı bağlamında ele alınacağı ve en yakın çeldiricinin 3, en uzak çeldiricinin

ise 1 olacak şekilde her bir maddenin çeldiricilerinin 1, 2, ve 3 şeklinde puanlanması gerektiği belirtilmiştir. Ayrıca bu puanlamaların Yüksek lisans tez çalışması kapsamı doğrultusunda kullanılacağına yer verilmiştir. Uzmanlardan çeldiricileri doğru yanıtla yakınlığı açısından derecelendirmeleri istenmiştir. İlgili uzman değerlendirme formu ve matematik alt testi maddeleri ekte sunulmuştur (Bakınız EK 2 ve EK 3). Uzmanlardan alınan görüşlere göre her bir maddenin kategorileri doğru yanıtla en yakın ve en uzak olan çeldiriciler olarak belirlenmiştir. Bu doğrultuda çoklu puanlama modelinin analizleri için seçenekler 1, 2, 3, 4 olacak şekilde kodlamalar yapılmıştır. Bu düzenlemeler ile veri seti analizler için hazır hale getirilmiştir.

Ölçüm Güvenirliği

Bu çalışma, çalışma grubundaki öğrencilerin 2012 SBS Matematik alt testinde yer alan maddelere verdikleri yanıtlara dayalı olarak gerçekleştirilmiştir. Çalışma verileri sadece MTK bağlamında incelendiği için, söz konusu ölçümlerin MTK bağlamında güvenirligine ilişkin kanıtlar elde edilmiştir. MTK altında söz konusu ölçümlerin güvenirligini belirlemek amacıyla marjinal güvenirlilik katsayısı hesaplanmıştır.

Bu katsayı, gerçek puan modeline dayandırılmakta (Lord ve Novick, 1968) ve koşullu standart hata ortalamasına dayalı olarak bir testin genel güvenirliginin kestirimi olarak ifade edilmektedir. Marjinal güvenirlilik katsayısının teorik alt yapısı ile klasik güvenirlilik katsayısının teorik alt yapısı karşılaştırılabilir olduğundan, elde edilen katsayılar da benzer şekilde yorumlanabilmektedir (Dimitrov, 2003). Tablo 1’de, bu çalışmada test edilen MTK modellerine dayalı olarak hesaplanan marjinal güvenirlilik katsayıları yer almaktadır:

Tablo 1. Marjinal Güvenirlilik Katsayıları

1PLM	2PLM	3PLM	STM
0,800	0,892	0,856	0,883

Her bir model altında elde edilen marjinal güvenirlilik katsayısının alt sınırları KTK’da olduğu gibi yorumlanabilmektedir. Bu doğrultuda ele alındığında, testin 1PLM, 2PLM, 3PLM ve STM altında güvenilir ölçümler sağladığı sonucuna varılmaktadır.

Ölçüm Yorumlarının ve Kullanımlarının Geçerliliği

SBS, ortaöğretime geçişte adayların seviyelerinin belirlenmesi ve buna bağlı olarak uygun orta öğretim kurumlarına yerleştirilmeleri hakkında karar verilmesini amaçlayan bir sınav niteliği taşımaktadır. Analizlerin MTK modelleri kapsamında yapılması durumunda veri için en uygun model belirleneceğinden, daha sağlıklı seviye belirleme ve yerleştirme işlemlerinin yapılması öngörülmektedir. De Ayala (2009) standart geçerlik kanıtı elde etme yollarının MTK bağlamında da kullanılabildiğini belirtmektedir. Bu çalışmada da tek boyutluluk varsayımının test edilmesinde kullanılan açımlayıcı faktör ve doğrulayıcı faktör analizi sonuçları, aynı zamanda geçerlik kanıtı olarak sunulmaktadır (faktör analizi sonuçları, tek boyutluluk varsayımına ilişkin de bilgi sağladığı için bu analiz süreci ve sonuçları, verilerin çözümlenmesi ve yorumlanması altında ele alınmış ve detaylı olarak sunulmuştur). Burada testin amaçlandığı gibi “matematik alanındaki başarı” yı tek boyutlu bir yapı olarak ölçebildiği gözlenmiştir. Bunun yanı sıra test bilgi fonksiyonundan elde edilen bilgi miktarı alt, orta ve yüksek yetenek ranjı olmak üzere üç farklı aralıkta ele alınmış olup hangi aralıkta daha fazla bilgi sağladığı ortaya çıkarılmıştır. Bu doğrultuda modellerin etkililiği farklı yetenek düzeyleri de dikkate alınarak ortaya çıkarılmaya çalışılmıştır. Buradan elde edilen bulgulardan hareketle, her model altında sağlanan bilgi miktarının en yüksek olduğu yetenek ranjının orta yetenek olduğu, dolayısıyla ölçümlerin amacına hizmet ettiği ve bu yetenek ranjında en az hatayla güvenilir ölçümler elde edilebileceği sonucuna varılmıştır. Bu sınav seviye belirlemeye yönelik olup, orta yetenek düzeyindeki bireyler arası farklılığı ortaya çıkarmayı amaçlamaktadır. Dolayısıyla testin kullanım amacına yönelik olduğu söylenebilir. Bu doğrultuda bu alt test puanlarına dayalı olarak öğrenciler hakkında sağlıklı ve geçerli kararlar alınabileceği düşünülmektedir.

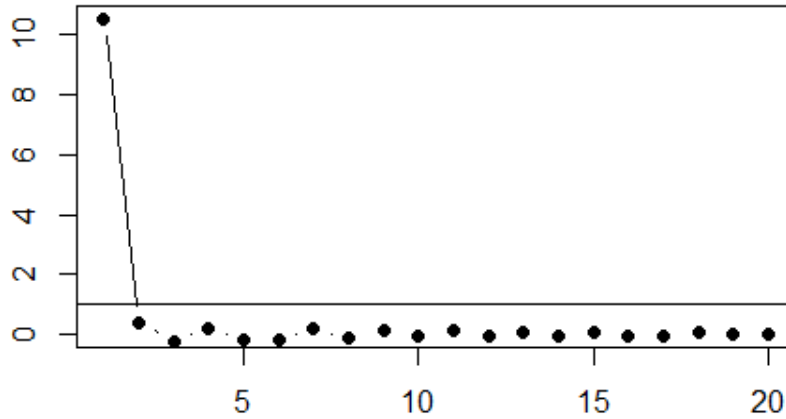
Verilerin Analizi

Bu çalışmada SBS-2012 çalışmasında yer alan Matematik alt testinden elde edilen verinin, MTK'nın 1-, 2-, 3- parametrelili lojistik modellerine ve sınıflamalı tepki modeline uyumu test edilmiş, bu model parametrelerinin kestirimlerinin doğruluğu ve bilgi fonksiyonları açısından karşılaştırılmıştır. Modelleri test etmenin öncesinde MTK varsayımlarının, ilgili veri seti için karşılanıp karşılanmadığı incelenmiş, MTK'nın en önemli avantajlarından biri olan değişmezlik özelliği madde ve birey parametreleri için kontrol edilmiştir.

Varsayımların Test Edilmesi

Tekboyutluluk

MTK'nın varsayımlarından biri tek boyutluluktur. Hambleton ve Swaminathan (1985) tek boyutluluk varsayımı için açıcı faktör analizinin (AFA) yapılabileceğini belirtmektedirler. Analiz sonucunda, eğer maddeler tek bir faktöre yükleniyorsa, tek boyutluluk varsayımının karşılandığı kabul edilmektedir. Bu çalışmada tek boyutluluk varsayımını test etmek için 1500 kişilik örneklemden elde edilen veri seti üzerinde AFA yapılmıştır. Bu doğrultuda 1-0 puan matrisine dayalı olarak tetrakorik korelasyon katsayıları hesaplanmış, bu korelasyon matrisine dayalı olarak AFA yapılmıştır. Faktör analizi için psych paketi ile R studio yazılım ortamından yararlanılmıştır. Rotasyonsuz “minimum residual” (unweighted least squares) faktör çıkartma tekniğine dayalı olarak maksimum olabilirlik yöntemi (maximum likelihood) ile gerçekleştirilen faktör analizi sonucunda özdeğeri 1'den büyük olan tek bir faktörün ortaya çıktığı görülmektedir. Tek boyutluluk varsayımı için alan yazında öz değerler eğrisi de incelenmektedir Burada yapılan AFA sonucunda oluşturulan öz değerler grafiği aşağıda Şekil 1'de sunulmaktadır.



Şekil 1. Matematik alt testine ilişkin yapılan AFA sonucu oluşturulan özdeğerler grafiği

Elde edilen scree plot grafiğindeki kırılma noktasının üzerinde tek nokta olması tek faktörlü yapıyı doğrulamaktadır Özdeğerler'in yanı sıra her bir maddeye ilişkin faktör yükleri elde edilerek ilgili faktöre yüklenip yüklenmediğine bakılmıştır. Bu faktör yüklerine ilişkin değerler Tablo 2'de sunulmuştur.

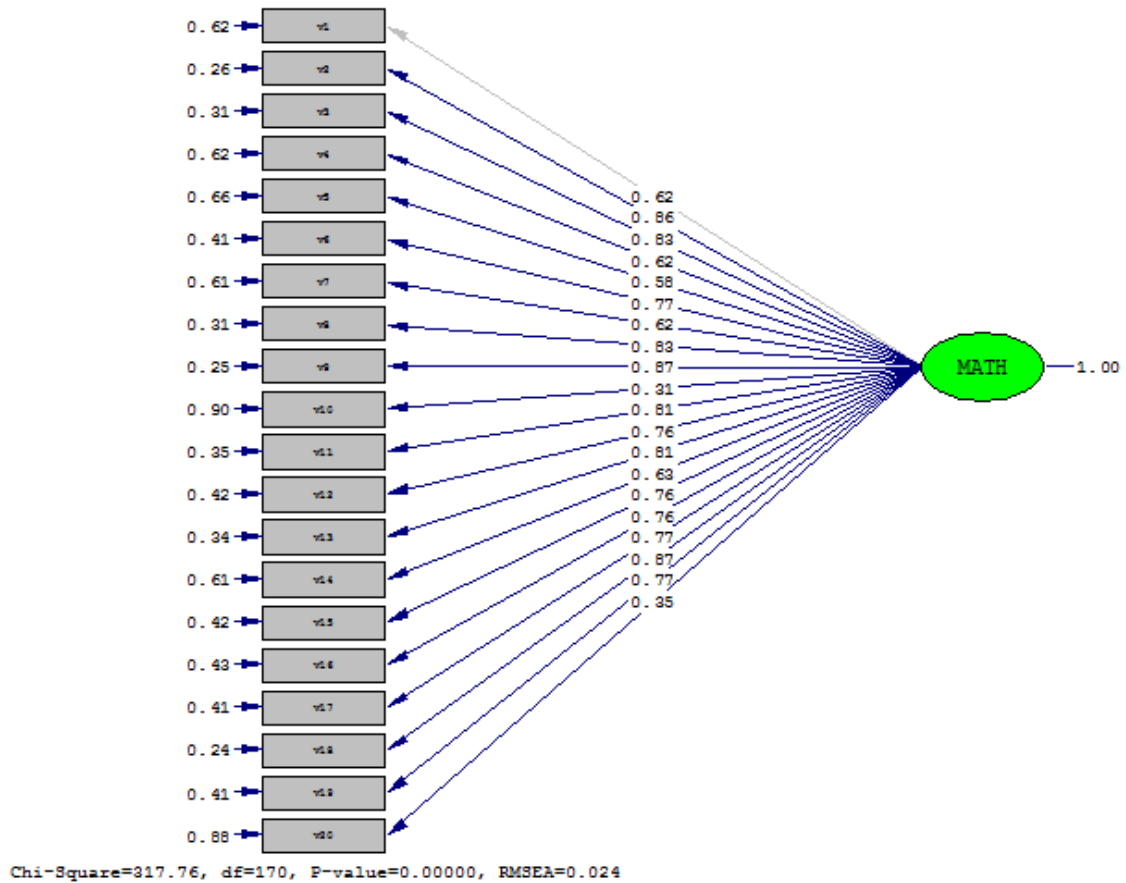
Tablo 2. Matematik Alt Testinden Elde Edilen Verilere Dayalı Yapılan Açımlyıcı Faktör Analizi

Maddeler	Faktör yükleri
m1	0,62
m2	0,86
m3	0,83
m4	0,62
m5	0,58
m6	0,77
m7	0,62
m8	0,83
m9	0,87
m10	0,31
m11	0,81
m12	0,76
m13	0,81
m14	0,63
m15	0,76
m16	0,76
m17	0,77
m18	0,87
m19	0,77
m20	0,35
öz deęer	10,53
açıklanan varyans(%)	0,53

Faktör yük deęerlerine bakıldığında, bütün maddelerin 0.30'un üzerinde olduęu yani maddelerin ilgili faktörün uygun birer göstergesi oldukları anlaşılmaktadır. Bu bulgular söz konusu maddelerin, "matematik başarısı" faktörünü yeterli düzeyde temsil edip ölçebildiğine işaret etmektedir.

Benzer şekilde tüm gruptan elde edilen veri seti üzerinde de AFA uygulanmış, bu veri seti için de tek bir faktörün ortaya çıktığı gözlenmiştir (Bakınız EK 4). Dolayısıyla, hem tüm grup (N=50.000) hem de çalışma grubu için (n=1500) Matematik alt testinin "Matematik başarısı" olarak adlandırılabilir tek faktörlü bir yapıyı ölçtüęü sonucuna ulaşılmıştır.

Açımlayıcı faktör analizinin yanı sıra, tek boyutluluk varsayımını test etmek üzere öncelikle ilgili ölçümlerin faktör yapısına ilişkin tek faktörlü bir ölçme modeli tanımlanmıştır. Bu ölçme modeli LISREL 8.7 Programında DFA uygulanarak test edilmiştir. Test maddelerine verilen (N=1500) yanıtlar 1-0 şeklinde puanlandığı için LISREL’de değişken türü ordinal olarak tanımlanmış ve parametre kestirim yöntemi olarak da “Robust Maksimum Likelihood” yöntemi kullanılmıştır. Gizil değişken ölçeğini tanımlamak amacıyla işaretçi gösterge yöntemi kullanılarak 1.madde için faktör yük değeri 1’e eşitlenerek doğrulayıcı faktör analizi yapılmıştır. Tek faktörlü ölçme modeline ilişkin hesaplanan uyum indeksleri (CFI = 1.00, NFI=1.00, GFI=0.92, AGFI=0.90) ve hata değeri (RMSEA=0.024) söz konusu modelin çalışma grubundan elde edilen veri setine çok iyi uyum sergilediğine işaret etmektedir (Brown, 2006). Bu ölçme modelinde yer alan göstergelere ilişkin hesaplanan faktör yük değerleri ve özgül varyanslar (standartlaştırılmış katsayılar) Şekil 2’de, modelin şekilsel gösterimi üzerinde sunulmuştur:



Şekil 2. SBS-2012 Matematik alt testinin faktör yapısına ilişkin tanımlanan tek faktörlü ölçme modeli

Şekil 2 incelendiğinde göstergelere ilişkin hesaplanan faktör yük değerlerinin $\lambda = 0.31 - \lambda = 0.87$ arasında, özgül varyans değerlerinin ise $\varepsilon = 0.24 - \varepsilon = 0.90$ arasında değiştiği görülmektedir. Bu bulgular söz konusu maddelerin “Matematik başarısı” olarak adlandırılabilir yapının uygun birer temsilcisi olduklarına ve ilgili yapıyı kabul edilebilecek şekilde ölçebildiğine işaret etmektedir. Benzer şekilde tüm gruptan (N=50.000) elde edilen veri üzerinde de DFA uygulanmış ve tüm grup için de tek faktörlü yapının doğrulandığı sonucuna ulaşılmıştır. Tüm gruba ait uyum DFA sonuçları ektedir (Bakınız EK 4). Dolayısıyla, hem evren hem de örneklemden elde edilen veri setinin söz konusu matematik başarısı faktör yapısını doğruladığı sonucuna varılmıştır.

Yerel Bağımsızlık

Alan yazında tek boyutluluk varsayımının karşılanmış olmasının yerel bağımsızlık varsayımı için de yeterli kanıt olduğu belirtilmektedir (Embretson ve Reise, 2000; Hambleton vd., 1991). Ayrıca alan yazında Yen (1984) tarafından geliştirilen Q3 istatistiği kullanılarak yerel bağımsızlık varsayımının test edildiği belirtilmektedir. MTK modelleri için artık değerlere (residuals) ilişkin bir korelasyon katsayısı olarak ifade edilen Q3 istatistiği, madde çiftleri arasındaki bağımlılığın derecesini gösteren bir indekstir. Bu indeks her bir madde çifti için (testte yer alan madde sayısının ikili kombinasyonları) elde edilmektedir. Q3 istatistiğinin mutlak değerinin 0.2 değerinden küçük olması, ilgili madde çiftlerinin birbirinden yerel bağımsız olduğuna işaret etmektedir (Chen ve Thissen, 1997; DeAyala, 2009). Bu doğrultuda, yerel bağımsızlık varsayımı olası tüm madde çiftleri için teker teker test edilmiştir. R studio programında, sirt R paketi kullanılarak elde edilen olası her bir madde çifti için (20 madde için 190 tane, 20 nin 2 li kombinasyonu) Q3 istatistik değerleri elde edilmiştir. Bu istatistiklerin mutlak değerinin 0.2 den küçük olması her bir madde çifti için yerel bağımsızlık varsayımının karşılandığına kanıt olarak gösterilmektedir (DeMars, 2010, s.50; DeAyala, 2009, s.134). 1PLM, 2PLM ve 3PLM altında 190 madde çifti için yerel bağımsızlık varsayımı test edilmiş ve her bir madde çifti için bu varsayımın karşılandığı görülmüştür. Hesaplanan Q3 istatistikleri, ekte tablo halinde sunulmuştur (Bakınız EK 5).

MKE'nin Monotonik Artış Göstermesi

MTK'nın bir diğer varsayımı madde karakteristik eğrisinin monotonik artış göstermesidir. Yetenek düzeyi arttıkça bireyin maddeyi doğru yanıtlama olasılığına işaret eden bu varsayım Multilog programı ile her bir model altında ve her bir madde için madde karakteristik eğrisi elde edilerek incelenmiştir (Hambleton vd., 1991). Her bir model altında elde edilen madde karakteristik eğrileri ekte verilmiştir (Bakınız EK6 ve EK 7). Bu incelemelere dayalı olarak her bir madde için örtük değişken ile madde yanıtları arasında doğrusal bir ilişki olduğu, yani yetenek düzeyi arttıkça maddeyi doğru yanıtlama olasılığının da arttığı gözlenmiştir.

Parametrelerin Değişmezliğinin İncelenmesi

MTK'nın geçerli sonuçlar vermesi için parametrelerin değişmezliğinin sağlanması gerekmektedir. Bu doğrultuda parametrelerin değişmezliği, madde ve birey parametrelerinin değişmezliği olmak üzere iki farklı şekilde incelenmiştir.

Madde Parametrelerinin Değişmezliği

Madde parametrelerinin değişmezliğini incelemek amacıyla çalışma grubundaki öğrenciler 750'şer kişilik iki alt gruba seçkisiz olarak atanmıştır. Her bir modele dayalı olarak, bu iki grup için madde parametreleri kestirilmiş, bu alt gruplara ait kestirimler arasındaki ilişkiler Pearson Momentler Çarpımı Korelasyonu ile incelenmiştir. Ayrıca bu alt gruplara ait kestirimler arasındaki ilişkinin doğrusal olup olmadığını incelemek amacıyla SPSS 20.0 programı ile korelasyon katsayıları hesaplanmış ve saçılma grafiği oluşturulmuştur. Hesaplanan korelasyon katsayıları Tablo 3'te sunulmuştur:

Tablo 3. İkili Kategorili Modeller Altında İki Farklı Gruba İlişkin Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Hesaplanan Korelasyon Katsayıları

Alt gruplar	parametre	1PLM	2PLM	3PLM
Grup1-Grup2	(b)	0.979	0.995	0.992
Grup1-Grup2	(a)		0.966	0.876
Grup1-Grup2	(c)			0.753

Tablo 3'ten 1PLM, 2PLM ve 3PLM altında kestirilen b parametreleri arasındaki korelasyon değerlerine bakıldığında, bu değerlerin her bir model için yüksek düzeyde ilişki sergilediği gözlenmektedir. Benzer şekilde 2PLM ve 3PLM altında kestirilen a parametreleri arasındaki korelasyon değerlerine bakıldığında, bu değerlerin her iki model için de yüksek düzeyde ilişki sergilediği gözlenmektedir. 3PLM altında kestirilen c parametresine ilişkin korelasyon katsayısının ise orta düzeyde olduğu görülmektedir. Bütün bu korelasyon katsayıları incelendiğinde 1PLM, 2PLM ve 3PLM için madde parametrelerinin değişmezliğinin sağlandığı sonucuna ulaşılmaktadır. Ayrıca saçılma grafikleri ile her bir model altında kestirilen madde parametreleri arasındaki ilişkiler incelenmiş ve bu ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır. İki ayrı gruptan elde edilen madde parametreleri arasındaki ilişkilere yönelik saçılma grafikleri Ek'te yer almaktadır (Bakınız EK 7)

Sınıflamalı tepki modeli için parametre değişmezliğini incelemek amacıyla, her bir kategoriye dayalı olarak iki alt gruptan (Grup1(750 kişi)–Grup2 (750 kişi)) elde edilen a ve b parametre kestirimleri arasındaki ilişkilere yönelik Pearson Momentler Çarpımı korelasyon katsayıları hesaplanmıştır. Ayrıca bu alt gruplardan elde edilen verilere dayalı olarak yapılan kestirimler arasındaki ilişkinin doğrusal olup olmadığını incelemek amacıyla SPSS 20.0 programı ile korelasyon katsayıları hesaplanmış ve saçılma grafiği oluşturulmuştur. Hesaplanan korelasyon katsayıları Tablo 4'te sunulmuştur:

Tablo 4. STM Altında İki Farklı Gruba İlişkin Madde Parametreleri Arasındaki İlişkilere Yönelik Korelasyon Katsayıları

Sınıflamalı Tepki Modeli				
Alt Gruplar	a		b	
Grup1-Grup2	(a ₁)	0.712	(b ₁)	0.992
Grup1-Grup2	(a ₂)	0.550	(b ₂)	0.963
Grup1-Grup2	(a ₃)	0.853	(b ₃)	0.981
Grup1-Grup2	(a ₄)	0.960	(b ₄)	0.981

Elde edilen madde parametreleri arasındaki korelasyon değerlerine bakıldığında, bu değerlerin her bir model için yüksek düzeyde ilişki sergilediği gözlenmektedir. Bu bulgular madde parametreleri açısından değişmezlik varsayımının karşılandığına işaret etmektedir (Hambleton ve Swaminathan, 1985). Ayrıca saçılma grafikleri ile parametrelerin arasındaki

ilişkiler incelenmiş ve bu ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır. İki ayrı gruptan elde edilen madde parametreleri arasındaki ilişkilere yönelik saçılma grafikleri Ek'te yer almaktadır (Bakınız EK 8).

Birey Parametrelerinin Değişmezliği

Birey parametrelerinin değişmezliğini incelemek amacıyla test maddeleri, tek maddeler ve çift maddeler olmak üzere 10'ar maddelik iki alt teste ayrılmıştır. Her bir model altında ve her bir alt test için birey parametreleri kestirilerek bu alt testlere dayalı yetenek parametreleri kestirimleri arasındaki ilişkiler Pearson Momentler Çarpımı Korelasyon Tekniği ile incelenmiştir. Ayrıca bu alt testlere dayalı kestirilen birey parametreleri arasındaki ilişkinin doğrusal olup olmadığını incelemek amacıyla SPSS 20.0 programı ile korelasyon katsayıları hesaplanmış ve saçılma grafiği oluşturulmuştur. Hesaplanan korelasyon katsayıları Tablo 5'te sunulmuştur:

Tablo 5. Bütün Modeller Altında İki Farklı Gruba İlişkin Yetenek Parametreleri Arasındaki İlişkilere Yönelik Hesaplanan Korelasyon Katsayıları

	1PLM	2PLM	3PLM	STM
Tek-çift maddeler	0.813	0.820	0.885	0.820

Tablo 5'teki değerler incelendiğinde, her bir model için, alt testlere dayalı yetenek parametrelerinin kestirimleri arasındaki ilişkilerin yüksek olduğu görülmektedir. Bu katsayılar, birey parametreleri açısından değişmezlik varsayımının karşılandığına işaret etmektedir. Ayrıca saçılma grafikleri ile birey parametreleri arasındaki ilişkiler incelenmiş ve bu ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır. İki alt testte dayalı kestirilen yetenek parametreleri arasındaki ilişkilere yönelik saçılma grafikleri Ek'te yer almaktadır (Bakınız EK 9).

Model-veri Uyumu

MTK'nın etkili ve geçerli sonuçlar vermesi için sağlanması gereken bir diğer özellik model ile veri arasındaki uyumun yeterli düzeyde olmasıdır ve bu uyum farklı şekillerde ele alınmaktadır. Model veri uyumu, her bir model için genel model uyumu (overall fit) olarak

ve madde düzeyinde ele alınmıştır. Model-veri uyumuna ilişkin bütün istatistikler, R studio ortamında mirt R paketi kullanılarak elde edilmiştir. Öncelikle her bir modele ilişkin uyum istatistikleri ve hata değerleri (CFI, TLI, RMSEA,SRMRS) elde edilmiştir. Genel model uyumu bazında bütün modellerin veriye iyi uyum gösterdiği gözlenmiştir. Bunun yanı sıra iki kategorili MTK modellerine ilişkin uyum istatistikleri aynı paket kullanılarak varyans analizi tekniği (Anova) ile karşılaştırılarak, modellerin uyum düzeylerinin birbirinden manidar bir şekilde farklılaşıp farklılaşmadığı incelenmiştir. Bu karşılaştırma her bir model altında elde edilen -2loglikelihood değerlerinin arasındaki farka dayalı olarak yapılmaktadır. Bu farkın manidar olması, daha kompleks olan modelin veriye daha iyi uyum sergilediğine işaret etmektedir (DeAyala, 2009, s.140). -2loglikelihood değerlerinin karşılaştırılabilmesi için bu değerlerin aynı veri setine dayalı olarak elde edilmesi gerekmektedir. Bu çalışmada 0-1 şeklinde ikili puanlanan MTK modelleri ve çoklu puanlanan STM test edildiğinden ikili karşılaştırma sadece iki kategorili MTK modelleri için mümkün olmaktadır. Genel model veri uyumunun yanı sıra, madde düzeyindeki uyumu değerlendirmek için mirt paketi kullanılarak $S-X^2$ istatistiği hesaplanmıştır. Bu yöntem ile istatistiksel hesaplamalar için gruplama işlemleri yapılmakta ve kesme puanları kullanılmaktadır. Geleneksel yöntemlerde kesme puanları örneklem ve modele bağımlıyken, $S-X^2$ istatistiğine dayalı yöntemde test puanına temellendirilmiştir. Bu istatistik ile hem iki kategorili MTK modelleri, hem de çok kategorili MTK modelleri madde uyumu bazında ele alınabilmektedir.

Model Parametrelerinin Kestirimi ve Bilgi Fonksiyonlarının Hesaplanması

MTK varsayımlarının ve değişmezlik özelliğinin sağlanıp sağlanmadığı test edilip, her bir model altında model-veri uyumu incelendikten sonra modeller R studio yazılım ortamında ve Multilog programında test edilmiştir. 1PLM, 2PLM ve 3PLM altında madde parametreleri sirt paketiyle R studio yazılım ortamında kestirilmiştir. Parametre kestiriminde pratik anlamlılık için ayırt edicilik parametresi (0-2) aralığına, güçlük parametresi ise ((-2)-2) aralığına sınırlandırılarak kestirilmiştir (Hambleton vd., 1991). STM'ye ilişkin temel parametreler olan zeta ve lambda parametreleri mcIRT paketiyle R studio yazılım ortamında kestirilmiş, bu parametrelerin doğrusal dönüşümü sonucu elde edilen a ve b parametreleri ise Multilog programında kestirilmiştir. Madde parametre kestirimlerinin tamamında Marjinal Maksimum Olabilirlik kestirim yöntemi kullanılmıştır. 1-, 2, 3PLM ve STM'ye

ilişkin madde ve test bilgi fonksiyonlarının tamamı Multilog programında hesaplanmıştır. Yetenek parametreleri bütün modeller için mirt paketiye R studio yazılım ortamında Maksimum Olabilirlik kestirim yöntemiyle kestirilmiştir. Farklı modeller altında yetenek parametrelerinin kestirimlerinin doğruluğunu karşılaştırmak amacıyla her bir modele dayalı elde edilen standart hata değerleri, ilişkili örneklemeler için tek yönlü varyans analizi ile SPSS paket programı kullanılarak karşılaştırılmıştır (Büyüköztürk, 2012; Field, 2012; Pallant, 2007). Varyans analizi sonuçları incelendiğinde modeller arasındaki farklılığın manidar olduğu gözlenmiş, elde edilen madde çiftleri için manidar farklılığın yönünü belirlemek amacıyla standart hata ortalamaları incelenmiştir. Ayrıca, modeller arasındaki manidar düzeyde olduğu gözlenen farklılığın derecesini belirlemek amacıyla ise eta kare (η^2) etki büyüklüğü hesaplanmıştır. Elde edilen etki büyüklükleri Cohen (1988)'in kriterine göre değerlendirilmiştir. Cohen (1988)'in kriterine göre $\eta^2 = 0.1$ küçük etki, $\eta^2 = 0.3$ orta büyüklükte etki ve $\eta^2 = 0.5$ büyük etki olarak yorumlanmaktadır.



BÖLÜM III

BULGULAR, SONUÇ VE TARTIŞMA

Bulgular ve Yorum

Bu bölümde araştırma sorularına yanıt vermek amacıyla 1PLM, 2PLM, 3PLM ve STM'ye ilişkin model veri uyumu, madde parametrelerinin kestirimlerinin doğruluğu, yetenek parametrelerinin kestirimlerinin doğruluğu, madde bilgi fonksiyonları ve test bilgi fonksiyonlarının test edilmesi amacıyla yapılan analizlere ilişkin bulgulara ve bu bulgulara ilişkin yorumlara yer verilmiştir. Araştırma bulguları aşağıda araştırma soruları doğrultusunda sırasıyla verilerek yorumlanmıştır.

“1PLM, 2PLM, 3PLM VE STM'den hangisi ilgili test verisine daha fazla uyum sergilemektedir?” araştırma sorusuna yanıt vermek amacıyla model-veri uyum düzeyi test edilmiştir. Genel model veri uyumu ve madde bazındaki model veri uyumu olmak üzere iki şekilde ele alınan model-veri uyumu için öncelikle genel model veri uyumu ele alınmıştır. Tablo 6'da her bir modele ilişkin hesaplanan uyum istatistik değerleri sunulmuştur:

Tablo 6. Model Veri Uyumuna İlişkin Hesaplanan Uyum İstatistikleri

	RMSEA	TLI	CFI	SRMRS
1PLM	0,072317	0,949685	0,949949	0,123093
2PLM	0,020914	0,995792	0,996235	0,034145
3PLM	0,016637	0,997337	0,997898	0,022987
STM	0,022834	0,976346	0,985808	

Tablo 6’da verilen uyum istatistikleri ve hata değerleri incelendiğinde bütün modellerin veriye iyi uyum sergilediği görülmektedir. İki kategorili modeller için elde edilen SRMRS değerlerine göre en fazla uyum gösteren model 3PLM, en az uyum gösteren model ise 1PLM’dir. SRMRS değerleri sadece sıralı (ordinal) veriye dayalı olarak elde edildiği için iki kategorili modelleri kendi aralarında karşılaştırmak amacıyla kullanılmıştır (Chalmers, 2012). Bütün modeller ele alındığında en az hata değerine (RMSEA) ve en yüksek model uyum istatistik değerlerine (CFI, TLI) göre model uyumu sıralaması, 3PLM, 2PLM, STM ve 1PLM şeklindedir. Burada 2PLM ile STM için RMSEA değerleri arası fark çok az olmakla birlikte CFI açısından biraz daha yüksek fark vardır. Bunun yanı sıra modellerin uyum düzeylerini karşılaştırmak amacıyla -2loglikelihood değerleri elde edilmiş ve bu değerler tek yönlü varyans analizi ile ikili olarak karşılaştırılmıştır. Modellerin -2 loglikelihood değerleri ve ikili model karşılaştırmasına ilişkin olasılık değerleri (p-values) Tablo 7’de yer almaktadır:

Tablo 7. İki Kategorili Modellerin Uyum Düzeylerinin Karşılaştırılmasına Yönelik Varyans Analizi Sonuçları

	1PLM	2PLM	3PLM
1PLM (-16666,3)		0 (p<0.01)	0 (p<0.01)
2PLM (-16123,6)			0 (p<0.01)
3PLM (-15632,9)			

*p<.01

R studio programı, bu varyans analizine dayalı olarak ikili karşılaştırmaya ilişkin olasılık değeri (p) vermektedir. Program bu değeri p=1 ya da p=0 olarak vermektedir. p=0 değerinin elde edilmesi, modellerin uyum düzeyleri arasındaki farkın istatistiksel olarak manidar olduğuna, p=1 olması ise farklılığın manidar olmadığına işaret etmektedir. -2loglikelihood değerlerine ilişkin olasılık değerleri incelendiğinde 1PLM’in veriye uyum düzeyi ile 2PLM’in uyum düzeyi arasında ve 2PLM’in veriye uyum düzeyi ile 3PLM’in uyum düzeyi arasında manidar farklılıkların bulunduğu görülmektedir. Farklılığın yönünü belirlemek için -2loglikelihood değerleri doğrudan incelendiğinde 3PLM’in, 1PLM ve 2PLM’ye göre daha iyi uyum gösterdiği sonucuna varılmaktadır. Ancak STM ile iki kategorili MTK modellerine ilişkin veri setleri farklılaştığından bu modeller arasında varyans analizine dayalı ikili karşılaştırma yapılamamaktadır.

Model bazında uyumun yanı sıra, madde bazında da uyum istatistikleri incelenmiştir. Her bir modele ilişkin hesaplanan madde uyum $S-X^2$ istatistikleri Tablo 8’de verilmiştir:

Tablo 8. Madde Uyumuna İlişkin Hesaplanan $S-X^2$ İstatistikleri

	1PLM	2PLM	3PLM	STM
m1	0,0099	0,2925	0,1575	0,6992
m2	0	0,0395	0,7926	0,2819
m3	0	0,7971	0,9474	0,9356
m4	0,1113	0,1024	0,1254	0,1148
m5	0	0,1786	0,5943	0,0706
m6	0,0972	0,4014	0,3103	0,0333
m7	0,0038	0,4081	0,2424	0,0292
m8	0,0904	0,0963	0,1011	0,1531
m9	0,0516	0,9618	0,905	0,0149
m10	0	0,0144	0,0348	0,1528
m11	0,0006	0,0375	0,4227	0,0058
m12	0,0007	0,0187	0,2197	0
m13	0,0001	0,0007	0,3034	0,0592
m14	0	0,1232	0,1824	0,0195
m15	0,2323	0,5297	0,7054	0,107
m16	0	0,0365	0,0311	0,5488
m17	0,0003	0,109	0,1957	0,0935
m18	0,0002	0,0473	0,0235	0,0102
m19	0,0112	0,282	0,1117	0,3778
m20	0	0,0001	0,0159	0,6384

Not: Anlamlı olan p değerleri kalın yazılmıştır.

Tablo 8’deki değerler incelendiğinde 3PLM altında bütün maddelerin veriye uyum sergilediği, 2PLM ve STM altında 18 maddenin veriye uyum sergilediği, 1PLM altında ise sadece 6 maddenin veriye uyum sergilediği gözlenmiştir. Uyum sergilemeyen maddelerin yüzdelere bakıldığında; 2PLM ve STM altında maddelerin %10’unun uyum sergilemediği, 1PLM altında maddelerin % 60’ının uyum sergilemediği görülmektedir. 1PLM ile 2PLM karşılaştırılacak olursa 1PLM altında 6 madde uyum sergilerken 2PLM altında bu 6

maddenin tamamını kapsayan 18 maddenin uyum sergilediği gözlenmiştir. 2PLM ve STM altında uyum sergilemeyen madde sayısının eşit olması dikkat çekmektedir. Bu maddelerden 16 tanesi ortaktır ancak 2 tanesi modellere göre farklılık göstermektedir. Detaylı olarak ele alınacak olursa, 2PLM altında 13. ve 14.maddelerin, STM altında ise 11. ve 12.maddelerin uyum sergilemeyen maddeler olduğu gözlenmiştir.

Madde bazındaki uyum istatistikleri ve genel model uyum istatistikleri birlikte değerlendirildiğinde, 3PLM'nin hem genel model uyumu hem de madde bazında veriye en iyi uyum sergileyen model olduğu sonucuna varılmıştır.

Model uyumu değerlendirildikten sonra "1PLM, 2PLM, 3PLM ve STM'e göre madde parametrelerinin kestiriminin doğruluğu farklılaşmakta mıdır?" araştırma sorusuna yanıt vermek amacıyla 1PLM, 2PLM ve 3PLM altında madde parametreleri kestirilmiştir. Bu parametre kestirimlerine ilişkin standart hata değerleri hesaplanıp, bunların aritmetik ortalaması elde edilmiştir. Tablo 9'da 1PLM, 2PLM ve 3PLM'ye ilişkin madde parametre kestirimleri, madde parametre kestirimlerinin standart hatası ve her bir parametre kestiriminin standart hata değerlerinin ortalaması yer almaktadır.

Tablo 9 incelendiğinde, 1PLM altında kestirilen b parametrelerinin, $b_i=-0,83$ ile $b_i=1,59$ arasında değiştiği ve bu parametrenin standart hatalarının ortalamasının $\bar{X}_{SHb}=0,062$ olduğu görülmektedir. 2PLM'ye ilişkin b parametreleri $b_i=-0,58$ ile $b_i=2$ arasında değişmektedir ve bu parametrelerin kestirimine ilişkin standart hata değerlerinin ortalaması $\bar{X}_{SHb}=0,049$ 'dur. Bu model altında kestirilen a parametresinin $a_i=0,46$ ile $a_i=2.00$ arasında değişim gösterdiği ve bu parametrenin kestirimine ilişkin standart hata değerlerinin ortalamasının $\bar{X}_{SHa}=0,084$ olduğu görülmektedir. 3PLM altında; kestirilen b parametrelerinin $b_i=-0,07$ ile $b_i=2$ arasında değiştiği ve bu parametrenin standart hata değerlerinin ortalamasının $\bar{X}_{SHb}=0,056$ olduğu, a parametresinin $a_i=0,46$ ile $a_i=2.00$ arasında değişim gösterdiği, ve bu parametrenin kestirimine ilişkin standart hata değerlerinin ortalamasının $\bar{X}_{SHa}=0,129$ olduğu ve c parametresinin $c_i=0$ ile $c_i=0,29$ arasında değişim gösterdiği ve bu parametrenin kestirimine ilişkin standart hata değerlerinin ortalamasının $\bar{X}_{SHc}=0,013$ olduğu görülmektedir.

Tablo 9. İki Kategorili Puanlanan MTK Modelleri Altında Kestirilen Madde Parametreleri ve Madde Parametre Kestirimlerine İlişkin Standart Hata (SH) Değerleri

	1PLM		2PLM				3PLM					
	b	SH	b	SH	a	SH	b	SH	a	SH	c	SH
m1	0,08	0,06	0,06	0,05	1,21	0,07	0,76	0,06	2	0,17	0,25	0,01
m2	0,24	0,06	0,05	0,03	2	0,1	0,18	0,04	2	0,11	0,04	0,01
m3	0,1	0,06	-0,03	0,03	2	0,1	0,09	0,04	2	0,11	0,04	0,01
m4	-0,83	0,06	-0,58	0,04	1,62	0,1	-0,05	0,05	2	0,16	0,29	0,02
m5	0,65	0,06	0,56	0,06	1,02	0,06	1,2	0,06	2	0,19	0,21	0,01
m6	0,35	0,06	0,16	0,04	1,75	0,09	0,53	0,05	2	0,13	0,13	0,01
m7	-0,1	0,06	-0,08	0,05	1,12	0,07	0,76	0,06	2	0,18	0,3	0,02
m8	0,45	0,06	0,2	0,04	1,88	0,09	0,47	0,04	2	0,12	0,08	0,01
m9	0,57	0,06	0,27	0,04	1,96	0,09	0,52	0,04	2	0,12	0,07	0,01
m10	1,59	0,07	2	0,12	0,52	0,03	2	0,12	0,5	0,03	0	0,01
m11	0,28	0,06	0,08	0,03	2	0,1	0,42	0,05	2	0,13	0,11	0,01
m12	0,24	0,06	0,08	0,04	1,81	0,09	0,49	0,05	2	0,14	0,15	0,01
m13	0,38	0,06	0,17	0,04	1,76	0,09	0,58	0,05	2	0,13	0,13	0,01
m14	0,57	0,06	0,49	0,06	1,04	0,06	1,14	0,06	2	0,18	0,21	0,01
m15	-0,4	0,06	-0,31	0,04	1,7	0,1	0,14	0,05	2	0,15	0,22	0,02
m16	-0,38	0,06	-0,31	0,03	2	0,11	-0,07	0,04	2	0,13	0,13	0,02
m17	0,12	0,06	-0,01	0,03	1,93	0,1	0,36	0,05	2	0,13	0,14	0,01
m18	0,87	0,07	0,45	0,04	2	0,09	0,67	0,04	2	0,11	0,04	0,01
m19	-0,17	0,06	-0,18	0,03	1,9	0,11	0,11	0,04	2	0,13	0,13	0,02
m20	1,24	0,07	2	0,13	0,46	0,03	2	0,13	0,46	0,03	0	0,02
SH _{ort}		0,062		0,049		0,084		0,056		0,129		0,013

Modellere ilişkin parametrelerin kestirimlerinin doğruluğu karşılaştırmalı olarak incelendiğinde b_i parametresinin en az hatayla, sırasıyla, 2PLM, 3PLM ve 1PLM altında kestirildiği görülmektedir. a_i parametresinin en az hatayla kestirildiği modeller sırasıyla 2PLM ve 3PLM şeklindedir. Ayrıca maddelerin güçlük ve ayırt edicilik düzeyleri her bir model altında ele alınacak olursa, genel olarak 1PLM altında maddelerin orta güçlükte olduğu ancak 3PLM’de madde güçlük düzeyinin yükseldiği görülmektedir. Ayırt edicilik parametresinin de benzer şekilde 1PLM’den 3PLM’ye doğru yükseldiği ve 3PLM’de maddelerin çoğunun maksimum ayırt edici düzeye ulaştığı görülmektedir. Bu durum c_i parametresinin bu veri seti için önemli derecede etkili olduğunu göstermektedir.

parametresi sadece 3PLM altında kestirildiğinden bu parametrenin model bazında karşılaştırması yapılamamaktadır. Ancak bu parametrenin standart hata ortalaması a ve b parametrelerinin standart hata ortalamalarıyla karşılaştırıldığında, c parametresinin a ve b parametrelerine göre oldukça düşük standart hata değerleri ile kestirildiği görülmektedir.

STM'de kestirilen parametreler ile iki kategorili MTK modelleri altında kestirilen parametreler farklılaşmaktadır. STM'de öncelikle zeta (ζ) ve lambda (λ) parametreleri kestirilmekte, daha sonra ise kestirilen bu parametrelere doğrusal dönüşüm uygulanarak parametrelerin pratik anlamlılığı sağlanmaktadır (Ostini ve Nering, 2006).

İki kategorili MTK modelleri ile çoklu puanlanan STM, kestirilen parametrelerin anlamlarının ve ele alınan kategori sayısının farklılaşmasından dolayı, iki kategorili MTK modelleri kendi içerisinde karşılaştırılmış olup, STM'ye ilişkin parametreler madde ve kategori bazında kendi içinde değerlendirilmiştir. STM altında her bir tepki kategorisi için ayrı ayrı zeta (ζ) ve lambda (λ) parametreleri kestirilmiştir. Zeta parametreleri ve bu parametrelerin kestirimine ilişkin standart hata değerleri, kategori bazında Tablo 10'da verilmiştir.

Elde edilen zeta parametre değerlerine göre, 1.kategorinin (doğru yanıt) popularitesinin en yüksek 4.maddede ($\zeta_1=1,59$ (f=944)), en düşük ise 20.maddede ($\zeta_1=0,14$ (f=426)) olduğu gözlenmiştir. Bu bulgulardan hareketle bu testteki en fazla doğru yanıtın 4.maddeye, en az doğru yanıtın ise 20.maddeye verildiği sonucuna ulaşılmaktadır. Bunun yanı sıra bu maddeler kendi içerisinde diğer kategorileriyle (çeldiricilerle) birlikte ele alındığında 4.madde için tepki kategorilerinin popularitesinin sırasıyla $\zeta_1 = 1,59$ (f=944), $\zeta_3 = -0,17$ (f=237), $\zeta_4 = -0,69$ (f=161) ve $\zeta_2 = -0,73$ (f=158) olduğu görülmektedir. Dolayısıyla işaretlenme popularitesi en yüksek olan kategorinin ζ_1 kategorisi (doğru yanıt), işaretlenme popularitesi en düşük olan kategorinin ise ζ_2 kategorisi olduğu sonucuna ulaşılmaktadır.

20.madde için kategorilerin popularitesinin sırasıyla $\zeta_1 = 0,14$ (f=426), $\zeta_3 = 0,06$ (f=382), $\zeta_2 = -0,02$ (f=356) ve $\zeta_4 = -0,17$ (f=336) olduğu, işaretlenme popularitesi en yüksek olan kategorinin ζ_1 kategorisi, işaretlenme popularitesi en düşük olan kategorinin ise ζ_4 kategorisi olduğu görülmektedir.

Tablo 10. STM altında kestirilen Zeta Parametreleri ve Bu parametre kestirimlerine İlişkin Standart Hata Değerleri

	ζ_1	SH1	ζ_2	SH2	ζ_3	SH3	ζ_4	SH4
m1	0,81	0,05	-0,26	0,06	-0,09	0,06	-0,46	0,07
m2	0,86	0,06	-0,48	0,09	-0,22	0,08	-0,17	0,08
m3	0,99	0,06	-0,22	0,08	-0,12	0,08	-0,65	0,1
m4	1,59	0,06	-0,73	0,11	-0,17	0,09	-0,69	0,11
m5	0,43	0,05	0,29	0,05	-0,31	0,06	-0,41	0,06
m6	0,65	0,05	0,08	0,06	-0,14	0,07	-0,59	0,08
m7	1,1	0,05	0,54	0,06	-0,92	0,1	-0,72	0,1
m8	0,59	0,05	0,03	0,06	-0,52	0,08	-0,1	0,07
m9	0,44	0,05	-0,17	0,06	-0,22	0,07	-0,05	0,06
m10	0,24	0,05	0,84	0,04	-0,52	0,07	-0,56	0,08
m11	0,72	0,05	-0,23	0,07	-0,04	0,07	-0,45	0,08
m12	0,77	0,05	0,11	0,06	-0,16	0,07	-0,72	0,09
m13	0,63	0,05	0,12	0,06	-0,44	0,07	-0,32	0,07
m14	0,45	0,05	0,12	0,05	-0,24	0,06	-0,33	0,06
m15	1,24	0,05	-0,34	0,08	-0,55	0,09	-0,35	0,08
m16	1,39	0,06	-0,53	0,1	-0,26	0,09	-0,61	0,1
m17	0,85	0,05	-0,49	0,08	-0,3	0,08	-0,06	0,07
m18	0,29	0,06	0,71	0,06	-0,74	0,09	-0,26	0,07
m19	1,14	0,06	0,04	0,07	-0,73	0,1	-0,44	0,09
m20	0,14	0,04	-0,02	0,05	0,06	0,05	-0,17	0,05

(Not: Zeta parametresi R studio yazılım ortamında kestirildiğinden 1.kategori en yüksek kategori olarak ele alınmaktadır. Dolayısıyla zeta parametresine ilişkin 1.kategori doğru yanıtı göstermektedir.)

Zeta parametrelerinin kestirimlerine ilişkin standart hata değerlerinin tüm maddeler için SH=0,04 ile SH=0,11 arasında değiştiği ve bu değerlerin düşük olduğu, dolayısıyla da zeta parametrelerinin çok az hata ile kestirilmiş olduğu görülmektedir. Ele alınan 4.madde ve 20.maddenin kategorilerine ilişkin standart hata değerleri incelendiğinde, 4.madde için kategorilerin standart hata sıralamasının SH1 = 0,06, SH3 = 0,09, SH4 = 0,11 ve SH2 = 0,11 şeklinde olduğu görülmektedir. Kategorilerin standart hata sıralaması 20.madde için ise SH1 = 0,04, SH3 = 0,05, SH2 = 0,05 ve SH4 = 0,05 seklindedir. Bu bulgular doğrultusunda genel

olarak bakıldığında popolaritesi en yüksek olan tepki kategorisi için standart hata değerinin daha düşük olduğu, popolarite değerinin düşmesiyle ise standart hata değerinin artış gösterdiği görülmektedir. Zeta parametre kestirimlerine bakıldığında; 10.madde hariç bütün maddelerde birinci kategori (doğru yanıt) için bütün değerlerin söz konusu maddenin diğer kategorilerine (çeldiricilere) göre en yüksek olduğu görülmektedir. Ancak 10.maddede birinci kategori doğru yanıt olmasına rağmen popolaritesi ($\zeta_1 = 0,24$) ikinci kategoriden ($\zeta_2 = 0,84$) oldukça düşüktür. Bu bulgu, 10.maddede bu çeldiricinin doğru yanıtta daha fazla işaretlendiğini göstermektedir. Bu madde için popolaritesi yüksek olan ikinci kategorinin (çeldiricinin) standart hata değeri de ($SH=0,04$) birinci kategoriye (doğru yanıt; $SH=0,05$) göre düşüktür. Bu durum söz konusu maddede çeldiricinin daha fazla işaretlendiğini göstermektedir. Genel olarak kategorilerin popolaritesi ve standart hataları eş zamanlı olarak değerlendirildiğinde, popolaritesi en yüksek olan çeldiricilerin işaretlenmesinin diğer çeldiricilerin işaretlenmesine oranla daha düşük standart hata ile sonuçlandığı, dolayısıyla çeldiricilerin doğru yanıtta yakınlık derecesinin artmasının ölçmenin doğruluğunu arttırabileceği sonucuna ulaşılmaktadır.

STM altında kestirilen lambda parametreleri ve bu parametrelerin kestirimine ilişkin standart hata değerleri Tablo 11’de yer almaktadır. Lambda parametresi doğrusal dönüşüme göre $\lambda=a$ olduğu için, doğrudan ayırt edicilik parametresi olarak yorumlanmıştır.

Tablo 11’den 1.kategoriye (doğru yanıt) ilişkin lambda değerleri incelendiğinde en ayırt edici maddenin, 2.madde ($\lambda_1=2.01$) olduğu görülmektedir. Bu maddenin lambda değerleri sırasıyla, $\lambda_1 = 2,01$, $\lambda_2 = -0,37$, $\lambda_4 = -0,74$ ve $\lambda_3 = -0,9$ şeklinde olup en yakın çeldiriciden en uzak çeldiriciye doğru maddenin ters yöndeki ayırt ediciliğinin arttığı görülmektedir. Birinci kategoriye (doğru yanıt) ilişkin lambda değeri en düşük yani en az ayırt eden madde ise 20.madde olarak görülmektedir. Bu maddenin lambda değerleri sırasıyla $\lambda_1 = 0,33$, $\lambda_2 = 0,18$, $\lambda_3 = -0,06$ ve $\lambda_4 = -0,45$ şeklindedir. Ancak bu maddede 2.kategorinin de ayırt ediciliği pozitifdir, dolayısıyla başarı düzeyi yüksek olan öğrencilerde bu çeldiriciyi işaretleme eğilimindedir. Ele alınan 2.madde ve 20.maddeye ilişkin standart hata değerleri incelendiğinde, lambda parametrelerine ilişkin standart hata değerlerinin 2.madde için sırasıyla $SH_2 = 0,12$ ve $SH_1 = SH_3 = SH_4 = 0,11$ şeklinde olduğu görülmektedir. Lambda parametre kestirimlerinin standart hatası 20.madde için ise $SH_4 = 0,06$, $SH_2 = SH_3 = 0,05$ ve $SH_1 = 0,04$ şeklindedir. Bu maddelere ilişkin standart hata değerlerine bakıldığında ayırt ediciliği yüksek olan kategoriden ayırt ediciliği düşük olan kategoriye doğru standart hata

değerlerinin arttığı, yani kestiriminin doğruluğunun yüksek ayırt edicilikten düşük ayırt ediciliğe doğru azaldığı gözlenmektedir. Lambda parametrelerinin kestirimlerine ilişkin standart hata değerlerinin tüm maddeler için ise SH= 0,04 ile SH=0,13 arasında değiştiği ve bu değerlerin düşük olduğu, dolayısıyla da lambda parametrelerinin çok az hata ile kestirilmiş olduğu görülmektedir. Lambda değerleri genel olarak ele alındığında, 10.madde hariç bütün maddelerde 1.kategori (doğru yanıt) ayırt edicilik değerlerinin pozitif ve diğer kategorilere göre en yüksek olduğu, bu doğrultuda standart hata değerlerinin de diğer kategorilere göre en düşük olduğu görülmektedir. Ancak 10.maddenin lambda değerleri sırasıyla $\lambda_2 = 0,73$, $\lambda_1 = 0,56$, $\lambda_3 = -0,49$ ve $\lambda_4 = -0,8$ şeklinde olup 2.kategorinin (çeldiricinin) ayırt ediciliği daha yüksektir. Maddeye ilişkin standart hata değerlerinin de SH4 = SH3 = 0,09 ve SH2 = SH1 = 0,06 olup ayırt ediciliği yüksek olan kategoriden ayırt ediciliği düşük olan kategoriye doğru standart hata değerlerinin arttığı, yani kestiriminin netliğinin yüksek ayırt edicilikten düşük ayırt ediciliğe doğru azaldığı burada da gözlenmektedir. Genel olarak bakıldığında, maddelerin çeldiricilerine ilişkin kestirilen ayırt edicilik parametrelerinin negatif değerler aldığı ancak beş maddenin (m1, m7, m10 ve m20) çeldiricilerine ilişkin kestirilen ayırt edicilik parametrelerinin pozitif değerler aldığı görülmektedir. Bu bulgular, bu 5 maddede başarı düzeyi yüksek olan öğrencilerin de bu çeldiricileri işaretleme eğiliminde olduklarını göstermektedir.

Tablo 11. Lamda Parametreleri ve Standart Hata Değerleri

	λ_1	SH1	λ_2	SH2	λ_3	SH3	λ_4	SH4
m1	0,98	0,06	0,2	0,07	-0,45	0,08	-0,72	0,09
m2	2,01	0,11	-0,37	0,12	-0,9	0,11	-0,74	0,11
m3	1,91	0,11	-0,51	0,11	-0,89	0,11	-0,51	0,13
m4	1,36	0,09	-0,54	0,13	-0,3	0,1	-0,51	0,13
m5	0,79	0,05	-0,23	0,06	-0,07	0,07	-0,49	0,08
m6	1,38	0,07	-0,15	0,08	-0,47	0,09	-0,75	0,1
m7	1,09	0,07	0,2	0,07	-0,41	0,12	-0,87	0,11
m8	1,54	0,08	0,01	0,08	-0,89	0,1	-0,66	0,09
m9	1,48	0,08	-0,33	0,08	-0,44	0,09	-0,71	0,08
m10	0,56	0,06	0,73	0,06	-0,49	0,09	-0,8	0,09
m11	1,54	0,08	-0,18	0,09	-0,56	0,09	-0,8	0,1
m12	1,44	0,08	-0,16	0,08	-0,54	0,09	-0,74	0,11
m13	1,4	0,07	-0,17	0,08	-0,5	0,1	-0,73	0,09
m14	0,79	0,05	-0,28	0,06	-0,23	0,07	-0,29	0,07
m15	1,36	0,08	-0,29	0,1	-0,61	0,11	-0,46	0,1
m16	1,78	0,1	-0,5	0,12	-0,67	0,11	-0,6	0,13
m17	1,46	0,08	-0,35	0,1	-0,72	0,1	-0,39	0,09
m18	1,61	0,08	-0,5	0,07	-0,61	0,11	-0,51	0,09
m19	1,54	0,09	-0,21	0,09	-0,52	0,12	-0,81	0,11
m20	0,33	0,04	0,18	0,05	-0,06	0,05	-0,45	0,06

(Not: Lambda parametresi R studio yazılım ortamında kestirildiğinden 1.kategori en yüksek kategori olarak ele alınmaktadır. Dolayısıyla lambda parametresine ilişkin 1.kategori doğru yanıtı göstermektedir.)

Lambda ve zeta parametrelerine doğrusal dönüşümler uygulanarak bu parametreler MTK'nın bilindik a ve b parametrelerine dönüştürülmektedir. Ancak, Lamda parametresi maddelerin kategorilerinin eğimine işaret etmekte olup doğrudan ayırt edicilik parametresi olarak yorumlanabilmektedir. Zeta parametresine ise doğrusal dönüşüm uygulandıktan sonra, pratikte madde yer parametresi (item location) olarak yorumlanmaktadır. Bu dönüşüm sonucu elde edilen a ve b parametreleri Tablo 12'de yer almaktadır:

Tablo 12. STM'ye İlişkin a ve b Parametreleri

	a1	a2	a3	a4	b1	b2	b3	b4
M1	-0,72	-0,46	0,19	0,98	-0,4	-0,05	-0,28	0,73
M2	-0,73	-0,9	-0,37	2	-0,1	-0,14	-0,45	0,68
M3	-0,5	-0,9	-0,52	1,92	-0,61	-0,04	-0,17	0,82
M4	-0,5	-0,31	-0,54	1,35	-0,65	-0,14	-0,68	1,47
M5	-0,49	-0,07	-0,23	0,8	-0,37	-0,31	0,31	0,36
M6	-0,75	-0,48	-0,16	1,39	-0,53	-0,09	0,09	0,53
M7	-0,87	-0,43	0,2	1,09	-0,64	-0,88	0,52	1
M8	-0,66	-0,89	0	1,55	-0,04	-0,44	0,03	0,46
M9	-0,71	-0,45	-0,33	1,49	0,01	-0,19	-0,14	0,32
M10	-0,81	-0,49	0,74	0,56	-0,49	-0,48	0,78	0,2
M11	-0,8	-0,56	-0,18	1,54	-0,38	0,01	-0,22	0,59
M12	-0,75	-0,54	-0,16	1,45	-0,65	-0,11	0,12	0,64
M13	-0,73	-0,5	-0,18	1,41	-0,25	-0,4	0,14	0,51
M14	-0,28	-0,23	-0,29	0,8	-0,31	-0,22	0,15	0,38
M15	-0,46	-0,6	-0,3	1,36	-0,31	-0,49	-0,32	1,12
M16	-0,59	-0,68	-0,49	1,76	-0,55	-0,2	-0,47	1,22
M17	-0,39	-0,71	-0,35	1,46	-0,02	-0,24	-0,46	0,72
M18	-0,51	-0,62	-0,49	1,63	-0,22	-0,68	0,75	0,15
M19	-0,8	-0,52	-0,22	1,54	-0,37	-0,69	0,05	1,01
M20	-0,45	-0,06	0,19	0,33	-0,14	0,06	-0,03	0,11

(Not: a ve b parametreleri Multilog programında kestirildiğinden 4.kategori en yüksek kategori olarak ele alınmaktadır. Dolayısıyla bu parametrelere ilişkin 4.kategori doğru yanıtı göstermektedir.)

Lamba doğrudan a parametresine işaret ettiğinden (doğrusal dönüşüme göre $\lambda = a$ olduğu için) lambda parametresi, ayırt edicilik olarak yorumlanmıştır. Madde bazında karşılaştırma yapıldığında en yüksek ve en düşük ayırt ediciliğe sahip maddelerin iki ayrı programdan elde edilen sonuçlara göre aynı olduğu görülmüştür.

Zeta parametresinin pratik anlamlılığı için doğrusal dönüşümün gerçekleştirilmesi (eşitlik 13 ve eşitlik 14) gerekmektedir. Bu nedenle, zetanın doğrusal dönüşüm sonucu elde edilen b parametresi pratik anlamlılık açısından ele alınmıştır. Yer parametresi olarak ele alınan b

parametresine ilişkin kategori deęerleri doęru yanıt kategorisi (en yüksek kategori) bazında ele alınmıştır. Doęru yanıt kategorisinin b deęerleri incelendięinde bu parametrenin 4.maddede en yüksek, 20.maddede ise en düşük deęerleri aldığı görölmektedir. Bu bulgu zeta parametresinin popolaritesine ilişkin bulguyla aynıdır. Bunun yanı sıra 4.madde ve 20.maddenin kategorilerine ilişkin b parametre deęerleri sırasıyla b_2 , b_3 ve b_1 olup bu sıralama da zeta parametresinin kategori sıralamasıyla benzerlik göstermiştir. Parametre deęerleri karşılaştırıldığında b parametresinin zeta parametresine göre daha düşük deęerler aldığı, ancak madde ve kategorilerin sıralamasına göre dönüştürölmüş ve dönüştürölmemiş parametrelerin örtüştüğü görölmektedir.

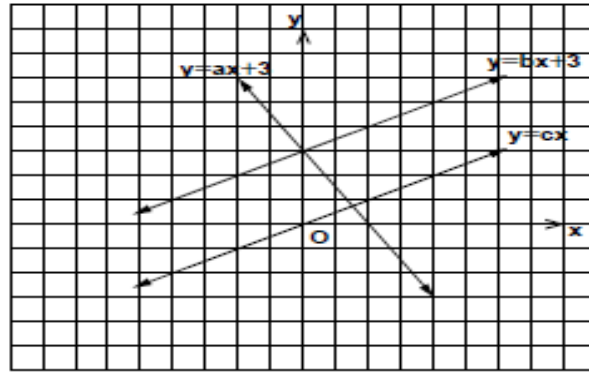
Çeldiriciler açısından a ve b parametreleri kategori karakteristik eęrileriyle birlikte deęerlendirildięinde genel olarak, çeldiricilerin düşük yetenek düzeyindeki bireyler tarafından işaretlendięi, ancak bazı maddelerin tepki kategorilerinin dięer maddelerdekine göre farklılık gösterdięi görölmektedir.

Yetenek ölçeęi boyunca (-3, 3) maddeler tepki kategorileriyle birlikte incelendięinde, bazı maddelerin tepki kategorilerinin işlevlerine ilişkin bir takım farklılıklar görölmektedir. Birinci maddenin tepki kategorileri incelendięinde (-3, -1) yetenek aralıęında bu maddenin, 3.kategori ve 4.kategori (doęru yanıt)'ye ilişkin işaretleme olasılıklarının ve bu kategorilerin eęrilerinin eęimlerinin birbirine yakın olduęu, dolayısıyla doęru yanıt ile söz konusu çeldiricinin bu yetenek aralıęında benzer şekilde çalıştığı görölmektedir. Dördüncü maddenin (-3, -1) yetenek aralıęında 1., 2. ve 3.kategorilerin yani bütün çeldiricilerin benzer eęime sahip olduęu ve işaretleme olasılıklarının da birbirine çok yakın olduęu görölmektedir. Dokuzuncu maddeye bakıldığında 2. ve 3.kategorilere ilişkin eęrilerin eęimlerinin bütün yetenek ölçeęi boyunca (-3, 3) benzer olduęu ve işaretlenme olasılıklarının da paralellik gösterdięi görölmektedir. Onuncu madde incelendięinde bu maddenin 3. ve 4. (doęru yanıt) tepki kategorilerinin işaretlenme olasılıklarının yetenek ölçeęinde üst yetenek düzeyine doęru gidildikçe arttığı, yani çeldirici olan 3.kategorinin de doęru yanıt gibi çalıştığı görölmektedir. On dördüncü maddeye bakıldığında 1. ve 2.kategorilerin; on altıncı maddede ise 1. ve 3.kategorilerin, bütün yetenek ölçeęi boyunca eęimlerdeki ve işaretlenme olasılıklarındaki deęişimin paralellik gösterdięi görölmektedir. Yirminci madde incelendięinde 3. ve 4.kategorilerin eęrilerinin eęimlerinin (-3, 0) aralıęında benzer olduęu, yetenek ölçeęinin üst kısmına doęru ise hem eęimin hem de işaretlenme olasılıklarının arttığı görölmektedir.

Bütün bu maddeler ve tepki kategorileri incelendiğinde 10. ve 20.maddelerin tepki kategorilerinin çeldirici ve doğru yanıt bağlamında işlevselliğinin uygun olmadığı görülmektedir. 10.madde ve çeldiricileri şöyledir: Aşağıdakilerden hangisi, $0,\overline{45}$ ile $0,\overline{452}$ devirli ondalık kesirleri arasında yer alır? A) 0,451 B) 0,453 C) 0,455 D) 0,457

Onuncu madde incelendiğinde, bu maddenin matematiğin alt öğrenme alanlarından “ölçme” ye yönelik öğrencinin hesaplama yapmasını gerektiren bir madde olduğu görülmektedir. Bu maddede A şıkkı 3.tepki kategorisi olarak, B şıkkı ise 4.tepki kategorisi yani doğru yanıt olarak analize alınmıştır. Analiz sonucuna göre 3.tepki kategorisi doğru yanıt gibi çalışmaktadır. Bu madde matematiksel hesaplama yönelik bir maddedir ve bu maddede öğrencilerin devirli ondalık kesri, ondalıklı kesre dönüştürerek bir aralık elde edip bu aralıkta hangi sayıların yer aldığını belirlemeleri beklenilmektedir. Ancak öğrencilerin bu maddede işlem hatası yapma olasılığı yüksek olduğundan aralığı belirlemede hata yapmış ve çeldiricilere daha fazla yönelmiş olabilirler. Dolayısıyla bu maddede çeldirici doğru yanıt gibi çalışmaktadır. Şekil 3’te 20.madde seçenekleriyle birlikte sunulmuştur:

20.



Yukarıda denklemleri ile verilen doğruların grafiklerine göre, aşağıdakilerden hangisi doğrudur?

- A) $a < b$ ve $b = c$
- B) $b = c$ ve $c < a$
- C) $c < b < a$
- D) $a < b < c$

Şekil 3. 2012-SBS Matematik alt testi A kitapçığı-20.madde

Şekil 3’te yer alan 20.maddenin A şıkkı 4.tepki kategorisi yani doğru yanıt olarak, B şıkkı ise 3.tepki kategorisi olarak analize alınmıştır. Analiz sonuçlarına göre 4.kategorinin işaretleme olasılığı 3.kategoriden daha yüksek olmasına rağmen genel olarak 3.kategoriyle benzer bir şekilde çalıştığı, kategori eğrilerinin eğimlerinin dolayısıyla ayırt edicilik değerlerinin de oldukça benzer olduğu görülmüştür. Bu kategorilerin içeriği ele alındığında her iki kategoride de ortak olarak yer alan “ b = c” ifadesinin bulunduğu görülmektedir. Analitik geometri konusuna yönelik bu maddede paralel ve dik kesişen doğrular olmak üzere madde iki farklı şekilde ele alınabileceğinden öğrenci sadece kısmi bilgisine dayalı olarak maddeyi yanıtlamış olabilir. Söz konusu kısmi bilgisine göre (b = c) ifadesinin doğruluğuna karar veren bir öğrenci 3. ve 4. kategoriye yöneleceğinden bu kategoriler benzer bir şekilde çalışmış olabilir.

Madde parametre kestirimleri incelendikten sonra, “1PLM, 2PLM, 3PLM ve STM altında yetenek parametresinin kestiriminin doğruluğu modeller arasında manidar bir şekilde farklılaşmakta mıdır?” araştırma sorusuna yanıt vermek amacıyla her bir model altında çalışma grubundaki bireyler için yetenek kestirimleri ve bu kestirimlere ilişkin standart hata değerleri hesaplanmıştır. Kestirimin doğruluğu modeller bazında karşılaştırmalı olarak ele alındığından, her bir model altında standart hata değerlerine ilişkin betimsel istatistikler elde edilmiş ve Tablo 13’te sunulmuştur:

Tablo 13.Yetenek Parametre Kestirimine İlişkin Standart Hatalara Yönelik Betimsel İstatistikler

Modeller	N	\bar{X}	SS	Çarpıklık	Basıklık
1PLM	1500	0,57	0,11	2,44	7,50
2PLM	1500	0,34	0,16	16,70	45,98
3PLM	1500	0,36	0,24	3,31	44,122
STM	1500	0,32	0,19	3,41	16,691

Tablo 13 incelendiğinde en yüksek hata ortalamasına sahip modelin 1PLM (0,57) olduğu, en düşük hata ortalamasına sahip modelin ise STM (0,32) olduğu, genel olarak modellere ilişkin standart hata ortalamalarının birbirinden farklılık gösterdiği görülmektedir. Modeller arası bu farklılığın manidar olup olmadığını test etmek amacıyla ilişkili örneklem için tek

faktörlü Varyans Analizi yapılmıştır. Varyans analizi sonucunda elde edilen değerler Tablo 14’te sunulmuştur.

Tablo 14. Farklı Modeller Altındaki Yetenek Kestirimlerinin Standart Hatalarına Ait ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Anlamlı Fark
Modeller	54,37	3	18,12	791,80*	STM<2PLM<3PLM<1PLM
Hata	102,93	4497	0,023		

*p<0,05

Analiz sonuçları, farklı modeller (1PLM, 2PLM, 3PLM ve STM) altındaki yetenek kestirimlerine ilişkin standart hata değerleri arasında manidar düzeyde bir farklılık olduğunu göstermektedir ($F(3, 4497) = 791,80; p<0,05$). Bu bulgu, bireylere ait yetenek kestiriminin doğruluğunun, modellere göre farklılaştığına işaret etmektedir. Gözlenen bu farklılığın, hangi modellere ilişkin standart hatalar arasında olduğunu araştırmak için yapılan ikili karşılaştırmalar, olası bütün model karşılaştırmaları arasında anlamlı bir fark olduğunu göstermektedir. Dolayısıyla Tablo 14’deki standart hata ortalamaları dikkate alındığında ve STM ile iki kategorili MTK modelleri karşılaştırıldığında, bu farklılığın STM’nin lehine olduğu gözlenmektedir. Ayrıca iki kategorili MTK modelleri arasında da manidar farklılığın olduğu, 2PLM’nin 1PLM ve 3PLM’ye göre daha doğru, 3PLM’nin ise 1PLM’ye göre daha doğru yetenek kestirimleri ile sonuçlandığı gözlenmektedir. Yetenek kestiriminin doğruluğu bağlamında da 2PLM ile STM tutarlı sonuçlar vermiştir. Standart hata ortalamaları anlamlı olarak farklılaşan model çiftlerine ilişkin etki büyüklükleri Tablo 15’te yer almaktadır:

Tablo 15. İkili Karşılaştırmalardaki Anlamlı Farklara İlişkin Etki Büyüklüğü η^2

	2PLM – 1PLM	2PLM – 3PLM	STM – 3PLM
η^2	0,752	0,005	0,019

Tablo 15 incelendiğinde 2PLM – 3PLM (0,005) ve STM – 3PLM (0,019) model çiftleri arasındaki karşılaştırmaya ilişkin hesaplanan etki büyüklüğünün oldukça düşük olduğu görülmektedir. Bu bulgu 2PLM ve 3PLM için modelin (model seçiminin), yetenek parametresinin kestirimlerinin doğruluğunu önemli ölçüde etkilemediğine işaret etmektedir.

1PLM – 2PLM (0,752) model çiftleri arasındaki karşılaştırmaya ilişkin hesaplanan etki büyüklüğünün ise oldukça yüksek olduğu görülmektedir. Bu bulgu ise 1PLM ve 2PLM için model seçiminin, yetenek parametre kestirimini önemli ölçüde etkilediğine, yani model türünün yetenek parametresi kestirimine ilişkin doğruluk derecesindeki farklılaşmanın %75'ini açıklayabildiğine işaret etmektedir.

Yetenek kestirimlerinin standart hata ortalamaları incelendiğinde 2PLM ile STM'nin benzer sonuçlar gösterdiği ve bu iki model altında 1PLM ve 3PLM'ye göre daha doğru yetenek kestirimleri elde edildiği görülmektedir. Bu bulgu, model-veri uyumu ve madde bilgi fonksiyonları bulguları ile paralellik göstermektedir.

Hangi model altında yetenek parametresinin daha doğru bir şekilde kestirildiği incelendikten sonra, "1PLM, 2PLM, 3PLM ve STM'ye dayalı olarak hesaplanan madde bilgi fonksiyonlarının sağladığı bilgi miktarı farklılaşmakta mıdır?" araştırma sorusuna yanıt vermek amacıyla her bir model altında madde bilgi fonksiyonları hesaplanmıştır. Tablo 16'da her bir model altında elde edilen madde bilgi fonksiyonlarının sağladığı bilgi miktarı yer almaktadır.

Madde bilgi fonksiyonlarına genel olarak bakıldığında test maddelerinin 2PLM ve STM altında sağladıkları bilgi miktarının birbirine yakın olduğu, ancak maddelerin en fazla 3PLM altında bilgi sağladıkları görülmektedir. Bu bulgu model-veri uyumunun test edilmesindeki sonuçlarla tutarlılık göstermektedir. Madde bazındaki model-veri uyumu bulgularına göre 2PLM ve STM altında elde edilen uyumlu madde sayıları eşittir. Dolayısıyla madde bilgi fonksiyonlarının bulguları ile madde bazındaki model veri uyumu bulguları birbirini desteklemektedir.

Modellerin sağladıkları bilgi açısından modeller arasında karşılaştırma yapıldığında; (a) sadece bir maddenin (m10) 1PLM altında en fazla bilgi sağladığı ancak bu model altında sağlanan bilgi miktarının (6,901), STM altında maddenin sağladığı bilgi miktarına (6,757) çok yakın olduğu, (b) en fazla 2PLM altında bilgi sağlayan hiçbir maddenin bulunmadığı, (c) 15 maddenin en fazla 3PLM altında bilgi sağladığı ve (d) iki maddenin de en fazla STM altında bilgi sağladığı görülmektedir.

Tablo 16. 1- PLM, 2PLM, 3PLM ve STM'ye Dayalı Olarak Hesaplanan Madde Bilgi Fonksiyonları

Maddeler	1PLM	2PLM	3PLM	STM
m1	7,128	5,958	8,96	7,572
m2	7,124	13,315	25,956	14,014
m3	7,129	12,878	18,154	13,466
m4	7,07	8,347	8,961	8,967
m5	7,099	4,799	7,372	5,177
m6	7,118	8,916	13,792	9,713
m7	7,129	5,425	6,595	8,742
m8	7,113	9,652	12,111	11,463
m9	7,101	10,023	14,436	10,37
m10	6,901	0,172	0,199	6,757
m11	7,123	10,355	21,757	10,924
m12	7,123	9,249	29,321	10,028
m13	7,119	9,095	38,555	9,934
m14	7,104	4,947	9,556	4,902
m15	7,115	8,695	11,154	9,17
m16	7,118	11,296	16,572	11,89
m17	7,126	9,921	19,999	10,133
m18	7,068	10,752	17,25	10,705
m19	7,125	9,756	11,069	10,912
m20	7	0,919	1,293	2,251

1PLM altında en fazla bilgi sağlayan ilk üç madde sırasıyla m3 (7,129), m7 (7,129) ve m1 (7,128)'dir. 2PLM altında en fazla bilgi sağlayan ilk üç madde sırasıyla m2 (13,315), m3 (12,878) ve m16 (11,296)'dir. 3PLM altında en fazla bilgi sağlayan ilk üç madde sırasıyla m13 (38,555), m12 (29,321) ve m2 (25,956)'dir. STM altında en fazla bilgi sağlayan ilk üç madde sırasıyla m2 (14,014), m3 (13,466) ve m16 (11,89)'dir. Karşılaştırmalı olarak en fazla bilgi sağlayan ilk üç madde incelendiğinde 2PLM ve STM'ye göre en fazla bilgi sağlayan üç maddenin (m2, m3, m16) ve bu maddelerin sıralamasının (m2 > m3 > m16) aynı olduğu gözlenmiştir. Ancak maddelerin sağladığı bilgi miktarı her üç madde için de STM altında

2PLM'ye göre daha fazladır. Bunun yanı sıra, en fazla bilgi sağlayan maddeler arasında yer alan m2, 2PLM ve STM'de birinci sıradayken, 3PLM'de üçüncü sırada yer almaktadır.

Bütün modellere göre madde bilgi fonksiyonları karşılaştırıldığında ise, en fazla bilgi miktarının 13.madde (38,555) ile 3PLM'de, en az bilgi miktarının ise 10.madde (0,172) ile 2PLM'de sağlandığı gözlenmektedir. Ayrıca 3PLM altında elde edilen bilgi miktarının m4, m7, m10 ve m20 dışındaki bütün maddelerin modeller arasındaki en yüksek değerlere ulaştığı gözlenmektedir. Madde bilgi fonksiyonlarına ilişkin grafiklerden de bu bulgular gözlenmektedir (Bakınız EK 10).

Madde bilgi fonksiyonlarının hesaplanmasının ardından, "1PLM, 2PLM, 3PLM ve STM'ye dayalı hesaplanan test bilgi fonksiyonlarının sağladığı bilgi miktarı düşük, orta ve yüksek yetenek düzeylerindeki bireyler için farklılaşmakta mıdır?" araştırma sorusuna yanıt vermek amacıyla her bir model altında ve üç farklı yetenek düzeyinde test bilgi fonksiyonları kestirilmiştir.

Tablo 17'de her bir model altında düşük, orta ve yüksek olmak üzere üç farklı yetenek düzeyinde test bilgi fonksiyonunun sağladığı bilgi miktarı yer almaktadır:

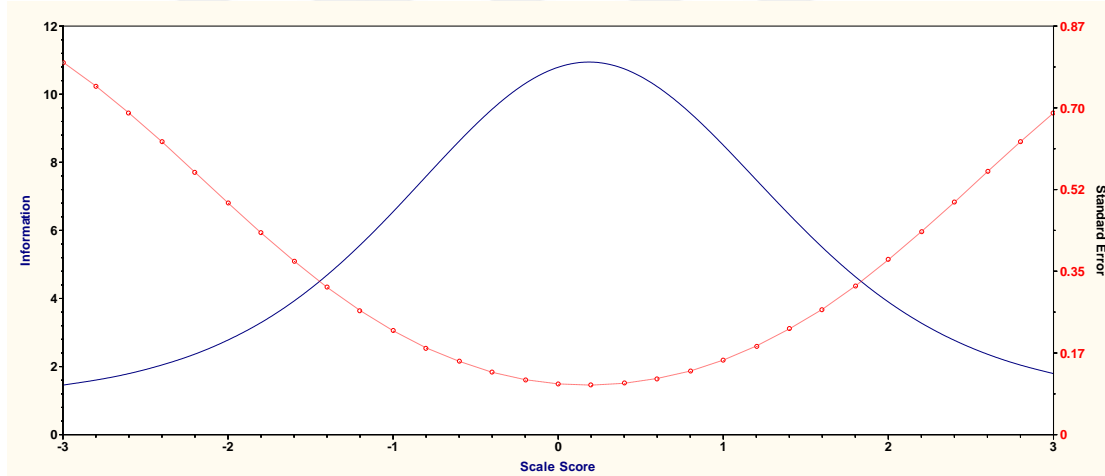
Tablo 17. Test Bilgi Fonksiyonları

Model/ θ ölçeği				
yetenek aralıkları	[-3, -1]	(-1, 1)	[1, 3]	[-3,3]
1PLM	29,479	103,238	40,2	172,917
2PLM	28,617	137,116	29,74	195,473
3PLM	10,338	290,793	22,928	324,059
STM	37,06	148,425	32,592	218,077

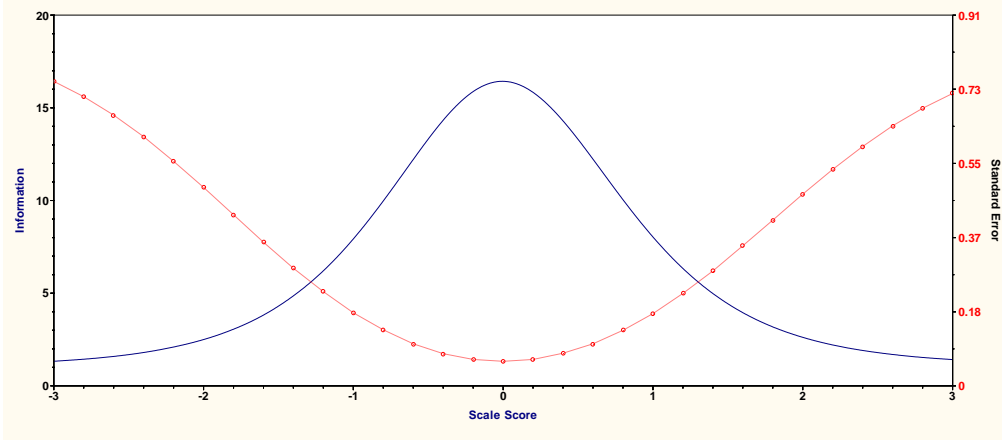
Tablo 17 incelendiğinde Matematik alt testinin tüm bu MTK modelleri altında yetenek kestirimi açısından en fazla bilgiyi orta yetenek düzeyindeki bireyler için sağladığı görülmektedir. Ancak testin yetenek kestirimi açısından en az bilgi sağladığı yetenek düzeylerine bakıldığında, iki kategorili MTK modelleri altında düşük yetenek düzeyindeki bireyler için en az bilgi sağlanırken, STM altında en az bilginin yüksek yetenek düzeyindeki bireyler için sağlandığı gözlenmiştir. Tüm yetenek ölçeği ($\theta=-3$, $\theta=+3$) boyunca sağladığı bilgi miktarına bakıldığında, testin yetenek kestirimi konusunda en fazla bilgiyi 3PLM altında, en az bilgiyi ise 1PLM altında sağladığı anlaşılmaktadır.

STM ile iki kategorili MTK modelleri karşılaştırıldığında, STM'nin orta yetenek ranjında, 3PLM'ye göre daha az bilgi sağladığı ancak, 1PLM ve 2PLM'den daha fazla bilgi sağladığı gözlenmiştir. Düşük yetenek ranjı ele alındığında, STM'nin 1, 2 ve 3PLM'den daha fazla bilgi sağladığı, yüksek yetenek ranjında ise STM'nin 2 ve 3PLM'ye göre daha fazla bilgi sağladığı ancak 1PLM'ye göre daha az bilgi sağladığı gözlenmiştir. Bütün bu bunlardan hareketle düşük yetenek ranjında kullanılabilir en iyi modelin STM, orta yetenek ranjında 3PLM ve yüksek yetenek ranjında ise 1PLM olduğu gözlenmiştir. Düşük yetenek ranjında STM'nin en fazla bilgi sağlıyor olmasının nedeni, düşük yetenek düzeyindeki bireylerin daha fazla çeldiricilere yöneliyor olmasından kaynaklanıyor olabilir.

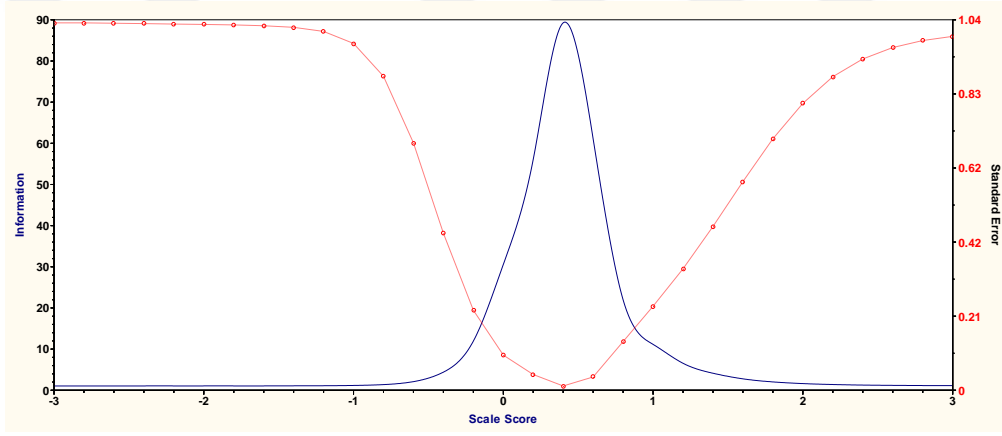
Aşağıda yer alan Şekil 4, Şekil 5, Şekil 6 ve Şekil 7'de her bir model altındaki test bilgi fonksiyonuna ilişkin grafikler yer almaktadır:



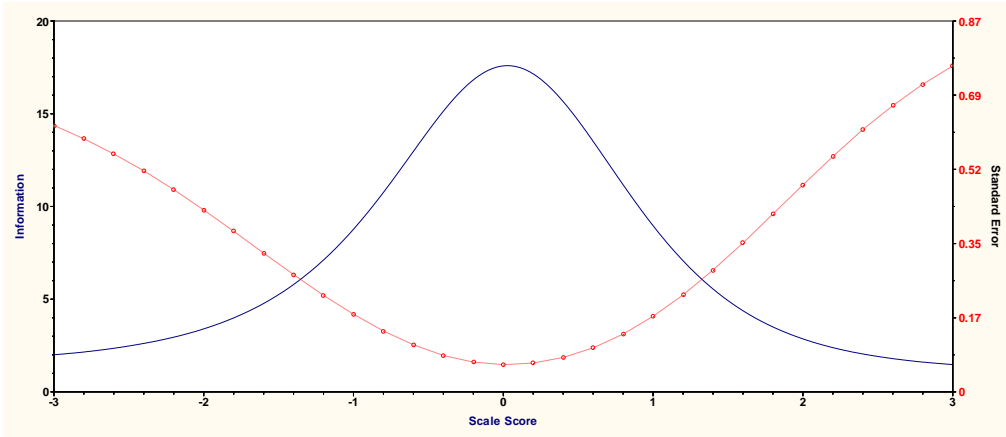
Şekil 4. 1PLM'ye dayalı test bilgi fonksiyonu grafiği



Şekil 5. 2PLM'ye dayalı test bilgi fonksiyonu grafiği



Şekil 6. 3PLM'ye dayalı test bilgi fonksiyonu grafiği



Şekil 7. STM'ye dayalı test bilgi fonksiyonu grafiği

Test bilgi fonksiyonlarına ait grafikler incelendiğinde her bir model altında hesaplanan fonksiyona ilişkin oluşturulan eğrinin, en fazla bilginin sağlandığı yetenek düzeyini veren, tepe noktasının -1 ile +1 aralığında yer aldığı gözlenmektedir. Ancak 3PLM altında elde edilen eğrinin tepe noktası, yetenek ölçeğinin daha sağına yaklaşmaktadır. Fonksiyon eğrilerinin tepe noktalarındaki değerler ve standart hataya ilişkin eğriler incelendiğinde en fazla bilginin sağlayan modelin 3PLM altında sağlandığı ve bilginin maksimum olduğu noktada standart hata değerinin de en düşük değeri aldığı görülmektedir.

Modellere ait eğrilerin 1PLM'den 3PLM'ye doğru giderek daha sivri bir dağılım gösterdiği ve aynı zamanda maksimum bilgi sağlanan yetenek aralığının da daraldığı gözlenmektedir. 3PLM altında sağlanan bilgi miktarının en yüksek ancak bu yetenek aralığının en dar olduğu açık bir şekilde gözlenmektedir. Bunun yanı sıra 1PLM altında elde edilen fonksiyon grafiği diğer modeller altında elde edilen fonksiyon grafiklerine göre oldukça basıktır ve sağlanan bilgi miktarı düşük olmasına rağmen testin yetenek ölçeği boyunca daha geniş bir aralıkta bilgi sağladığı görülmektedir.

Sonuç ve Tartışma

Bu çalışmada 1PLM, 2PLM 3PLM ve STM model-veri uyumu, madde ve yetenek parametre kestirimlerinin doğruluğu, yetenek kestiriminin kesinliği, madde ve test bilgi fonksiyonlarının sağladığı bilgi miktarı açısından karşılaştırmalı olarak incelenmiştir. Bu karşılaştırmalar yapılırken ilgilenilen nitelik açısından en iyi model belirlenmeye çalışılmıştır. Model-veri uyumu incelendiğinde en iyi uyum gösteren modellerin sırasıyla genel model uyumu açısından sırasıyla 3PLM, STM, 2PLM ve 1PLM olduğu; madde uyumu açısından ise 3PLM, STM ve 2PLM, 1PLM olduğu görülmektedir. Uyum istatistikleri, genel uyum açısından bütün modellerin veriye iyi uyum gösterdiğine işaret etmektedir. Ayrıca -2 loglikelihood değerlerine göre yapılan varyans analizinden elde edilen bulgulara dayalı olarak 3PLM'nin 1PLM ve 2PLM'den genel uyum bazında anlamlı olarak farklılaştığı ve daha iyi uyum gösterdiği sonucuna ulaşılmıştır. Bu bulgu Teker, Kelecioğlu ve Eroğlu (2013) ve Önder (2007)'in çalışmalarından elde edilen bulgular ile benzerlik göstermektedir. Bu çalışmalarda iki kategorili MTK modelleri -2loglikelihood değerlerine göre karşılaştırılmış ve 3PLM'in en iyi uyum sergileyen model olduğu gözlenmiştir. Ayrıca, Çelik (2001) çalışmasında 1PLM, 2PLM ve 3PLM'nin matematik alt testinden elde edilen

veriye uyumunu incelemiş ve elde ettiği χ^2 istatistiklerine göre en iyi uyum sergileyen modelin 3PLM olduğu sonucuna ulaşmıştır. Hem kullanılan alt testin türü hem de uyum gösteren model açısından sonuçların paralellik gösterdiği gözlenmektedir. Genel uyumun ardından, her bir model altında hesaplanan madde uyum istatistiklerine göre modellerin uyumu madde bazında değerlendirilmiştir. Bu istatistiklere göre, madde uyumunu parametre sayısı arttıkça artış göstermektedir. Buna paralel olarak en iyi uyumun 3PLM için sağlandığı, STM ve 2PLM altında 18 madde veriye uyum gösterirken, 1PLM altında sadece 6 maddenin uyum sergilediği gözlenmiştir. Bu bulgu Önder (2007)'in çalışmasından elde edilen bulgu ile benzerlik göstermektedir. Önder (2007) çalışmasında 1PLM'den 3PLM'ye doğru uyum sergilemeyen madde sayısının azaldığını gözlemlemiştir. Bu bulgulara göre, 1PLM, 2PLM ve 3PLM'ye ilişkin madde uyumu değerlendirildiğinde en iyi uyumun sırasıyla 3PLM ve 2PLM şeklinde olduğu, 1PLM'nin ise en az uyumu sergilediği sonucuna ulaşılmıştır. Daha detaylı ele alınacak olursa, 1PLM altında 6 madde uyum sergilerken, 2PLM altında bu 6 maddenin tamamını kapsayan 18 madde uyum göstermiştir. Bu durum söz konusu maddeler için ayırt edicilik parametrelerinin de ele alınmasının önemli olduğunu göstermektedir. 2PLM ve STM altında uyum sergileyen maddeler incelendiğinde 16 maddenin ortak olduğu, geriye kalan uyum sergileyen iki maddenin modellere göre farklılık gösterdiği gözlenmektedir. Bu durum, çoklu puanlama yapmanın madde uyumunu etkilediğini göstermektedir. Genel olarak uyum sergileyen madde sayıları değerlendirildiğinde, 2PLM ve STM'nin benzer sonuç sergilediği görülmektedir. 1PLM ve STM'ye ilişkin bulguların benzerlik göstermesinin nedeni olarak, söz konusu modellerde dikkate alınan parametre türünün benzer olması (a ve b parametreleri), ve parametre türünün de test sonuçlarında etkili bir faktör olması sunulabilir.

Yetenek kestirimlerinin doğruluğunun seçilen modele göre manidar olarak farklılaştığı görülmüştür. STM altında, 1PLM, 2PLM ve 3PLM'ye göre daha doğru yetenek kestirimlerinin elde edildiği görülmektedir. Bu bulgu Bock (1972)'in ve Ching-Fung (2002)'un çalışmasında elde ettiği bulgular ile benzerlik göstermektedir. Bock (1972) çalışmasında orta yetenek düzeyinin altındaki bireylerin yetenek kestiriminin doğruluğunun çok kategorili modellerde iki kategorili modellere göre oldukça önemli ölçüde artış gösterdiğini belirtmektedir. Ching-Fung (2002) ise çalışmasında iki kategorili MTK modelleriyle çoklu puanlanan MTK modellerini farklı kombinasyonlar altında ele almış ve ele aldığı bütün kombinasyonlarda çoklu puanlanan MTK modellerinin ikili puanlanan MTK

modellerine göre daha doğru yetenek kestirimleri ile sonuçlandığını gözlemlemiştir. Ayrıca DeAyala ve Koch (1987) çalışmasında STM altında ve 3PLM altında CAT uygulamalarıyla elde ettiği yetenek kestirimlerinin STM’de daha düşük standart hata ile sonuçlandığını gözlemlemiştir. Bu çalışmadan elde edilen bulgular da araştırma sonuçlarını desteklemektedir.

İki kategorili MTK modellerine ilişkin madde parametreleri incelendiğinde güçlük parametresinin en az hatayla kestirildiği modellerin sırasıyla 2PLM, 3PLM ve 1PLM olduğu; ayırt edicilik parametresinin en az hatayla kestirildiği modellerin sırasıyla 2PLM, 3PLM olduğu gözlenmiştir. Ortak olarak kestirilen parametrelerin yanı sıra sadece 3PLM altında kestirilen şans parametresi de oldukça düşük standart hata değerleri ile kestirilmiştir. Madde parametre kestirimlerine ilişkin standart hata değerlerinin ortalamalarına ilişkin bulgular genel olarak incelendiğinde en doğru parametre kestirimlerinin sırasıyla 2PLM, 3PLM ve 1PLM altında yapıldığı sonucuna ulaşılmaktadır. 3PLM altında şans parametresinin kestirimine ilişkin standart hata değerleri oldukça düşük olmasına rağmen, ayırt edicilik ve güçlük parametrelerinin kestirimine ilişkin standart hata değerleri yüksek çıkmıştır. Bu duruma sebep olarak madde parametrelerinin sınırlandırılması gösterilebilir. Madde parametreleri bu model altında serbestçe kestirildiğinde oldukça yüksek çıkmıştır. Pratik anlamlılık için bu parametreler, Hambleton ve Swaminathan (1991)’in ifade ettiği sınır değerlere göre sınırlandırılarak kestirimler yeniden yapılmış, ancak bu durum parametre kestirimlerinin standart hata değerlerini yükseltmiştir. Ayrıca ayırt ediciliği yüksek olan maddeler pozitif hata içermeye meyillidir ve hatanın belli bir kaynağı olmayıp random hata olabilir (Hambleton ve Jones, 2009). Bu çalışmada 3PLM altında kestirilen ayırt edicilik parametreleri de yüksek değerlere ulaştığından standart hata değerlerinin de yüksek olması beklenen bir durumdur.

Madde bilgi fonksiyonlarından elde edilen bulgular incelendiğinde, 3PLM altında maddelerin sağladığı bilgi miktarının en yüksek değerlere ulaştığı görülmüştür. 2PLM ve STM altında maddelerin sağladığı bilgi miktarı ele alındığında ise genel olarak maddelerin STM’de 2PLM’ye göre daha yüksek bilgi miktarı ile sonuçlandığı, ancak sağlanan bilgi miktarı açısından maddelerin sıralamasının benzer olduğu gözlenmiştir. 3PLM altında kestirilen madde ayırt edicilik parametrelerinin yüksek olması, maddenin daha bilgi verici olmasını sağlamaktadır (Gierl, Henderson, Jodoin ve Klinger, 2001). Dolayısıyla bu model altında kestirilen ayırt edicilik parametreleri incelendiğinde, testin 3PLM altında en fazla

bilgiyi sağlaması beklenen bir durumdur. 2PLM ve STM altında elde edilen sonuçların benzerliği ise madde parametrizasyonu (item parametrization)'nun yani modellerde ele alınan parametre sayısı ve çeşidinin benzerlik göstermesinin bir sonucu olabilir. Çünkü STM 2PLM'nin bir uzantısıdır ve her iki modelde ele alınan parametre türü (a ve b parametreleri) benzerlik göstermektedir (Baker, 1992). Modellerde ele alınan parametreler analiz sonuçlarını etkilediğinden, 2PLM ile STM'ye ilişkin bu bulgular da benzerlik göstermektedir.

Test bilgi fonksiyonlarından elde edilen bulgulara göre en fazla bilgiyi sağlayan modeller sırasıyla 3PLM, STM, 2PLM ve 1PLM şeklindedir. Standart hata değerlerinin düşük olması, testin sağladığı bilgi miktarının artmasını sağlar (Hambleton ve Swaminathan, 1985, s.25-50). Test bilgi fonksiyonlarını sağladığı bilgi miktarı ve yetenek kestirimlerinin doğruluğu birlikte ele alındığında, STM'nin 1PLM ve 2PLM'ye göre daha yüksek test bilgi fonksiyonları ve daha doğru yetenek kestirimleri ile sonuçlandığı görülmüştür. Ancak, STM 3PLM'ye göre daha doğru yetenek kestirimleri ile sonuçlanmasına rağmen STM'nin sağladığı bilgi miktarı 3PLM'den düşüktür. Bu bulgu madde ayırt edicilik parametrelerine dayandırılarak açıklanabilir. 3PLM altında kestirilen ayırt edicilik parametreleri incelendiğinde, yaklaşık bütün maddeler için bu parametre kestirimlerinin maksimum değerlerine ulaştığı dolayısıyla maddelerin sağladığı bilgi miktarının da oldukça yüksek değerlere ulaştığı gözlenmektedir. Hambleton ve Jones (2009) tarafından yapılan çalışmada yüksek ayırt ediciliğe sahip maddeler için bilgi miktarının olduğundan daha fazla kestirildiği sonucuna ulaşılmıştır. 3PLM'nin yetenek kestiriminin doğruluğunun STM'ye göre daha düşük olmasına rağmen, 3PLM'nin daha yüksek test bilgi fonksiyonları ile sonuçlanmasının nedeni olarak ayırt edicilik parametre değerleri sunulabilir. Ayrıca, bu bulgu ayırt edicilik parametresi ile yetenek ölçeği üzerinde ele alınan bölge arasındaki ilişkisi ile de açıklanabilir. Ayırt edicilik parametresi arttıkça maddelerin bireyleri ayırma gücü de artmaktadır. Ancak, ayırt edicilik parametresinin kestiriminin yüksek olması ilgili maddenin yetenek ölçeği üzerinde ayırt ettiği yetenek ranjını daraltmaktadır. (Harris, 1989). Embretson ve Reise (2000)'e göre yüksek ayırt ediciliğe sahip madde tepe yapmış bir bilgi eğrisi sunar ve bu tür maddeler dar bir yetenek aralığında çok bilgi sağlar. Tersine düşük ayırt ediciliğe sahip maddeler daha düz ve yayılmış bir şekilde bilgi verir. Test bilgi fonksiyonlarına ait grafikler incelendiğinde 1PLM'den 3PLM'ye doğru fonksiyon eğrisinin daraldığı, özellikle 3PLM'ye ait grafiğin sağladığı bilgi miktarı oldukça yüksek olmasına rağmen bilgi sağlanan

yetenek aralığının diğer modellere göre oldukça dar olduğu görülmektedir. Yani testin bilgi sağladığı yetenek aralığı daralmaktadır. Ayrıca yapılan çalışmalar (DeAyala, 1989; Borgatto, Azevedo, Pinherio ve Andrade, 2015) kuramsal bilgilerle doğrudan örtüşmemesine rağmen, test bilgi fonksiyonunun yetenek ölçeği üzerinde düşük olduğu durumda daha doğru yetenek kestirimlerin yapıldığını göstermektedir. Borgatto, Azevedo, Pinherio ve Andrade (2015) çalışmasında yetenek kestirimi yapmış ve elde ettiği sonuçlara göre test bilgi fonksiyonunun sağladığı bilgi miktarının en düşük olduğu yetenek ranjında daha doğru kestirimler yapıldığını gözlemlemiştir. DeAyala (1989) çalışmasında STM altında hesapladığı test bilgi fonksiyonuna ilişkin grafiğin pozitif kayıslı olduğunu, yani üst yetenek düzeyindeki bireyler için daha fazla bilgi sağlandığı sonucuna ulaşmıştır. STM ve 3PLM altında hesaplanan test bilgi fonksiyonları ele alındığında, 3PLM'in orta yetenek düzeyinde sağladığı bilgi miktarının STM'de sağlanan bilgi miktarının benzer olduğu gözlenmiştir. Dolayısıyla bu yetenek düzeyinde STM 3PLM'ye göre daha doğru kestirimler yapacağından yetenek kestiriminin doğruluğunu etkilemiş olabilir. Bu bulgular Borgatto, Azevedo, Pinherio ve Andrade (2015) ve DeAyala (1989)'nın çalışmalarından elde edilen bulgular ile de desteklenmektedir.

Öneriler

Bu çalışmaya katılan öğrenciler ilgili yetenek düzeyleri bakımından heterojen bir yapı sergilemektedir. Araştırmacılara, yetenek düzeyi bakımından homojen gruplar ile STM'nin etkililiğini incelemesi, bu incelemede kullanılacak farklı yetenek gruplarının oluşturulması için ise bireylerin yeteneklerinin önceki ölçümlere dayandırılarak yapılması ve STM'nin bu tür ölçümler için test edilmesi önerilebilir. Çünkü bireylerin yetenek düzeyinin, test uygulanmadan önce belli ölçümlere dayalı olarak belirlenerek grupların oluşturulması kestirimlerin daha sağlam (rough) olmasını sağlamaktadır (de la Torre, 2010). STM'nin farklı yetenek düzeyleri için çoklu puanlanan MTK modelleri ile de karşılaştırmalı olarak incelenmesi önerilmektedir.

STM, maddeye ilişkin kategorileri ele aldığından değişen çeldirici fonksiyonunu (differential distractor function) ortaya çıkarmak amacıyla kullanılabilir (Penfield, 2008). Bu doğrultuda maddenin kategori bazında analizleri yapılacağından daha işlevsel

maddeler geliştirilebilecektir. Madde havuzu oluşturmak amacıyla başarı testi maddelerinin STM ile analiz edilmesi önerilmektedir.

STM'ye ilişkin test bilgi fonksiyonunun sağladığı bilgi miktarı iki kategorili MTK modellerine ilişkin test bilgi fonksiyonlarının sağladıkları bilgi miktarı ile düşük, orta ve yüksek olmak üzere üç farklı yetenek düzeyinde karşılaştırmalı olarak incelendiğinde; düşük yetenek düzeyindeki bireyler için en fazla bilgi sağlayan modelin STM olduğu, diğer yetenek düzeylerinde ise iki kategorili modellerden 3PLM'nin daha fazla bilgi sağladığı gözlenmiştir. Bu nedenle STM'nin düşük yetenek düzeyindeki bireyler için kullanılması önerilmektedir.

Bu çalışmada madde parametreleri marjinal maksimum olabilirlik kestirim yöntemi, yetenek parametreleri ise maksimum olabilirlik kestirim yöntemi kullanılarak kestirilmiştir. Farklı parametre kestirim yöntemleri altında madde nitelikleri ve öğrencilerin yetenek düzeyleri incelenebilir ve farklı kestirim yöntemleri altında modellerin etkililiği karşılaştırılabilir.

Bireyler hakkında birçok alanda önemli kararlar vermeye yönelik yapılan ulusal ve uluslararası geniş ölçekli sınavlarda, objektif puanlama yapılmasını sağlayan çoktan seçmeli test maddelerinin kullanımı önemli bir yer tutmaktadır. Ancak, çoktan seçmeli test maddelerindeki çeldiricilerin puanlamaya dahil edilmemesi ve analizde hepsinin tek bir kategoride ele alınması, bilgi kaybına neden olabileceği yönünde eleştirilmektedir. STM altında çeldiriciler de dikkate alınarak kestirimler yapıldığından, bireylerin sahip olduğu kısmi bilgi de dikkate alınmaktadır. Bu nedenle, STM birçok alanda sıklıkla kullanılan çoktan seçmeli test maddelerinin değerlendirilmesinde daha ayrıntılı bilgiler sağlayacağından uygulamaya katkısı olacağı düşünülmektedir. Ayrıca, STM altında maddelerin tepki kategorileri yetenek ölçeği boyunca incelenmekte ve çeldiricilerin işlevselliği ortaya konulabilmektedir. Bu doğrultuda çoktan seçmeli test maddelerinde önemli bir husus olan çeldiricilerin etkililiği incelenerek maddelerin her yönüyle test edilip geliştirilmesine katkıda bulunacaktır. Böylelikle, STM'nin uygulamada madde bankası oluşturma çalışmalarına katkı sağlaması beklenilmektedir.

KAYNAKLAR

- Adedoyin, O. O. (2010). Investigating the Invariance of Person Parameter Estimates Based on Classical Test and Item Response Theories. *Int J Edu Sci*, 2(2), 107-113.
- Alberto Maydeu-Olivares & Harry Joe (2014) Assessing Approximate Fit in Categorical Data Analysis. *Multivariate Behavioral Research*, 49(4), 305-328
- Baker, E. L. (1988). *Can We Fairly Measure the Quality of Education*. CSE Technical Report. U.S.A.
- Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Technique*. New York: Marcel Dekker.
- Baker, F. B. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Baykul, Y. (1979) *Örtük özellikler ve klasik test kuramları üzerine bir karşılaştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem.
- Berberoğlu, G. (1988). *Seçme amacıyla kullanılan testlerde Rasch modelinin katkıları*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29–51.

- Bock, R. D. (1997a). The Nominal Categories Model. van der Linden, W. J., & Hambleton, R. K. (Ed.), *Handbook of Modern Item Response Theory* içinde (s. 33-49). New York: Springer.
- Bock, R. D. (1997b). A brief history of item response theory. *Educational Measurement: Issues and Practice*, 16(4), 21–33.
- Borgatto, A. F., Azevedo, C. L. N., Pinheiro, A., & Andrade, D. F. (2015). Comparison of ability estimation methods using irt for test with different degrees of difficulty. *Communications in Statistics – Simulation and Computation*, 44(2), 474-488.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Brown, C. J., Templin, J., & Cohen, A. (2015). Comparing the two- and three-parameter logistic models via likelihood ratio tests: A commonly misunderstood problem. *Applied Psychological Measurement*, 39, 335-348.
- Büyüköztürk, Ş. (2012). *Sosyal Bilimler için Veri Analizi El Kitabı*. Ankara: Pegem.
- Can, S. (2003). *The analyses of secondary education institutions student selection and placement test's verbal section with respect to item response theory models*. Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. URL <http://www.jstatsoft.org/v48/i06/> .
- Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265–289.
- Chen, W.H., & Thissen, D. (1999). Estimation of item parameters for the three-parameter logistic model using the marginal likelihood of summed scores. *British Journal of Mathematical and Statistical Psychology*, 52(1), 19-37.
- Ching-Fung, B. S. (2002). *Ability estimation under different item parametrization and scoring models*. Doktora Tezi, Kuzey Teksas Üniversitesi, Teksas.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cox, F. & Godfrey, J. (1997). The importance of assessment procedures to student learning outcomes in religious education. *Australian Journal of Teacher Education*. 22(2), 46-56.
- Crocker, L., & Algina, J. (1986). *Introduction to modern test theory*. California: Thomson Learning.
- Çakıcı-Eser, D. (2015). *Çok boyutlu madde tepki kuramının farklı modellerinden çeşitli koşullar altında kestirilen parametrelerin incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Çalışkan, M. (2000). *The Fit of one, two and three-parameter models of item response theory (IRT) to the ministry of National Education-Educational Research and Development Directorates's (MNE-ERDO) science achievement test data*. Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çelen, Ü. (2008). *Klasik test kuramı ve madde tepki kuramına dayalı olarak geliştirilen iki testin psikometrik özelliklerinin karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Çelik, D. (2001). *The Fit of one, two and three-parameter models of item response theory (IRT) to the ministry of National Education secondary school institutions student selection and placement test data*. Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, Mantel Haenszel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çıtak, G. G. (2007). *Klasik test ve madde tepki kuramlarına göre çoktan seçmeli testlerde farklı puanlama yöntemlerinin karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order irt model approach. *Applied Psychological Measurement*, 34(4), 267–285.

- DeMars, C. E. (2008). *Scoring Multiple Choice Items: A Comparison of IRT and Classical Polytomous and Dichotomous Methods*. Center for Assessment and Research Studies, James Madison University, Harrisonburg.
- DeMars, C. E. (2010). *Item response theory*. New York: Oxford University Press.
- De Ayala, R.J. & Koch, W.R. (1987, Nisan). Computerized adaptive testing : A comparison of the nominal response model and the three parameter model. Yıllık Ulusal Eğitimde Ölçme Kurul Görüşmesinde sunulmuş bildiri, Washington, D.C.
- De Ayala, R.J. & Koch, W.R. (1988, Nisan). Computerized adaptive testing : A comparison of the nominal and graded response model in computerized testing. Yıllık Amerikan Araştırmaları Derneği Görüşmesinde sunulmuş bildiri, New Orleans.
- De Ayala, R.I. (1989). Computerized adaptive testing: A comparison of the nominal response model and the three-parameter model. *Educational and Psychological Measurement*, 49, 789-805.
- De Ayala, R. J. (1989). A comparison of the nominal response model and the three-parameter logistic model in computerized adaptive testing. *Educational and Psychological Measurement*, 49, 789-805.
- De Ayala, R. J. (1992). The nominal response model in computerized adaptive testing. *Applied Psychological Measurement*, 16, 327-343.
- De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, 25, 172- 189.
- De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. U. S. A.
- Demirtaşlı, N., Yalçın, S. & Ayan, C. (2014). Ölçme ve Değerlendirme Dersine Yönelik Tutum Ölçeğinin Çok Kategorili Madde Tepki Kuramı Modellerine Dayalı Olarak Ölçeklenmesi. IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde sunulmuş bildiri, Hacettepe Üniversitesi, Ankara.
- Dimitrov, D. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement*, 27, 440-458.

- Dolma, S. (2009). *Çok ihtimalli Rach modeli ile derecelendirilmiş yanıt modelinin örtük özellikleri tahminleme performansı açısından simülasyon yöntemiyle karşılaştırılması*. Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response models to multiple-choice tests. *Applied Psychological Measurement, 19*, 143-165.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University Press
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists. Multivariate applications book series* (Vol. 4). New Jersey: Lawrence Erlbaum Associates.
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item/Person Statistics. *Educational and Psychological Measurement, 58*(3), 357-381.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Singapore: Sage
- Foley, B. P. (2010). *Improving IRT parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique*. Doktora Tezi, The College of Education and Human Sciences, Nebraska
- Fraenkal, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: Mc Graw Hill.
- Gierl, M. J., Henderson, D. L., Jodoin, M. G., & Klinger, D. A. (2001). Minimizing the influence of item parameter estimation errors in test development: A comparison of three selection procedures. *Journal of Experimental Education, 69*, 261-279.
- Gültekin, S. (2011). *Çoktan seçmeli, açık uçlu ve karma testlerin psikometrik özelliklerinin Madde Tepki Kuramı'na dayalı olarak değerlendirilmesi*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Hambleton, R. K. & Van Der Linden, W. J. (1982). Advances in item response theory and applications: An Introduction. *Applied Psychological Measurement. Special Issue: Advances in item response theory and applications. 6*(4), 373-378.

- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. *Contemporary Sociology* (Vol. 21). London: Sage.
- Hambleton, R. K. & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Instructional Topics in Educational Measurement*. 12(2), 38-47.
- Hambleton, R. K., & Jones, R. W. (1993). Item parameter estimation errors and their influence on test information functions. *Applied Measurement in Education*, 7(3), 171-186.
- Han-Ho, T. (2010). *A Comparison of Item Selection Procedures Using Different Ability Estimation Methods in Computerized Adaptive Testing Based on the Generalized Partial Credit Model*. Doctoral Thesis, The University of Texas, Austin
- Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Instructional Topics in Educational Measurement*, 8(1), 35-41.
- Harvey, R. J., & Hammer, A. L. (1999). Item Response theory. *Counseling Psychologist*, 27(3), 353-374.
- Khalid, M. N. (2009). *IRT model fit from different perspectives*. Doktora Tezi, Twente Üniversitesi, Enschede.
- Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X² item fit index for the graded response model. *Asia Pacific Education Review*, 12(1), 89-96.
- Kılıç, İ. (1999). *The fit of one- two- and three- parameter models of item response theory to the student selection test of the student selection and placement center*. Doktora Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Koğar, H. (2014). *Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüğü ve test uzunluğu açısından karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

- Korkmaz, M. (2005). *Madde cevap kuramına dayalı olarak çok kategorili maddelerde madde ve test yanlılığının (işlevsel farklılığının) incelenmesi*. Doktora Tezi. Ege Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Lord, F.M. ve Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McDonald, R.P. (1982). Linear Versus Models in Item Response Theory Applied Psychological Measurement, 6, 379-396.
- MEB. (2009). *İlköğretim matematik dersi 6-8. sınıflar öğretim programı ve kılavuzu*. 15 Temmuz 2014 tarihinde <http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72> sayfasından erişilmiştir.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of $S-X^2$: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27, 289-298.
- Ostini, R. & Nering, M. L. (2006). *Polytomous Item Response Model*. U.S.A.
- Önder, İ. (2007). An investigation of goodness of model data fit. *Hacettepe University Journal of Education*, 32, 210-220.
- Özkurt, S. (2002). *The Fit of one-, two- and three-parameter models of item response theory (IRT) to An English proficiency achievement test data*. Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis using SPSS for Windows*. New York: Open University Press.
- Partchev, I. (2004). *A visual guide to item response theory*. Friedrich-Schiller-University Jena.


- Penfield, R. D. (2008). An Odds Ratio Approach for Assessing Differential Distractor Functioning Effects under the Nominal Response Model. *Journal of Educational Measurement, 45*(3), 247-269.
- Penfield, R. D. (2014). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues and Practice, 33*(1), 36-48.
- Reise, S.P., Ainsworth, A.T. and Haviland, M.G. (2005). Item Response Theory. Fundamentals, Applications, and Promise in Psychological Research. *Current Directions in Psychological Science, 14* (2), 95–101.
- Robitzsch (2015). sirt: Supplementary Item Response Theory Models. *R package version 1.6-0*. <http://CRAN.R-project.org/package=sirt>
- Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: a comparison of traditional and alternative procedures. *Journal of Educational Measurement, 40*, 331-352.
- Stone, A. (2006). *A Psychometric Analysis of The Statistics Concept Inventory*. Doctoral Thesis. University of Oklahoma, U. S. A.
- Şahin, A. (2012). *Madde tepki kuramında test uzunluğu ve örneklem büyüklüğünün model veri uyumu, madde parametreleri ve standart hata değerlerine etkisinin incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- U.S.A. Department of Education, Office of Educational Research and Improvement. (1988). *Can We Fairly Measure the Quality of Education*. CSE Technical Report. (Grant No:G0086-003)
- Tay, L., Huang, Q., & Vermunt, J. K. (2015). Item response theory with covariates (IRT-C): Assessing item recovery and item functioning for the three-parameter logistic model. *Educational and Psychological Measurement,*
- Teker, G. T., Kelecioğlu H., & Eroğlu, M. G. (2013). An investigation of goodness of model data fit. *Social and Behavioral Sciences, 106*, 394-400.
- Thissen, D. & Steinberg, L. (1986). Taxonomy of Item Response Models. *Psychometrika, 51*(4), 567-577.

- Thissen, D. (1991). MULTILOG: Multiple, categorical item analysis and test scoring using item response theory (Version 7.03). [Computer software and manual]. Chicago, IN: Scientific Software. [Computer Program.]
- Tomkowicz, J. T. (2000). *Comparison of Ability Estimates From Dichotomously and Nominally Scored Testwise Susceptible and Non-Susceptible Items*. Doctoral Thesis. University of Alberta, Edmonton.
- Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5, 245–262.
- Yen, M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145.

EKLER



EK 1. İzin Yazısı



T.C.
MİLLÎ EĞİTİM BAKANLIĞI
Ölçme, Değerlendirme ve Sınav Hizmetleri
Genel Müdürlüğü

Sayı : 57750415/480.01/1990640
Konu: Veri Talebi

23/02/2015

T.C.
GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜNE
(Eğitim Bilimleri Enstitüsü)

İlgi : 17/12/2014 tarih ve 80287700-302.08.01/6153 sayılı yazınız.

Gazi Üniversitesi Eğitim Bilimleri Enstitüsü Eğitiminde Ölçme ve Değerlendirme Bilim Dalı yüksek lisans programı öğrencisi Serpil ÇELIKTEN' in tez çalışması için ilgi yazı ile 2012 SBS A kitapçığı Matematik, Türkçe, Fen ve Teknoloji, Sosyal Bilgiler ve İngilizce alt testlerine ait ham veriler talep edilmektedir.

Söz konusu öğrenci cevapları versiyi hazırlanarak yazımız ekinde gönderilmiştir. Bilgilerinizi ve gereğini rica ederim.

Salih ALTUNTAŞ
Bakan a.
Genel Müdür V.

EK:
CD (1 adet)

Teknikokullar/ANKARA
Elektronik Ağ: www.meb.gov.tr
e-posta: mustafacengiz@meb.gov.tr

Ayrıntılı bilgi için: Mustafa CENGİZ (V.H.K.İ)
Tel: (0 312) 296 94 00 / 9595
Faks: (0 312) 296 94 57

Bu evrak güvenli elektronik imza ile imzalanmıştır. <http://evraksorgu.meb.gov.tr> adresinden CBC-F952-3F08-8a26-bbbc kodu ile teyit edilebilir.

EK 2. Uzman Deęerlendirme Formu

UZMAN DEęERLENDİRME FORMU

Sayın.....

Yüksek lisans tez çalışmam kapsamında 2012 SBS Matematik alt testi puanlarının madde tepki kuramı modellerinden Sınıflamalı Tepki Modeline uyum düzeyini test edeceğim. Bu konuda sizin maddeleri deęerlendirmenize ihtiya duymaktayım. Ařaęıda 2012 SBS matematik alt testi sınav soruları yer almaktadır. Her bir soru ise, 1 i doęru yanıt, dięer 3 ü ise eldirici olmak üzere toplamda 4 seenekten oluşmaktadır. Bilindięi üzere, öęrencilerin verdięi yanıtların puanlaması yapılırken, doęru yanıt olarak nitelendirilen seenek işaretlendięi takdirde öęrenciye “1” puan, eldiriciler işaretlendięi takdirde ise “0” puan verilmektedir. Ancak, Madde Tepki Kuramına göre eldiriciler de öęrenciler hakkında bilgi içermektedir, ve bu bilginin dikkate alınması kestirimlerin daha net yapılmasına katkıda bulunarak, ölçme hatasını azaltacaktır. Bu amaç doęrultusunda bu çalışmada çoklu puanlama modellerinden olan sınıflamalı tepki modeli kullanılacaktır. Bu modelin kullanılması için doęru yanıt dışındaki seeneklerin puanlama için derecelendirilmesi gerekmektedir. Dolayısıyla eldiriciler derecelendirilirken/sıralanırken doęru yanıtta en yakın (en doęru, ya da en yüksek puan alması gereken eldirici) ve en uzak (en az doęru, ya da en düşük puan alması gereken eldirici) olacak şekilde sıralanmalıdır.

Bu dereceleme yi dikkate alarak, ařaęıdaki test maddelerinin eldiricilerini her bir test maddesini ayrı ayrı dikkate alarak nasıl derecelendirirsiniz? (doęru yanıtta en yakın eldiriciyi 1 ve doęru yanıtta en uzak eldiriciyi 3 olacak şekilde dereceleyiniz.)

Katkılarınız için teşekkür ederim

Arş. Gör. Serpil ELİKTEN

8. SINIF

MATEMATİK TESTİ

A

1. Türkiye genelinde bir yılda $8,1 \times 10^5$ adet çam ağacının kurtarılması hedeflenmektedir.



Yukarıdaki afişe göre, bu hedefe ulaşmak için bir yılda kaç ton kullanılan kâğıt geri dönüşüme kazandırılmalıdır?

- A) $4,5 \times 10$ B) $4,5 \times 10^2$
C) $4,5 \times 10^3$ D) $4,5 \times 10^4$

2. $2\sqrt{10}$, $\sqrt{17}$, $3\sqrt{3}$ sayılarının küçükten büyüğe doğru sıralanışı, aşağıdakilerden hangisidir?

- A) $\sqrt{17} < 3\sqrt{3} < 2\sqrt{10}$
B) $3\sqrt{3} < \sqrt{17} < 2\sqrt{10}$
C) $\sqrt{17} < 2\sqrt{10} < 3\sqrt{3}$
D) $3\sqrt{3} < 2\sqrt{10} < \sqrt{17}$

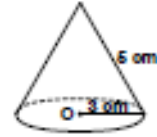
3. Şekildeki beş katlı düğün pastasında her kattaki pastanın yançap uzunluğu, bir üstündekinin yançap uzunluğunun 2 katıdır.



En üstteki pastanın yançap uzunluğu 2^2 cm olduğuna göre, en alttaki pastanın yançap uzunluğu kaç santimetredir?

- A) 2^{10} B) 2^6 C) 2^7 D) 2^8

4. Şekildeki O noktası, verilen dik dairesel koninin taban merkezidir.



Şekil üzerindeki verilere göre bu koninin açısını aşağıdakilerden hangisi olabilir?

- A) B)
C) D)

5. Bir mimar bir kenar $3a$ metre olan kare şeklindeki arsada, her birinin kenar uzunluğu $2b$ metre olan belli sayıda karesel bölge belirlemiştir. Arsada geri kalan bölgenin alanı $(3a - 4b)(3a + 4b)$ metrekare olduğuna göre, belirlenen karesel bölgelerin sayısı kaçtır?

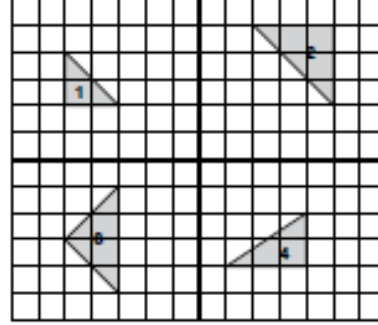
A) 8 B) 4 C) 3 D) 2

6. Ahmet, kenarlarından ikisinin uzunluğu 17 m ve 10 m olan üçgen biçimindeki bahçesinin çevresini çit ile çevirmiştir.

Ahmet'in kullandığı çitin uzunluğu, metre olarak aşağıdakilerden hangisi olabilir?

A) 30 B) 34 C) 40 D) 54

7. Aşağıdaki şekilde dört farklı bahçede yer alan havuzların kuşbakışı çizimleri verilmiştir.



Bu havuzlardan hangisi diğer üç havuzla benzer değildir?

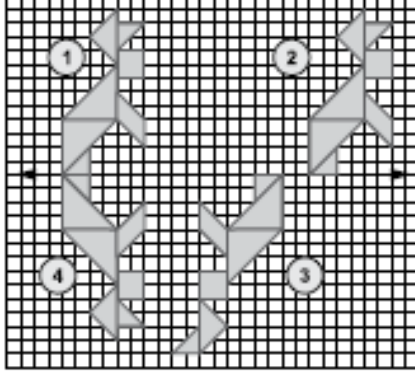
A) 1 B) 2 C) 3 D) 4

8. Seda'nın matematik dersi dönem sonu puanı, ilk sınavda aldığı puanın iki katından 4 eksiktir. Dönem sonu puanı 50 ve üzeri olan öğrenci, o dersten başarılı olmaktadır.

Matematikten dönem sonunda başarılı olan Seda'nın ilk sınavdaki puanı en az kaçtır?

A) 28 B) 27 C) 26 D) 25

9. Aşağıdaki numaralandırılmış tangram şekillerinden hangi ikisi, birbirinin ötelemele yanmasıdır?

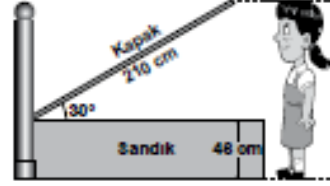


- A) 2 ve 4 B) 1 ve 4
C) 1 ve 2 D) 2 ve 3

10. Aşağıdakilerden hangisi, $0,\overline{45}$ ile $0,\overline{452}$ devirli ondalık kesirleri arasında yer alır?

- A) 0,451 B) 0,453
C) 0,455 D) 0,457

11. Ceren, açılıp kapanabilir bir sandıklı yatağın 210 cm uzunluğundaki kapağını aşağıdaki gibi açığında, kapağın yerden yüksekliği Ceren'in boyuna eşit olmaktadır.



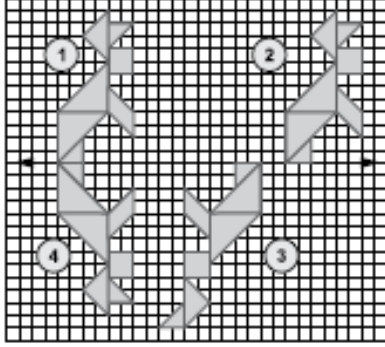
Sandığın yüksekliği 46 cm olduğuna göre, Ceren'in boyu kaç santimetredir? ($\sin 30 = 0,5$)

- A) 155 B) 151 C) 128 D) 105

12. Aşağıdakilerden hangisi yapıldığında, taban ayırının uzunluğu 8 cm, yüksekliği 12 cm olan kare dik prizma şeklindeki kutunun hacmi yarıya iner?

- A) Bütün ayır uzunlukları ikişer santimetre azaltıldığında
B) Taban ayır uzunlukları dörder santimetre azaltıldığında
C) Bütün ayır uzunlukları yarıya indirildiğinde
D) Yüksekliği 6 cm azaltıldığında

9. Aşağıdaki numaralandırılmış tangram şekillerinden hangi ikisi, birbirinin öteleme ile yanmasıdır?

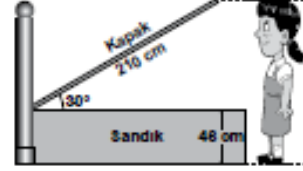


- A) 2 ve 4
B) 1 ve 4
C) 1 ve 2
D) 2 ve 3

10. Aşağıdakilerden hangisi, $0,\overline{45}$ ile $0,\overline{452}$ devirli ondalık kesirleri arasında yer alır?

- A) 0,451
B) 0,453
C) 0,455
D) 0,457

11. Ceren, açılıp kapanabilir bir sandıklı yatağın 210 cm uzunluğundaki kapağını aşağıdaki gibi açtığında, kapağın yerden yüksekliği Ceren'in boyuna eşit olmaktadır.



Sandığın yüksekliği 48 cm olduğuna göre, Ceren'in boyu kaç santimetredir? ($\sin 30 = 0,5$)

- A) 156
B) 151
C) 128
D) 105

12. Aşağıdakilerden hangisi yapıldığında, taban ayrıntının uzunluğu 8 cm, yüksekliği 12 cm olan kare dik prizma şeklindeki kutunun hacmi yarıya iner?

- A) Bütün ayrıntı uzunlukları ikişer santimetre azaltıldığında
B) Taban ayrıntı uzunlukları dörtör santimetre azaltıldığında
C) Bütün ayrıntı uzunlukları yarıya indirildiğinde
D) Yüksekliği 6 cm azaltıldığında

13. Eser, kare dik prizma biçiminde ve ayrıtlarından birinin uzunluğu 30 cm olan yandaki teneke kutunun yan yüzlerini boyadıktan sonra bir yüzüne "çöp" yazısını yazıyor.



Boyanan yüzeyin alanı 1680 cm^2 olduğuna göre, bu çöp kutusunun diğer farklı ayrıtlarının uzunluğu kaç santimetredir?

- A) 14 B) 21 C) 25 D) 28

14. Bir silindir, diğeri küp şeklinde olan iki kap, tamamen su ile doludur. Silindir şeklindeki kabın taban çapının ve yüksekliğinin uzunlukları, küpün bir ayrıtlarına eşittir. Buna göre, silindir şeklindeki kaptaki su bulunan suyun kaçta kaç kadar su bulunmaktadır? (π yerine 3 alınız.)

- A) $\frac{3}{4}$ B) $\frac{2}{3}$ C) $\frac{1}{2}$ D) $\frac{1}{8}$

15. Bir torbadaki özdeş topların 11'i kırmızı, 8'i beyaz, 9'u mavi ve 12'si siyahtır.

En az kaç top çıkarılırsa, torbada kalan topların renklerine göre çekilme olasılıkları eşit olur?

- A) 7 B) 8 C) 9 D) 10

16. Bir sınıftaki öğrencilerin günlük kitap okuma süreleri aşağıdaki tabloda verilmiştir.

Tablo: Öğrencilerin Kitap Okuma Süreleri

Kitap okuma süresi (dakika)	Öğrenci sayısı
20 - 24	2
25 - 29	6
30 - 34	5
35 - 39	3
40 - 44	8
45 - 49	1
50 - 54	3
55 - 59	12
60 - 64	6
65 - 69	2

Tabloya göre kaç kişi günde 1 saatten daha az süre kitap okumaktadır?

- A) 8 B) 12 C) 40 D) 46

17. Üç yol ağzında bulunan bir tavşan, bu yollardan rastgele birine yöneliyor. Bu üç yolun her biri, devamında iki dar yola ayrılıyor. Tavşan yol ayrımına geldiğinde iki dar yoldan birini rastgele seçerek yoluna devam ediyor.

Tavşanın, dar yollardan birinde bekleyen kaplumbağa ile karşılaşma olasılığı nedir?

- A) $\frac{1}{2}$ B) $\frac{1}{3}$ C) $\frac{1}{6}$ D) $\frac{1}{9}$

18. Bir tel, şekildeki gibi kıvrılarak üç tane kare oluşturuluyor.



Karelerin sınırladıkları bölgelerin alanları 25 cm^2 , 49 cm^2 ve 64 cm^2 olduğuna göre, bu tel ile oluşturulabilecek en büyük karenin sınırladığı bölgenin alanı kaç santimetrekare olur?

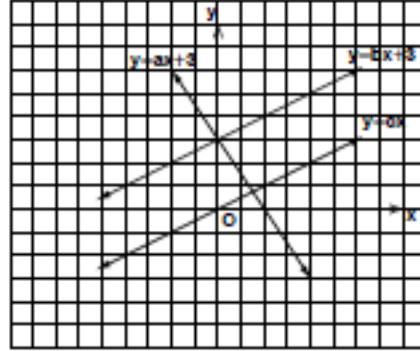
- A) 100 B) 138 C) 225 D) 400

19. Bir yarışma programında, verilen her doğru cevaba + 3 puan, her yanlış cevaba - 2 puan verilmektedir.

Bu yarışmaya katılan Aysun, sorulan 5 sorunun tümünü cevaplamıştır. Yarışma sonunda 10 puan aldığına göre, Aysun kaç soruyu doğru cevaplamıştır?

- A) 2 B) 3 C) 4 D) 5

- 20.

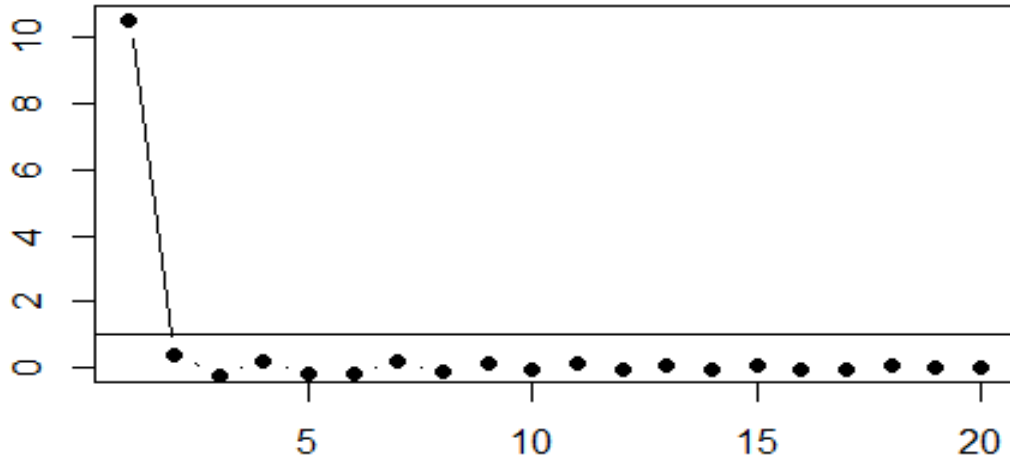
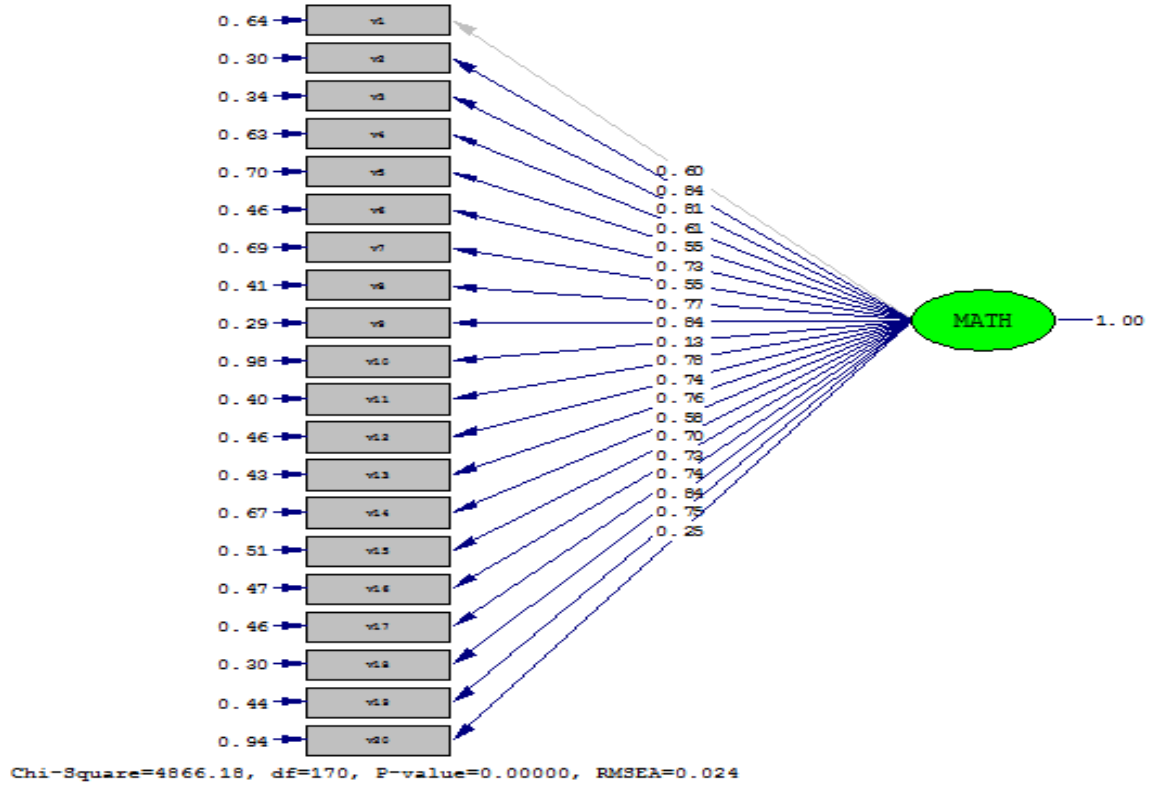


Yukarıda denklemleri ile verilen doğruların grafiklerine göre, aşağıdakilerden hangisi doğrudur?

- A) $a < b$ ve $b = c$
 B) $b = c$ ve $c < a$
 C) $c < b < a$
 D) $a < b < c$

MATEMATİK TESTİ BİTTİ.
 FEN VE TEKNOLOJİ TESTİNE GEÇİNİZ.

EK 4. Araştırma Kapsamında Yapılan AFA ve DFA Sonuçları – (N=50.000)



AFA sonucunda elde edilen kırılma grafiği

EK 5. 1PLM, 2PLM, 3PLM için Yerel Bağımsızlık Varsayımı Testi Sonuçları

Madde	Çiftleri	1 PLM	2 PLM	3 PLM
Madde1	Madde2	Q3	Q3	Q3
m1	m2	-0.0607984791	-0.053876250	-0.0365026265
m1	m3	-0.0122135632	0.004567514	0.0261702627
m1	m4	-0.1247919268	-0.099216359	-0.0832141560
m1	m5	-0.0308716993	-0.003569557	-0.0071978274
m1	m6	-0.0115786385	0.005939213	0.0118866338
m1	m7	-0.0377163956	-0.007659610	-0.0071328324
m1	m8	-0.0718505040	-0.054446870	-0.0451312629
m1	m9	-0.0840593162	-0.071900709	-0.0638371199
m1	m10	-0.0424522736	0.003884855	0.0012144859
m1	m11	-0.0575666619	-0.046588483	-0.0360911639
m1	m12	-0.0497766157	-0.031993668	-0.0264296071
m1	m13	-0.0306162419	-0.018487051	-0.0182317486
m1	m14	0.0211096571	0.046283355	0.0372692342
m1	m15	-0.1074477327	-0.079409489	-0.0642028905
m1	m16	-0.0851068165	-0.060731281	-0.0376893769
m1	m17	-0.0734686331	-0.054094437	-0.0449552310
m1	m18	-0.0330819798	-0.028890346	-0.0241944191
m1	m19	-0.1007857558	-0.076236082	-0.0590524405
m1	m20	-0.0398330352	0.005780379	0.0062922150
m2	m3	0.1033797877	-0.029960892	-0.0248744744
m2	m4	0.0468053584	-0.007071141	-0.0125803143
m2	m5	-0.0946477208	-0.056075342	-0.0387426763
m2	m6	-0.0142419529	-0.076186521	-0.0635898374
m2	m7	-0.1151631564	-0.103295115	-0.0893198366
m2	m8	0.0029302824	-0.079186592	-0.0723991982
m2	m9	0.0382904591	-0.049484347	-0.0385387857
m2	m10	-0.1557522980	-0.011881352	-0.0131600063
m2	m11	0.0228169508	-0.073795878	-0.0526637667

m2	m12	-0.0509333276	-0.120623553	-0.0976565385
m2	m13	0.0145950547	-0.048537296	-0.0303861394
m2	m14	-0.1216154368	-0.084696478	-0.0710808446
m2	m15	0.0232188975	-0.036955800	-0.0313532679
m2	m16	-0.0124241682	-0.118344217	-0.1147342640
m2	m17	0.0143015102	-0.070797501	-0.0522134064
m2	m18	0.0798564382	-0.018849844	-0.0037715827
m2	m19	-0.0641646621	-0.150474326	-0.1521736698
m2	m20	-0.1415414949	-0.018627841	-0.0236206192
m3	m4	0.0229968583	-0.024255095	-0.0301908063
m3	m5	-0.1346678485	-0.097003566	-0.0749831510
m3	m6	-0.0407101154	-0.093700472	-0.0807827351
m3	m7	-0.0695766563	-0.048191016	-0.0300150325
m3	m8	0.0129929699	-0.055686597	-0.0510104424
m3	m9	0.0231375677	-0.052768239	-0.0442688245
m3	m10	-0.1058050748	0.031309738	0.0298836560
m3	m11	-0.0459442707	-0.136250385	-0.1164126087
m3	m12	-0.0327926077	-0.088629118	-0.0657818995
m3	m13	-0.0483812821	-0.105046162	-0.0862732829
m3	m14	-0.0717653613	-0.029161397	-0.0099378817
m3	m15	-0.0420857516	-0.097236001	-0.0917146584
m3	m16	-0.0136344772	-0.109512033	-0.1094263842
m3	m17	-0.0453625588	-0.122953879	-0.1053860056
m3	m18	0.0267702666	-0.061751836	-0.0501450261
m3	m19	0.0554379145	-0.009821211	-0.0142002855
m3	m20	-0.1341698936	-0.021016358	-0.0267057757
m4	m5	-0.0591659080	-0.029454180	-0.0091008203
m4	m6	-0.0118235942	-0.023866559	-0.0147297676
m4	m7	-0.0512038458	-0.025164328	-0.0082511825
m4	m8	-0.0711199660	-0.090910822	-0.0894299853
m4	m9	-0.0813320546	-0.106798437	-0.1046752948

m4	m10	-0.1012993122	-0.033665989	-0.0356591762
m4	m11	-0.0751194714	-0.108253845	-0.0978841918
m4	m12	-0.0357229715	-0.050014955	-0.0353240111
m4	m13	-0.0411351609	-0.058124242	-0.0479732781
m4	m14	-0.0977931209	-0.067271280	-0.0518343045
m4	m15	0.0049165163	-0.003442642	0.0017286537
m4	m16	0.0014127693	-0.029415439	-0.0271635393
m4	m17	-0.0449815255	-0.066558935	-0.0584127630
m4	m18	-0.0504813307	-0.081965790	-0.0814314037
m4	m19	-0.0205704814	-0.038767251	-0.0389169466
m4	m20	-0.0894241514	-0.026640229	-0.0308750488
m5	m6	-0.0650789513	-0.033393935	-0.0289037848
m5	m7	-0.0314578789	-0.006792303	-0.0079092554
m5	m8	-0.1007091479	-0.065826703	-0.0563731502
m5	m9	-0.0858871805	-0.051884137	-0.0451469580
m5	m10	-0.0513222183	-0.042013776	-0.0453011162
m5	m11	-0.0726742280	-0.039818950	-0.0310516767
m5	m12	-0.1101859929	-0.076716641	-0.0732941115
m5	m13	-0.0587764143	-0.029505428	-0.0315289650
m5	m14	-0.0303216540	-0.011562884	-0.0287591127
m5	m15	-0.0335424658	0.001348177	0.0193975987
m5	m16	-0.0487318959	-0.008610877	0.0164625592
m5	m17	0.0166376863	0.054899057	0.0641107969
m5	m18	-0.0628445878	-0.030719470	-0.0288232887
m5	m19	-0.0738006398	-0.037557520	-0.0179123814
m5	m20	-0.0086167836	0.006556889	0.0063787699
m6	m7	-0.0723547610	-0.050443676	-0.0442303476
m6	m8	-0.0020892795	-0.028705058	-0.0206004947
m6	m9	-0.0560902860	-0.091690342	-0.0820124624
m6	m10	-0.0890013606	0.009033275	0.0078157519
m6	m11	0.0530322927	0.018572864	0.0316942493

m6	m12	-0.0069502806	-0.027935354	-0.0157904641
m6	m13	-0.0824301864	-0.108539513	-0.1001472339
m6	m14	-0.0643044325	-0.034265212	-0.0339368816
m6	m15	-0.0642564259	-0.076933733	-0.0649107875
m6	m16	-0.0695408756	-0.100989184	-0.0857459610
m6	m17	-0.0244244580	-0.051795468	-0.0398370500
m6	m18	0.0217062498	-0.024836156	-0.0154747140
m6	m19	-0.0141631637	-0.034945849	-0.0249440420
m6	m20	-0.1146532918	-0.029506316	-0.0300053429
m7	m8	-0.0088233316	0.014896616	0.0229381512
m7	m9	-0.0520800267	-0.032597863	-0.0263072834
m7	m10	-0.0559065071	-0.023517493	-0.0262057246
m7	m11	-0.0747861061	-0.058634838	-0.0492881945
m7	m12	-0.0405257893	-0.019001369	-0.0136981497
m7	m13	-0.0561402593	-0.040379093	-0.0402234536
m7	m14	-0.0767940881	-0.052167970	-0.0598208311
m7	m15	-0.0186848441	0.010773406	0.0259696052
m7	m16	-0.0521685462	-0.023653884	-0.0018325357
m7	m17	-0.0926966275	-0.068972987	-0.0603604139
m7	m18	-0.0887029350	-0.076591966	-0.0742419949
m7	m19	-0.0470315658	-0.018637126	-0.0020288019
m7	m20	-0.0583885899	-0.023743391	-0.0235256768
m8	m9	-0.0061302943	-0.051765439	-0.0444969225
m8	m10	-0.1480466424	-0.042662424	-0.0431187341
m8	m11	-0.0503669739	-0.102132402	-0.0884103320
m8	m12	-0.0386011755	-0.069872286	-0.0561491220
m8	m13	-0.0485970342	-0.083027489	-0.0728399824
m8	m14	-0.0440558781	-0.010591197	-0.0038410403
m8	m15	-0.0422796449	-0.062790057	-0.0565441765
m8	m16	-0.0282390627	-0.072183239	-0.0657232050
m8	m17	0.0118620847	-0.026171392	-0.0159667856

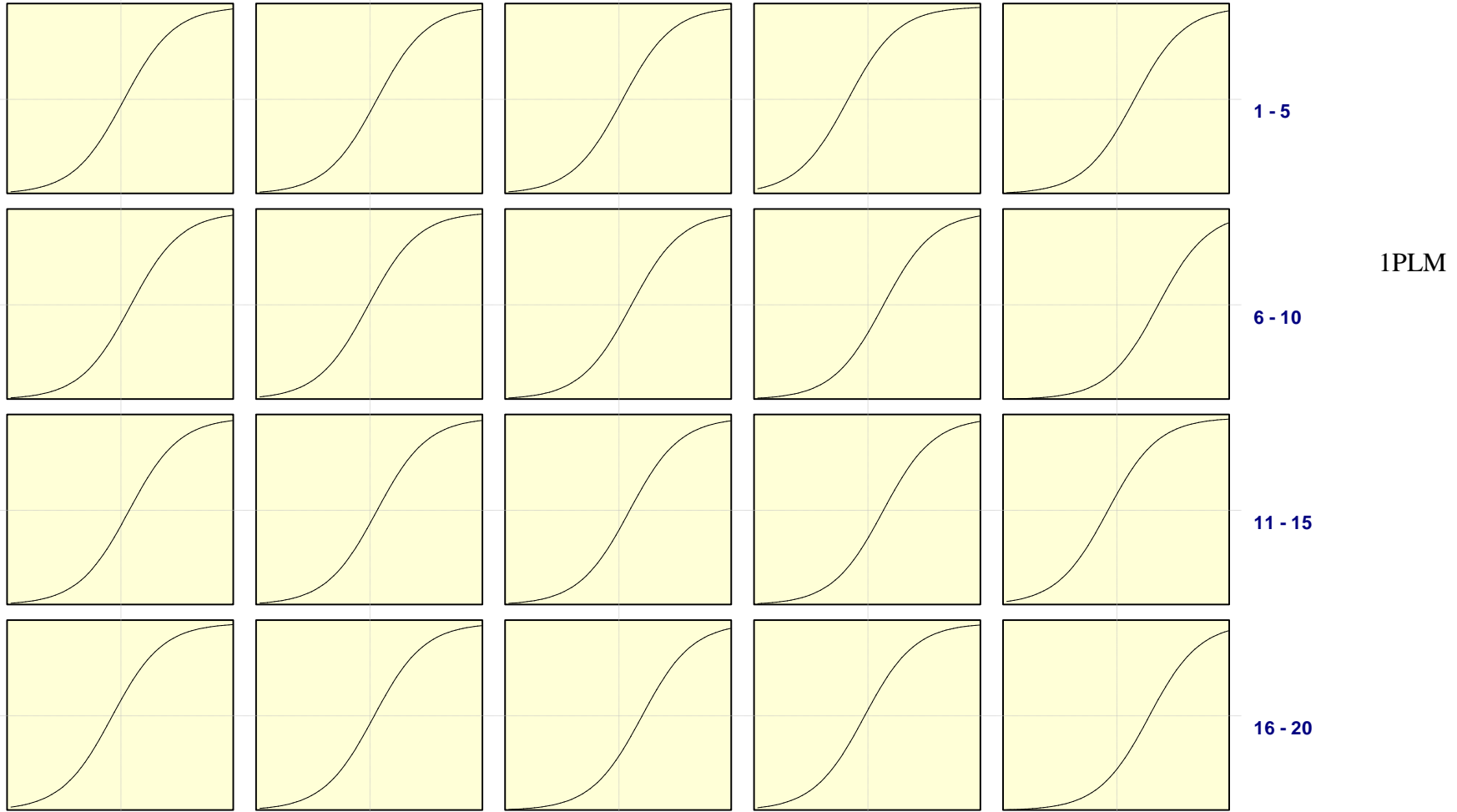
m8	m18	0.0035094773	-0.057898479	-0.0483474856
m8	m19	0.0376971133	0.008046711	0.0105446210
m8	m20	-0.0658525196	0.033014331	0.0310331440
m9	m10	-0.0562205396	0.070652568	0.0700540920
m9	m11	0.0389369957	-0.016344973	-0.0015772132
m9	m12	0.0160661214	-0.021094122	-0.0067855849
m9	m13	-0.0768350811	-0.120579266	-0.1086401096
m9	m14	-0.0529815516	-0.021326968	-0.0171694579
m9	m15	-0.0465386494	-0.073282580	-0.0661771423
m9	m16	0.0059331875	-0.041955208	-0.0340374627
m9	m17	0.0047322436	-0.041445617	-0.0292405529
m9	m18	0.0506237863	-0.018275992	-0.0050879653
m9	m19	-0.0186967863	-0.056706307	-0.0524526235
m9	m20	-0.1401172587	-0.036317003	-0.0376382324
m10	m11	-0.1602082483	-0.046734269	-0.0470413615
m10	m12	-0.1038255322	0.001345868	0.0001772215
m10	m13	-0.0990392927	0.006014171	0.0046281914
m10	m14	-0.0326001090	-0.011557826	-0.0151656177
m10	m15	-0.0971751300	-0.016377590	-0.0179641161
m10	m16	-0.1038034405	0.003241465	0.0011225462
m10	m17	-0.1334925796	-0.024318862	-0.0253984114
m10	m18	-0.1473124017	-0.007468845	-0.0065088796
m10	m19	-0.0763087647	0.019836243	0.0184057314
m10	m20	0.0666825832	-0.032454303	-0.0331176989
m11	m12	0.0050963721	-0.034528585	-0.0155162931
m11	m13	0.0518649940	0.012172580	0.0261470491
m11	m14	-0.0553268911	-0.023644180	-0.0190125786
m11	m15	-0.0472764869	-0.080755711	-0.0657449227
m11	m16	-0.0154698733	-0.072887885	-0.0541263345
m11	m17	0.0416851177	-0.006709993	0.0111805673
m11	m18	0.0295634078	-0.042048873	-0.0246514730

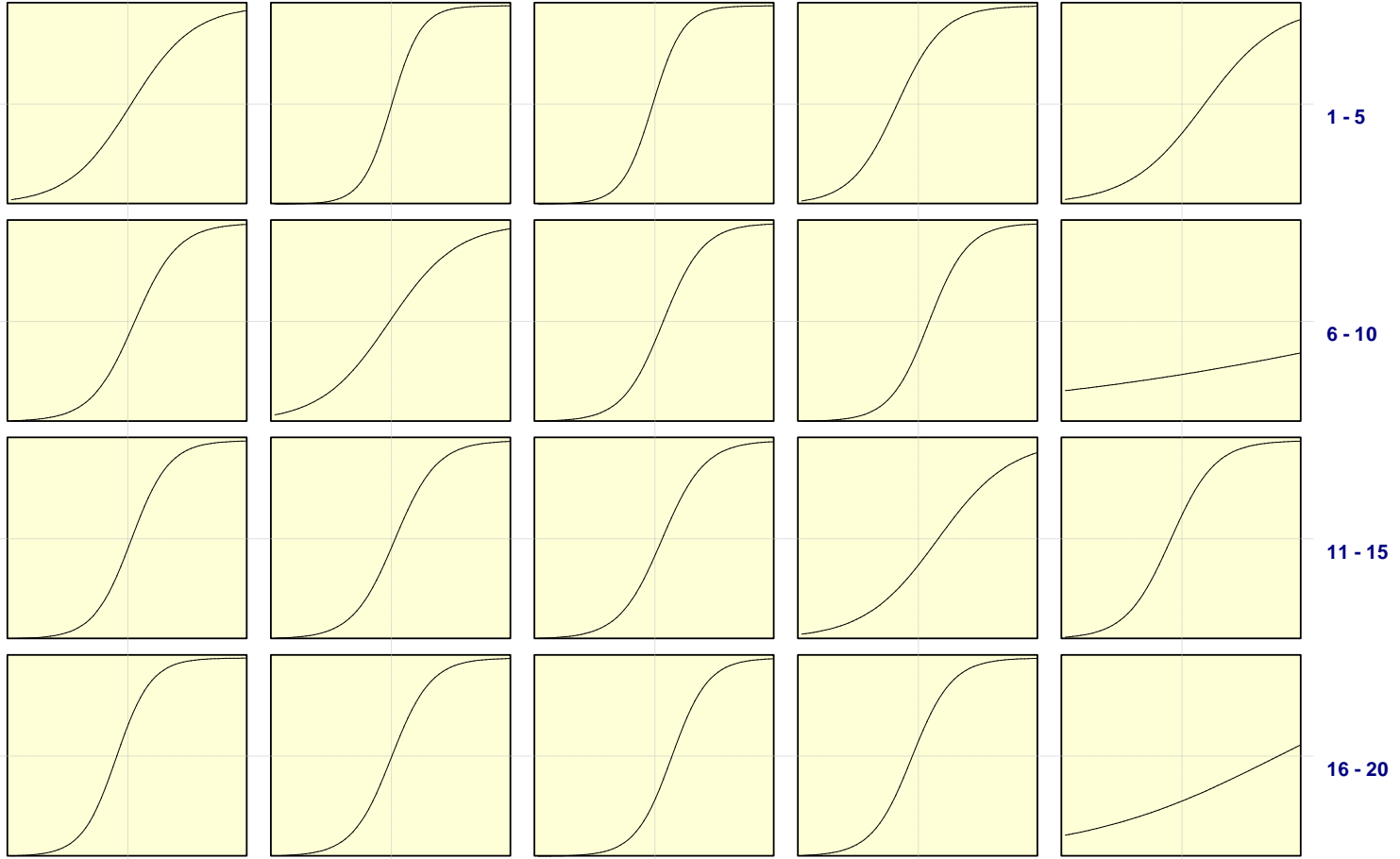
m11	m19	-0.0666263539	-0.113251419	-0.1001432624
m11	m20	-0.1478481725	-0.048499009	-0.0485730307
m12	m13	0.0927592315	0.068921370	0.0782577367
m12	m14	-0.0274026270	0.004793242	0.0033411164
m12	m15	-0.0280228637	-0.042244348	-0.0255022894
m12	m16	-0.0675732870	-0.100647106	-0.0773136181
m12	m17	-0.0426014531	-0.073160814	-0.0560257216
m12	m18	0.0177550414	-0.034717695	-0.0192605589
m12	m19	-0.1054992263	-0.130818627	-0.1136403965
m12	m20	-0.1113450651	-0.019707949	-0.0189207884
m13	m14	-0.1443741948	-0.117770495	-0.1253015854
m13	m15	-0.0166100850	-0.033226195	-0.0213938982
m13	m16	-0.0516683481	-0.084087346	-0.0662895193
m13	m17	0.0042836943	-0.027196310	-0.0154406339
m13	m18	0.0260697705	-0.029617440	-0.0163559506
m13	m19	-0.0260329233	-0.051170072	-0.0388310276
m13	m20	-0.1222935622	-0.032610895	-0.0322269722
m14	m15	-0.0419881675	-0.006694592	0.0067924415
m14	m16	-0.0537302238	-0.010989336	0.0096031507
m14	m17	-0.0627369276	-0.026078863	-0.0214803858
m14	m18	-0.0862215221	-0.060524638	-0.0600280375
m14	m19	-0.0852058377	-0.048263454	-0.0324653428
m14	m20	0.0002778298	0.023804472	0.0237955592
m15	m16	0.0356324621	0.005978598	0.0162736159
m15	m17	-0.0443797561	-0.066894349	-0.0542673193
m15	m18	-0.1167400968	-0.156122379	-0.1491508396
m15	m19	-0.0530001895	-0.071919103	-0.0657459191
m15	m20	-0.1173099875	-0.044404150	-0.0468008014
m16	m17	-0.0800109452	-0.129429778	-0.1116123546
m16	m18	-0.0418112214	-0.100774420	-0.0928914055
m16	m19	0.0353657007	-0.008309401	-0.0062726515

m16	m20	-0.0788849293	0.015507051	0.0106014798
m17	m18	0.0370252531	-0.021572648	-0.0083428026
m17	m19	-0.0591532793	-0.094165746	-0.0832251272
m17	m20	-0.1133841357	-0.015969668	-0.0167937870
m18	m19	-0.0038677464	-0.050707113	-0.0465438801
m18	m20	-0.1086867365	0.015323328	0.0145257049
m19	m20	-0.0690928997	0.016220129	0.0121992653



EK 6. 1PLM, 2PLM, 3PLM'ye ilişkin madde karakteristik eğrileri ve STM'ye ilişkin Kategori Karakteristik Eğrileri





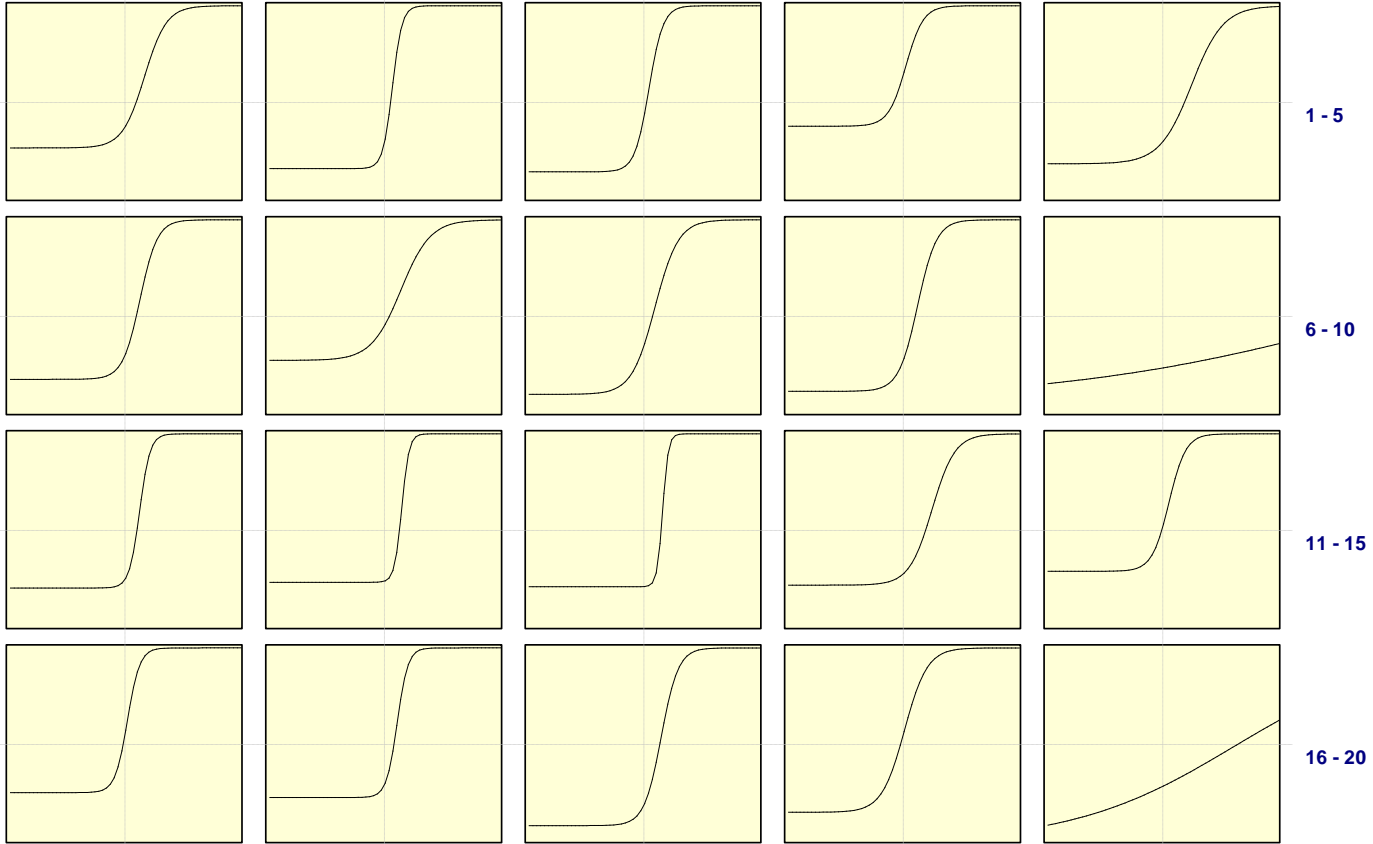
1-5

6-10

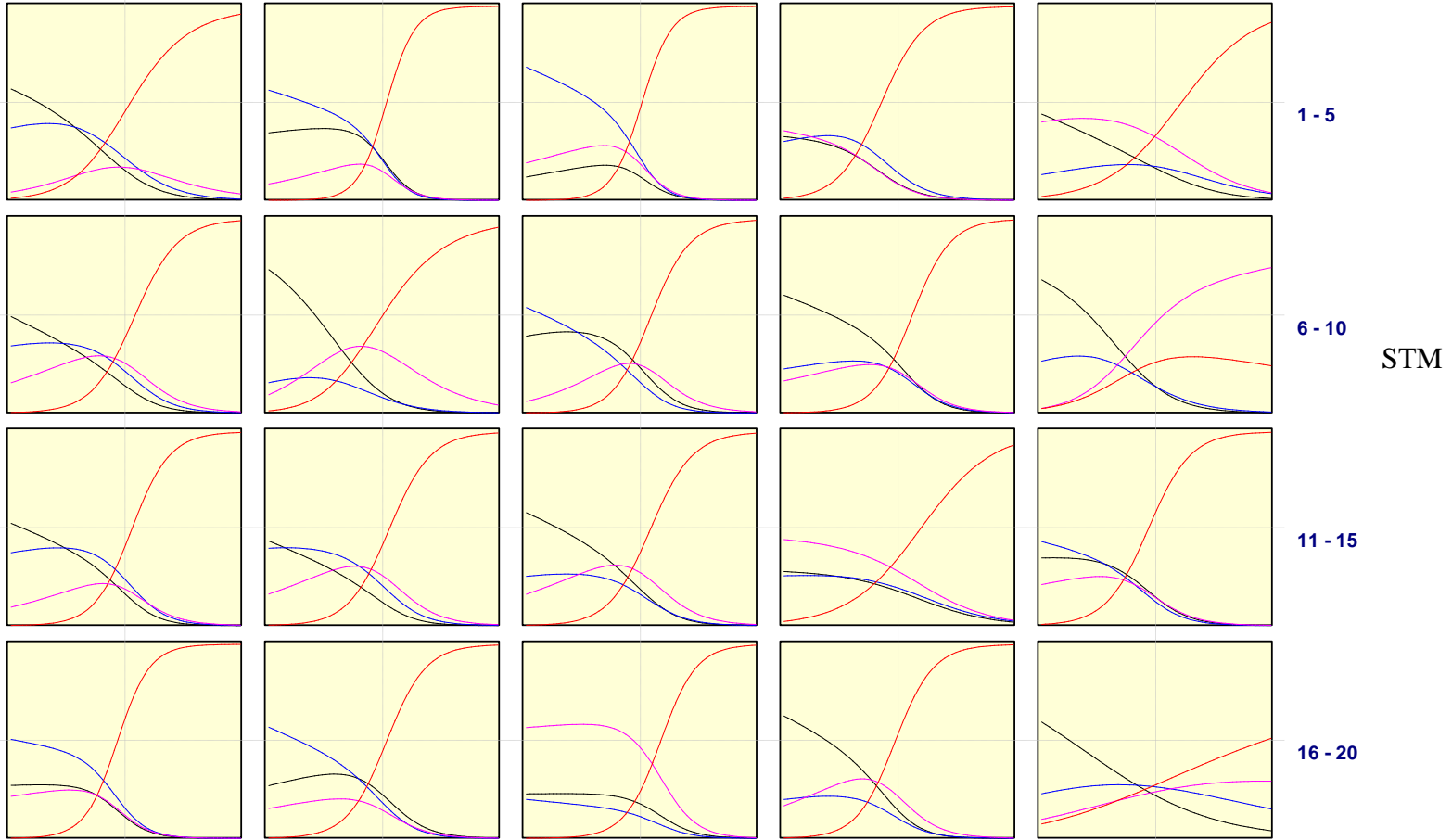
11-15

16-20

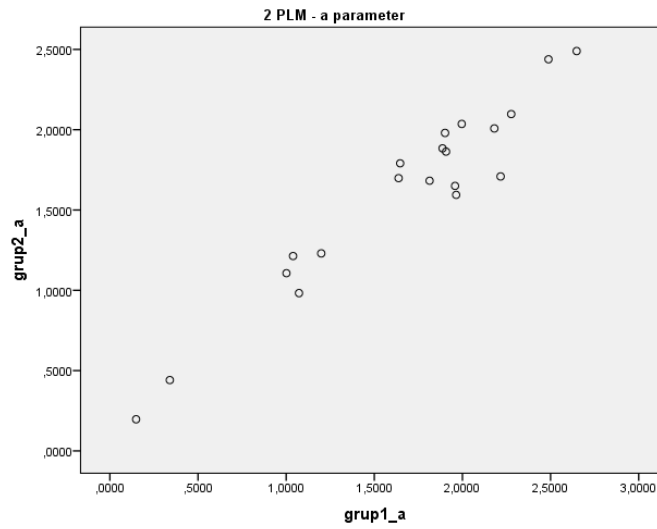
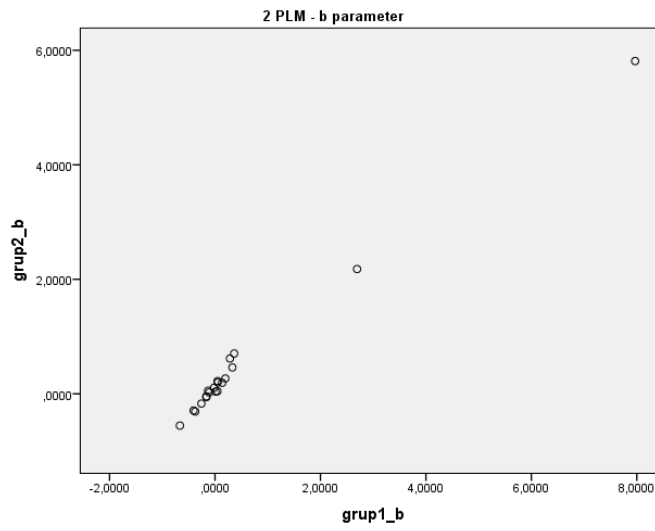
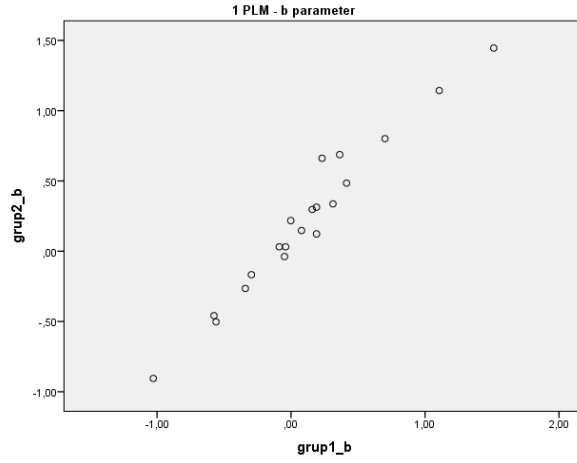
2PLM

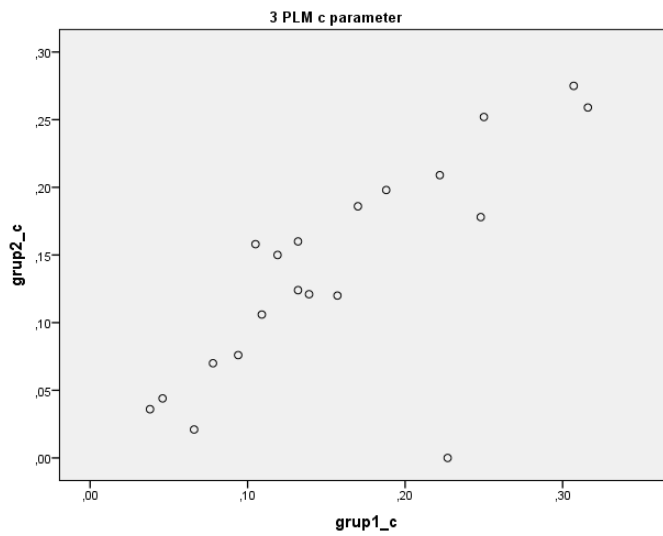
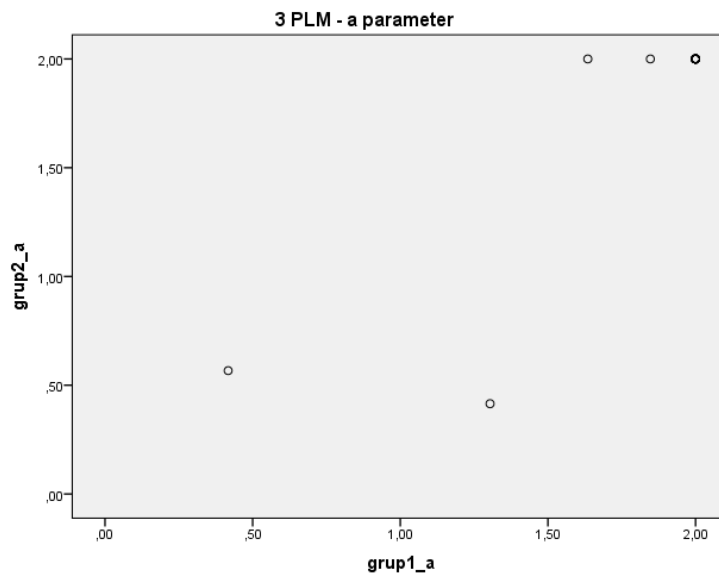
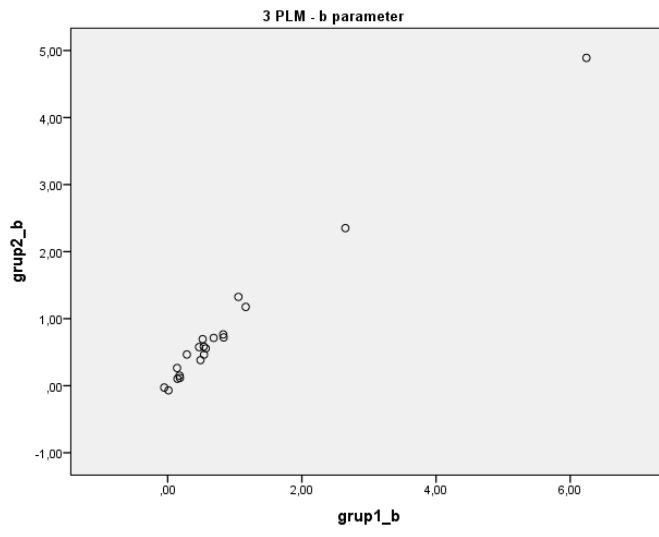


3PLM

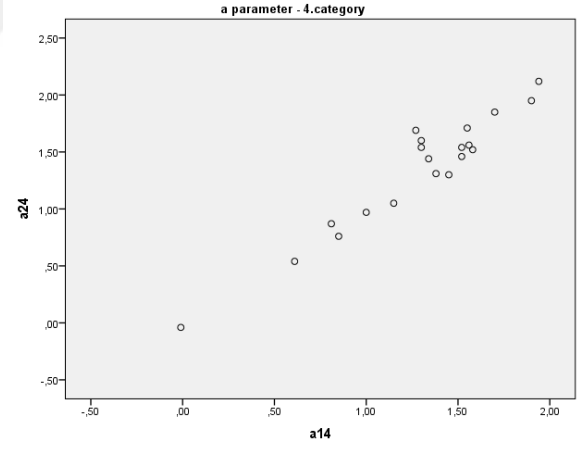
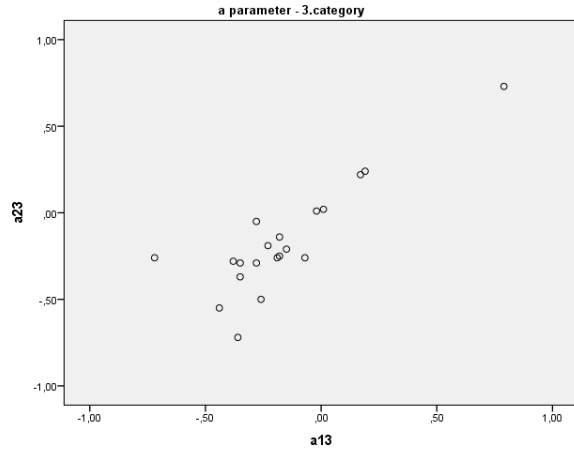
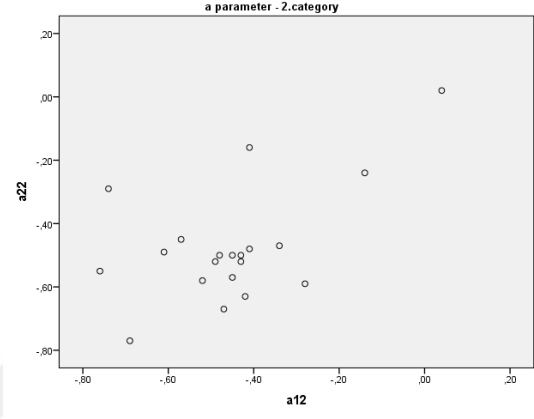
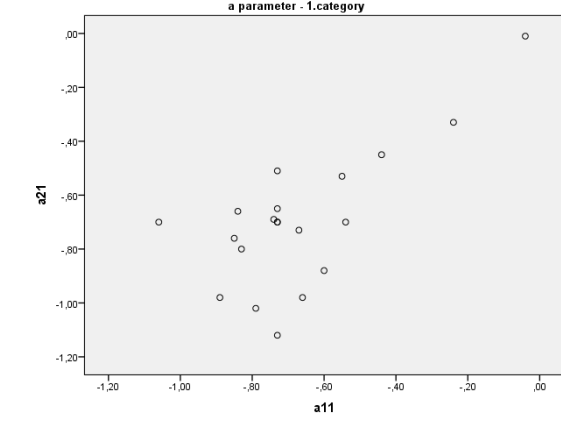


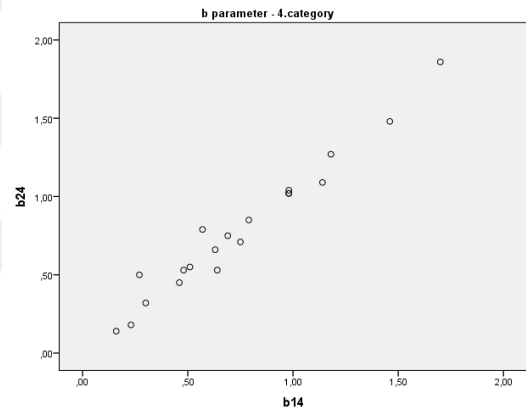
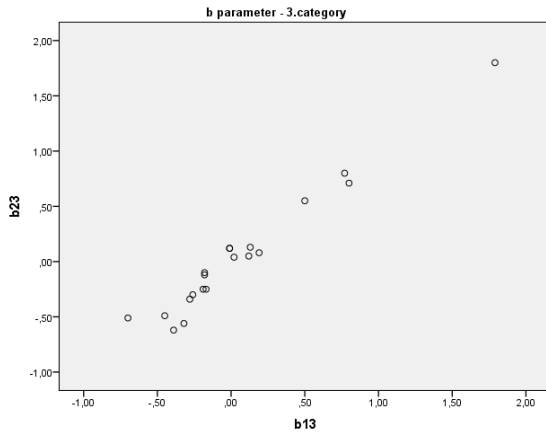
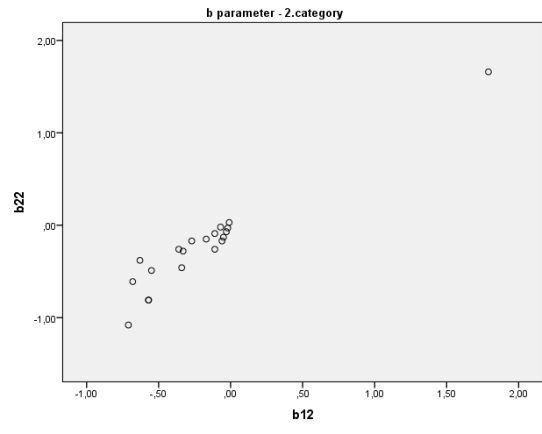
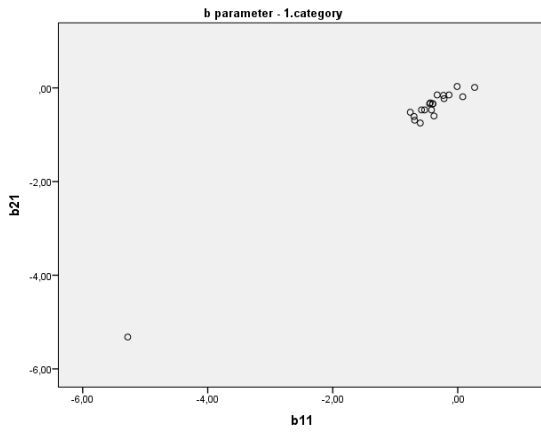
EK 7. 1PLM, 2PLM, 3PLM Altında İki Farklı Gruptan Elde Edilen Verilere Dayalı Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Saçılma Grafikleri



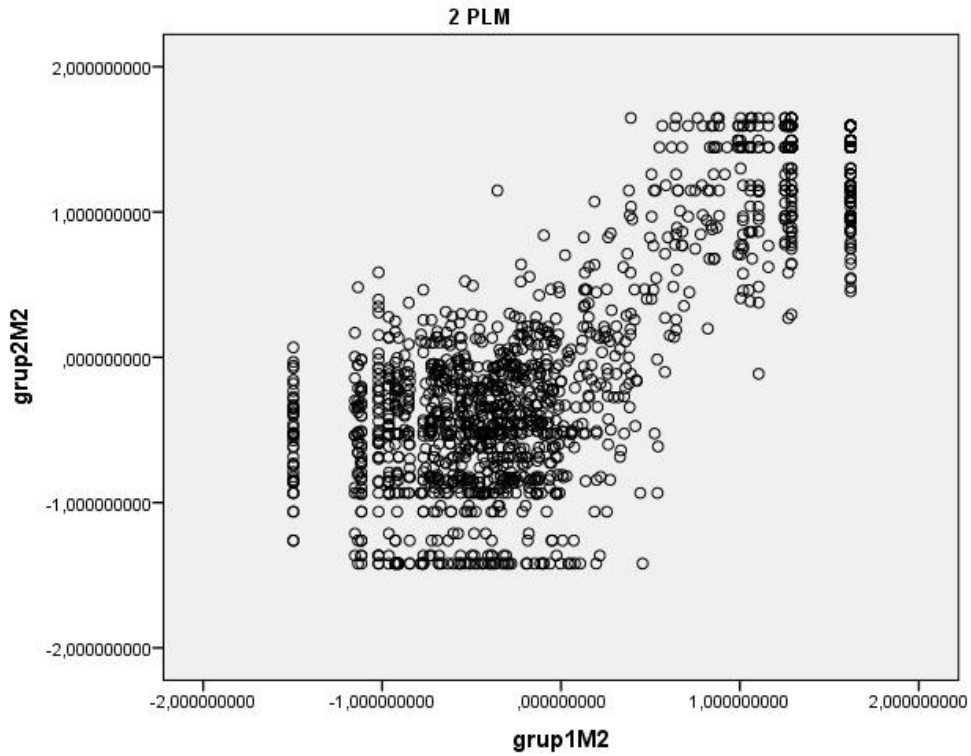
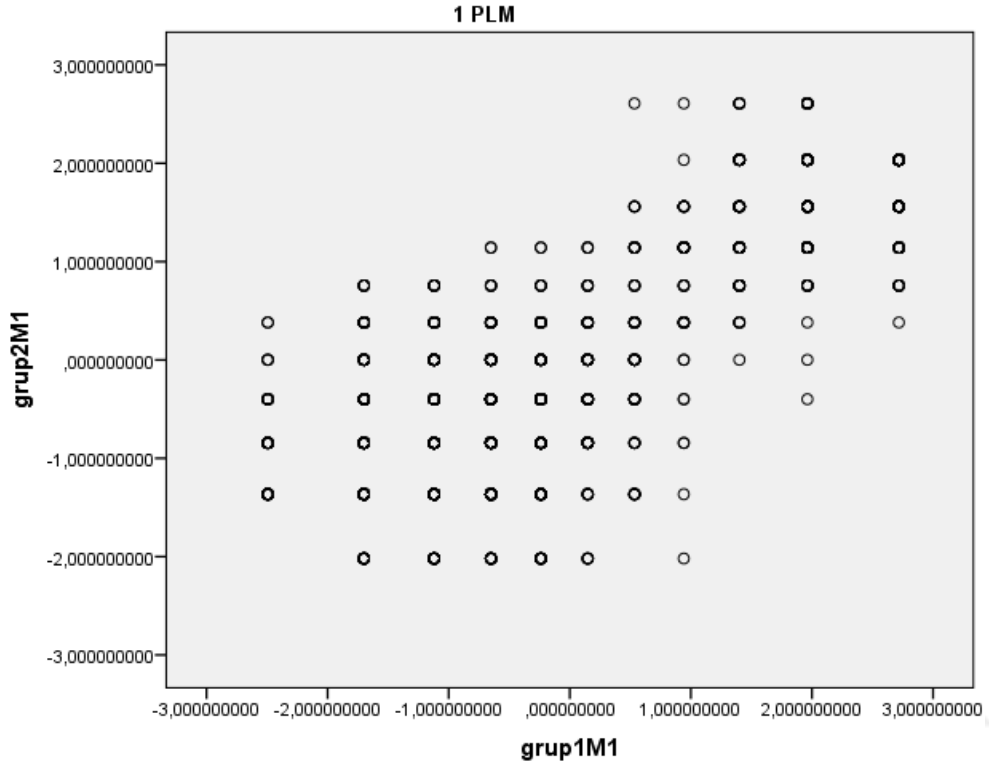


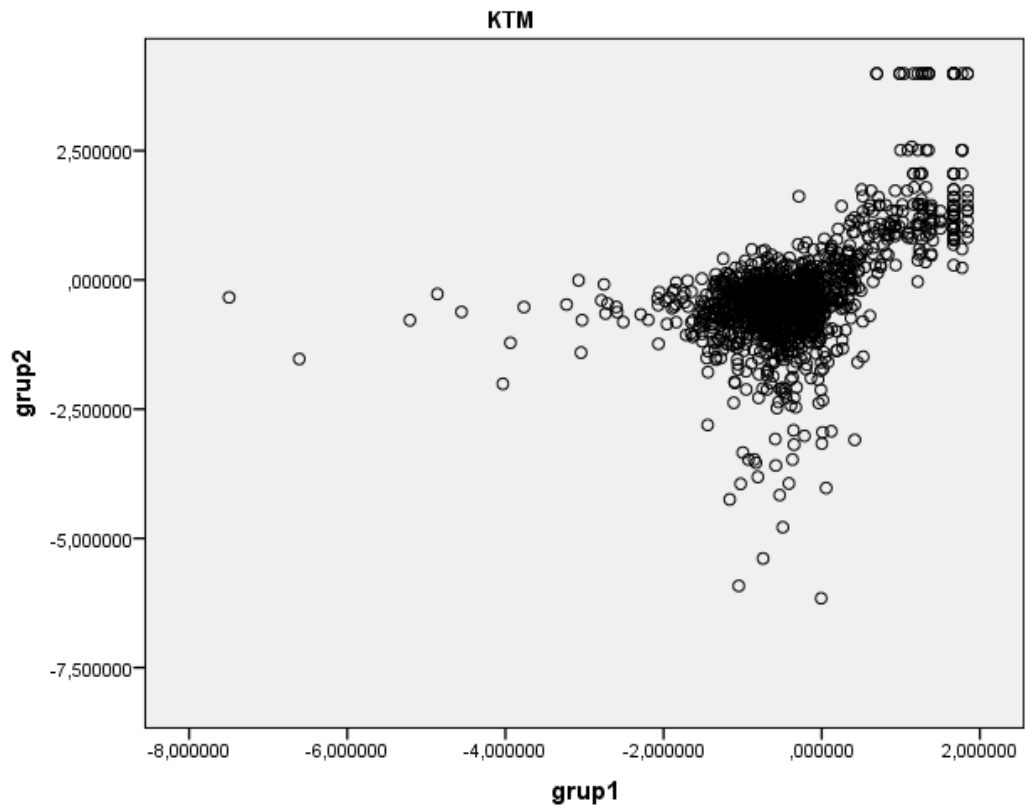
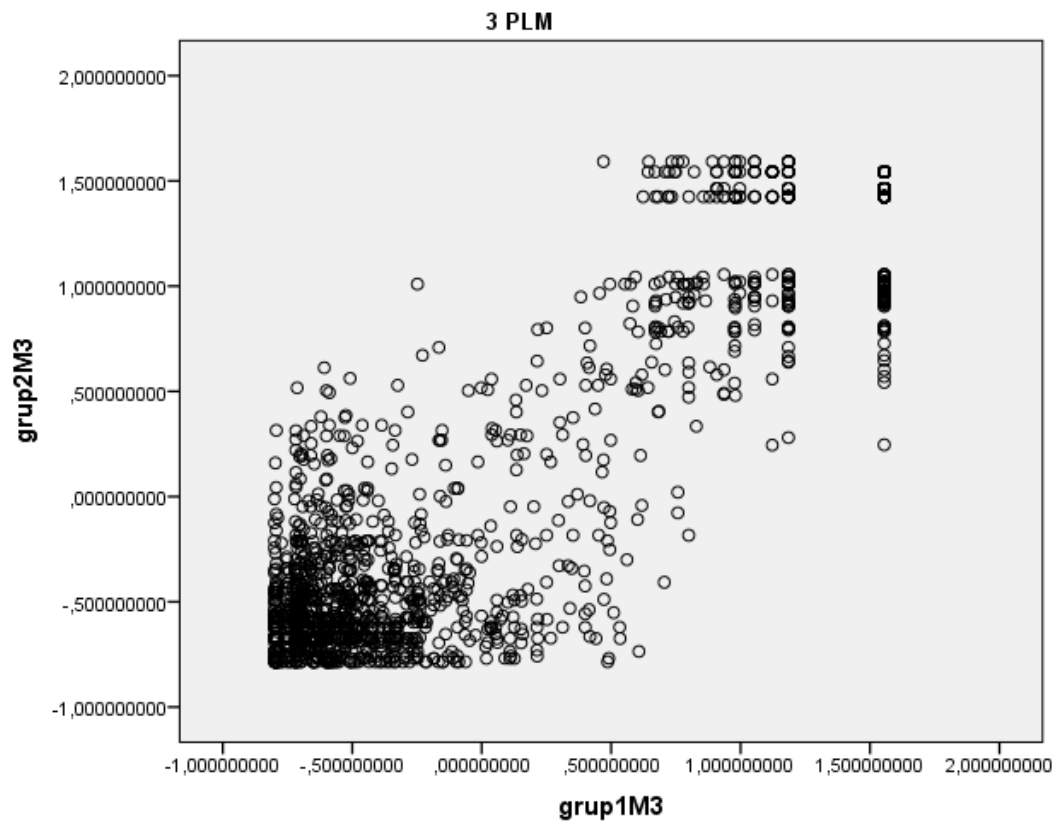
EK 8. STM Altında İki Farklı Gruptan Elde Edilen Verilere Dayalı Kestirilen Madde Parametreleri Arasındaki İlişkilere Yönelik Saçılma Grafikleri





EK 9. İki Farklı Gruptan Elde Edilen Verilere Dayalı Olarak Her Bir Model Altında Kestirilen Yetenek Parametreleri Arasındaki İlişkiler

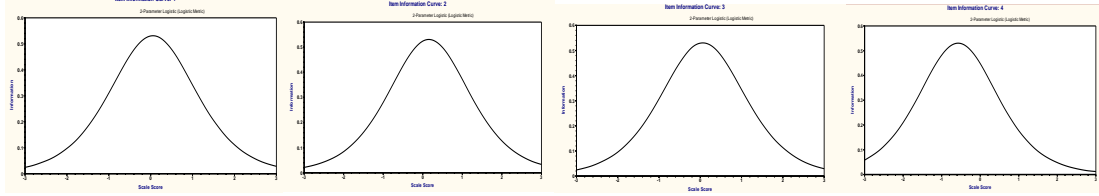




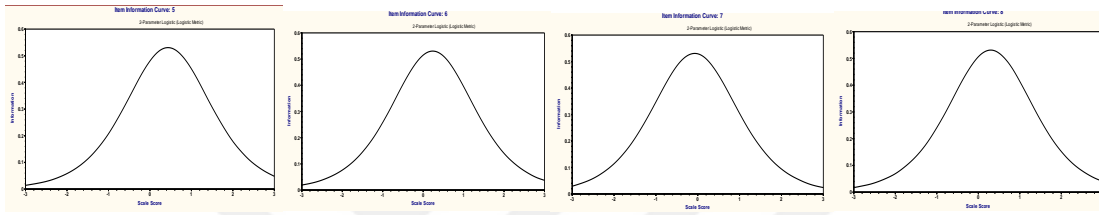
EK 10. 1PLM, 2PLM, 3PLM ve STM'ye Dayalı Hesaplanan Madde Bilgi Fonksiyonları

1 Parametrel Lojistik Model (Normal Metric)

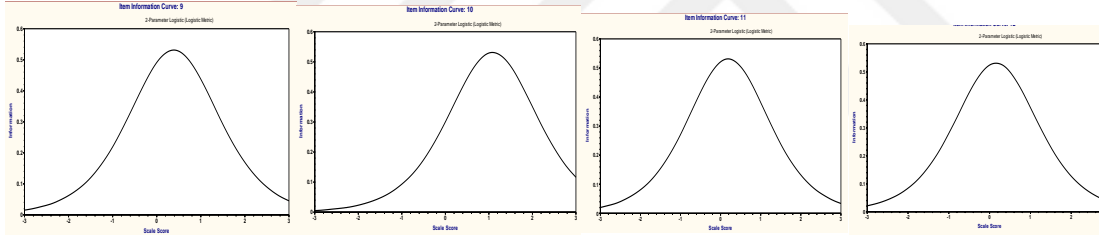
(madde 1-madde 4)



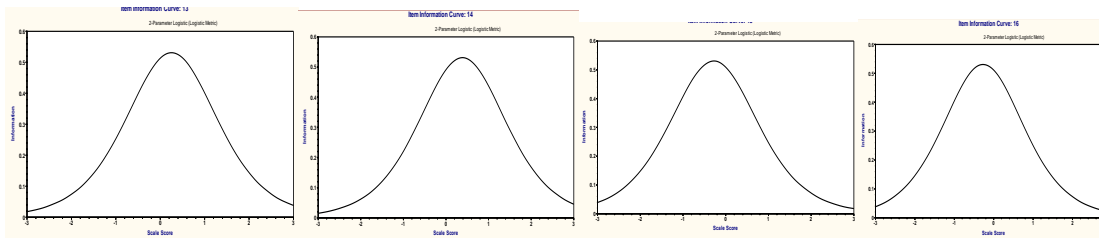
(madde 5-madde 8)



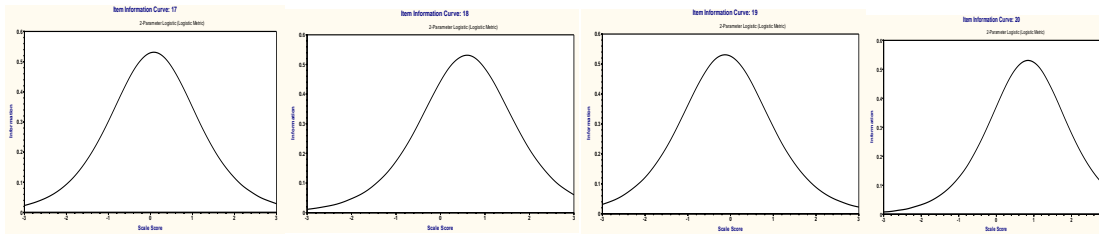
(madde 9-madde 12)



(madde 13- madde 16)

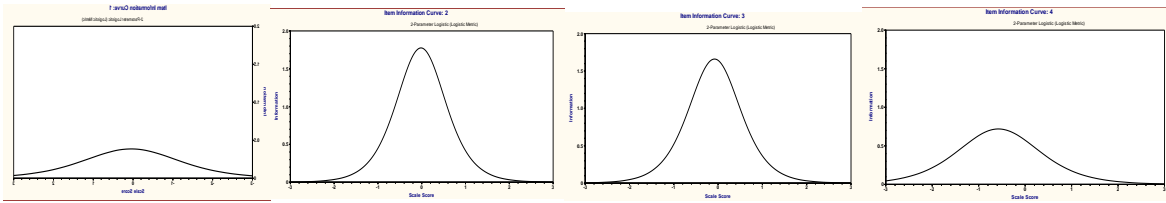


(madde 17-madde 20)

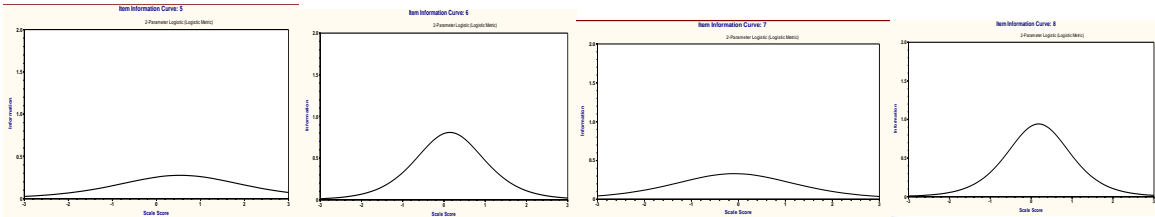


2 Parametrelili Lojistik Model (Logistic Metric)

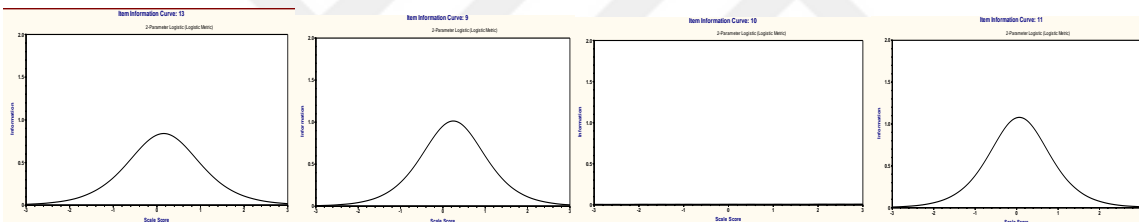
(madde 1-madde 4)



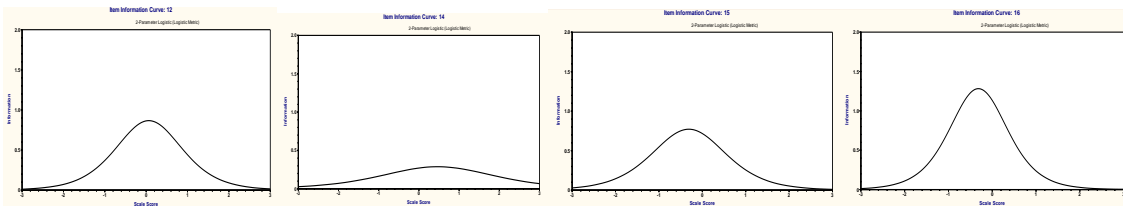
(madde 5-madde 8)



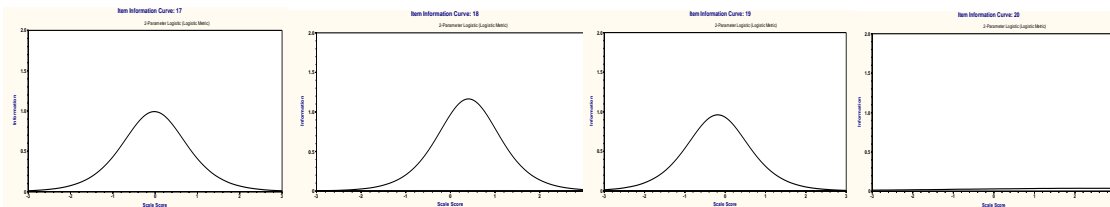
(madde9-madde12)



(madde 13-madde 16)

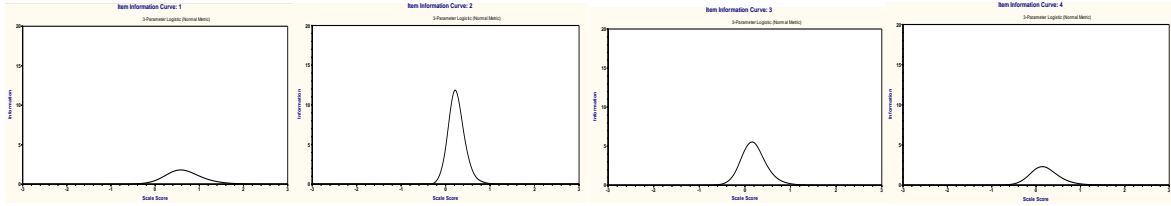


(madde 17-madde 20)

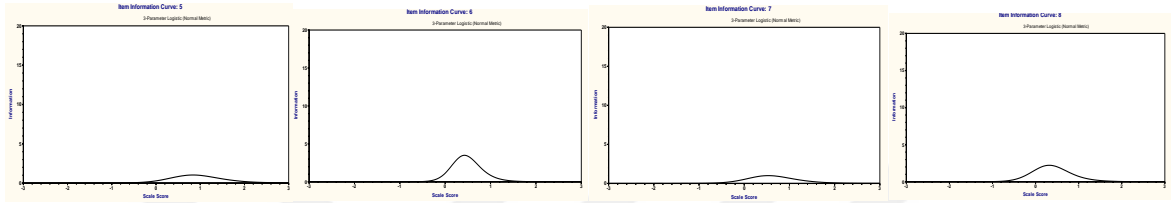


3 Parametrelili Lojistik Model (Logistic Metric)

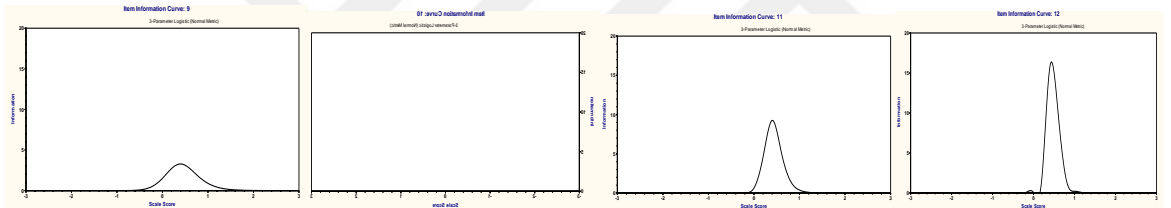
(madde 1-madde 4)



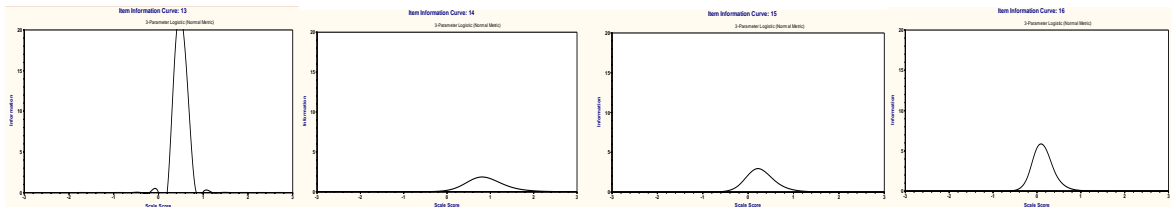
(madde 5-madde 8)



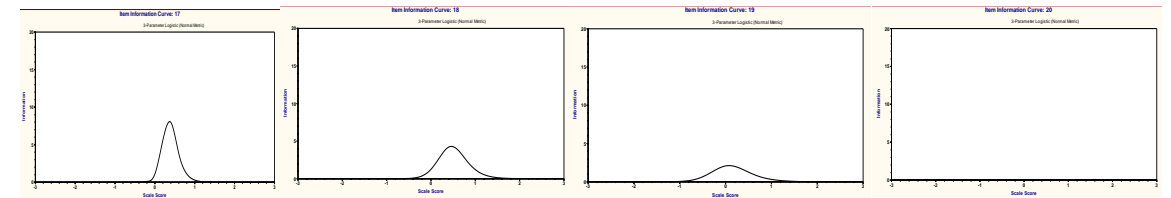
(madde 9-madde 12)



(madde 13-madde 16)

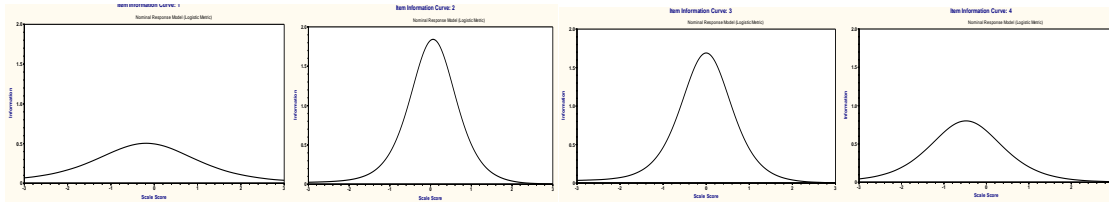


(madde 17-madde 20)

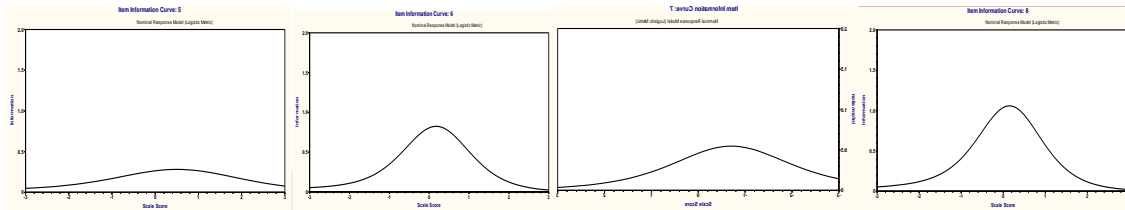


Sınıflamalı Tepki Model

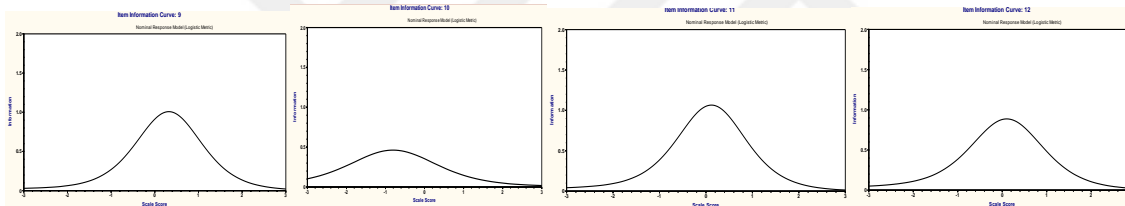
(madde 1-madde 4)



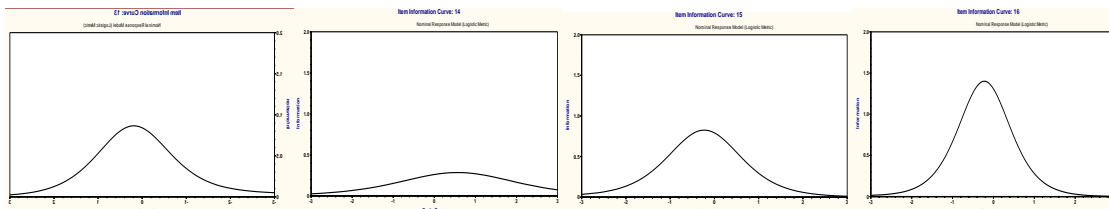
(madde 5-madde 8)



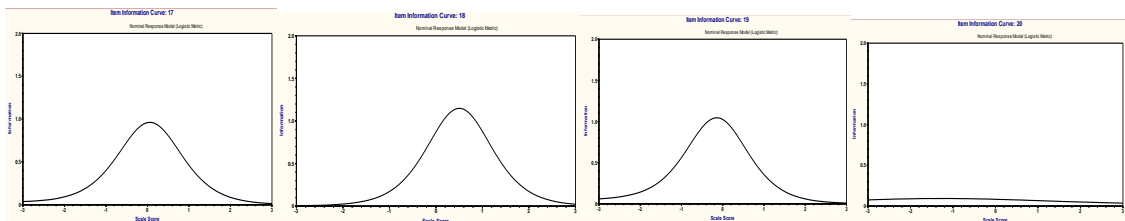
(madde 9-madde 12)



(madde 13-madde 16)



(madde 17-madde 20)



EK 11. Uzman Görüşleri

M1				M2				M3				M4				
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P1	2	1	3	4	4	3	1	2	1	3	4	2	1	3	2	4
P2	1	2	3	4	4	2	1	3	2	3	4	1	2	1	3	4
P3	1	2	3	4	4	2	1	3	2	3	4	1	3	1	2	4
P4	1	2	3	4	4	3	1	2	2	3	4	1	1	3	2	4
P5	1	2	3	4	4	2	3	1	2	3	4	1	3	2	1	4
	1	2	3	4	4	3	1	2	2	3	4	1	3	1	2	4
M5				M6				M7				M8				
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P1	3	4	1	2	1	3	4	2	1	2	3	4	3	4	2	1
P2	1	4	2	3	1	2	4	3	1	2	3	4	3	4	2	1
P3	3	4	1	2	3	4	1	2	1	2	3	4	3	4	2	1
P4	2	4	1	3	1	2	4	3	2	1	3	4	3	4	2	1
P5	1	4	3	2	1	3	4	2	3	2	1	4	2	4	3	1
	3	4	1	2	1	2	4	3	1	2	3	4	3	4	2	1
M9				M10				M11				M12				
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P1	4	1	2	3	3	4	2	1	1	4	2	3	1	2	3	4
P2	4	1	2	3	3	4	2	1	2	4	1	3	1	2	3	4
P3	4	2	3	1	3	4	2	1	1	4	2	3	2	1	3	4
P4	4	3	1	2	3	4	2	1	1	4	2	3	2	1	3	4
P5	4	1	2	3	3	4	1	2	3	4	2	1	1	2	3	4
	4	1	2	3	3	4	2	1	1	4	2	3	1	2	3	4
M13				M14				M15				M16				
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P1	4	2	1	3	4	3	1	2	3	4	2	1	2	1	4	3
P2	4	3	1	2	4	3	2	1	3	4	2	1	2	1	4	3
P3	4	3	1	2	4	2	3	1	3	4	2	1	1	3	4	2
P4	4	2	1	3	4	3	2	1	2	4	3	1	2	1	4	3
P5	4	1	2	3	4	1	3	2	1	4	3	2	1	2	4	3
	4	2	1	3	4	3	2	1	3	4	2	1	2	1	4	3
M17				M18				M19				M20				
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P1	1	2	4	3	3	2	1	4	2	1	4	3	4	3	2	1
P2	3	2	4	1	1	3	2	4	1	2	4	3	4	3	1	2
P3	3	1	4	2	2	3	1	4	2	3	4	1	4	3	1	2
P4	1	3	4	2	2	3	1	4	2	3	4	1	4	3	1	2
P5	1	2	4	3	2	1	3	4	2	3	4	1	4	3	2	1
	1	2	4	3	2	3	1	4	2	3	4	1	4	3	1	2