



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eđitim Bilimleri Anabilim Dalı

Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**MADDE TEPKİ KURAMI'NDA TEST UZUNLUĐU VE ÖRNEKLEM
BÜYÜKLÜĐÜNÜN MODEL VERİ UYUMU, MADDE PARAMETRELERİ
VE STANDART HATA DEĐERLERİNE ETKİSİNİN İNCELENMESİ**

Alper ŞAHİN

Doktora Tezi

Ankara, 2012

MADDE TEPKİ KURAMI'NDA TEST UZUNLUĐU VE ÖRNEKLEM
BÜYÜKLÜĐÜNÜN MODEL VERİ UYUMU, MADDE PARAMETRELERİ VE
STANDART HATA DEĐERLERİNE ETKİSİNİN İNCELENMESİ

ALPER ŞAHİN

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eđitim Bilimleri Anabilim Dalı
Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

Doktora Tezi

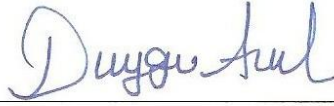
Ankara, 2012

KABUL VE ONAY

Alper Şahin tarafından hazırlanan “Madde Tepki Kuramında Test Uzunluğu ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi” başlıklı bu çalışma, 28.12.2012 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



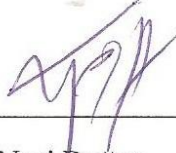
Prof. Dr. Selahattin Gelbal (Başkan)



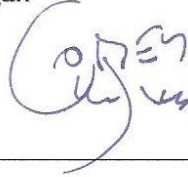
Doç. Dr. Duygu Anıl (Danışman)



Doç. Dr. Hülya Kelecioğlu



Doç. Dr. Nuri Doğan



Yrd. Doç. Dr. Ömer Kutlu

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf Çelik

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun (2) yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

28.12.2012



Alper Şahin

TEŞEKKÜR

Öncelikle doktora tezimi yazarken sürekli desteğini, dostluğunu ve en önemlisi bilgi birikimini benden esirgemeyen tez danışmanım değerli hocam Doç. Dr. Duygu Anıl'a; doktor ünvanını aldığım bu doktora programına beni kabul eden ve hayatımın en önemli adımlarından birisini atmamı sağlayan değerli hocam Prof. Dr. Selahattin Gelbal'a, değerli yorumları ve önerileri ile bu çalışmaya katkıda bulunan sayın Yrd. Doç. Dr. Ömer Kutlu'ya ve değerli hocam Doç. Dr. Nuri Doğan'a ve üzerimde emeği geçen tüm hocalarıma;

Ayrıca Eğitimde Ölçme ve Değerlendirme alanına yönelmem için bana ilham veren ve bu araştırmanın verilerinin toplanmasında büyük katkıları olan yüksek lisans tez danışmanım, değerli hocam Yrd. Doç. Abdullah Ertaş'a; verilerin toplanması aşamasında zamanlarını ayırıp benimle birlikte çalışan Gazi Üniversitesi Yabancı Diller Yüksekokulu Modern Diller Birimi Başkanı sayın Can Gür ve idari koordinatörü İbrahim Pür'e;

Tezimin bulgularını bir bilgisayar programına çevirirken bana sınırsız destek sağlayan arkadaşım Laszlo Menczel'e;

Tez çalışmamda parametre kestirimi yaptığım ticari yazılımı ücretsiz kullanmamı sağlayarak bir anlamda tezimin yazılım sponsorluğunu üstlenen Dr. David Weiss ve ekibi Assessment Systems Corporation'a;

Tezimi yazdığım en stresli günlerde, bana ve iş arkadaşlarıma stresten uzak, mutlu bir iş ortamı sağlayan Birim Koordinatörüm Meral Kızrak'a;

Son olarak akademik hayatım boyunca benden desteğini hiç esirgemeyen, bana her zaman ilham kaynağı olan sevgili eşim Dr. İclal Şahin'e; doktora çalışmalarımı yürüttüğüm son iki sene zarfında hayatının en güzel zamanlarında onunla daha çok vakit geçiremediğim ama bunu her zaman anlayışla karşılayan canım oğlum Tuna Şahin'e; hayatım boyunca her daim yanımda olup, bana sürekli destek olan ve her zaman beni karşılıksız seven babam Mehmet Şahin ve annem Aysel Şahin'e teşekkür ediyorum.

ÖZET

ŞAHİN, Alper. *Madde Tepki Kuramı'nda test uzunluğu ve örneklem büyüklüğünün model veri uyumu, madde parametreleri ve standart hata değerlerine etkisinin incelenmesi*, Doktora Tezi, Ankara, 2012.

Bu araştırmanın amacı, Madde Tepki Kuramı'nda test uzunluğu ve örneklem büyüklüğünün model veri uyumu, madde parametreleri ve standart hata değerlerine etkisinin incelenmesidir. Bu amaçla çalışma evreninden tabakalı seçkisiz örnekleme yöntemiyle küçük (150, 250, 350, 500, 750) ve büyük (1000, 2000, 3000, 5000) örneklemler oluşturulmuş ve bu örneklemlerde 10, 20 ve 30 maddelik üç farklı test uzunluğunda madde parametreleri kestirimi yapılmıştır. Bulgular model veri uyumu, madde parametreleri ve standart hata değerleri açısından karşılaştırılmıştır.

Örneklem büyüklüğü ve test uzunluğunun model veri uyumu üzerine etkisinin incelenmesi amacıyla χ^2/sd oranı, standartlaştırılmış artık analizi, G^2 olabilirlik oranı, test karakteristik eğrileri incelenmesi şeklinde dört farklı yöntemle farklı örneklem ve test uzunluklarında model veri uyumu incelenmiştir. Elde edilen bulgular doğrultusunda, küçük örneklemlerde kullanılan Standartlaştırılmış artık analizi (SAA) yönteminin örneklem büyüklüğünden ve kısmen de olsa test uzunluğundan etkilendiği; ancak bu yöntemin 10 maddelik bir testte 500, 20 ve 30 maddelik testlerde ise 750 kişilik örneklemlerden fazla örneklemlerde örneklem büyüklüğünden etkilenmediği belirlenmiştir. χ^2/sd yönteminin ise test uzunluğu ve örneklem büyüklüğünden kısmen etkilendiği; 20 maddelik testlerde örneklem büyüklüğünden etkilenmeyen bu yöntemin, madde sayısı 10 olduğu durumlarda, 3000 kişilik örneklem büyüklüğünden sonra; madde sayısı 30 olduğu durumlarda ise 5000 kişilik örneklem büyüklüğünden sonra örneklem büyüklüğünden etkilenmeye başladığı ve dolayısıyla bu yöntemin 10 maddelik testlerde en çok 3000 kişilik, 30 maddelik testlerde ise en çok 5000 kişilik örneklemlerle kullanılmasının uygun olacağı sonucuna varılmıştır. G^2 Olabilirlik Oranı, test karakteristik eğrilerinin incelenmesi yöntemlerinin test uzunluğu ve örneklem büyüklüğünden etkilenmediği ve dolayısıyla çalışmada kullanılan farklı test uzunlukları ve örneklem büyüklüklerinde kullanılmalarının uygun olabileceği sonucuna varılmıştır.

Test uzunluğu ve örneklem büyüklüğünün madde parametreleri üzerindeki etkisinin araştırılması amacıyla çalışma evreninden elde edilen madde parametreleri ile küçük ve büyük örneklemde elde edilen madde parametreleri arasındaki korelasyon katsayıları incelenmiştir. Elde edilen bulgular doğrultusunda, 150 kişilik örneklemde, 10, 20 veya 30 maddelik test uzunluklarında; MTK'nın 1PL modeli kullanılarak kestirilen "b" parametreleri ile çalışma evreninden elde edilen "b" parametreleri arasında 0,94 ile 0,96 arasında değişen oranlarda yüksek korelasyon katsayıları elde edilmiş ve dolayısıyla 10, 20 ve 30 maddelik testlerde 1PL modelde parametre kestirimi yapabilmek için 150 kişilik örneklemin yeterli olabileceği sonucuna varılmıştır. 2PL modelde uygun örneklem büyüklüğünün belirlenmesinde "a" parametresi kestirimlerinin önemli rol oynadığı, "b" parametresinin her örneklem büyüklüğü ve test uzunluğunda yüksek korelasyon katsayıları ($r_b=0,94-1,00$) ile kestirilebildiği belirlenmiş; dolayısıyla uygun örneklem büyüklüğünün belirlenmesi için "a" parametrelerinden elde edilen korelasyon değerleri referans alınmıştır. Bu bilgiler ışığında, 2PL modelde madde parametresi kestirimi yapılabilmesi için, 10 maddelik bir test için en az 750 ($r_a=0,706$, $r_b=0,999$); 20 maddelik bir test için en az 500 ($r_a=0,795$, $r_b=0,998$) ve 30 maddelik bir test için en az 250 ($r_a=0,830$, $r_b=0,978$) kişilik örneklemelerin yeterli olabileceği sonucuna varılmıştır. 3PL modelde parametre kestiriminde örneklem büyüklüğünün belirlenmesinde 10 ve 20 maddelik testlerde "a" parametresi, 30 maddelik testte ise "c" parametresi belirleyici rol oynamıştır. Bu bilgiler ışığında MTK'da madde parametreleri kestiriminde 10 maddelik bir test için en az 750 ($r_a=0,901$, $r_b=0,998$, $r_c=0,922$), 20 maddelik bir test için en az 750 ($r_a=0,840$, $r_b=0,995$, $r_c=0,829$) ve 30 maddelik bir test için en az 350 ($r_a=0,749$, $r_b=0,987$, $r_c=0,734$) kişilik örneklemelerin yeterli olabileceği sonucuna varılmıştır.

Test uzunluğu ve örneklem büyüklüğünün standart hata değerlerine etkisinin incelenmesi amacıyla farklı test uzunluklarında ve örneklem büyüklüklerinde elde edilen madde parametrelerine ait ortalama standart hata değerleri birbirleriyle karşılaştırılmıştır. Araştırma bulguları ışığında, 10, 20 ve 30 maddelik testlerde en yüksek hata değerlerinin "c" parametresinden elde edildiği; tüm parametrelere ait ortalama standart hata değerlerinin örneklem büyüklüğünün artmasından olumlu etkilendiği ve örneklem büyüklüğünün artmasından en olumlu etkilenen parametrenin

“c” parametresi olduđu sonucuna varılmıřtır. Ancak kk rneklemlerde madde sayısı 30’a ulařtıđında 20 maddelik testten elde edilen ortalama standart hata deđerlerinden ok daha yksek ortalama standart hata deđerleri elde edilmiřtir. Dolayısıyla madde sayısının artmasının kk rneklemlerde ortalama standart hata deđerlerini 20 maddeden sonra olumsuz etkilediđi, bu olumsuz etkinin bk rneklemlerde kk rneklemlere oranla olduka sınırlı olduđu grlmektedir. Bu nedenle kk rneklemlerde madde sayısının en fazla 30 ile sınırlandırılmasının ortalama standart hata deđerlerinin kabul edilebilir dzeylerde kalmasını sađlamak aısından nemli olduđu sonucuna varılmıřtır.

Anahtar Szckler

Madde Tepki Kuramı, Model Veri Uyumu, rnekleme, rnekleme Byklđ, Test Uzunluđu, Madde Sayısı, Standart Hata Deđerleri

ABSTRACT

SAHIN, Alper. *An Investigation on the effects of test length and sample size in Item Response Theory on model-data fit, item parameters and standard error values*, Doctoral Dissertation, Ankara, 2012.

The purpose of this study was to investigate the effect of test length and sample size in Item Response Theory (IRT) on model-data fit, item parameters and standard error values. For this purpose, small (150, 250, 350, 500, 750) and large (1000, 2000, 3000, 5000) samples were drawn from the research population through stratified random sampling method. Item parameter estimation has been done in these samples in three (10, 20 30 item) different test lengths. The findings were compared in terms of model-data fit, item parameters, and standart error values.

Model-data fit in different sample sizes and test lengths was analyzed through four different methods, namely, χ^2 /sd ratio, standardized residual analysis, G^2 likelihood ratio, investigation of the test characteristics curves so as to assess the effect of sample size and test length on model-data fit. Based on the findings of the study, it was concluded that standardized residuals analysis (SRA) method was affected by the sample size and partially from the test length. In addition, this method was found not to be affected by the sample size above the sample size of 500 for a 10 item test and above the sample size of 750 in 20 or 30 item tests. Moreover, it was concluded that χ^2 /sd method was partially affected from test length and sample size; although it wasn't affected by the sample size in 20 item test data; it was found that it was affected by the sample size in the sample sizes of above 3000 with 10 item test data and in sample sizes of above 5000 with 30 item test data. Accordingly, it was concluded that it would be better to use this method in sample size of at most 3000 with a 10 item test and a sample size of 5000 with a 30 item test. It was also concluded that the methods of G^2 likelihood ratio and analysis of test characteristics curves were found not to be affected by the sample size and test length; and accordingly, using these methods in different test lengths and sample sizes employed in the study could be appropriate.

In order to investigate the effect of test length and sample size on item parameters, correlation coefficients between the item parameters obtained from the research

population and obtained from the small and large sample sizes were examined. Based on the findings, in a sample size of 150 people and in 10, 20, or 30 item test lengths, high correlation coefficients ranging between 0,94 and 0,96 were obtained between "b" parameters estimated through 1PL model of IRT and "b" parameters estimated from the research population. Accordingly, it was concluded that a sample of 150 people was thought to be sufficient in estimating parameters in 10, 20, and 30 item tests in 1PL model. It was found that "a" parameter estimates played a crucial role in determining an appropriate sample size in 2PL model; "b" parameter was able to be estimated with high correlation coefficients ($r_b=0,94-1,00$) in all sample sizes and test lengths; so, correlation coefficients obtained from "a" parameters were used as reference to identify the appropriate sample size for 2PL model. Based on these findings, it was concluded that to estimate parameters in 2PL model, at least a sample size of 750 for a 10 item test ($r_a=0,706$, $r_b=0,999$); 500 for a 20 item test ($r_a=0,795$, $r_b=0,998$); and 250 for a 30 item test ($r_a=0,830$, $r_b=0,978$) could be sufficient. In the identification of the appropriate sample size for parameter estimation in 3PL model, "a" parameter played a significant role in 10 and 20 item tests and "c" parameter played a crucial role in 30 item tests. Based on these findings, it was concluded that at least a sample of 750 for a 10 item test ($r_a=0,901$, $r_b=0,998$, $r_c=0,922$); 750 for a 20 item test ($r_a=0,840$, $r_b=0,995$, $r_c=0,829$); and 350 for a 30 item test ($r_a=0,749$, $r_b=0,987$, $r_c=0,734$) could be sufficient to estimate item parameters in IRT.

In order to investigate the effect of test length and sample size on standard error values, average standard error values obtained from different test lengths and sample sizes were compared with each other. Based on the findings, it was concluded that the highest error value was obtained from "c" parameter in 10, 20, and 30 item tests; the average standard error value of all parameters was positively affected by the increase in the sample size; and it was the "c" parameter which was the most positively affected parameter from the increase in the sample size. However, when the number of items reached to 30 in small sample sizes, higher average standard error values were obtained compared to the standard error value obtained in 20 item test. Accordingly, an increase in number of the items was found to affect standard error values negatively in small sample sizes in test lengths above 20 items, and this negative effect was found to be quite limited in large

sample sizes compared to the small sample sizes. Therefore, it was concluded that limiting the number of items to 30 in small sample sizes was necessary in order to hold standard error values at reasonable levels.

Keywords

Item Response Theory, Model-Data Fit, Sampling, Sample Size, Test Length, Number of Items, Standard Error Values

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	vii
İÇİNDEKİLER	x
KISALTMALAR	xii
TABLolar DİZİNİ	xiii
ŞEKİLLER DİZİNİ	xvi
BÖLÜM I	1
GİRİŞ	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı ve Önemi	17
1.3. Problem Cümlesi	18
1.4. Sayıtlar	19
1.5. Sınırlılıklar	19
1.6. İlgili Araştırmalar	20
BÖLÜM II	25
YÖNTEM	25
2.1. Araştırmanın Türü	25
2.2. Evren ve Örneklem	25

2.3. Araştırma Verilerinde MTK Varsayımlarının Sınanması	29
2.4. Verilerin Çözümlemesi	34
BÖLÜM III	38
BULGULAR VE YORUMLAR	38
3.1. Birinci Alt Probleme İlişkin Bulgular ve Yorumlar	38
3.2. İkinci Alt Probleme İlişkin Bulgular ve Yorumlar	86
3.3. Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar	98
BÖLÜM IV	108
SONUÇLAR VE ÖNERİLER	108
4.1. Sonuçlar	108
4.2. Öneriler	115
KAYNAKÇA	116
Ek 1 : Küçük Örneklemelere Ait Betimsel İstatistikler.....	123
Ek 2 : Büyük Örneklemelere Ait Betimsel İstatistikler.....	124
Ek 3 : Çalışmada Kullanılan Test Maddelerine Ait Faktör Yükleri	125
Ek 4 : Kestirilen Ortalama Madde Parametreleri Değerleri	126
Ek 5 : χ^2/sd yöntemine göre Madde-Model Uyumu	127
Ek 6 : Tetrakorik Korelasyon Matrisine Dayalı Faktör Analizi için	
SPSS Syntax'leri	128
Ek 7 : Büyük ve Küçük Örneklemelere Ait Histogram Grafikleri.....	129
Ek 8 : Araştırmanın Alt Problemlerine İlişkin Toplu Bulgular	132

KISALTMALAR

1PL	:	Bir-Parametrelili Lojistik
2PL	:	İki-Parametrelili Lojistik
3PL	:	Üç-Parametrelili Lojistik
SAA	:	Standardlaştırılmıř Artık Analizi
KTK	:	Klasik Test Kuramı
MTK	:	Madde Tepki Kuramı
MKE	:	Madde Karakteristik Eğriři
TKE	:	Test Karakteristik Eğriři
BEYOM	:	Birleřik En Yüksek Olabilirlik Metodu
MEYOM	:	Marjinal En Yüksek Olabilirlik Metodu
ŞEYOM	:	Şartlı En Yüksek Olabilirlik Metodu

TABLOLAR DİZİNİ

Tablo 2.1.	Çalışma Evrenine Ait Veride Öğrencilerin Fakülteleere Göre Dağılımı	27
Tablo 2.2.	Küçük Örneklemeler ve Çalışma Evrenine Ait Karşılaştırmalı Fakülte Dağılım Yüzdeleri.....	28
Tablo 2.3.	Büyük Örneklemeler ve Çalışma Evrenine Ait Karşılaştırmalı Fakülte Dağılım Yüzdeleri.....	29
Tablo 2.4.	Çalışmada Kullanılan Veri Dosyaları	34
Tablo 3.1.	Küçük Örneklemelerde 10 Maddelik Teste Ait χ^2 /sd Değerleri	39
Tablo 3.2.	Küçük Örneklemelerde 20 Maddelik Teste Ait χ^2 /sd Değerleri.....	41
Tablo 3.3.	Küçük Örneklemelerde 30 Maddelik Teste Ait χ^2 /sd Değerleri	43
Tablo 3.4.	Küçük Örneklemelerde 10 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler.....	45
Tablo 3.5.	Küçük Örneklemelerde 20 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler.....	46
Tablo 3.6.	Küçük Örneklemelerde 30 maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler	47
Tablo 3.7.	Küçük Örneklemelerde 10 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	48
Tablo 3.8.	Küçük Örneklemelerde 20 maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	49
Tablo 3.9.	Küçük Örneklemelerde 30 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	49

Tablo 3.10. Küçük Örneklerde 10 Maddelik Teste Ait Model Veri Uyumu	60
Tablo 3.11. Küçük Örneklerde 20 Maddelik Teste Ait Model Veri Uyumu	61
Tablo 3.12. Küçük Örneklerde 30 Maddelik Teste Ait Model Veri Uyumu	62
Tablo 3.13. Büyük Örneklerde 10 Maddelik Teste Ait x^2/sd Değerleri.....	63
Tablo 3.14. Büyük Örneklerde 20 Maddelik Teste Ait x^2/sd Değerleri	65
Tablo 3.15. Büyük Örneklerde 30 Maddelik Teste Ait x^2/sd Değerleri	66
Tablo 3.16. Büyük Örneklerde 10 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler.....	68
Tablo 3.17. Büyük Örneklerde 20 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler.....	69
Tablo 3.18. Büyük Örneklerde 30 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler.....	70
Tablo 3.19. Büyük Örneklerde 10 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	71
Tablo 3.20. Büyük Örneklerde 20 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	72
Tablo 3.21. Büyük Örneklerde 30 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri	73
Tablo 3.22. Büyük Örneklerde 10 Maddelik Teste Ait Model Veri Uyumu	83
Tablo 3.23. Büyük Örneklerde 20 Maddelik Teste Ait Model Veri Uyumu	84
Tablo 3.24. Büyük Örneklerde 30 Maddelik Teste Ait Model Veri Uyumu	85
Tablo 3.25. Küçük Örnekler ve Çalışma Evreninden 10 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	87

Tablo 3.26. Küçük Örneklem ve Çalışma Evreninden 20 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	89
Tablo 3.27. Küçük Örneklem ve Çalışma Evreninden 30 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	90
Tablo 3.28. Büyük Örneklem ve Çalışma Evreninden 10 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	92
Tablo 3.29. Büyük Örneklem ve Çalışma Evreninden 20 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	94
Tablo 3.30. Büyük Örneklem ve Çalışma Evreninden 30 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi	96
Tablo 3.31. Küçük Örneklemde 10 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	99
Tablo 3.32. Küçük Örneklemde 20 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	101
Tablo 3.33. Küçük Örneklemde 30 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	102
Tablo 3.34. Büyük Örneklemde 10 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	103
Tablo 3.35. Büyük Örneklemde 20 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	105
Tablo 3.36. Büyük Örneklemde 30 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri	106

ŞEKİLLER DİZİNİ

Şekil 1.	Madde Karakteristik Eğrisi Örneği	3
Şekil 2.	20 Maddelik Bir Test İçin Örnek Test Karakteristik Eğrisi	4
Şekil 3.	Madde Tepki Kuramı'nda Madde Parametrelerinin Değişmezliği	6
Şekil 4.	Kolay Ve Zor İki Maddeye Ait Örnek Madde Karakteristik Eğrileri	8
Şekil 5.	Ayrırdediciliği Yüksek (I) Ve Düşük(II) İki Maddeye Ait Örnek Madde Karakteristik Eğrileri	9
Şekil 6.	Farklı "c" Parametrelerine Sahip İki Maddeye Ait Örnek Madde Karakteristik Eğrileri	11
Şekil 7.	30 Maddelik Testte Ait Özdeğer Grafiği	31
Şekil 8.	20 Maddelik Testte Ait Özdeğer Grafiği	32
Şekil 9.	10 Maddelik Testte Ait Özdeğer Grafiği	33
Şekil 10.	Küçük Örneklerde 10 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	51
Şekil 11.	Küçük Örneklerde 10 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	52
Şekil 12.	Küçük Örneklerde 10 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	53
Şekil 13.	Küçük Örneklerde 20 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	54

Şekil 14. Küçük Örneklerde 20 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	55
Şekil 15. Küçük Örneklerde 20 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	56
Şekil 16. Küçük Örneklerde 30 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	57
Şekil 17. Küçük Örneklerde 30 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	58
Şekil 18. Küçük Örneklerde 30 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	59
Şekil 19. Büyük Örneklerde 10 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	74
Şekil 20. Büyük Örneklerde 10 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	75
Şekil 21. Büyük Örneklerde 10 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	76
Şekil 22. Büyük Örneklerde 20 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	77
Şekil 23. Büyük Örneklerde 20 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	78
Şekil 24. Büyük Örneklerde 20 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	79

Şekil 25. Büyük Örneklerde 30 Maddelik Testten 1PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	80
Şekil 26. Büyük Örneklerde 30 Maddelik Testten 2PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	81
Şekil 27. Büyük Örneklerde 30 Maddelik Testten 3PL Modelde Elde Edilen Test Karakteristik Eğrileri.....	82

BÖLÜM I

GİRİŞ

Bu bölümde, araştırmanın problem durumuna, problem cümlesine ve alt problemlerine, amacı ve önemine, sayıltılarına, sınırlılıklarına ve ilgili araştırmalara yer verilmiştir.

1.1. PROBLEM DURUMU

Klasik Test Kuramı (KTK), test geliştirme ve test puanlarının yorumlanması çalışmalarında ölçme ve değerlendirme uzmanlarına uzun yıllar başarıyla hizmet etmiştir (Hambleton, Swaminathan ve Rogers, 1991; Embretson ve Reise, 2000). Ancak KTK, test geliştirme, test eşitleme ve yanlılık analizi gibi araştırmaların gereksinimlerine tatmin edici çözümler üretmekte başarısız olmuştur (Hambleton ve Swaminathan, 1985). Bu sorunlara çözüm olarak 1930'lu yılların sonunda Madde Tepki Kuramı (MTK) geliştirilmiştir. MTK sayesinde araştırmacılar, madde ve birey parametrelerini birbirinden bağımsız olarak kestirme imkanına kavuşmuşlardır. Özellikle, Lawley'in 1943 yılında yaptığı çalışma, maddeyi esas alan bu teorinin başlangıcını oluşturmuştur (Baker, 2004). Bu çalışmayı Lord'un (1952) çalışması takip etmiş ve MTK günümüzde test yayımcıları, eğitim kurumları, silahlı kuvvetler ve sanayi kuruluşları tarafından test geliştirme, test eşitleme ve yanlı maddelerin belirlenmesi çalışmalarında yaygın olarak kullanılan bir kuram haline gelmiştir (Hambleton, Swaminathan ve Rogers, 1991).

Test geliştirme süreçleriyle ilgili sorunlara çözümler sunan MTK, özellikle yalnızca küçük örneklemelere ulaşabilen araştırmacıları sınırlayan önemli bir sorunu da beraberinde getirmiştir. Bu sorun, MTK'da madde ve yetenek parametreleri kestirimi için büyük örneklemelere ihtiyaç duyulmasıdır (Hambleton, 1989; Lenth, 2001). Bu sorundan dolayı MTK temel alınarak yapılan araştırmalarda örneklem büyüklüğü kritik bir unsur olarak gösterilmektedir (Orlando, 2004). Bu unsur, madde ölçümleme çalışmalarını etkileyecek en önemli faktörlerin başında gelmektedir (Torre ve Hong,

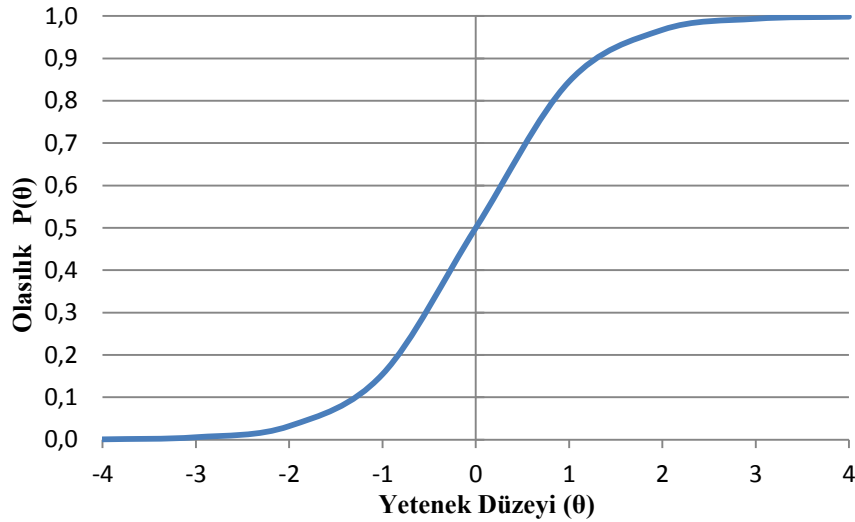
2010). Ancak bu kadar önemli bir konu üzerine şaşırtıcı bir şekilde oldukça az çalışma vardır (Lenth, 2001). Hatta, MTK'nın işe koşulduğu çalışmalarda, hangi örneklem büyüklüklerinin yeterli olduğuna yönelik, araştırmacıların üzerinde anlaştığı bir yönerge yoktur (De Ayala, 2009; Kirisci et al.,2001; Reise ve Yu, 1990). Şöyle ki, MTK'nın gelişmesine önemli katkıları olan Lord (1968), yıllardır birçok araştırmacı tarafından referans alınan çalışmasında; parametre kestirimini Birleşik En Yüksek Olabilirlik Metodu (BEYOM) kullanarak yapmış ve madde sayısı 50'ye, örneklem büyüklüğü 1000'e ulaşmadan kestirilen parametrelerin hata oranlarının oldukça yüksek çıktığı sonucuna varmıştır. Alanyazında yaygın olarak referans alınan ve parametre kestiriminde yine BEYOM kullanılan bir diğer çalışmada, Hulin, Lissak ve Drasgow (1982), parametre kestirimi için 2000 kişilik bir örneklem büyüklüğünden fazlasının gereksiz olduğu sonucuna varmıştır. Bu iki temel araştırmanın yanısıra, alanyazındaki bazı araştırmalarda örneklem büyüklüğünün, araştırmada kullanılan modele göre farklılık gösterebileceği belirlenmiştir. Örneğin bir-parametrelili lojistik model (1PL model) için 200 (Wright ve Stone, 1979), 500 (Hulin, Lissak ve Drasgow, 1982) veya iki-parametrelili lojistik model (2PL model) için 1000 (Ree ve Jensen, 1980) ve üç parametrelili lojistik model (3PL model) için 1000 (Lord, 1968) veya 10.000 ve daha fazla (Thissen ve Wainer, 1982) kişiden oluşan örneklem büyüklüklerinin yeterli olacağını belirten araştırmalar yapılmıştır. Bu durumu Hambleton (1989) bu alanda araştırmaya olan ihtiyacın bir işareti olarak kabul etmektedir. Daha sonra yapılan araştırmalar da bu konudaki belirsizliği devam ettirecek düzeydedir. Örneğin, Mislevy (1986) çalışmasında 1000 kişilik bir örneklem kullanmasına rağmen, Scherbaum ve diğerleri (2006) "Lider-üye etkileşimi" adlı çalışmalarında 445 kişilik bir örneklem kullanmışlardır. Ayrıca bazı araştırmacılar (De Ayala, 2009; Russell, 2002) 250 veya 500 kişilik örneklem büyüklüklerinin parametre kestirimi için yeterli olduğunu düşünürken, bazı araştırmacılar tahmin parametresinin kestirilmesinin güçlüğünden dolayı, 3PL model için 3000 kişiyi aşan örneklem büyüklüklerinin kullanılmasını (Drasgow ve diğerleri, 1995) önermişlerdir.

Tüm bu bilgiler ışığında, MTK'nın avantajlarından en önemlisi olan parametre kestiriminde değişmezliğin sağlanması için yeterli büyüklükteki örneklemin kaç kişiden

oluşması gerektiği hususundaki belirsizliğin, araştırmalarında MTK'yı kullanacak özellikle sadece küçük örneklem büyüklüklerine ulaşabilen araştırmacıların önünde büyük bir engel oluşturmakta olduğu söylenebilir. Bu nedenle MTK'da değişmez parametre kestirimi için en az kaç kişilik örneklem büyüklüğünün yeterli olabileceğinin belirlenmesi ve araştırmacıların bu konudaki sorunlarını hafifletecek geniş kapsamlı bir araştırmaya ihtiyaç olduğu düşünülmektedir. Bu sorunun daha ayrıntılı olarak anlaşılabilmesi için MTK'nın bazı temel özelliklerinin bilinmesinde yarar vardır.

1.1.1. Madde Tepki Kuramı (MTK)

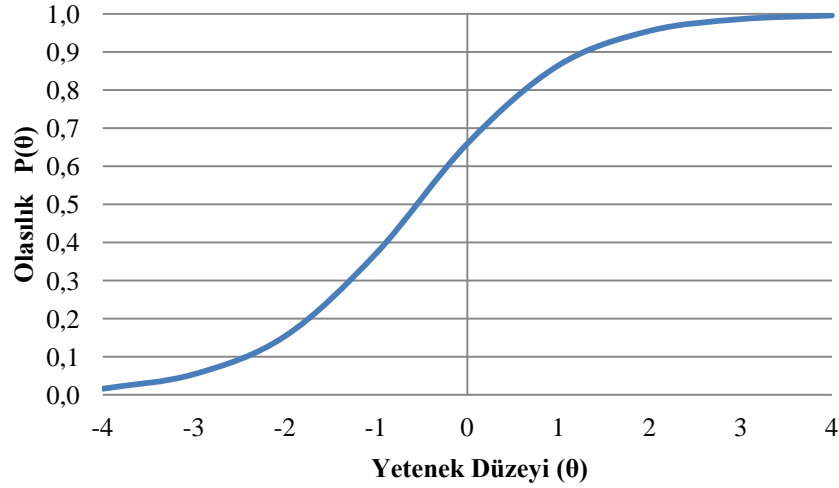
Madde Tepki Kuramı, bireyin bir maddeye verdiği tepki ile o bireyin yetenek veya örtük özellik olarak da adlandırılan özelliği arasındaki ilişkiyi açıklayan bir modeldir (Reeve, 2006). Bu örtük özellik teta (θ) ile gösterilir. Steinberg ve Thissen (1995)'e göre teta maddeler arası kovaryansı açıklayan tek boyutlu ve sürekli bir yapıdır (aktaran; Reeve, 2006). Teta yetenek düzeyi olarak da isimlendirilir. Yüksek teta düzeyine sahip bireylerin maddelere doğru yanıt verme olasılığı da daha fazladır. Örtük özellikler teorisinde, madde puanının teta vektörü üzerindeki regresyonu “madde karakteristik fonksiyonu” ya da “madde karakteristik eğrisi” olarak adlandırılır ve $P(\theta)$ ile gösterilir. Şekil 1’de örnek bir madde karakteristik eğrisi (MKE) gösterilmiştir.



Şekil 1. Madde karakteristik eğrisi örneği.

Şekil 1’de görülen MKE her bir madde için ayrı ayrı hesaplanır ve Baker’a (2001) göre bu eğri MTK’nın temel taşıdır. MKE, belirli bir yetenek düzeyine (θ) sahip bireylerden seçkisiz olarak belirlenmiş bir bireyin o maddeyi doğru cevaplama olasılığını göstermektedir. Şekil 1’de görüldüğü gibi MKE üzerinde yetenek düzeyi azaldıkça, o maddeye doğru yanıt verme olasılığı sıfıra yaklaşmakta; yetenek düzeyi arttıkça maddeye doğru yanıt verme olasılığı 1’e yaklaşmaktadır.

Madde karakteristik eğrileri test karakteristik eğrilerini oluşturmak için toplanabilir (Demars, 2010). Test karakteristik eğrisi, her yetenek düzeyinde maddelerin doğru cevaplanma olasılıklarının ortalamasını göstermektedir. 20 maddelik bir test için test karakteristik eğrisi örneği Şekil 2’de gösterilmiştir.



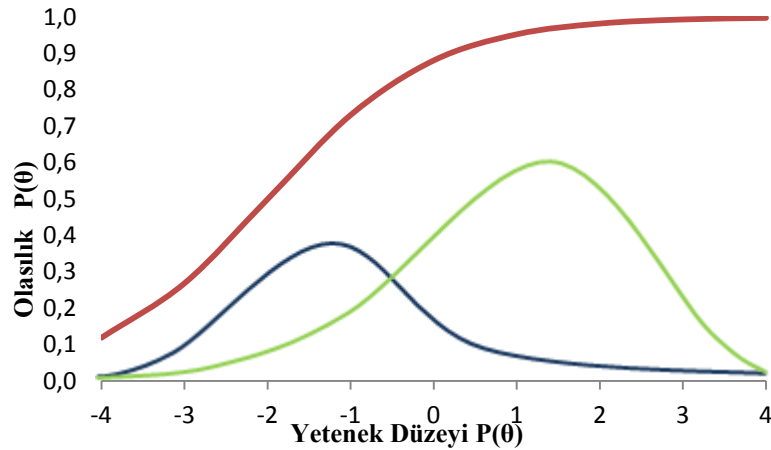
Şekil 2. 20 maddelik bir test için örnek test karakteristik eğrisi.

Şekil 2 incelendiğinde, $\theta_A = 1,0$ yetenek düzeyine sahip A deneğinin testin tümünü doğru cevaplama olasılığı 0,8’dir. Bu durumda bu kişinin 20 maddelik bir testte yaklaşık 16 maddeyi doğru cevaplayabileceği söylenebilir. Baker’a (2001) göre test karakteristik eğrilerinin şekli madde sayısına, kullanılan modele ve kestirilen madde parametrelerine bağlıdır.

Madde ve test karakteristik eğrileri yanlılık analizi ve test eşitleme çalışmalarında yanlı maddelerin belirlenmesi ve eşitlenen iki testte farklı teta düzeylerinde maddelere doğru

yanıt verme olasılıklarının eşit olup olmadığının belirlenmesi amacıyla kullanılmaktadırlar. Bu, MTK'nın en belirgin özelliği olan, madde ve yetenek parametrelerinin değişmezliği ilkesi ile mümkün olmaktadır. Bu özellik, MTK'nın KTK'ya tercih edilmesinin en önemli nedenidir. Çünkü, KTK'da madde parametreleri ile bireylerin karakteristik özelliklerini birbirinden ayırmak mümkün değildir (Hambleton, Swaminathan ve Rogers, 1991). KTK'nın aksine MTK, model veri uyumu sağlandığında, yetenek parametrelerinin madde parametrelerinden bağımsız bir şekilde kestirilmesini sağlamaktadır. Çünkü bu maddelerin, aynı örtük özelliği (θ) ölçen maddelerden oluşan oldukça geniş bir madde havuzundan seçildiği ve bireylerin yetenek düzeylerinin, seçilen farklı madde setlerinden bağımsız olarak kestirildiği varsayılmaktadır (Hambleton ve Swaminathan, 1985). Bu özellik, yetenek parametrelerinin aynı özelliği ölçen farklı maddelere verilen cevaplardan bağımsız olarak kestirilmesini mümkün kılmaktadır.

KTK'da elde edilen madde parametreleri de, o maddeyi cevaplayan grubun özelliğinden büyük oranda etkilenmektedir. Örneğin; aslında orta güçlüğü sahip bir madde, o madde ile ölçülen yetenek düzeyi yüksek bireylerden oluşan bir gruba uygulandığında, maddeye doğru yanıt veren denek sayısı artacağı için kolay bir madde; yetenek düzeyi düşük bir gruba uygulandığında ise maddeye doğru yanıt veren deneklerin sayısı azalacağından dolayı, zor bir madde olarak yorumlanmaktadır. Bu KTK'nın önemli bir sınırlılığıdır. Ancak MTK'da model veri uyumu sağlandığında, bir testteki maddelere ait madde parametreleri de testin uygulandığı grubun yetenek düzeylerinden bağımsız olarak kestirilmektedir. Çünkü testi yanıtlayan bireylerin sonsuz sayıda denek içeren bir evrenden seçildiği ve madde parametrelerinin, bu denek evreninden seçilen farklı örneklerdeki bireylerin maddelere verdiği yanıtlardan bağımsız bir şekilde kestirildiği varsayılır (Hambleton ve Swaminathan, 1985). Bu, aynı testin farklı bireylere uygulanması sonucu elde edilen madde parametrelerinin değişmeyeceği anlamına gelmektedir. Bu durum Şekil 3'teki MKE üzerinde gösterilmiştir.



Şekil 3. Madde Tepki Kuramı'nda madde parametrelerinin değişmezliği.

Şekil 3'te mavi ve yeşil renklerle gösterilen iki farklı örnekleme ait yetenek düzeyi frekans dağılımları incelendiğinde, mavi renkle gösterilen grupta “-1,5” yetenek düzeyindeki bireylerin sayıca diğer yetenek düzeylerine sahip bireylerden daha fazla olduğu, yeşil ile gösterilen dağılımda ise “1,5” yetenek düzeyindeki bireylerin sayıca diğer yetenek düzeylerine sahip bireylerden daha fazla olduğu görülmektedir. Şekil 3'te kırmızı ile gösterilen bu iki gruba uygulanmış maddeye ait madde karakteristik eğrisi incelendiğinde, iki örneklemin farklı yetenek düzeylerine sahip bireylerden oluşmasına rağmen, maddenin her iki gruba uygulanmasının sonucunda aynı madde karakteristik eğrisinin elde edildiği görülmektedir. Bu durum maddenin farklı yetenek düzeylerine sahip bireylerden oluşan iki farklı gruba uygulanmasına rağmen bu iki grupta aynı yetenek düzeyine sahip bireylerce doğru cevaplama olasılıklarının değişmediğini göstermektedir. Bu durum, madde parametrelerinin değişmezliğinin bir göstergesi şeklinde yorumlanmaktadır (Hambleton, Swaminathan ve Roger, 1991).

MTK'da madde ve yetenek parametrelerinin kestirimi için kullanılan tek boyutlu MTK modellerinden bu noktada bahsetmek faydalı olacaktır. MTK'da parametre kestiriminde kullanılan üç adet tek boyutlu model vardır. Bunlar “bir-parametrelili lojistik model”, “iki-parametrelili lojistik model” ve “üç-parametrelili lojistik model”dir.

Bir-parametrelili Lojistik Model (1PL model): Alanyazında Rasch Model olarak da anılan 1PL model ile Rasch Model arasında küçük bir fark vardır. Rasch modelde “a” parametresi değeri sabit $a = 1.0$ kabul edilirken, 1PL modelde “a” parametresi bu değeri

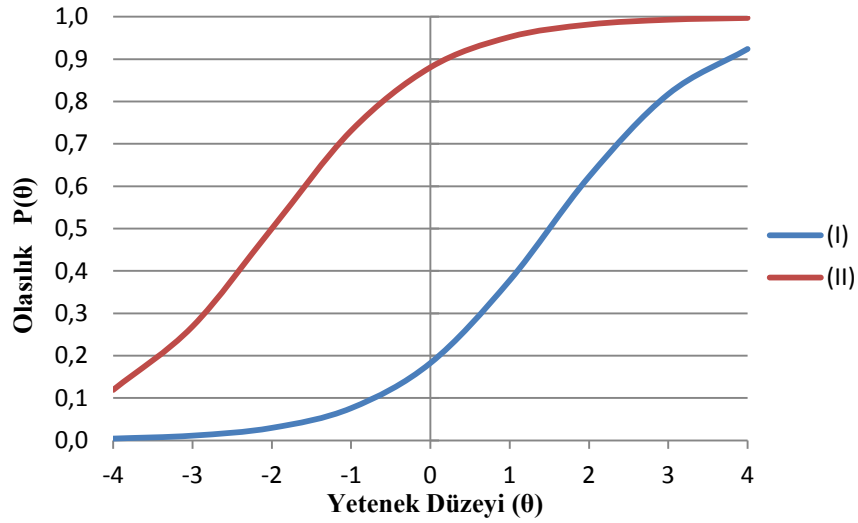
almak zorunda değildir (De Ayala, 2009). 1PL modelde bu sabit değer yerine maddelerin “a” parametrelerinin ortalaması sabit “a” değeri olarak alınmaktadır. Bu küçük farka rağmen her iki model matematiksel olarak birbirlerine denktir.

Rasch’ın (1966) 1PL modeli aslında diğer MTK modellerinden bağımsız geliştirildiği halde, madde karakteristik eğrisinin bir parametrelili lojistik fonksiyon olduğu bir MTK modeli olarak görülebilir (Hambleton ve Swaminathan, 1985). Bu durumda 1PL model, Birnbaum’un 3PL modelinin özel bir durumudur denilebilir. Bu özel modelde tüm maddelerin eşit ayırdedicilik gücüne sahip olduğu ve şans başarısının olmadığı varsayılmaktadır. Bu yüzden sadece “b” parametresi, yani madde güçlüğü parametresi bu model ile kestirilebilmektedir. 1PL modelde madde karakteristik eğrileri Eşitlik 1 (Hambleton, Swaminathan ve Rogers, 1991) kullanılarak elde edilmektedir.

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

Eşitlik 1, $P_i(\theta)$, belirli bir yetenek düzeyine sahip bireylerden seçkisiz olarak belirlenmiş bir bireyin “i” maddesini doğru cevaplama olasılığını; b_i , madde güçlüğü parametresini; n , testteki madde sayısını ifade etmektedir.

MTK modellerinde kestirilen “b” parametresi deneklerin maddeye doğru yanıt verme olasılıklarının %50 olduğu noktayı işaret etmektedir (Demars, 2010). Bu yüzden eşik (threshold) parametresi şeklinde de anılmaktadır. 1PL modelde kestirilen “b” parametresi teorik olarak $(-\infty, +\infty)$ aralığında değerler alsa da, pratikte -2 ile +2 arası değerler almaktadır. Madde kolaylaştıkça sıfırdan küçük değerler alarak -2 değerine yaklaşırken, madde zorlaştıkça sıfırdan büyük değerler alarak +2 değerine yaklaşmaktadır. Bir başka deyişle, kolay bir madde en çok bilgiyi alt yetenek düzeylerinde sağlarken, zor bir madde en çok bilgiyi üst yetenek düzeylerinde sağlamaktadır (Baker, 2001). Şekil 4’te zor (I) ve kolay (II) iki maddeye ait madde karakteristik eğrileri tek düzlem üzerinde gösterilmiştir.



Şekil 4. Kolay ve zor iki maddeye ait örnek madde karakteristik eğrileri.

Şekil 4 incelendiğinde, kırmızı renkle gösterilen madde karakteristik eğrisinin düşük yetenek düzeylerinde daha fazla bilgi verdiği ($b=-2$) ve dolayısıyla çok kolay bir madde olduğu; yeşil renkle gösterilen madde karakteristik eğrisinin ise yüksek yetenek düzeylerinde daha fazla bilgi verdiği ($b=+1,5$) ve dolayısıyla zor bir madde olduğu söylenebilir.

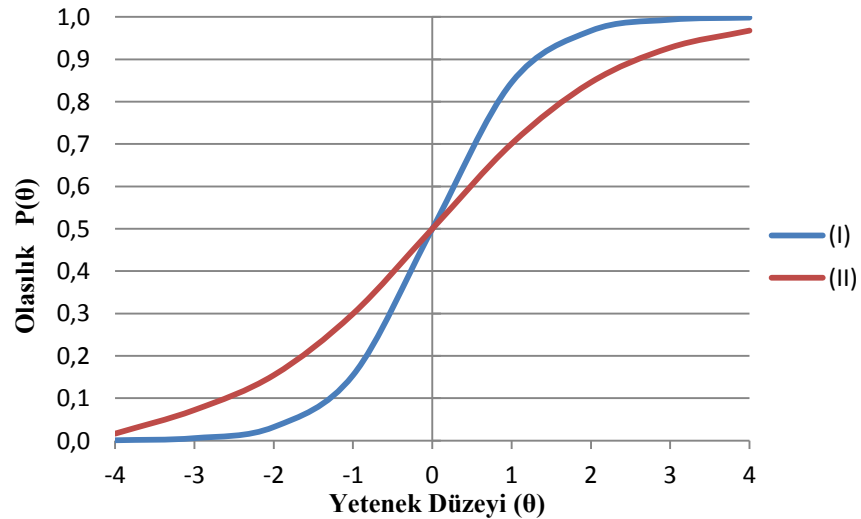
1PL modelde tüm maddelerin eşit ayırdedicilik gücüne sahip olduğu oldukça sınırlayıcı bir varsayımdır (Hambleton ve Swaminathan, 1985). Eğer maddeler özel olarak bu doğrultuda seçilmediyse, bu varsayımın sağlanması zordur. Bu yüzden madde ayırdediciliklerini sabit kabul etmeyen bir modele duyulan ihtiyaç iki-parametrelili lojistik model ile giderilmiştir.

İki-parametrelili Lojistik Model (2PL Model): 2PL modelin temelleri Lord (1952) tarafından atılmış ve 1968 yılında Birnbaum tarafından daha da geliştirilmiştir (Hambleton, Swaminathan ve Rogers, 1991). 2PL modelde, madde güçlüğü (b) parametresine ek olarak, madde ayırdediciliği (a) parametresi modele eklenmiştir. Modele eklenen bu parametre Eşitlik 2'de (Hambleton, Swaminathan ve Rogers, 1991) gösterilmiştir .

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

Eşitlik 2 incelendiğinde, Eşitlik 1’den farklı olarak “a” parametresinin modele eklendiği görülmektedir. Modele eklenen bir diğer unsur D ölçekleme faktörüdür. Bu ölçekleme faktörü hesaplamalarda 1,7 sabit değerini alır.

Ayırdedicilik (a) parametresi, o maddenin “b” parametresi ile belirlenen eşiğin (threshold) alt ve üst yetenek düzeylerine sahip bireyleri birbirinden ne düzeyde ayırdebildiği hakkında bilgi verir. Madde karakteristik eğrilerinde eğimi “a” parametresi değerleri belirlemektedir. Bu nedenle eğim (slope) parametresi olarak da anılan bu parametrenin de “b” parametresi gibi yine teorik olarak $(-\infty, +\infty)$ aralığında değerler aldığı kabul edilse de pratikte $(0, +2)$ arası değerler almaktadır. Bu değer sıfıra yaklaştıkça maddenin ayırdedicilik gücü düşmekte; +2’ye yaklaştıkça ayırdedicilik gücü yükselmektedir. Yüksek ayırdediciliğe sahip maddelerde, madde karakteristik eğrileri alt yetenek düzeylerinden üst yetenek düzeylerine geçiş yaparken doğrunun eğimi artmakta yani daha dik bir şekil almakta; düşük ayırdedicilik değerlerine sahip maddelerde ise eğim azalmakta ve doğru daha yatay bir şekil almaktadır. Yüksek (I) ve düşük (II) ayırdedicilik (a) değerlerine sahip iki maddeye ait madde karakteristik eğrileri Şekil 5’te gösterilmiştir.



Şekil 5. Ayırdediciliği yüksek (I) ve düşük (II) iki maddeye ait örnek madde karakteristik eğrileri.

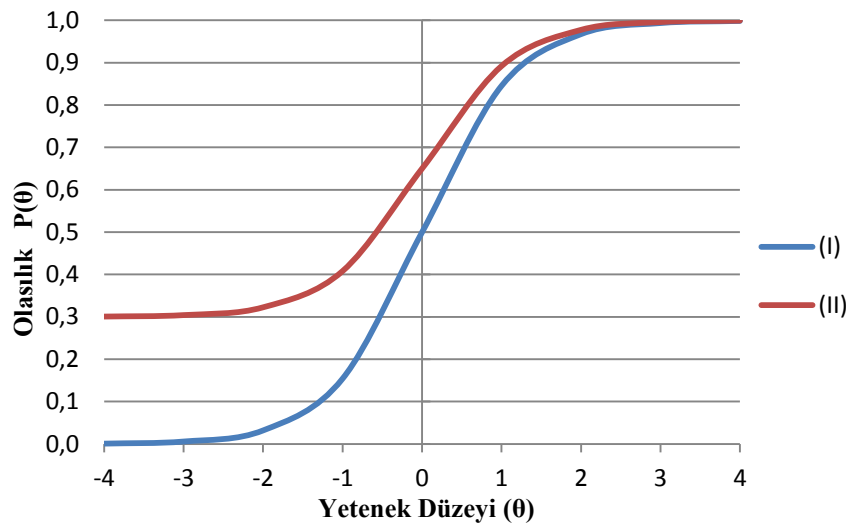
Şekil 5 incelendiğinde, her iki maddeye ait “b” değerleri eşit olmasına rağmen (b=0), mavi renkle gösterilen maddeye ait madde karakteristik eğrisinin alt yetenek düzeylerinden üst yetenek düzeylerine geçerken daha dik bir şekil aldığı ve dolayısıyla daha yüksek ayırdedicilik değerine sahip olduğu (a=1,0); kırmızı renkle gösterilen maddeye ait madde karakteristik eğrisinin daha yatay bir şekil aldığı (a=0,5) ve bu yüzden daha düşük ayırdedicilik değerine sahip olduğu söylenebilir.

Bu modelde kestirilen madde güçlüğü (b) parametreleri, ayırdedicilik (a) parametresinin 1PL modelde tüm maddeler için sabit bir değer olarak alınırken bu modelde her madde için ayrı olarak hesaplanabilmesinden dolayı 2PL modelde kestirilen “b” parametreleri, 1PL modelde kestirilen “b” parametrelerinden farklıdır (Embretson ve Reise, 2000). Modelde dikkat çeken bir diğer nokta şans başarısı parametresi olan “c” parametresinin modelde olmamasıdır. Bu parametreyi içeren tek boyutlu bir MTK modeli daha vardır. Bu model üç-parametrelili lojistik modeldir.

Üç-parametrelili Lojistik Model (3PL Model): 3PL model, 2PL modele üçüncü bir parametre olan şans başarısı (c) parametresi eklenerek elde edilebilir (Hambleton ve Swaminathan, 1985). Bu parametre ile alt yetenek düzeylerindeki bireylerin maddeyi şans başarısıyla doğru olarak cevaplama olasılığını kestirmek mümkündür. 3PL model bu özelliğiyle 1PL ve 2PL modellerden ayrılmaktadır. Çünkü bu iki modelde alt yetenek düzeylerindeki bireylerin bir maddeyi şans başarısıyla doğru cevaplama olasılığı sıfır olarak kabul edilmektedir. Ayrıca maddeleri yazan kişilerin bir maddenin ölçtüğü özellik açısından yanlış ama albenisi yüksek çeldiricilerin alt yetenek düzeylerindeki bireyleri doğru cevabı tahmin etmeye çalışmak yerine maddeyi yanlış yanıtlamaya yönlendirmelerinden dolayı “c” parametresi olduğundan daha düşük değerlerle kestirilmektedir (Hambleton ve Swaminathan, 1985). 3PL model Eşitlik 3’te (Hambleton, Swaminathan ve Rogers, 1991) gösterilmiştir.

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

Eşitlik 3 incelendiğinde, şans başarısı (c) parametresinin modele eklendiği açıkça görülmektedir. Modele eklenen “ c ” parametresi 0 ve 1 arası değerler almaktadır ancak testlere madde seçiminde genellikle 0,3’ten küçük “ c ” parametresi değerlerine sahip maddeler tercih edilmektedir. Eğer “ c ” parametresi bu değerlerin üzerinde bir değer alırsa, o maddede şans başarısının etkisinin çok yüksek olduğu kabul edilir. Şans başarısı parametresinin sıfırdan farklı değerler alması 3PL modelden elde edilen madde karakteristik eğrilerinin, 1PL ve 2PL modellerden elde edilenlerden farklılaşmasına neden olmaktadır. Şöyle ki, şans başarısı parametresinin sıfırdan büyük bir değer alması, madde karakteristik eğrisinin en düşük yetenek düzeylerinde bile, X ve Y düzlemlerinin orijininin değil, bu parametrenin aldığı değerden başlamasına neden olmaktadır. Şans başarısı parametresi sıfıra eşit (I) ve sıfırdan farklı ($c=0,3$) olan (II) iki maddeye ait örnek madde karakteristik eğrileri Şekil 6’da gösterilmiştir.



Şekil 6. Farklı “ c ” parametrelerine sahip iki maddeye ait örnek madde karakteristik eğrileri.

Şekil 6 incelendiğinde, kırmızı ve mavi renklerle gösterilen iki ayrı maddeye ait “ a ” ($a=1$) ve “ b ” ($b=0$) parametreleri her iki madde için de aynı olmasına rağmen .kırmızı renkle gösterilen (II) numaralı maddenin “ c ” parametresi değerinin sıfırdan farklı olduğu ($c=0,3$); dolayısıyla düşük yetenek düzeyine sahip bir bireyin bu maddeyi doğru cevaplama olasılığının da sıfırdan farklı olduğu; mavi renkle gösterilen (I) numaralı maddede ise bu olasılığın sıfır ($c=0$) olduğu söylenebilir.

Araştırmalarda, MTK'nın sağladığı avantajlardan yararlanılabilmesi için bazı varsayımların karşılanması gerekmektedir. Bu varsayımlar tek boyutluluk ve yerel bağımsızlıktır. Demars (2010) tek boyutluluğu; modelin her bir birey için sadece bir yetenek düzeyine (θ) sahip olması; maddelere verilen tepkileri etkileyen diğer faktörlerin tesadüfi hatadan kaynaklandığı ve sadece özel olarak o maddeye ait bir durum olduğu; ya da diğer maddelerde bu faktörlerin etkili olmadığı şeklinde açıklamaktadır. Hambleton ve Swaminathan (1985) ise tek boyutluluğu, bireylerin test performanslarının açıklanması için sadece bir gizil yeteneğin etkili olduğunun varsayımı şeklinde açıklamaktadır. Bu varsayımın tam olarak karşılanması zordur. Çünkü bilişsel, kişisel ve testin uygulanmasıyla ilgili birçok faktör en azından bir oranda test performansını etkilemektedir (Hambleton, Swaminathan ve Rogers, 1991). MTK'nın bir diğer varsayımı olan yerel bağımsızlık, test performansını etkileyen yetenek düzeyleri sabitlendiğinde testi alan bireylerin herhangi bir soru çiftine verdiği cevapların istatistiki olarak birbirinden bağımsız olmasıdır (Hambleton, Swaminathan ve Rogers, 1991). Diğer bir deyişle, bir bireyin testteki maddelere verdiği tepkilerin tek belirleyicisinin o bireyin yetenek düzeyi olduğunun varsayılmasıdır. Tek boyutlulukla yerel bağımsızlık birbirleriyle oldukça ilgilidir. Şöyle ki, tek boyutluluk varsayımı karşılandığında, yerel bağımsızlık varsayımının da karşılandığı düşünülmektedir (Hambleton ve Swaminathan, 1985). Ancak bu durumun tersi doğru değildir. Yani yerel bağımsızlık varsayımının karşılanması tek boyutluluk şartının da karşılandığı anlamına gelmemektedir.

1.1.2. Madde Tepki Kuramı'nda Madde Parametreleri Kestirim Metodları

MTK'nın test verisine yönelik uygulamalarının en önemlisi, seçilen MTK modelinin karakterini oluşturan madde parametrelerinin kestirimidir (Hambleton, Swaminathan ve Rogers, 1994). Madde parametrelerinin kestiriminde N sayıdaki denek n sayıda maddeye sahip bir testi aldığımda madde parametreleri, kullanılan MTK modeline göre değişiklik gösterir. 1PL model için madde parametreleri n sayıdadır. Parametre sayısı 2PL model için $2n$ ve 3PL model için $3n$ şeklindedir (Hambleton ve Swaminathan, 1985). Bu durum, bir testteki madde sayısı arttığında kestirilecek madde parametresi

sayısının da orantısal olarak artmakta olduğu anlamına gelmektedir. Bu, madde ve yetenek parametreleri birlikte kestirildiğinde potansiyel bir kestirim sorununa neden olmaktadır (Hambleton ve Swaminathan, 1985). Bu sorun madde ve yetenek parametrelerinin birbirinden bağımsız olarak kestirimini sağlayan en yüksek olabilirlik kestirim metodları (Maximum Likelihood Estimation Methods) ile aşılmıştır. Bu nedenle en yüksek olabilirlik kestirim metodları; yetenek düzeylerinin bilinmediği durumlarda; madde parametreleri kestiriminde sıklıkla kullanılmaktadır. En yüksek olabilirlik metodunu kullanan 3 farklı kestirim metodu vardır. Bunlar, “Birleşik En Yüksek Olabilirlik Metodu (BEYOM)”, “Marjinal En Yüksek Olabilirlik Metodu (MEYOM)” ve “Şartlı En Yüksek Olabilirlik Metodu (ŞEYOM)”dur.

1.1.2.1. Birleşik En Yüksek Olabilirlik Metodu (BEYOM)

Birnbaum tarafından 1968 yılında formülleştirilen madde parametresi kestirimine yaklaşım öncelikle BEYOM ile olmuştur (Baker ve Kim, 2004). BEYOM’da öncelikle bilinmeyen yetenek düzeyleri, tahmini yetenek düzeyleri kullanılarak bilinen değerler haline getirilmektedir. Daha sonra, bu değerler kullanılarak madde parametreleri kestirilmekte ve daha sonra kestirilen madde parametreleri kullanılarak yetenek düzeyi parametrelerinde düzeltmeler yapılmaktadır (Embretson ve Reise, 2000). Görüldüğü gibi BEYOM aslında madde ve yetenek parametrelerinin eşzamanlı kestirimini temel almaktadır (Hambleton, Swaminathan ve Rogers, 1991).

Embretson ve Reise’ye (2000) göre BEYOM’un avantajları, algoritmasının kolay programlanabilir olması; 2PL ve 3PL modeller de dahil olmak üzere birçok MTK modeline uygulanabilir olması ve uygulanması için hızlı bilgisayarlar gerektirmemesi olarak sıralanmıştır. Diğer taraftan, Hambleton ve Swaminathan’a (1985) göre MKE’nin doğrusal olmayan yapısı bu metodun uygulanmasında birçok soruna neden olmaktadır. MKE’nin doğrusal olmaması en yüksek olabilirlik denklemlerinin de doğrusal olmamasına sebep olmaktadır. Bu durum, özellikle denek sayısı arttıkça kestirilecek parametre sayısının da arttığı 2PL ve 3PL modellerde matematiksel sorunlara neden olmaktadır.

BEYOM'un dezavantajları ise şu şekilde sıralanmıştır. Öncelikle madde parametreleri, istenen özelliklere sahip değildir. Çünkü her ne kadar düzeltme faktörleri bu sorunu oldukça aza indirgese de bu metotla kestirilen parametreler yanlıdır. Ayrıca sabit uzunluktaki testlerden alınan veriye uygulanan BEYOM ile kestirilen madde parametreleri tutarsızdır. Yani madde sayısını sabit tutarak çalışmaya yeni denekler eklemek BEYOM'da daha iyi parametreler elde etmek anlamına gelmemektedir (Embretson ve Reise, 2000). Çünkü bu, daha fazla parametre hesabı yapılması anlamına gelmektedir. İkinci olarak, BEYOM ile elde edilen standart hataların anlamının sorgulanabilir olduğunu düşünmektedir (Embretson ve Reise, 2000). Çünkü kestirilen madde parametreleri, yanlı ve tutarsızdır ve standart hataları hesaplamada en uygun kestirim yöntemini sunmamaktadır. Üçüncü olarak BEYOM'un MTK modelleri üzerine araştırmalarda kullanımının sorgulanması gereken bir durumdur (Embretson ve Reise, 2000). Çünkü olabilirlik oranları çok yüksektir (Holland, 1990). Dördüncü olarak BEYOM'da maddelerin hepsine doğru veya hepsine yanlış yanıt vermiş bireyler için yetenek parametreleri; ya da herkes tarafından doğru veya yanlış yanıtlanan maddeler için de madde parametreleri kestirimi bu metotla yapılamamaktadır.

BEYOM tarafından kestirilen madde parametrelerine ait standart hata değerleri diğer en yüksek olabilirlik metodlarına göre düşük çıkmaktadır. Holland (1990) bu durumu bilinmeyen yetenek düzeyi parametreleri sorununun yönetiliş biçiminden dolayı hata değerlerinin muhtemelen olması gerekenden küçük kestirildiği şeklinde açıklamaktadır. Yukarıda bahsedilen eleştiriler ve Holland'ın görüşü BEYOM'a güven duyulmamasının bir işareti olarak yorumlanabilir.

1.1.2.2. Şartlı En Yüksek Olabilirlik Metodu (ŞEYOM)

Rasch, 1960 yılında bazı formüllerin şartlı olasılıklar için kestirim amaçlı kullanılabileceğini belirtmiş ve bu konuda oldukça genel bir yaklaşım ve bir uygulama yayınlamıştır; ancak bunların ikisi de uygulamada tatmin edicilikten uzaktır (Baker ve Kim, 2004). Sonraki yıllarda Andersen (1972) şartlı kestirim prosedürleri uygulanarak tutarlı madde parametreleri elde edilebileceğini göstermiştir. ŞEYOM'da madde parametreleri, tepki örüntülerinin (response patterns) olasılık değerleri, yetenek

düzeylerini hesaba katmadan kestirilmiş ve bu sayede bilinmeyen yetenek düzeyleri sorununa çözüm bulunmaya çalışılmıştır (Embretson ve Reise, 2000). Bu metod, sadece 1PL model uygulamalarında kullanılabilir (Hambleton, Swaminathan ve Rogers, 1991; Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985) ve olabilirlik fonksiyonu doğru cevap sayısına bağlıdır (Hambleton, Swaminathan ve Rogers, 1991).

Embretson ve Reise (2000), ŞEYOM'un avantajlarını tahmini bir yetenek düzeyi dağılımı gerektirmemesi; madde parametreleri yetenek düzeyi dağılımından ya da yetenek düzeylerinden bağımsız kestirildiğinden madde parametrelerinin değişmezliği prensibini yansıtmaması; birçok farklı durumda madde parametrelerinin tutarlı olması, normal dağılım göstermesi ve etkili olması; madde parametrelerinin standart hata oranları BEYOM'la karşılaştırıldığında daha açıklanabilir olması şeklinde sıralamaktadır.

Embretson ve Reise (2000) ŞEYOM'un dezavantajlarını ise sadece Rasch model ve çeşitlerine uygulanabilir olması; herkes tarafından doğru ya da yanlış cevaplanan maddelere ait parametre kestiriminde ve tüm soruları doğru ya da yanlış cevaplayan bireyler için yetenek parametreleri kestirimi yapamaması nedeniyle veri kaybına sebep olması ve son olarak ilk bilgisayar programlarında ŞEYOM'la yapılan kestirimlerin 30 maddeden sonra başarısızlıkla sonuçlanmasından dolayı uzun testlerde bazı sayısal sorunlara sebep olması şeklinde sıralamaktadır.

1.1.2.3. Marjinal En Yüksek Olabilirlik Metodu (MEYOM)

MEYOM'da bilinmeyen yetenek düzeyleri sorunu, bireylere ait tepki örüntülerinin (Response patterns) olasılık değerleri, evrendeki yetenek düzeyi dağılımından beklenen yetenek düzeyleri şeklinde ifade edilerek aşılacaktır (Embretson ve Reise, 2000). Yani, MEYOM'da yaklaşım, testi alan kişilerin bir evrenden alınmış tesadüfi bir örnekleme temsil ettiklerinin varsayılması şeklindedir (Baker ve Kim, 2004; Bock ve Lieberman, 1970). Bu nedenle MEYOM'da, bilinmeyen yetenek düzeylerini birey bazında tahmin etmekten ziyade yetenek düzeyi dağılımı tahmini yaparak madde

parametreleri kestirilmektedir. Bu dağılımın da önceden bilinmesine gerek yoktur. Yeterli büyüklükte bir örneklem olması durumunda, yetenek düzeyi dağılımları bu veriden kestirilebilir (Embretson ve Reise, 2000). Bu durumda MEYOM, artan örneklem büyüklüğü ile daha fazla yetenek düzeyi parametresi kestirimini gerektirmeyeceğinden herhangi bir büyüklükteki örneklemle madde parametreleri kestirimini de mümkün kılmaktadır (Baker ve Kim, 2004).

Bock ve Aitkin (1981), MEYOM için EM (Expectation / Maximization) algoritması geliştirmişlerdir (aktaran; Embretsen ve Reise, 2000). Bu algoritmaya bu ismin verilmesinin sebebi, tahmin (Expectation) ve azamileştirme (Maximization) adlı iki aşamadan oluşmasıdır. Tahmin (Expectation) aşamasında her yetenek düzeyinde kaç bireyin bulunacağı ve bu kişilerden kaçının her bir soruya doğru cevap vereceğinin beklenen değerleri hesaplanır. Daha sonra azamileştirme (Maximization) aşamasında olabirlikleri azami değerlere ulaştırmak için kestirim eşitliklerindeki beklenen değerleri kullanarak, madde parametreleri kestirilir (Embretsen ve Reise, 2000).

MEYOM'un Rasch modeldeki aşamaları ŞEYOM ile karşılaştırılabilir sonuçlar vermektedir. Ancak karmaşık temel simetrik işlevler (Complex elementary symmetric functions) gerektirmediğinden MEYOM'un ŞEYOM'dan daha etkili olduğu görülmektedir (Hambleton ve Swaminathan, 1985).

Embretson ve Reise (2000) tarafından MEYOM'un avantajları, çok boyutlu modeller de dahil tüm MTK modellerine uygulanabilmesi; hem uzun hem de kısa testlerde etkili olması; elde edilen madde standart hata değerlerinin, beklenen örnekleme varyansının (expected sampling variance) başarılı bir kestirimi olması; mükemmel puanlar için kestirimler sağlayabilmesi; ve son olarak verilerinin hipotez testi ve model veri uyumu indisleri için kullanılabilir olmasından dolayı G^2 olabirlik oranı testi gibi yöntemlerle model veri uyumunun sınanmasına imkan vermesi şeklinde sıralanmıştır.

Embretson ve Reise'ye (2000) göre MEYOM'un bazı dezavantajları da vardır. Öncelikle, MEYOM kestirimi için etkili bir algoritma oluşturmak oldukça zordur.

Ancak bilgisayar teknolojisindeki gelişmeler yeni ve daha karmaşık programların MEYOM'u kullanmasına imkan tanımaktadır (Adams ve diğerleri, 1997; Wilson, 1989). İkinci olarak MEYOM'da madde parametrelerinin kestirimi için yetenek düzeyleri ile ilgili bir dağılımın varsayılması gerekmektedir. Bu da madde parametrelerinin doğruluğunun bu varsayılan yetenek düzeyi dağılımının doğruluğu şartına bağlı olması demektir.

1.2. ARAŞTIRMANIN AMACI VE ÖNEMİ

Bu bilgiler ışığında bu araştırma; test uzunluğu ve örneklem büyüklüğünün MTK'nın 1PL, 2PL ve 3PL modellerinde model-veri uyumu, madde parametreleri ve madde kestirimlerine ait standart hata değerleri üzerine etkisinin araştırılması amacını taşımaktadır. MTK temel alınarak yapılan çalışmalarda yeterli örneklem büyüklüğünün ne olması gerektiği konusundaki tartışmalara önemli katkıları olacağı düşünülen bu çalışmayı alanyazındaki diğer çalışmalardan ayıran en önemli özellik hem küçük, hem de büyük örneklem içeren bu kadar kapsamlı bir çalışmaya alanyazında yapılan taramalarda rastlanmamış olmasıdır. Araştırmada kullanılacak toplam dokuz farklı örneklem büyüklüğünün bu tarz bir çalışmada ilk kez kullanılıyor olmasının araştırmanın önemini artırdığı düşünülmektedir. Bu araştırmayı diğer araştırmalardan ayıran bir diğer özellik ise araştırmada kullanılan verinin bilgisayar ortamında oluşturulmuş simülasyon verisi yerine gerçek veri olmasıdır. Bu konuda daha önceki yıllarda yapılmış ve günümüzde sıklıkla referans gösterilen araştırmalarda (Hulin, Lissak ve Drasgow, 1982; Yen, 1987; Baker, 1998; Gao ve Chen, 2005; Thissen ve Wainer, 1982) gerçek veriler yerine bilgisayar ortamında oluşturulmuş simülasyon verileri kullanılmış olmasının, gerçek verilerin kullanıldığı sınırlı sayıdaki araştırmalardan birisi olacak bu araştırmayı daha da önemli kıldığı düşünülmektedir. Çünkü gerçek test verisi yerine bilgisayar ortamında oluşturulan verilerin kullanıldığı araştırmalarda, bu verilerin gerçek verilerin karakteristik özelliklerini yansıtıp yansıtmadığı bilinmemektedir (Sireci, 1991). Bunların yanısıra alanyazında yaygın olarak referans alınan çalışmaların (Lord, 1968; Hulin, 1982) oldukça eski yıllarda yapılmış olmaları; bu araştırmalarda kullanılan kestirim metodunun (BEYOM) daha

önce sıralandığı şekliyle günümüz şartlarında eleştirilen bir metod olması; ve güçlü bilgisayarların olmadığı o dönemlerde elle veya basit bilgisayar programlarıyla yapılmış hesaplamalarda hata yapılmış olabileceği ihtimali bu tür kapsamlı bir çalışmanın günümüzdeki ileri teknolojik imkanlardan yararlanarak yapılmasını gerekli kıldığı düşünülmektedir. Araştırmanın, bu özelliğiyle de alana oldukça önemli katkılar sağlayacağı düşünülmektedir. Ayrıca, ülkemizde MTK'yı temel alarak yapılan araştırmalar tarandığında, MTK'yı bu yönü ile ele alan ulusal çalışmalara rastlanılmaması çalışmanın alana katkısını artırmakla birlikte, önemini de artırmaktadır. Yukarıda belirtilen özelliklerinden dolayı, araştırmanın alana yapacağı katkının yadsınamaz boyutlarda olduğu değerlendirilmektedir. Araştırmanın yukarıda belirtilen amacına ulaşabilmesi için problem cümlesi aşağıdaki şekilde ifade edilmiştir.

1.3. PROBLEM CÜMLESİ

Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, Madde Tepki Kuramı'nın 1PL, 2PL ve 3PL modellerinden kestirilen madde parametreleri nasıl değişmektedir?

Alt Problemler

1. Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında model veri uyumu nasıl değişmektedir?
 - a. Küçük örneklemelerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlere ait verilerin model veri uyumu örneklem büyüklüğüne ve test uzunluğuna göre değişmekte midir?
 - b. Büyük örneklemelerde (1000, 2000, 3000, 5000) ve çalışma evreninde 10, 20 ve 30 maddelik testlere ait verilerin model veri uyumu örneklem büyüklüğüne ve test uzunluğuna göre değişmekte midir?
2. Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametreleri, çalışma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerine göre nasıl değişmektedir?

- a. Küçük örneklerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametreleri, çalışma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerine göre nasıl değişmektedir?
 - b. Büyük örneklerde (1000, 2000, 3000, 5000) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametreleri, çalışma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerine göre nasıl değişmektedir?
3. Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıl değişmektedir?
- a. Küçük örneklerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıldır?
 - b. Büyük örneklerde (1000, 2000, 3000, 5000) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıldır?

1.4. SAYILTILAR

Bu araştırmada;

1. Kullanılan verilerin normal dağılım gösterdiği;
2. Araştırmada kullanılan testlerin kısaltılmasının madde özelliklerini etkilemediği varsayılmıştır.

1.5. SINIRLILIKLAR

Bu araştırma;

1. Gazi Üniversitesi'nde 2011-2012 Eğitim-Öğretim yılı güz döneminde ING 101 dersi final sınavına katılan 9 farklı fakülte ve yüksekokuldan öğrenciler ile;

2. MTK'nın tek boyutlu modelleri ve kullanılan model veri uyumu sınaama metodları ile;
3. Marjinal en yüksek olabilirlik metodu (MEYOM) ile kestirilen madde parametreleri ile;
4. En az 150 kişilik örneklem büyüklüğüyle sınırlıdır.

1.6. İLGİLİ ARAŞTIRMALAR

Lord (1968) çalışmasında SAT verisi üzerinden Birleşik En Yüksek Olabilirlik Metodu'na (BEYOM) benzer tekrarlamalı bir metod kullanarak 3PL modele uygun kestirim yapmıştır. Sonuç olarak madde sayısı 50'ye ve örneklem büyüklüğü 1000'e ulaşmadan örnekleme hatasının yüksek değerler aldığını belirtmiştir. Bu çalışma literatürde referans alınan 2 önemli araştırmadan birisidir.

Hulin ve diğerleri (1982) yine 3 PL model için BEYOM kullanarak parametre kestirimi yaptıkları Monte Carlo Simülasyon çalışmasında bilgisayar yardımıyla oluşturdukları 200, 500, 1000 ve 2000 kişilik örneklem grupları ve 15, 30 ve 60 maddelik test maddesi cevapları üzerinden tekrarlamalı simülasyon denemeleri yapmış ve 30 maddelik ve 500 kişilik örneklem grubunun bazı çalışmalar için uygun olabileceğini tespit etmiştir. 1000 kişilik örneklem ve 60 maddeden oluşan testin daha isabetli kestirim için gerekli olduğunu belirtmişlerdir. Bu çalışma da literatürde referans gösterilen ikinci önemli araştırmadır.

Thissen ve Wainer (1982), Serbest En Yüksek Olabilirlik (Unrestricted Maximum Likelihood) metodunu kullandıkları araştırmalarında, madde parametrelerine ait standart hata değerlerinin değişimini 1PL, 2PL, 3PL modellerde, farklı ayırdecilik ve güçlük değerlerine sahip maddeler için farklı örneklem büyüklükleri kullanarak sınımışlardır. Çalışmada 500, 1.000, 2.500 ve 10.000 kişilik örneklere ait bulgularını rapor eden araştırmacılar, 1PL modelde $a \sim 1,00$ değeri alan maddelerde, 2500 kişilik örnekleme ulaşıldığında oldukça düşük ($SE_a < 0,5$) hata değerleri elde edildiğini tespit etmişlerdir. Yine aynı maddelerde 1PL modelde standart hata oranları yaklaşık 0,1 değeri referans

alındığında 1PL model için 500 kişilik örneklemin yeterli olacağını; yine 1PL modelde $a \approx 0.5$ değeri alan maddelerde 0,1 düzeyinde hata değerlerine ulaşabilmek için en az 1000 kişiye ihtiyaç duyulduğunu; madde parametreleri uç değerlere yaklaştığında standart hata değerlerinin arttığını; ve 3PL modelde 10.000 kişilik bir örnekleme bile çok yüksek standart hata değerlerine ulaşıldığını tespit etmişlerdir.

Goldman ve Raju (1986) 3000 kişinin bir davranış ölçeğine verdikleri tepkileri kullanarak yaptıkları çalışmada, 3000 kişinin tümünden elde ettikleri yetenek ve madde parametreleri ile bu 3000 kişiden elde ettikleri 250, 500 ve 1000 kişilik örneklemlerden yetenek ve madde parametreleri kestirimi yapmış ve bu iki kestirimin korelasyonlarını 1PL ve 2PL modellerde incelemişlerdir. Yetenek parametreleri için tüm örneklemlerde, çalışma evreninden elde edilen yetenek parametreleriyle mükemmel korelasyon katsayılarına ulaşan araştırmacılar, madde parametreleri kestirimi için 1PL modelde 250 kişilik, 2PL modelde 500 kişilik örneklemin yeterli olduğunu; ancak madde ayırdediciliği (a) parametresinin doğru kestirimi için 1000 kişilik örnekleme ihtiyaç duyulduğunu tespit etmişlerdir.

Yen (1987) gerçekleştirdiği simülasyon çalışmasında 3PL model için parametre kestirimlerini 1000 kişilik bir örneklem grubu ve bir tane 10, dört tane 20, dört tane 40 madde içeren test kullanarak madde parametreleri kestirimi yapmıştır. Kestirdiği bu parametreler üzerinden BILOG ve LOGIST programlarının parametre kestirimindeki başarısını karşılaştırmıştır. Araştırma sonucunda LOGIST programının en yüksek olabilirlik kestirimlerini oluşturma açısından BILOG'dan biraz daha hızlı olduğunu ancak BILOG'un her zaman madde parametrelerini kestirmede daha başarılı olduğunu belirlemiştir. BILOG'un 10 maddelik testte daha başarılı olduğunu, 20 ve 40 maddelik testlerde ise iki programın da eşit bir şekilde isabetli kestirimler yaptığını belirtmektedir.

Sireci (1991) MTK'nın küçük boyuttaki örnekleme çalışmaya uygun olup olmadığını kontrol etmek için yaptığı çalışmada MTK madde parametrelerinin değişmezliğini aynı veriden elde ettiği KTK madde parametreleriyle karşılaştırarak değerlendirmiştir.

Araştırma aynı testin yılda bir kez olmak üzere üç yıl süreyle üç kez farklı örneklemelere uygulanmasını içermektedir. Çalışmada kullanılan test, 1988 yılında 173, 1989 yılında 149 ve 1990 yılında 106 kişiye uygulanmıştır. Aynı veriye KTK güvenilirlik ve çift serili korelasyon analizleri ve MTK'nın 1PL, 2PL ve 3PL modellerinde analizler uygulanmıştır. Araştırma sonucunda modelin veriye uyması durumunda, 1 PL model ve 2 PL modelde küçük örneklemelerden elde edilen madde güçlüğü parametrelerinin değişmez olduğu tespit edilmiştir. Hem MTK ve KTK'da "a" parametreleri açısından değişmezlik sağlanamamıştır. Ayrıca yine 1PL ve 2PL modellere göre madde güçlüğü parametrelerinin KTK'da daha iyi kestirildiği sonucuna varmıştır.

Patsula ve Gessaroli (1995) yaptıkları çalışmada farklı örneklem büyüklüğü ve test uzunluklarında BILOG ve TESTGRAF adlı programların 3PL modelde parametre kestirimindeki başarılarını karşılaştırmıştır. Bu çalışmada 100, 250, 500 ve 1000 kişilik örneklem gruplarını kullanmış ve madde sayısı olarak 20 ve 40 madde kullanmışlardır. Araştırmanın sonunda TESTGRAF yazılımının genel olarak BILOG'dan daha iyi veya en az onun kadar iyi kestirimlerde bulunduğunu tespit etmişlerdir. TESTGRAF'ın sadece 1000 kişilik örneklem büyüklüğünde "a" parametresini ve diğer tüm örneklem büyüklüklerinde "c" parametresini kestirmede BILOG'dan biraz daha az isabetli kestirim yaptığı sonucuna varmışlardır.

Baker (1998) BILOG programındaki parametre kestirim yaklaşımı ve Gibbs'in örnekleme yaklaşımını karşılaştırdığı çalışmasında farklı örneklem büyüklükleri, test uzunluğu ve yetenek düzeyi dağılımlarını iki-parametrelili normal ogive model için incelemiştir. Çalışmada 30, 60, 120 ve 500 kişilik örneklem gruplarını kullanan araştırmacı ayrıca 10, 20, 30 ve 50 maddelik testler kullanmıştır. Sonuç olarak 50 madde ve 500 kişilik örneklem büyüklüğünde BILOG'da elde edilen madde parametrelerinin mükemmel sonuçlar verdiğini tespit etmiştir. Ayrıca BILOG programının madde parametreleri kestiriminin kısa testler ve küçük örneklemelerde Gibbs'in örnekleme metoduna göre üstün olduğu sonucuna varmıştır.

Gao ve Chen (2005) gerçekleştirdikleri geniş ölçekli simülasyon çalışmasında bilgisayar yardımıyla 100, 500 ve 2000 kişilik örneklem grupları ve 10, 30 ve 60 maddelik testlere verilen tepkiye ait veriler oluşturmuşlardır. Bu veri setleri üzerinden Marjinal En Yüksek Olabilirlik Metodu (MEYOM) ile Bayes Kestirimi metodlarını 3PL modelde karşılaştırmışlardır. Ayrıca Bayes Kestiriminde kullanılan farklı öncüllerin etkisi de bu çalışmada araştırılmıştır. Araştırmanın sonucunda genel olarak Bayes Kestirimi'nin MEYOM'dan daha isabetli madde parametre kestirimi yaptığı sonucuna varmışlardır. Ayrıca farklı öncüllerin Bayes Kestirimi üzerindeki etkisinin örneklem büyüklüğü çok küçük olmadığı sürece önemsiz düzeyde kaldığı sonucuna varmışlardır.

Köse (2010), geliştirdiği Türkçe testini 1516 kişiye uygulamış ve elde ettiği veriler üzerinden MTK'nın tek boyutlu modellerinden 2PL model ve çok boyutlu MTK modellerinden Telafisel Tipte Çok Boyutlu Model kullanarak farklı örneklem büyüklükleri (500-1000-1500) ve test uzunluklarında (12, 24) model veri uyumunu karşılaştırmış ve çok boyutlu MTK'da modellerin örneklem büyüklüğü arttıkça daha düşük hata değerlerine ulaştığını ve dolayısıyla örneklem büyüklüğünün artmasının model veri uyumunu artırdığı sonucuna varmıştır. Ayrıca örneklem büyüklüğünün artmasının tek boyutlu MTK'da model veri uyumu ve hata parametreleri üzerine bir etkisi olmadığını belirlemiştir. Bunun nedenini ise kullanılan Türkçe testinin MTK'nın tek boyutluluk varsayımını karşılamaması şeklinde belirtmiştir.

Chuah, Drasgow ve Luecht (2010) CAST (Computer Adaptive Sequential Test) uygulamalarında yeni maddelerin deneme uygulamalarında parametre kestirimi için yeterli örneklem büyüklüğünü tespit etmeyi amaçladıkları araştırmalarında, simülasyon ile ürettikleri veriden elde ettikleri 20 maddelik testten madde parametrelerini 300, 500, 1000 kişilik örneklemelerde 3PL modelde kestirmiş ve bu parametre değerlerini kullanarak CASTISEL programında yetenek parametrelerini kestirmişlerdir. Çalışmalarının sonucunda tüm örneklemelerde kestirilen yetenek düzeyi parametreleri ile gerçek yetenek düzeyi parametreleri arasında oldukça yüksek korelasyon değerleri elde etmiş ve madde parametrelerinin yetenek parametrelerinin kestirimi için kullanılması durumunda 300 kişilik küçük örneklemelerde bile yetenek parametrelerinin

kestirilebileceği sonucuna varmışlardır. Ancak bu örneklem büyüklüğünde kestirilen madde parametrelerini, özellikle ayırdedicilik (a) parametresinin olduğundan fazla kestirilmesinden dolayı tatmin edici bulmamışlardır.

Doğan (2002) gerçek veri kullanarak gerçekleştirdiği çalışmasında, 552725 kişilik çalışma evrenine ulaşmış ve bu çalışma evreninden tesadüfi (27501), sağa çarpık (23531), sola çarpık (23531), basık (24691) ve normal dağılım gösteren (29244) örneklem elde ederek bu örneklemde KTK ve MTK'da madde parametrelerinin değişmezliğini karşılaştırmıştır. Çalışmasının sonucunda MTK ve KTK'da farklı örneklemelerden elde edilen madde istatistiklerinin örneklem değişimlerinden etkilendiği ve dolayısıyla madde istatistiklerinin de değiştiği; her iki kuramın da en çok dağılımların çarpıklığından etkilendiği ve sonuç olarak MTK'nın KTK'ya üstünlüğünün teorik düzey kaldığı şeklinde yargılara ulaşmıştır.

Yukarıda bahsedilen konu ile ilgili araştırmaların tümü bir bütün olarak incelendiğinde, yapılan çalışmalarda sınırlı sayıda örneklem büyüklüğünün çalışmalara dahil edildiği; gerçek veriden ziyade çoğunlukla simülasyon verisinin kullanıldığı, özellikle 1968 yılında Lord tarafından yapılan ve literatürde sıklıkla referans alınan çalışmada BEYOM benzeri bir metod kullanılarak madde parametreleri kestirimi yapıldığı dikkat çekmektedir. Bu bilgiler; MEYOM kullanılarak, büyük ve küçük birçok örneklem büyüklüğünü kapsayan gerçek veri ile yapılacak bir araştırmanın alana önemli katkılar sağlayacağını doğrular niteliktedir.

BÖLÜM II

YÖNTEM

Bu bölümde, araştırmanın türü, evren ve örneklem, araştırma verilerinde MTK varsayımlarının sınanması ve son olarak verilerin çözümlenmesi konularına yer verilmiştir.

2.1. ARAŞTIRMANIN TÜRÜ

Bu araştırma, var olan bir durumun halihazırdaki değişkenler üzerinde etkide bulunmadan olduğu gibi betimlenerek, belirlenen hipotezleri test etmek veya soruları cevaplamak için veri toplanmasını içermesi ve değişkenler arasında iki veya daha fazla özelliğin ilişkisinin ne olduğunu açıklamayı amaçlamasından dolayı betimsel bir araştırmadır (Karasar, 2005; Arlı ve Nazik, 2004).

2.2. EVREN VE ÖRNEKLEM

Araştırmanın evrenini Gazi Üniversitesi'nde 2011-2012 eğitim-öğretim yılı güz döneminde ING 101 dersi final sınavına katılan 9 farklı fakülte ve yüksekokuldan toplam 6288 öğrenci oluşturmaktadır. Araştırmada birey ve madde olmak üzere iki tür örnekleme yapılmıştır. Birey örneklemelerini çalışma evreninden elde edilen 150, 250, 350, 500, 750 kişilik küçük ve 1000, 2000, 3000, 5000 kişilik büyük örneklemeler oluşturmaktadır. Madde örneklemelerini ise 50 maddelik testten amaçlı örnekleme yöntemiyle bir araya getirilen 10, 20 ve 30 maddelik testler oluşturmaktadır.

Araştırmada kullanılacak örneklemelerin ve test uzunluklarının belirlenmesi için bir dizi ön çalışma yapılmıştır. Öncelikle ilgili alanyazında daha önceki yıllarda yapılan araştırmalarda en çok 10 (Yen, 1987; Baker, 1998; Gao ve Chen, 2005), 20 (Yen, 1987; Patsula ve Gessaroli, 1995; Baker, 1998; Gao ve Chen, 2005) ve 30 (Hulin, Lissak ve Drasgow, 1982; Baker, 1998; Gao ve Chen, 2005) maddelik test uzunluklarının ve 250 (Goldman ve Raju, 1986), 500 (Thissen ve Wainer, 1982; Hulin, Lissak ve Drasgow,

1982; Goldman ve Raju, 1986; Baker, 1998; Gao ve Chen, 2005), 1000 (Lord, 1968; Hulin, Lissak ve Drasgow, 1982; Thissen ve Wainer, 1982; Yen, 1987; Goldman ve Raju, 1986) ve 2000 (Hulin, Lissak ve Drasgow, 1982; Gao ve Chen, 2005) kişilik örneklemelerin kullanıldığı tespit edilmiş ve bu örneklemelerin ve test uzunluklarının çalışmada kullanılmasına karar verilmiştir. Sonraki aşamada fakülte dağılım yüzdelерinin çalışma evrenindeki dağılım yüzdelерini yansıtabilmesi açısından mümkün olan en küçük örneklem büyüklüğü olan 150 kişilik örneklem çalışmaya dahil edilmiştir. Bu örneklem büyüklüklerinin yanısıra küçük örneklem için 350 ve 750 kişilik; büyük örneklem için 3000 ve 5000 kişilik ara örneklem çalışmaya dahil edilmiştir. Böylelikle çalışmada kullanılacak madde örneklemelerinin 10, 20, 30 maddelik testlerden ve 150, 250, 350, 500, 750, 1000, 2000, 3000 ve 5000 kişilik örneklemelerden oluşmasına karar verilmiştir.

Araştırma verileri 2011-2012 eğitim-öğretim yılı güz döneminde Gazi Üniversitesi Yabancı Diller Yüksekokulu tarafından uygulanan ING 101 dersi final sınavından elde edilmiştir. İngilizce dil bilgisini ölçmeye yönelik 50 maddeden oluşan sınav şu aşamalardan geçmiştir. Öncelikle ING 101 dersi öğretim elemanları tarafından dersin amaçları doğrultusunda hazırlanan madde havuzunda, İngiliz Dili Eğitimi alanında yüksek lisans derecesine sahip Gazi Üniversitesi Yabancı Diller Yüksekokulu test hazırlama birimi başkanı ve iki test hazırlama birimi elemanından alınan uzman görüşleri doğrultusunda değişiklikler yapılmış ve Modern Diller Birimi başkanının geribildirimleri sonucu sınavda kullanılacak 50 madde seçilmiştir. Sınav uygulanmadan önce maddeler araştırmacı tarafından incelemeye tabi tutulmuş ve 50 madde üzerinde son değişiklikler yapılarak sınav asıl formuna ulaştırılmıştır. Uygulanmanın ardından elde edilen veride bir dizi düzenlemeler yapılmıştır. Veri, öğrenci adı, numarası, dersi veren öğretim elemanı bilgilerinden arındırılmış, kitapçık ve fakülte türünü işaretlemeyen öğrenciler veriden çıkarılmıştır. Veri dosyasında yalnızca fakülte kodu ve öğrenci cevaplarının kalması sağlanmıştır. Öğrenci sayısı 150'den az olan fakülteler de, küçük örneklemelerin oluşturulmasında sorun oluşturacağı için veriden çıkarılmıştır. Bu sayede 6288 kişilik çalışma evrenine ulaşılmıştır. Kitapçık türlerine göre iki ayrı dosyaya ayrılmış olan veriler doğru cevaba 1 yanlış cevaba 0 puan verilerek puanlanmış

ve ikili veri matrisi hazırlanmıştır. Daha önce A ve B kitapçığı olarak ayrılan veri dosyası tek dosyada birleştirilerek çözümlenmelere hazır hale getirilmiştir. Çalışma Evrenine ait veride öğrencilerin fakültelerine göre dağılım yüzdeleri Tablo 2.1’de gösterilmiştir.

Tablo 2.1

Çalışma Evrenine Ait Veride Öğrencilerin Fakülteleere Göre Dağılımı

Fakülte/Yüksekokul	<i>f</i>	%
Beden Eğitimi ve Spor Yüksekokulu	354	5,6
Fen Fakültesi	1346	21,4
Fen Edebiyat Fakültesi	189	3,0
Gazi Eğitim Fakültesi	1859	29,6
Hukuk Fakültesi	211	3,4
İktisadi ve İdari Bilimler Fakültesi	1421	22,6
İletişim Fakültesi	246	3,9
Teknoloji Fakültesi	227	3,6
Turizm Fakültesi	435	6,9
Toplam	6288	100,0

Tablo 2.1 incelendiğinde, çalışmaya katılan öğrencilerin %29,6’sının Gazi Eğitim Fakültesi, %22,6’sının İktisadi ve İdari Bilimler Fakültesi, %21,4’ünün Fen Fakültesi, %6,9’unun Turizm Fakültesi, %5,6’sının Beden Eğitimi ve Spor Yüksekokulu, %3,9’unun İletişim Fakültesi, %3,6’sının Teknoloji Fakültesi, %3,4’ünün Hukuk Fakültesi, %3’ünün ise Fen Edebiyat Fakültesi öğrencisi olduğu görülmektedir. Sonraki aşamada çalışma evrenine ait bu veriden tabakalı seçkisiz örnekleme yöntemiyle küçük ve büyük örneklem verileri oluşturulmuştur. Bu aşamada her bir örneklemin, çalışma evrenindeki fakülte dağılım yüzdelerini yansıtmasına özen gösterilmiştir. Ayrıca oluşturulan her bir örneklemin aynı deneği sadece bir kez içermesi için yerine koymadan örnekleme (sampling without replacement) yöntemi uygulanmış ve örneklemelerin oluşturulmasından sonra her bir örneklem verisi bilgisayar ortamında çift

(duplicate) veri kontrolüne tabi tutulmuştur. Oluşturulan küçük örneklem ve çalışma evrenine ait karşılaştırmalı fakülte dağılım yüzdeleri Tablo 2.2’de gösterilmiştir.

Tablo 2.2

Küçük Örneklem ve Çalışma Evrenine Ait Karşılaştırmalı Fakülte Dağılım Yüzdeleri

Fakülte / Yüksekokul	Küçük Örneklem										Çalışma Evreni	
	150		250		350		500		750		6288	
	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
BESYO.	8	5,3	14	5,6	19	5,4	28	5,6	42	5,6	354	5,6
Fen	32	21,3	53	21,2	75	21,4	107	21,4	161	21,5	1346	21,4
Fen Edebiyat	5	3,3	8	3,2	11	3,1	15	3,0	23	3,1	189	3,0
Gazi Eğitim	44	29,3	74	29,6	104	29,7	148	29,6	222	29,6	1859	29,6
Hukuk	5	3,3	8	3,2	11	3,1	17	3,4	26	3,5	211	3,4
İİBF.	34	22,7	57	22,8	79	22,6	113	22,6	169	22,5	1421	22,6
İletişim	6	4,0	10	4,0	14	4,0	20	4,0	29	3,9	246	3,9
Teknoloji	6	4,0	9	3,6	13	3,7	18	3,6	26	3,5	227	3,6
Turizm	10	6,7	17	6,8	24	6,9	34	6,8	52	6,9	435	6,9
Toplam	150	100	250	100	350	100	500	100	750	100	6288	100

Tablo 2.2 incelendiğinde, küçük örneklem ve çalışma evrenine ait karşılaştırmalı fakülte ve yüksek okul dağılım yüzdelerinde bazı küçük sapmalar olduğu görülmektedir. Bu küçük sapmalar düşük sayıda denek bulunan küçük örneklem ve çalışma evrenine ait fakülte ve yüksek okulda 1 denek eklenmesi veya çıkarılması durumunda dağılım yüzdesinde meydana gelen değişimin büyük olmasından kaynaklanmaktadır. Bu sapmaların örneklem ve çalışma evreni temsil yeteneğini büyük ölçüde etkilemeyecek derecede küçük sapmalar olmasına dikkat edilmiştir.

Bu aşamada oluşturulan büyük örneklemeler ve çalışma evrenine ait karşılaştırmalı fakülte dağılım yüzdeleri Tablo 2.3'te gösterilmiştir.

Tablo 2.3

Büyük Örneklemeler ve Çalışma Evrenine Ait Karşılaştırmalı Fakülte Dağılım Yüzdeleri

Fakülte / Yüksekokul	Büyük Örneklemeler								Çalışma Evreni	
	1000		2000		3000		5000		6288	
	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
BESYO.	56	5,6	112	5,6	167	5,6	282	5,6	354	5,6
Fen	214	21,4	428	21,4	643	21,4	1068	21,4	1346	21,4
Fen Edebiyat	30	3,0	60	3,0	89	3,0	148	3,0	189	3,0
Gazi Eğitim	296	29,6	592	29,6	887	29,6	1480	29,6	1859	29,6
Hukuk	34	3,4	68	3,4	103	3,4	169	3,4	211	3,4
İİBF.	226	22,6	452	22,6	677	22,6	1131	22,6	1421	22,6
İletişim	39	3,9	78	3,9	117	3,9	197	3,9	246	3,9
Teknoloji	36	3,6	72	3,6	109	3,6	179	3,6	227	3,6
Turizm	69	6,9	138	6,9	208	6,9	346	6,9	435	6,9
Toplam	1000	100	2000	100	3000	100	5000	100	6288	100

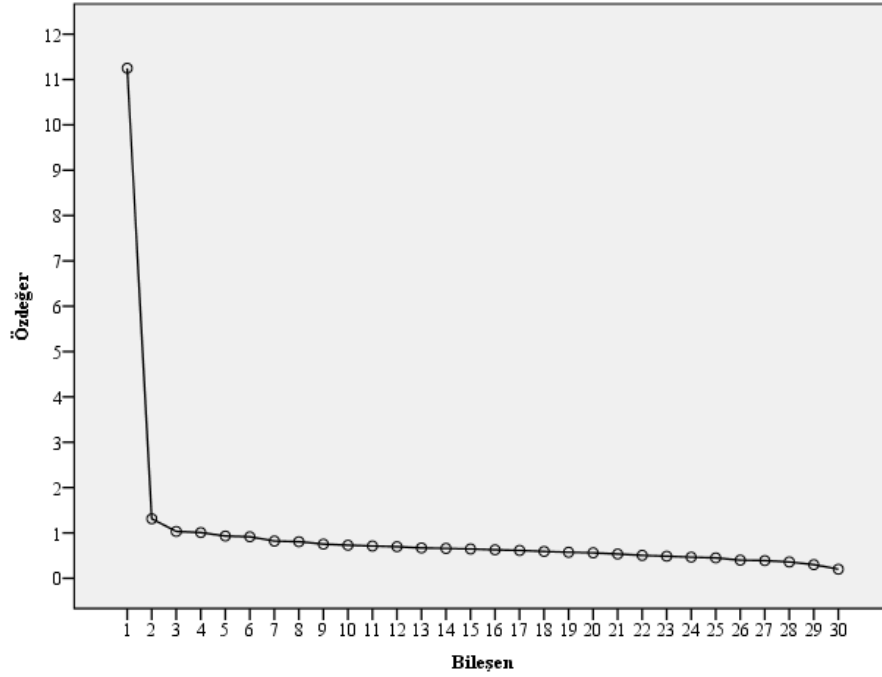
Tablo 2.3'te verilen büyük örneklemelere ait fakülte dağılım yüzdeleri çalışma evrenine ait dağılım yüzdeleriyle karşılaştırıldığında, yüzdelerin çalışma evreninin sahip olduğu fakülte ve yüksekokul dağılım yüzdeleri ile birebir aynı olduğu görülmektedir. Çalışma kapsamında elde edilen küçük ve büyük örneklemelere ait betimsel istatistikler Ek 1 ve Ek 2'de verilmiştir.

2.3. ARAŞTIRMA VERİLERİNDE MTK VARSAYIMLARININ SINANMASI

MTK'nın üstünlüklerinden yararlanılabilmesi için bazı varsayımların karşılanması gerektiğinden daha önce bahsedilmiştir. Araştırma verilerinin çözümlenmesine

geçmeden önce araştırmada kullanılacak 10, 20 ve 30 maddelik testlerin oluşturulması aşamasında tüm verilerde MTK varsayımlarının karşılanması sağlanmıştır.

Tek Boyutluluk : Bilindiği üzere, MTK'nın tek boyutlu lojistik modellerinde parametre kestirimi yapılabilmesi için tek boyutluluk varsayımının karşılanması gerekmektedir. Eğer bir tepki örüntüsünde tek boyutluluk sağlanırsa bir deneğin bir maddeyi cevaplamaında etkili olan şeyin bireyin maddelerin ölçtüğü örtük özelliğe yönelik yetenek düzeyi (θ) olduğu varsayılır. Bu çalışmada sadece MTK'nın tek boyutlu lojistik modelleri işe koşulacağından dolayı, çalışmada kullanılacak 10, 20 ve 30 maddelik testlerin MTK'nın tek boyutluluk varsayımını karşılayan maddelerden oluşması gerekmiştir. Bu amaçla, toplam 50 soruluk madde havuzundan MTK'nın tek boyutluluk varsayımını karşılayan 10, 20 ve 30 madde açımlayıcı faktör analizi yöntemi ile seçilmiştir (Edelen ve Reeve, 2007). Çalışma verilerinin ikili veri olduğu durumlarda faktör analizinde Pearson korelasyon matrisinin yerine tetrakorik korelasyon matrisi kullanılması istendik bir durumdur. Bu yüzden MTK'nın tek boyutluluk varsayımının test edilmesi için yapılacak olan açımlayıcı faktör analizinden önce 6288 kişilik çalışma evrenine ait 50 maddelik test verisinin tetrakorik korelasyon matrisi elde edilmiş ve aynı veriye tetrakorik korelasyon matrisine dayalı açımlayıcı faktör analizi uygulanmıştır. Açımlayıcı faktör analizi sonuçlarına göre 1. faktördeki yükleri .30 ve altında olan maddeler MTK'nın tek boyutluluk varsayımını karşılamadıklarından dolayı testten çıkarılmıştır (Fidelman, 2012). Kalan maddeler arasında 1. faktörde en yüksek faktör yüküne sahip 30 madde seçilerek çalışmada kullanılacak 30 maddelik test verisi oluşturulmuştur. Sonraki aşamada 30 maddelik test verisine ait tetrakorik korelasyon matrisi elde edilmiş ve bu matrise dayalı olarak 30 maddelik teste ait veriye tekrar açımlayıcı faktör analizi uygulanmış ve 30 maddelik test için Şekil 7'de görülen özdeğer grafiği elde edilmiştir.

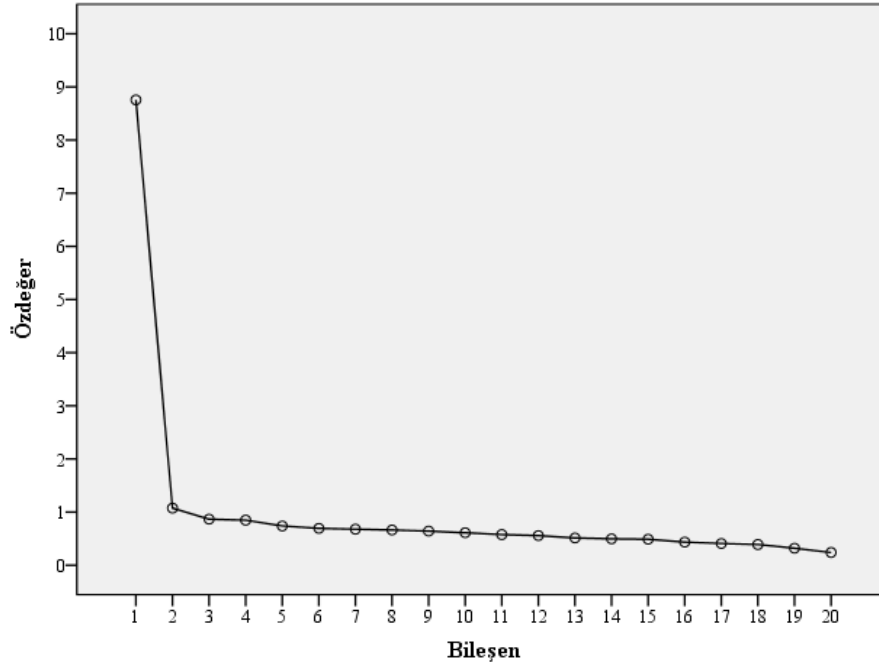


Şekil 7. 30 maddelik testte ait özdeğer grafiği.

Şekil 7 incelendiğinde, 30 maddenin en yüksek yük değerini 1. faktörde alması, birinci faktöre ait özdeğer ile ikinci faktöre ait özdeğer arasında önemli bir fark olması ($\lambda_1/\lambda_2=8,5$), birinci faktörden sonra grafikte keskin bir kırılma olması ve bu kırılmanın solunda sadece tek bir faktör bulunması 30 maddelik testin MTK'nın tek boyutluluk varsayımını karşıladığının kanıtı olarak kabul edilmiştir. Ayrıca 1. faktörün varyans açıklama oranının %37,50 gibi yüksek bir değer alması 30 maddelik testte tek boyutluluk varsayımının karşılanmasının bir diğer kanıtı olarak kabul edilmiştir. Çünkü birinci faktörde varyans açıklama oranının %30 ve üzeri olması tek boyutluluk için yeterli görülmektedir (Büyüköztürk, 2003). Testteki 30 maddenin 1. faktördeki yükleri 0,41 ile 0,75 arasında değişen değerler (Ek-3) almıştır.

Sonraki aşamada tek boyutluluk şartını karşılayan 30 maddelik test verisinden 20 maddelik test verisi elde edilmiştir. Bu amaçla, daha önce oluşturulan 30 maddelik testten 1. faktörde en yüksek faktör yüküne sahip 20 madde seçilmiş ve 20 maddelik test verisi oluşturulmuştur. Oluşturulan 20 maddelik test verisinden tetrakorik korelasyon matrisi elde edilmiş ve bu veriye de tetrakorik korelasyon matrisine dayalı açımlayıcı

faktör analizi uygulanmıştır. Açımlayıcı faktör analizi sonucu elde edilen özdeğer grafiği Şekil 8'de gösterilmiştir.

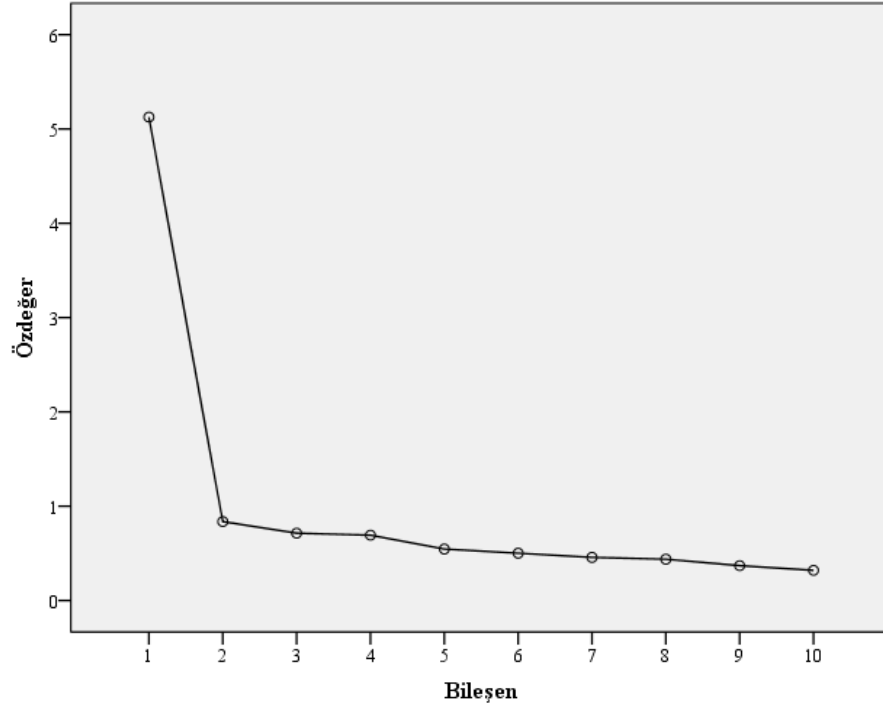


Şekil 8. 20 maddelik teste ait özdeğer grafiği.

Şekil 8 incelendiğinde, 20 maddenin en yüksek faktör yükünü 1. faktörde alması, ikinci faktöre ait özdeğer ile birinci faktöre ait özdeğer arasında önemli bir fark olması ($\lambda_1/\lambda_2=8,15$), birinci faktörden sonra grafikte keskin bir kırılma olması ve bu kırılmanın sağında sadece bir faktör bulunması, ikinci ve sonraki faktörlerin birbirine yakın değerler alması 20 maddelik testin de MTK'nın tek boyutluluk varsayımını karşılamasının kanıtı olarak kabul edilmiştir. Ayrıca 1. faktörün varyans açıklama oranının %43,78 gibi yüksek bir değer alması da tek boyutluluk varsayımının karşılanmasının bir diğer kanıtı olarak kabul edilmiştir. Testteki 20 maddenin 1. faktördeki yüklerinin 0,59 ile 0,75 arasında değişen değerler aldığı görülmüştür.

Çalışmada kullanılacak 30 ve 20 maddelik testlerin MTK'nın tek boyutluluk varsayımını karşılayacak maddelerden oluşturulmasından sonra 10 maddelik testin de tek boyutluluk varsayımını karşılayacak şekilde oluşturulması aşamasına geçilmiştir. Bu amaçla, 20 maddelik testten 1. faktörde en yüksek faktör yüküne sahip 10 madde

seçilmiştir. Seçilen 10 madde için tekrar tetrakorik korelasyon matrisi hazırlanmış ve veriye tetrakorik korelasyon matrisine dayalı açımlayıcı faktör analizi uygulanmıştır. Açımlayıcı faktör analizi sonucu elde edilen özdeğer grafiği Şekil 9'da gösterilmiştir.



Şekil 9. 10 maddelik teste ait özdeğer grafiği.

Şekil 9 incelendiğinde, 10 maddenin en yüksek yük değerini 1. faktörde alması, birinci faktöre ait özdeğer ile ikinci faktöre ait özdeğer arasında önemli bir fark olması ($\lambda_1/\lambda_2=6,1$), birinci faktörden sonra grafikte keskin bir kırılma olması ve bu kırılmanın solunda sadece tek bir faktör bulunması, 10 maddelik testin MTK'nın tek boyutluluk varsayımını karşıladığının kanıtı olarak kabul edilmiştir. Ayrıca 1. faktörün varyans açıklama oranının %51,29 gibi yüksek bir değer alması 10 maddelik testte tek boyutluluk varsayımının karşılanmasının bir diğer kanıtı olarak kabul edilmiştir. Oluşturulan 10 maddelik testteki maddelerin 1. faktördeki yüklerinin 0,68 ile 0,77 arasında değiştiği belirlenmiştir. Oluşturulan 10, 20 ve 30 maddelik testlerin tüm maddelerine ait faktör yükleri Ek-3'te verilmiştir.

Yerel Bağımsızlık: Yerel bağımsızlık varsayımı tek boyutluluk ile doğrudan ilişkilidir. Lord'a (1980) göre yerel bağımsızlık ek bir varsayım olmaktan çok, tek boyutluluk ile otomatik olarak elde edilen bir özelliktir. Bu bilgiler ışığında veri dosyalarında tek boyutluluğun sağlanması yerel bağımsızlığın da sağlandığının bir göstergesi olarak kabul edilmiştir.

Tek boyutluluk varsayımını karşıladığı belirlenen 10, 20 ve 30 madde kullanılarak her bir örneklem büyüklüğü için 10, 20 ve 30 maddelik üç ayrı madde seti belirlenmiş ve daha önce hazırlanmış olan 150, 250, 350, 500, 750, 1000, 2000, 3000, 5000 kişilik örneklem ve çalışma evrenine ait veri dosyaları sadece bu belirlenen 10, 20 ve 30 maddeyi içerecek şekilde Tablo 2.4'te görülen 3 ayrı dosyaya ayrılmıştır.

Tablo 2.4

Çalışmada Kullanılan Veri Dosyaları

Madde Sayısı	Veri dosyası										Toplam
	150	250	350	500	750	1000	2000	3000	5000	6288	
10	1	1	1	1	1	1	1	1	1	1	10
20	1	1	1	1	1	1	1	1	1	1	10
30	1	1	1	1	1	1	1	1	1	1	10
Toplam	3	3	3	3	3	3	3	3	3	3	30

Şekil 2.4 incelendiğinde, her bir örneklem verisi için 10, 20 ve 30 maddelik üç adet veri dosyası hazırlandıktan sonra çalışmada kullanılacak veri dosyası sayısının 30'a ulaştığı görülmektedir.

2.4. VERİLERİN ÇÖZÜMLENMESİ

Verilerin çözümlenmesi aşamasında verilere araştırmanın alt problemlerini yanıtlamak amacıyla bir dizi analiz uygulanmıştır. Öncelikle daha önce belirtilen 30 farklı veri

dosyası kullanılarak MTK'nın 1PL, 2PL ve 3PL modellerinde madde parametreleri kestirimi yapılmıştır. Tüm örneklem büyüklükleri ve test uzunluklarında elde edilen ortalama parametre değerleri Ek 4'te gösterilmiştir.

Araştırmanın birinci alt problemi olan “Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında model veri uyumu nasıl değişmektedir?” sorusunu cevaplamak amacıyla model veri uyumu, tüm araştırma durumlarında dört farklı yöntemle sınanmıştır. Çalışmada model veri uyumunun sınanmasında birçok örneklemin bulunması ve bu farklı örneklem büyüklüklerinde bazı maddelerin bazı örneklemlerde modele uyarken bazılarında uymamasından dolayı (EK 5), model veri uyumu maddelerin modele uyumu şeklinde değil genel olarak (overall) verinin modele uyumu şeklinde ele alınmıştır. Ancak kullanılan 4 yöntemden birisi (χ^2/sd) hem genel hem de maddelerin tek tek modele uyumunun incelenmesinde de kullanılmıştır. χ^2/sd yöntemi beklenen ve gözlenen frekanslar veya oranlar arasındaki farkın doğal sonucu olan (Rao ve Sinharay, 2007) kıkare (chi-square) istatistiklerine dayalıdır. Drasgow ve diğerleri (1995) kıkare istatistiğini serbestlik derecesi ile bu şekilde oranlayarak (χ^2/sd) bir tür düzeltme formülü geliştirmiş ve bu değerleri model veri uyumunun sınanmasında kullanmışlardır. Bu yöntemde χ^2/sd oranı 3 kritik değerinin altında kalan modellerin veriyle uyumlu olduğu kabul edilmektedir (Chernyshenko ve diğerleri, 2001). Çalışmada ayrıca standartlaştırılmış artık değer model veri uyumunun sınanmasında kullanılmıştır. Hambleton ve Swaminathan'a (1985) göre daha düşük standartlaştırılmış artık değer daha iyi bir model veri uyumuna işaret etmektedir. Bu nedenle standartlaştırılmış artık değer analizi de ikinci bir yöntem olarak model veri uyumunun sınanmasında kullanılmıştır. Model veri uyumunun sınanmasında kullanılan bir diğer yöntem -2loglikelihood değerlerine dayalı bir yöntem olan olabirlik oranı (G^2) testidir. Her bir modelden elde edilen G^2 değerleri incelenmiş ve değerler arasındaki fark manidarlık testine tabi tutulmuştur. Aralarındaki farkı manidar olan değerlerden daha fazla parametreye sahip olan model veriye daha uyumlu olarak kabul edilmiştir (Kang ve Cohen, 2007). Model veri uyumunun sınanmasında bir diğer yöntem olarak test karakteristik eğrileri incelenmiştir. Hambleton ve Swaminathan (1985)'a göre madde ve yetenek parametrelerinin değişmezliği model veri uyumu için önemli bir göstergedir.

Hambleton ve diğerleri (1991) düşük model veri uyumu söz konusu olduğunda madde ve yetenek parametrelerinde değişmezlikten söz edilmesinin mümkün olmadığını belirtmişlerdir. Van der Linden (1996) yetenek ve madde parametrelerinin değişmezliğinin model veri uyumu için değerli bir bilgi teşkil ettiğini söylemektedir. Bu bilgiler ışığında madde ve yetenek parametrelerinde değişmezliğin göstergesi olduğu düşünülen test karakteristik eğrileri incelenmiş ve farklı örneklem büyüklüklerinde test karakteristik eğrilerinin farklılaşıp farklılaşmadığı grafiksel olarak incelenmiş ve farklı örneklemelerde değişiklik göstermeyen test karakteristik eğrileri o modele uyumun kanıtı olarak alınmıştır. Madde parametrelerinin değişmezliği diğer araştırma probleminde irdeleneceğinden dolayı bu aşamada kullanılmamıştır. Sonuç olarak birinci araştırma problemini yanıtlayabilmek amacıyla model veri uyumu χ^2/sd oranı, standartlaştırılmış artık analizi (SAA), G^2 olabilirlik oranı, test karakteristik eğrilerinin incelenmesi şeklinde dört farklı yöntemle sınanmış ve model veri uyumunun bu farklı yöntemlerde örneklem büyüklüğüne göre nasıl değiştiği incelenmiştir.

Araştırma alt problemlerinden ikincisi olan “Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, MTK’nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametreleri çalışma evreninden MTK’nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerine göre nasıl değişmektedir?” sorusunu cevaplayabilmek amacıyla çalışma evreninden kestirilen madde parametreleri; küçük ve büyük örneklemelerde; 10, 20 ve 30 maddelik testlerde; 1PL, 2PL, 3PL modellerde kestirilen madde parametreleri ile karşılaştırılmıştır (Sireci, 1991; Goldman ve Raju, 1986). Bu değerler arasındaki ilişkinin düzeyini belirlemek için parametre setleri arasındaki korelasyon katsayıları hesaplanmıştır (Sireci, 1991; Gao ve Chen, 2005; Liu, Schulz ve Yu, 2008; Foley, 2010; Chuah, Drasgow ve Luecht, 2010). Hesaplanan korelasyon katsayıları tablolandırılmış ve yorumlanmıştır.

Üçüncü araştırma alt problemi olan “Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında MTK’nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıl değişmektedir?” sorusunu yanıtlayabilmek amacıyla küçük ve büyük örneklemelerden 10, 20 ve 30 maddelik testler için MTK’nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametrelerinin

standart hata deęerlerinin yanısıra alıřma evrenine ait veriden kestirilen madde parametrelerine ait standart hata deęerleri tablolařtırılmıř ve madde parametrelerine ait hata deęerlerinin rneklem byklę ve madde sayısına gre nasıl deęiřtięi tablolar zerinden incelenerek yorumlanmıřtır.

Arařtırma kapsamında gerekleřtirilen veri zmlerinde yetenek ve madde parametreleri ve bu parametrelere ait standart hata deęerlerinin kestirimi, X^2 deęerlerinin hesaplanması ve standart artık deęerlerin kestiriminde Xcalibre 4.1.5; tetrakorik korelasyon katsayılarının hesaplanmasında Statistica 8.0; alıřmada kullanılan rneklerin oluřturulması, histogram grafikleri ve Pearson korelasyon katsayılarının hesaplanması ve tetrakorik korelasyon matrisine dayalı aımlayıcı faktr analizi uygulamalarında, oluřturulan syntax (EK-6) yardımıyla SPSS 20; madde ve test karakteristik eęrilerinin iziminde ise MS Excel 2007 programı kullanılmıřtır.

BÖLÜM III

BULGULAR VE YORUMLAR

Bu bölümde araştırma problemi ve alt problemlerine cevap bulmak amacıyla model veri uyumu, madde parametreleri ve madde parametrelerine ait standart hata değerlerinin karşılaştırılmasına yönelik olarak yapılan istatistiksel çözümler sonucunda elde edilen bulgulara ve bu bulgulara ilişkin yorumlara yer verilmiştir.

3.1. BİRİNCİ ALT PROBLEME İLİŞKİN BULGULAR VE YORUMLAR

Birinci Alt Problem: Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında model veri uyumu nasıl değişmektedir?

- 1.a. Küçük örneklerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlere ait verilerin model veri uyumu örneklem büyüklüğüne ve test uzunluğuna göre değişmekte midir?
- 1.b. Büyük örneklerde (1000, 2000, 3000, 5000) ve çalışma evreninde 10, 20 ve 30 maddelik testlere ait verilerin model veri uyumu örneklem büyüklüğüne ve test uzunluğuna göre değişmekte midir?

3.1.1. Küçük Örneklerde (150, 250, 350, 500, 750) 10, 20 ve 30 Maddelik Testlere Ait Verilerin Model Veri Uyumu Örneklem Büyüklüğüne ve Test Uzunluğuna Göre Değişmekte midir?

Küçük örneklerde; 10, 20 ve 30 maddelik testlere ait verilerde model veri uyumu x^2/sd oranı, standartlaştırılmış artık analizi (SAA), G^2 olabilirlik oranı ve test karakteristik eğrilerinin (TKE) incelenmesi yöntemleri kullanılarak sınanmıştır.

3.1.1.1. Küçük Örneklerde X^2/sd Oranı İle Model Veri Uyumunun Sınanması

Bu incelemede x^2/sd oranı 3'ten küçük değer alan verilerde o modelin veriye uyum sağladığı kabul edilmiştir. 10 maddelik test için elde edilen değerler Tablo 3.1'de gösterilmiştir.

Tablo 3.1

Küçük Örneklerde 10 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
150	52,90	119	0,4	100	75,06	172	0,4	100	100,77	168	0,6	100
250	69,51	135	0,5	100	49,84	139	0,4	100	79,33	154	0,5	100
350	85,04	139	0,6	100	135,38	167	0,8	100	165,54	184	0,9	100
500	63,09	111	0,6	100	143,02	139	1,0	100	205,02	144	1,4	90
750	162,45	137	1,2	100	149,69	160	0,9	100	251,57	170	1,5	100

Tablo 3.1 incelendiğinde, 150 kişilik örnekleme ait x^2/sd değerlerinin 1PL ve 2PL modellerde 0,4 olduğu ve dolayısıyla verinin her iki modele de eşit oranda uyum gösterdiği görülmektedir. 3PL modelde ise x^2/sd değerinin 0,6 olduğu ve verinin bu modele de uyduğu belirlenmiştir.

250 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en düşük değer 0,4 ile 2PL modele ait olduğu ve dolayısıyla en iyi uyumun 2PL modelle olduğu görülmektedir. Ancak 1PL ve 3PL modellere ait x^2/sd değerleri incelendiğinde 0,5 değerini aldığı ve bu

nedenle yine kritik deęer olan 3'ün altında olduęu için verinin bu iki modele de uyumlu olduęu tespit edilmiřtir.

350 kiřilik örnekleme ait x^2/sd deęerleri incelendięinde, en düşük 0,6 deęerini alarak en iyi uyumu 1PL modelle saęladıęı görölmektedir. 2PL ve 3PL modellerde de sırasıyla 0,8 ve 0,9 gibi kritik deęer olan 3'ten daha düşük deęerler aldıęı dikkate alındıęında, 350 kiřilik örnekleme ait verinin de 250 ve 150 kiřilik örneklemlere ait verilerde olduęu gibi her 3 modele de uyum gösterdięi görölmektedir.

500 kiřilik örnekleme ait x^2/sd deęeri incelendięinde, verinin en iyi uyumu x^2/sd oranı 0,6 olan 1PL modelle saęladıęı görölmektedir. Bunun yanısıra, 2PL ve 3PL modellerde de sırasıyla 1,0 ve 1,4 deęerleri alarak yine 3 kritik deęerinin altında bir deęer aldıęı için bu iki modele de uyumlu olduęu belirlenmiřtir.

750 kiřilik örnekleme ait x^2/sd deęerleri incelendięinde, verinin en iyi uyumu x^2/sd deęeri 0,9 olan 2PL modelle gösterdięi belirlenmiřtir. x^2/sd deęerlerinin 1PL ve 3PL modeller için de sırasıyla 1,2 ve 1,5 deęerleri aldıęı gözlenmiřtir. Bu deęerlerin de kritik deęer olan 3'ün altında deęerler almasından dolayı verinin bu iki modelle de uyum gösterdięi belirlenmiřtir.

Tablo 3.1'de “%” ile gösterilen deęer x^2/sd oranınının 10 maddelik testte bulunan her madde için ayrı ayrı hesaplanarak maddelerin her birinin modele uyup uymadıęının incelenmesinin ardından testte modele uyduęu belirlenen madde yüzdesini göstermektedir. Table 3.1 bu açıdan incelendięinde küçük örneklemlerde testte bulunan 10 maddenin hepsinin 1PL, 2PL ve 3PL modellerle uyum saęladıęı ve dolayısıyla uyum yüzdelerinin tüm örnekleme büyüklüklerinde %100 olduęu görölmektedir.

Küçük örneklemlerde 20 maddelik teste ait x^2/sd deęerleri Tablo 3.2'de gösterilmiřtir.

Tablo 3.2

Küçük Örneklerde 20 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
150	219,47	410	0,5	100	219,47	410	0,5	100	175,98	366	0,5	100
250	225,21	432	0,5	100	152,01	330	0,5	100	168,62	365	0,5	100
350	215,87	413	0,5	100	167,92	384	0,4	100	185,86	407	0,5	100
500	177,83	393	0,5	100	139,58	351	0,4	100	123,62	310	0,4	100
750	332,47	459	0,7	100	208,03	395	0,5	100	183,52	389	0,5	100

Tablo 3.2 incelendiğinde, 150 kişilik örnekleme uygulanan 10 maddelik test verisine ait x^2/sd değerlerinin 1PL, 2PL ve 3PL modellerde 0,5 değeri olarak eşit düzeyde olduğu görülmektedir. Bu durum bu verinin her 3 modele eşit düzeyde uyduğu şeklinde yorumlanmıştır.

250 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, tüm modellerde yine 0,5 değerinin elde edildiği görülmektedir. Bu da 150 kişilik örnekleme olduğu gibi verinin 250 kişilik örnekleme de 1PL, 2PL ve 3PL modellerle eşit düzeyde uyum sağladığı şeklinde yorumlanmıştır.

350 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en iyi uyumun 0,4 değeri ile 2PL modelle sağlandığı belirlenmiştir. 2PL ve 3PL modellerde de 0,5 değeri elde edilerek kritik değer olan 3'ten daha düşük değerler aldığı dikkate alındığında, 350 kişilik örnekleme ait verinin de yine 250 ve 150 kişilik örneklemlerde olduğu gibi her 3 modelle de uyumlu olduğu görülmektedir.

500 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, verinin 0,4 değeri ile en iyi 2PL ve 3PL modellerle uyumlu olduğu gözlenmektedir. 1PL modele ait değer 0,5 ile yine kritik değer olan 3'ün altında olduğu için 500 kişilik örnekleme ait veride de her 3 modelle uyumun sağlandığı tespit edilmiştir.

750 kişilik örnekleme ait x^2/sd değerleri incelendiğinde verinin en iyi uyumu 0,5 değeri ile 2PL ve 3PL modellerle sağladığı belirlenmiştir. 1PL modelde elde edilen x^2/sd değerleri olan 0,7 değeri de yine kritik değer olan 3'ün altında olduğundan dolayı, verinin bu modelle de uyumlu olduğu sonucuna varılmıştır.

Table 3.2 madde-model uyum yüzdeleri açısından incelendiğinde küçük örneklerde 20 maddelik testte bulunan maddelerin hepsinin 1PL, 2PL ve 3PL modellerin üçüyle de uyum sağladığı ve dolayısıyla uyum yüzdelerinin tüm örneklem büyüklüklerinde %100 olduğu görülmektedir.

Küçük örneklere uygulanan 30 maddelik teste ait x^2/sd değerleri Tablo 3.3'te gösterilmiştir.

Tablo 3.3

Küçük Örneklerde 30 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
150	301,30	661	0,5	100	214,77	530	0,4	100	234,62	535	0,4	100
250	282,05	533	0,5	100	199,40	518	0,4	100	220,66	601	0,4	100
350	354,17	611	0,6	100	240,55	618	0,4	100	279,83	635	0,4	100
500	299,26	575	0,5	100	205,81	566	0,4	100	195,74	529	0,4	100
750	498,35	604	0,8	97	228,53	557	0,4	100	253,87	577	0,4	100

Tablo 3.3 incelendiğinde, 150 kişilik örnekleme uygulanan 30 maddelik test verisine ait x^2/sd değerlerinin 2PL ve 3PL modellerde 0,4 değeri aldığı ve bu iki modele de eşit düzeyde uyumlu olduğu; 1PL modelde ise 0,5 değeri alarak bu modele de uyumlu olduğu görülmektedir. Bu da verinin her 3 modelle de uyumlu olduğu anlamına gelmektedir.

250 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, 2PL ve 3PL modelde 0,4, 1PL modelde 0,5 değeri aldığı görülmektedir. Bu da 250 kişilik örneklem verisinin de 1PL, 2PL ve 3PL modellerle uyumlu olduğu anlamına gelmektedir.

350 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en iyi uyumun 0,4 değeri ile 2PL ve 3PL modellerle sağlandığı belirlenmiştir. 1PL modelde ise 0,6 değeri ile yine kritik değer olan 3'ten daha düşük bir değer elde edildiğinden, 350 kişilik örnekleme ait

verinin yine 150 ve 250 kişilik örneklemlerde olduğu gibi her üç modele de uyumlu olduğu belirlenmiştir.

500 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, verinin 0,4 değeri ile en iyi 2PL ve 3PL modellerle uyumlu olduğu gözlenmektedir. 1PL modele ait değer ise 0,5 ile yine kritik değer olan 3'ün altında olduğu için 500 kişilik örnekleme ait verinin de her 3 modelle uyumlu olduğu tespit edilmiştir.

750 kişilik örnekleme ait x^2/sd değerleri incelendiğinde verinin en iyi uyumu 0,4 değerleri ile 2PL ve 3PL modellerle sağladığı belirlenmiştir. 1PL modelde elde edilen x^2/sd değerleri olan 0,8 değeri de yine kritik değer olan 3'ün altında olduğundan dolayı verinin bu modelle de uyumlu olduğu sonucuna varılmıştır.

Table 3.3 uyum yüzdeleri açısından incelendiğinde neredeyse küçük örneklemlerin tümünde 30 maddelik testte bulunan maddelerin hepsinin 1PL, 2PL ve 3PL modellerin üçüyle de uyum sağladığı ve dolayısıyla uyum yüzdelerinin tüm örneklem büyüklüklerinde %100 olduğu görülmektedir. Ancak 750 kişilik örnekleme ait verideki 30 maddenin %97'sinin 1PL modele uyum sağladığı görülmektedir.

3.1.1.2. Ortalama Standartlaştırılmış Artık (Standardized Residual) Analizi İle Model Veri Uyumunun Sınanması

10 maddelik testten küçük örneklem büyüklüklerinde kullanılan modele göre elde edilen ortalama standartlaştırılmış artık değerler Tablo 3.4'te gösterilmiştir.

Tablo 3.4

Küçük Örneklerde 10 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model	2PL Model	3PL Model
	Ort. Std. Artık	Ort. Std. Artık	Ort. Std. Artık
150	0,978	0,393	0,414
250	1,167	0,392	0,360
350	1,519	0,405	0,419
500	1,915	0,608	0,552
750	2,118	0,661	0,627

Tablo 3.4 incelendiğinde, 150 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 2PL modelde ulaştığı ve dolayısıyla 150 kişilik örnekleme ait 10 maddelik veri için en uygun modelin 2PL model olduğu görülmektedir. Yine aynı tabloda 250 kişilik örnekleme ait 10 maddelik test verisinin ortalama standartlaştırılmış artık değerinin en düşük 3PL modelde elde edildiği ve dolayısıyla 250 kişilik örneklem için en uygun modelin 3PL model olduğu; 350 kişilik örneklem için en düşük değerin 2PL modelde elde edildiği ve 350 kişilik örneklem için en uygun modelin 2PL model olduğu; 500 ve 750 kişilik örneklem için ise en düşük standartlaştırılmış artık değerlerinin 3PL modelde elde edildiği ve bu iki örnekleme ait 10 maddelik test verisi için en uygun modelin 3PL model olduğu belirlenmiştir.

20 maddelik test için küçük örneklerde ve 1PL, 2PL ve 3PL modellerde ortalama standartlaştırılmış artık değerler Tablo 3.5'te gösterilmiştir.

Tablo 3.5

Küçük Örneklerde 20 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model	2PL Model	3PL Model
	Ort. Std. Artık	Ort. Std. Artık	Ort. Std. Artık
150	0,625	0,429	0,487
250	0,753	0,465	0,438
350	0,808	0,458	0,464
500	0,950	0,400	0,405
750	1,097	0,511	0,448

Tablo 3.5 incelendiğinde, 150 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 2PL modelde ulaştığı ve dolayısıyla bu örnekleme ait 20 maddelik veri için en uygun modelin 2PL model olduğu görülmektedir. Tablo 3.5'te ayrıca 250 kişilik örnekleme ait 20 maddelik test verisinin ortalama standartlaştırılmış artık değerinin en düşük 3PL modelde elde edildiği ve dolayısıyla bu örneklem için en uygun modelin 3PL model olduğu; 350 kişilik örneklem için en düşük değerin 2PL modelde elde edildiği ve bu örneklem için en uygun modelin 2PL model olduğu; 500 kişilik örneklem için en düşük değerin 2PL modelde elde edildiği ve bu örneklem için en uygun modelin 2PL model olduğu; 750 kişilik örneklem için ise en düşük ortalama standartlaştırılmış artık değerinin 3PL modelde elde edildiği ve bu örnekleme ait 20 maddelik test verisi için en uygun modelin 3PL model olduğu belirlenmiştir.

30 maddelik testten farklı örneklem büyüklüklerinde ve modellerde elde edilen ortalama standartlaştırılmış artık değerler Tablo 3.6'da gösterilmiştir.

Tablo 3.6

Küçük Örneklerde 30 maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model	2PL Model	3PL Model
	Ort. Std. Artık	Ort. Std. Artık	Ort. Std. Artık
150	0,694	0,495	0,505
250	0,743	0,465	0,457
350	0,858	0,565	0,620
500	1,020	0,474	0,469
750	1,160	0,550	0,495

Tablo 3.6 incelendiğinde, 150 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 2PL modelde ulaştığı ve dolayısıyla bu örnekleme ait 30 maddelik veri için en uygun modelin 2PL model olduğu; 250 kişilik örnekleme ait 30 maddelik test verisinin ortalama standartlaştırılmış artık değerinin en düşük 3PL modelde elde edildiği ve dolayısıyla bu örneklem için en uygun modelin 3PL model olduğu; 350 ve 500 kişilik örneklem için en düşük değerin, 2PL modelde elde edildiği ve bu örneklem için en uygun modelin 2PL model olduğu; 750 kişilik örneklem için ise en düşük ortalama standartlaştırılmış artık değerinin 3PL modelde elde edildiği ve bu örnekleme ait 30 maddelik test verisi için en uygun modelin 3PL model olduğu belirlenmiştir.

3.1.1.3. G^2 Olabilirlik Oranı (Likelihood Ratio) Testi İle Model Veri Uyumunun Sınanması

Olabilirlik oranı testi (G^2) bir kıkare (chi-square) istatistiğidir. Olabilirlik oranı testi karşılaştırılan iki model arasındaki sapmadan ibarettir. İki model arasındaki sapma kıkare istatistiği olarak dağılım göstermektedir. Bu nedenle hangi modelin daha iyi uyum sağladığı manidarlık testine tabi tutulabilmektedir (Andersen, 1973; Baker ve Kim,

2004; Bock ve Aitkin, 1981). Küçük örneklerde 10 maddelik teste ait G^2 değerleri Tablo 3.7’de gösterilmiştir.

Tablo 3.7

Küçük Örneklerde 10 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

Model	150		250		350		500		750	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	1130,43		1918,34		2664,03		3669,17		5743,48	
2PL	1111,39	19,05*	1874,26	44,08*	2607,66	56,37*	3619,82	49,35*	5628,58	114,91*
3PL	1119,06	7,67	1870,29	3,97	2632,39	24,73*	3622,31	2,49	5622,49	6,09

* $p (X^2_{df=10} > 18.30) < 0.05$

Tablo 3.7’de 150, 250, 350, 500 ve 750 kişilik örneklemere ait G^2 değerleri incelendiğinde, 2PL modele ait G^2 değerleri ile 1PL modele ait G^2 değerleri arasında küçük örneklemelerin tümünde manidar bir fark olduğu belirlenmiştir. Bu durum 150, 250, 350, 500, 750 kişilik örneklemere ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığını göstermektedir. Tablo 3.7 tekrar incelendiğinde, 150, 250, 350, 500 ve 750 kişilik örneklemere ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri arasında farkın manidar olmadığı görülmektedir. Bu durumda 2PL ve 3PL modellerde birbirinden manidar olarak farklı olmayan değerler elde edildiğinden verilerin her iki modele de uyduğu şeklinde yorumlanmıştır.

Küçük örneklerde 20 maddelik teste ait G^2 değerleri Tablo 3.8’de gösterilmiştir.

Tablo 3.8

Küçük Örneklerde 20 maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

Model	150		250		350		500		750	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	2663,99		4644,26		6449,30		9080,25		13826,55	
2PL	2626,27	37,72*	4590,01	54,26*	6393,02	56,28*	8999,94	80,32*	13687,54	139,01*
3PL	2643,68	17,41	4589,79	0,22	6388,41	4,61	9001,04	1,10	13679,49	8,05

*p ($X^2_{df=20} > 31.41$) < 0.05

Tablo 3.8’de 150, 250, 350, 500 ve 750 kişilik örneklemere ait G^2 değerleri incelendiğinde, 2PL modele ait G^2 değeri ile 1PL modele ait G^2 değeri arasında istatistiksel olarak manidar bir fark olduğu belirlenmiştir. Bu durum 150, 250, 350, 500, 750 kişilik örneklemere ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığı şeklinde yorumlanmaktadır. Tablo 3.8 tekrar incelendiğinde 150, 250, 350, 500 ve 750 kişilik örneklemere ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri arasında manidar bir fark olmadığı belirlenmiştir. Bu durum yine 10 maddelik test verisinde olduğu gibi verilerin her iki modele de uyumlu olduğu şeklinde yorumlanmıştır.

Küçük örneklerde 30 maddelik teste ait G^2 değerleri Tablo 3.9’da gösterilmiştir.

Tablo 3.9

Küçük Örneklerde 30 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

Model	150		250		350		500		750	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	4379,88		7466,30		10344,64		14777,99		22167,62	
2PL	4301,93	77,94*	7372,93	93,37*	10234,55	110,09*	14587,19	190,80*	21839,80	327,82*
3PL	4325,16	23,23	7381,02	8,09	10239,57	5,02	14582,81	4,38	21816,72	23,08

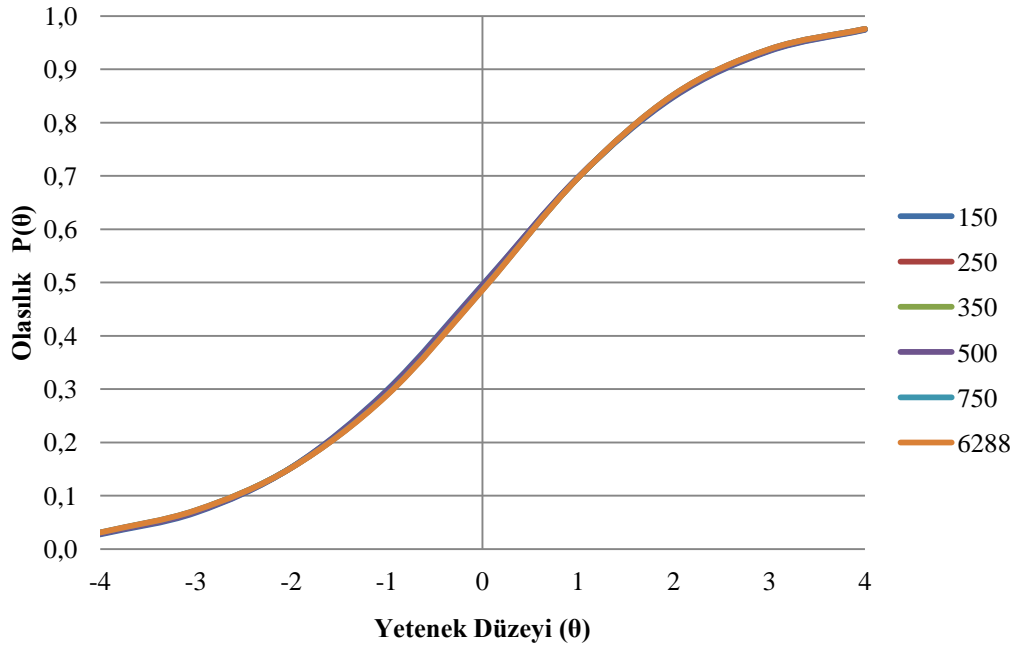
*p ($X^2_{df=30} > 43.77$) < 0.05

Tablo 3.9’da 150, 250, 350, 500 ve 750 kişilik örneklemlere ait G^2 değerleri incelendiğinde, 2PL modele ait G^2 değerleri ile 1PL modele ait G^2 değerleri arasında istatistiksel olarak manidar bir fark olduğu belirlenmiştir. Bu durum 150, 250, 350, 500, 750 kişilik örneklemlere ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığı şeklinde yorumlanmaktadır. Tablo 3.9 tekrar incelendiğinde 150, 250, 350, 500 ve 750 kişilik örneklemlere ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri arasında manidar bir fark olmadığı belirlenmiştir. Bu da verilerin belirtilen örneklem büyüklüklerinde 10 ve 20 maddelik testlerde olduğu gibi verilerin her iki modele de uyum sağladığı şeklinde yorumlanmıştır.

3.1.1.4. Test Karakteristik Eğrilerinin İncelenmesi İle Model Veri Uyumunun Sınanması

Madde ve yetenek parametrelerinin değişmezliği model veri uyumu için önemli bir kanıttır. Modelin veriye uyması durumunda MTK’nın bu özelliğinden yararlanılacağı düşünülür. Daha önce de belirtildiği gibi test karakteristik eğrileri tüm maddeler için ortalama doğru cevaplanma olasılıklarını örtük yapının (underlying construct) bir fonksiyonu olarak göstermektedir (Weiss, 1995). Bu nedenle model veri uyumu test karakteristik eğrileri ile incelenmiştir.

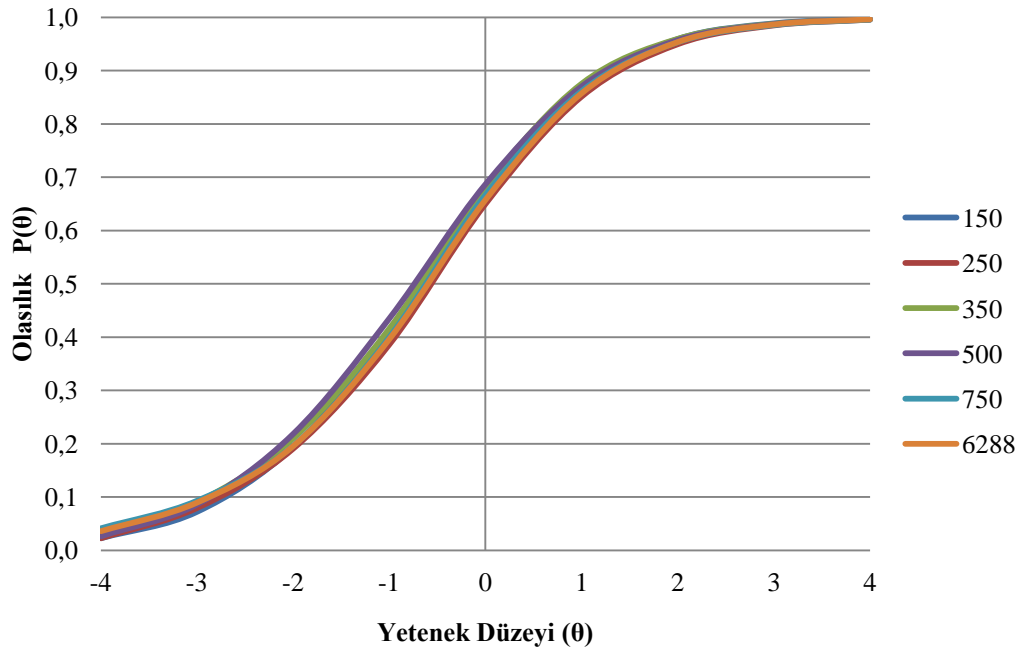
Küçük örneklemlerden 1PL modelde elde edilen test karakteristik eğrileri Şekil 10’da gösterilmiştir.



Şekil 10. Küçük örneklerde 10 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri.

Şekil 10 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 1PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden kestirilen test karakteristik eğrisiyle de neredeyse aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemere uygulanan 10 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak yorumlanmıştır.

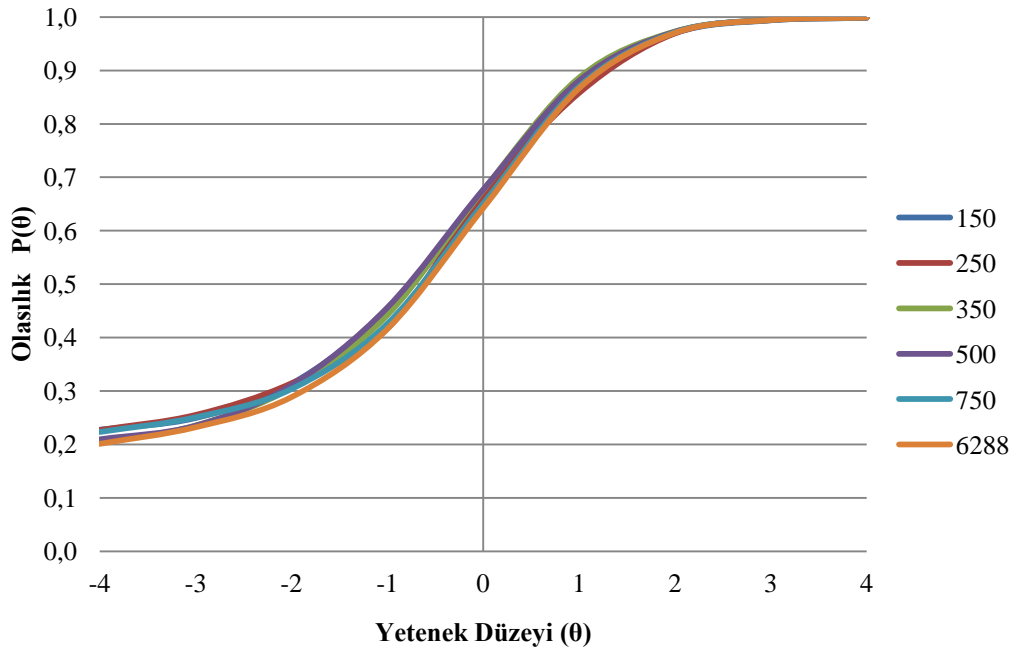
Küçük örneklemelerden 2PL modelde elde edilen test karakteristik eğrileri Şekil 11'de gösterilmiştir.



Şekil 11. Küçük örneklerde 10 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 11 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 2PL modelde elde edilen test karakteristik eğrileri ile çalışma evreninden kestirilen test karakteristik eğrisinin neredeyse aynı olduğu görülmüştür. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemelere uygulanan 10 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak yorumlanmıştır.

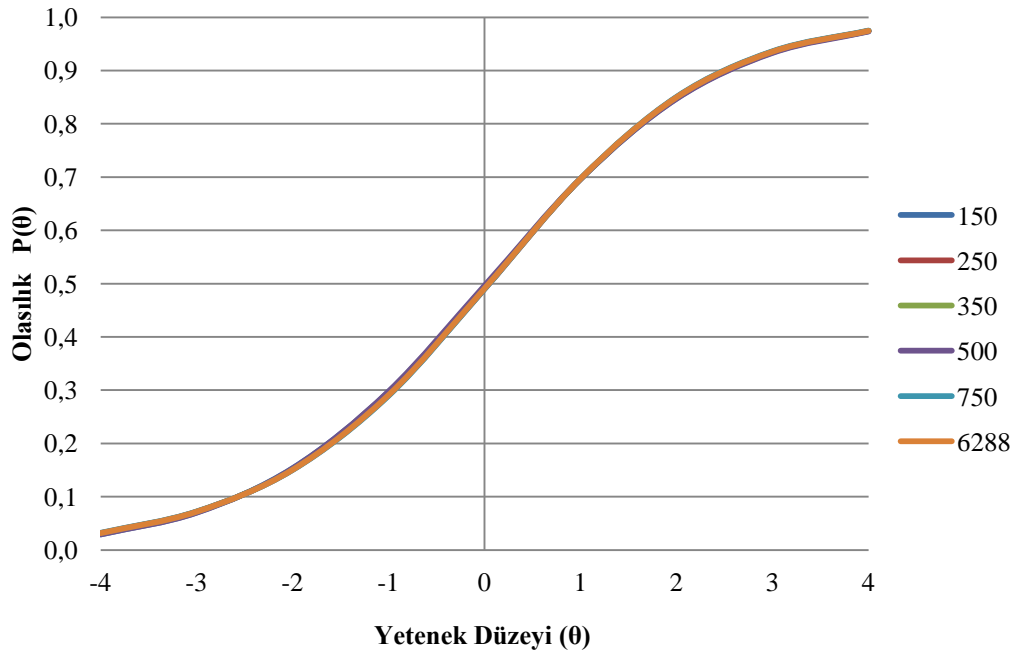
Küçük örneklerde 10 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri Şekil 12’de gösterilmiştir.



Şekil 12. Küçük örneklerde 10 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 12 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 3PL modelde elde edilen test karakteristik eğrileri ile çalışma evreninden kestirilen test karakteristik eğrisinin neredeyse aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve küçük örneklemelere uygulanan 10 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

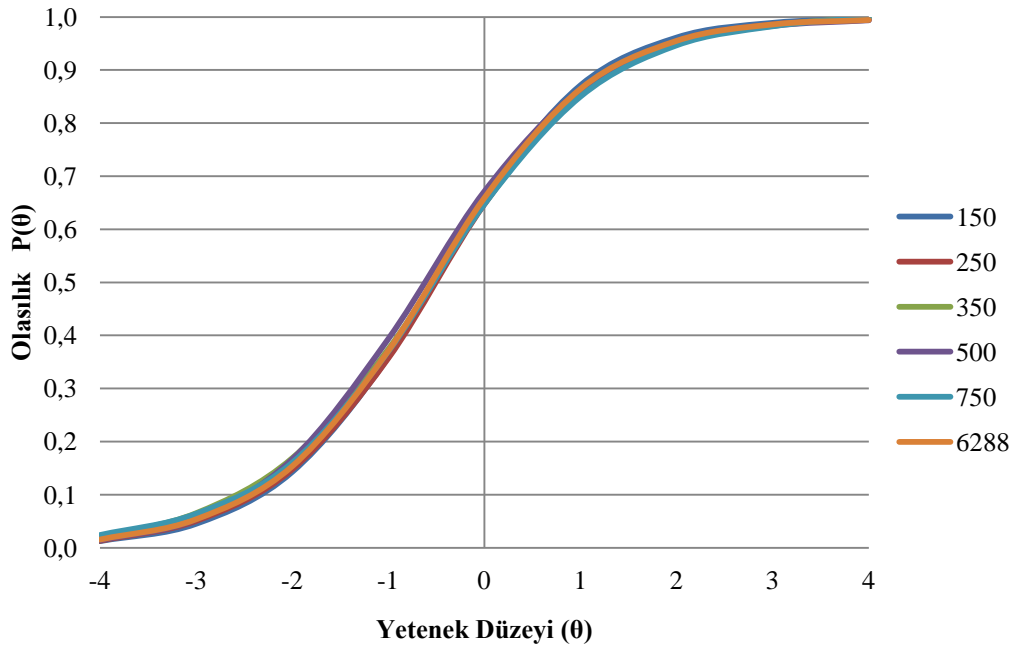
Küçük örneklerde 20 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri Şekil 13'te gösterilmiştir.



Şekil 13. Küçük örneklerde 20 maddelik testten 1PL modelde elde edilen test karakterististik eğrileri.

Şekil 13 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 1PL modelde elde edilen test karakterististik eğrilerinin çalışma evreninden elde edilen test karakterististik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemelere uygulanan 20 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak kabul edilmiştir.

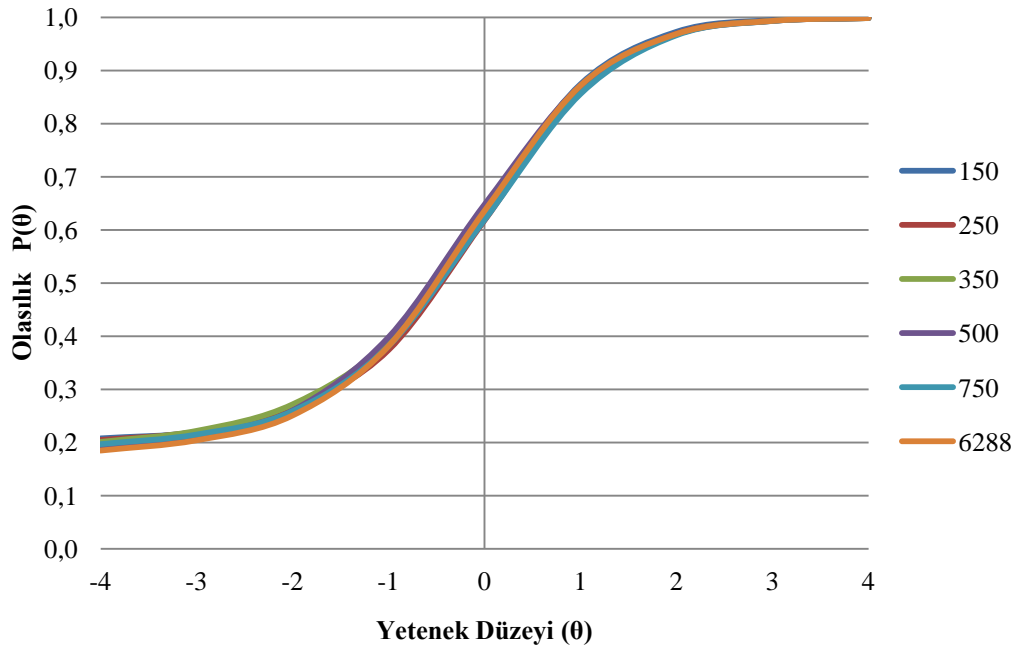
Küçük örneklerde 20 maddelik testten 2PL modelde elde edilen test karakterististik eğrileri Şekil 14'te gösterilmiştir.



Şekil 14. Küçük örneklerde 20 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 14 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 2PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle büyük oranda benzer olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemelere uygulanan 20 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

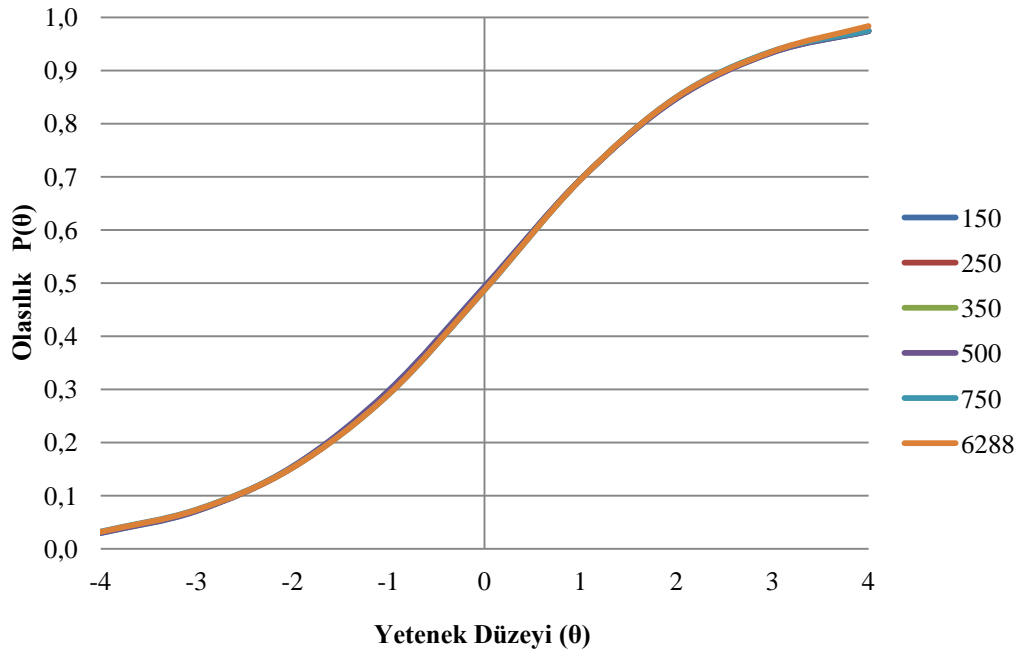
Küçük örneklerde 20 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri Şekil 15'te gösterilmiştir.



Şekil 15. Küçük örneklerde 20 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 15 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 3PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden kestirilen test karakteristik eğrisiyle aynı olması dikkat çekicidir. Bu durum veride değişmezliğin sağlandığının ve küçük örneklemelere uygulanan 20 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

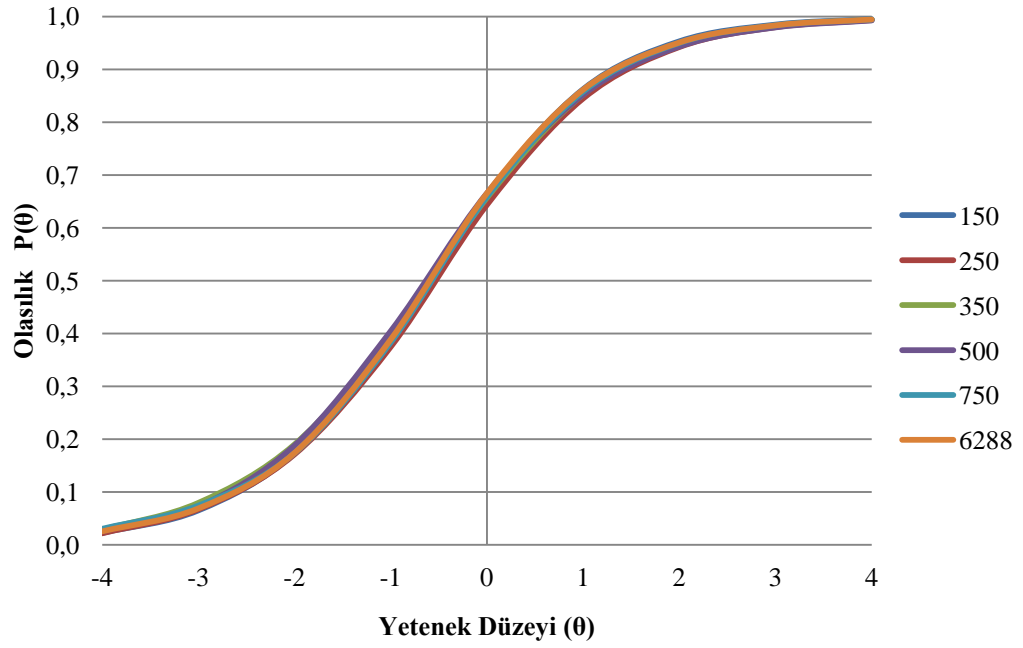
Küçük örneklemelerden 30 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri Şekil 16'da gösterilmiştir.



Şekil 16. Küçük Örneklerde 30 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri.

Şekil 16 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 1PL modelde elde edilen test karakteristik eğrilerinin birbirinden farklı olmadığı görülmektedir. Bu test karakteristik eğrilerinin çalışma evrenine ait test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemde uygulanan 30 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak kabul edilmiştir.

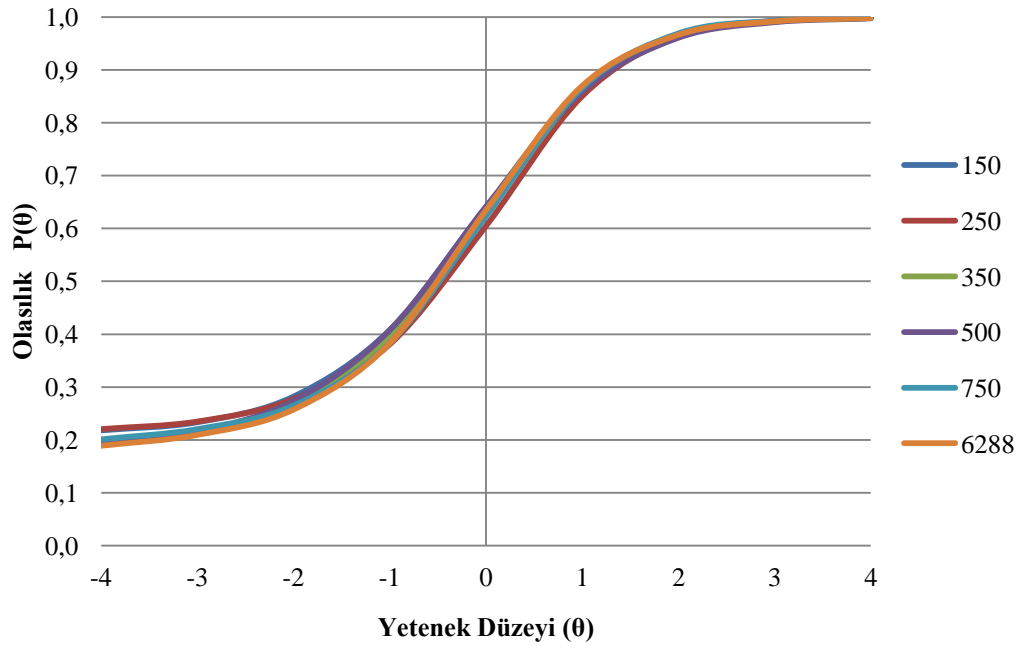
Küçük örneklerde 30 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri Şekil 17'de gösterilmiştir.



Şekil 17. Küçük örneklerde 30 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 17 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde 2PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden kestirilen test karakteristik eğrisiyle de büyük oranda benzer olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla küçük örneklemelere uygulanan 20 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

Küçük örneklerde 30 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri Şekil 18’de gösterilmiştir.



Şekil 18. Küçük örneklerde 30 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 18 incelendiğinde, sırasıyla 150, 250, 350, 500, 750 kişilik örneklerde, 3PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden kestirilen test karakteristik eğrisiyle büyük oranda benzerlik gösterdiği görülmektedir. Bu durum veride değişmezliğin sağlandığının ve küçük örneklemelere uygulanan 30 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

Sonuç olarak küçük örneklemelere uygulanan 10 maddelik test verisinde model veri uyumunun sınanmasında kullanılan farklı yöntemler ve uyumlu modeller Tablo 3.10'da gösterilmiştir.

Tablo 3.10

Küçük Örneklerde 10 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G^2	TKE
150	1,2,3	2	2,3	1,2,3
250	1,2,3	3	2,3	1,2,3
350	1,2,3	2	2,3	1,2,3
500	1,2,3	3	2,3	1,2,3
750	1,2,3	3	2,3	1,2,3

Tablo 3.10 incelendiğinde, küçük örneklerde 10 maddelik test verisinde model veri uyumunun sınanmasında kullanılan dört farklı yöntemden elde edilen sonuçlar birbiri ile farklılık gösterse de, model veri uyumunun sınanmasında kullanılan dört farklı yöntemden SAA dışındaki 3 yöntemde elde edilen sonuçlar model veri uyumunun örneklem büyüklüğünden etkilenmediğini göstermektedir. SAA modelinden elde edilen sonuçlar bu yöntemle model veri uyumunun sınanması durumunda farklı örneklerde farklı sonuçlar elde edilebileceğini göstermektedir.

Küçük örneklerde 20 maddelik test verisinde model veri uyumunun sınanmasında kullanılan farklı yöntemler ve uyumlu modeller Tablo 3.11’de gösterilmiştir.

Tablo 3.11

Küçük Örneklerde 20 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G^2	TKE
150	1,2,3	2	2,3	1,2,3
250	1,2,3	3	2,3	1,2,3
350	1,2,3	2	2,3	1,2,3
500	1,2,3	2	2,3	1,2,3
750	1,2,3	3	2,3	1,2,3

Tablo 3.11 incelendiğinde, model veri uyumunun sınanmasında kullanılan üç yöntemin örneklem büyüklüğünden etkilenmeden her bir örneklem büyüklüğünde aynı modellerde uyumu işaret ettiği görülmektedir. 10 maddelik testte olduğu gibi SAA yöntemi 20 maddelik testte de örneklem büyüklüğünden etkilenerek farklı örneklem büyüklüklerinde farklı modellerle uyumu işaret ettiği görülmektedir. Bu bilgiler ışığında 20 maddelik bir test verisinin model veri uyumunun örneklem büyüklüğüne göre değişmediği sonucuna varılmıştır.

Küçük örneklerde 30 maddelik test verisinde model veri uyumunun sınanmasında kullanılan farklı yöntemler ve uyumlu modeller Tablo 3.12’te gösterilmiştir.

Tablo 3.12

Küçük Örneklerde 30 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G^2	TKE
150	1,2,3	2	2,3	1,2,3
250	1,2,3	3	2,3	1,2,3
350	1,2,3	2	2,3	1,2,3
500	1,2,3	2	2,3	1,2,3
750	1,2,3	3	2,3	1,2,3

Tablo 3.12 incelendiğinde, model veri uyumunun sınanmasında kullanılan dört farklı teknikten SAA dışındaki 3 yöntemde elde edilen sonuçlar model veri uyumunun örneklem büyüklüğünden etkilenmediğini göstermektedir. Ancak SAA yöntemi 10 ve 20 maddelik testler verilerinde olduğu gibi 30 maddelik test verisinde de farklı örneklem büyüklüklerinde farklı modellerle uyumu işaret ettiği görülmektedir. Elde edilen bilgiler ışığında 30 maddelik bir test verisinde de model veri uyumunun örneklem büyüklüğüne göre değişmediği sonucuna varılmıştır.

Tüm sonuçlar genel olarak değerlendirildiğinde, model veri uyumunun genel olarak incelenmesini sağlayan yöntemlerden SAA yöntemi dışındaki diğer yöntemlerin küçük örneklerde model veri uyumunun genel olarak incelenmesinde kullanımı için daha uygun olabileceği şeklinde yorumlanmıştır.

3.1.2. Büyük Örneklerde (1000, 2000, 3000, 5000) ve çalışma evreninde 10, 20 ve 30 Maddelik Testlere Ait Verilerin Model Veri Uyumu Örneklem Büyüklüğüne ve Test Uzunluğuna Göre Değişmekte midir?

Büyük örneklerde ve çalışma evreninde 10, 20 ve 30 maddelik testlere ait verilerde model veri uyumu yine dört farklı yöntemle (x^2/sd oranı, SAA, G^2 , test karakteristik eğrilerinin incelenmesi) sınanmıştır.

3.1.2.1. X^2/sd Oranı İle Model Veri Uyumunun Sınanması

Bu yöntemde, x^2/sd oranı 3'ten küçük değerler alan verinin o modele uyum sağladığı kabul edilmiştir. Elde edilen x^2/sd değerleri Tablo 3.13'te gösterilmiştir.

Tablo 3.13

Büyük Örneklerde 10 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
1000	130,75	143	0,9	100	136,56	150	0,9	100	361,65	175	2,1	90
2000	317,82	148	2,1	90	291,60	174	1,7	90	547,80	200	2,7	90
3000	401,37	149	2,7	60	418,58	169	2,5	80	954,60	204	4,7	50
5000	606,91	150	4,0	60	555,73	167	3,3	80	1658,32	208	8,0	0
6288*	751,03	150	5,0	50	677,72	172	3,9	70	1964,84	208	9,4	0

*Çalışma Evreni

Tablo 3.13'teki x^2/sd değerleri incelendiğinde, 1000 kişilik örnekleme ait 10 maddelik test verisinin 1PL ve 2PL modeller için 0,9 değeri aldığı ve bu yöntemle göre iki modele de eşit oranda uyum gösterdiği görülmektedir. 3PL modelde ise 2,1 değeri ile kritik değer olan 3'ten düşük bir değer aldığı görülmektedir. Ayrıca madde uyum yüzdeleri incelendiğinde; testte bulunan 10 maddenin tümünün 1PL ve 2PL modeller ile %100, 3PL model ile %90 oranında uyumlu olduğu görülmektedir. Bu bilgiler ışığında 1000

kişilik örnekleme ait 10 maddelik test verisinin farklı yüzdelerle de olsa her 3 modele de uyumlu olduğu tespit edilmiştir.

2000 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en düşük değer 1,7 ile 2PL modele ait olduğu, 1PL ve 3PL modellere ait değerler incelendiğinde, sırasıyla 2,1 ve 2,7 ile yine kritik değer olan 3'ün altında değerler aldığı görülmektedir. Ayrıca madde uyum yüzdeleri incelendiğinde her üç modelde de uyum yüzdesinin %90 olduğundan dolayı verinin her üç modele de uyumlu olduğu belirlenmiştir.

3000 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en düşük 2,5 değeri ile 2PL modelle en iyi uyumu sağladığı, 1PL modelde ise 2,7 değeri aldığı ve bu modelle de uyumlu olduğu belirlenmiştir. 3PL modelde elde edilen x^2/sd değeri, 3 kritik değerinin üzerinde olduğundan dolayı verinin x^2/sd yöntemine göre 3PL model ile uyumlu olmadığı sonucuna varılmıştır. Ayrıca uyum yüzdesi %60'ta kaldığı için verinin 1PL model ile de uyum sağlamadığı belirlenmiştir.

5000 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, 1PL modelde 4,0, 2PL modelde 3,4 ve 3PL modelde ise 8,0 değerleri aldığı görülmektedir. Bu bilgiler ışında x^2/sd yöntemine göre 5000 kişilik verinin hiçbir model ile tam uyum sağlamadığı belirlenmiştir ancak madde uyum yüzdeleri incelendiğinde 2PL modelde maddelerin %80'inin modele uyum sağladığından dolayı verinin en iyi 2PL modelle uyumlu olduğu belirlenmiştir.

Çalışma evrenine ait x^2/sd değeri incelendiğinde, 1PL modelde 5,0, 2PL modelde 3,9 ve 3PL modelde ise 9,4 değerleri aldığı görülmektedir. Bu bilgiler ışında x^2/sd yöntemine göre 6288 kişilik verinin hiçbir model ile tam uyum sağlamadığı belirlenmiştir ancak madde uyum yüzdeleri incelendiğinde maddelerin %70'inin 2PL modele uyum sağladığından dolayı yine verinin en iyi 2PL modelle uyumlu olduğu kabul edilmiştir.

Büyük örneklemlerde 20 maddelik teste ait x^2/sd oranları Tablo 3.14'te gösterilmiştir.

Tablo 3.14

Büyük Örneklemelerde 20 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
1000	258,11	451	0,6	100	206,88	412	0,5	100	215,74	385	0,6	100
2000	475,39	510	0,9	100	319,77	438	0,7	100	334,20	426	0,8	100
3000	535,45	521	1,0	100	336,53	457	0,7	100	414,32	462	0,9	100
5000	746,38	527	1,4	95	662,38	472	1,4	90	571,54	451	1,3	95
6288*	890,67	532	1,7	90	757,89	477	1,6	80	736,37	476	1,5	95

*Çalışma evreni

Tablo 3.14'teki x^2/sd değerleri incelendiğinde, 1000 kişilik örnekleme ait 20 maddelik test verisinin 1PL ve 3PL modeller için 0,6 değeri aldığı ve bu yöntemle göre iki modele de eşit oranda uyum gösterdiği görülmektedir. 2PL modelde ise 0,5 değerini aldığı ve en iyi uyumu bu modelle gösterdiği görülmektedir. Bu bilgiler ışığında 1000 kişilik örnekleme ait 10 maddelik test verisinin her 3 modele de uyumlu olduğu belirlenmiştir.

2000 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en düşük değer 0,7 ile 2PL modele ait olduğu ve dolayısıyla verinin en iyi 2PL modelle uyum sağladığı görülmektedir. Ancak 1PL ve 3PL modellere ait değerler incelendiğinde sırasıyla 0,9 ve 0,8 ile verinin bu iki modelle de uyumlu olduğu belirlenmiştir.

3000 kişilik örnekleme ait değerler incelendiğinde, en düşük 0,7 değeri olarak, en iyi uyumu 2PL model ile sağladığı, 3PL modelde 0,9, 1PL modelde 1,0 değeri olarak verinin yine her 3 modelle de uyumlu olduğu tespit edilmiştir.

5000 kişilik örnekleme ait x^2/sd değeri incelendiğinde, 1,3 değeri ile en iyi uyumu 3PL modelle sağladığı, ayrıca 1PL ve 2PL modellerde 1,4 değeri olarak bu modellerle de uyumlu olduğu tespit edilmiştir.

Çalışma evrenine ait x^2/sd değeri incelendiğinde; 1PL modelde 1,7, 2PL modelde 1,6 ve 3PL modelde ise 1,5 değerleri aldığı görülmektedir. Bu bilgiler ışında x^2/sd yöntemine göre çalışma evrenine ait verinin her üç modelle de uyum sağladığı belirlenmiştir.

Maddelerin modellere uyum yüzdeleri genel olarak incelendiğinde 1000, 2000 ve 3000 kişilik örneklerde maddelerin tümünün 1PL, 2PL, 3PL modellerle uyumlu olduğu; dolayısıyla da uyum yüzdelerinin %100 olduğu; 5000 kişilik örneklerde ise maddelerin %95'inin 1PL ve 3PL modeller ile uyumlu olduğu, 2PL ile ise %90'ının uyumlu olduğu görülmektedir. Çalışma evrenine ait veride 20 maddenin %90'ının 1PL modele, %80'inin 2PL modele, %95'inin ise 3PL modele uyum sağladığı görülmektedir.

Büyük örneklerde 30 maddelik teste ait x^2/sd oranları Tablo 3.15'te gösterilmiştir.

Tablo 3.15

Büyük Örneklerde 30 Maddelik Teste Ait x^2/sd Değerleri

Örneklem	1PL Model				2PL Model				3PL Model			
	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%	x^2	sd	x^2/sd	%
1000	598,29	632	0,9	97	285,46	622	0,5	100	291,41	628	0,5	100
2000	925,02	721	1,3	90	320,81	655	0,5	100	413,79	702	0,6	100
3000	1497,61	783	1,9	87	439,38	718	0,6	100	498,50	717	0,7	100
5000	2333,13	808	2,9	83	586,66	747	0,8	100	599,31	743	0,8	100
6288*	2792,15	817	3,4	77	647,10	752	0,9	100	700,69	745	0,9	97

*Çalışma Evreni

Tablo 3.15'teki x^2/sd değerleri incelendiğinde, 1000 kişilik örnekleme ait 30 maddelik test verisinin 2PL ve 3PL modeller için 0,5 değeri aldığı, 1PL modelde ise 0,9 değerini aldığı ve bu modelle de uyumlu olduğu görülmektedir. Bu bilgiler ışığında 1000 kişilik

örnekleme ait 30 maddelik test verisinin her 3 modele de uyumlu olduğu tespit edilmiştir.

2000 kişilik örnekleme ait x^2/sd değerleri incelendiğinde, en düşük değer 0,5 ile 2PL modele ait olduğu ve dolayısıyla verinin en iyi 2PL modelle uyum sağladığı görülmektedir. Ancak 1PL ve 3PL modellere ait değerler incelendiğinde sırasıyla 1,3 ve 0,6 değerleri ile verilerin bu iki modelle de uyumlu olduğu belirlenmiştir.

3000 kişilik örnekleme ait değerler incelendiğinde, en düşük 0,6 değeri ile en iyi uyumu 2PL model ile sağladığı, 3PL modelde 0,7 ve 1PL modelde ise 1,9 değeri olarak verilerin her 3 modelle de uyumlu olduğu tespit edilmiştir.

5000 kişilik örnekleme ait x^2/sd değeri incelendiğinde, 2PL ve 3PL modellerde 0,8 değeri ile her iki modele de eşit oranda uyumlu olduğu belirlenmiştir. Ayrıca 1PL modelde elde edilen 1,9 değeri de verinin 1 parametrelili modelle de uyumlu olduğunu göstermektedir.

Çalışma evrenine ait x^2/sd değerleri incelendiğinde, 1PL modelde 3,4 değeri aldığı belirlenmiştir. Bu modele 30 maddelik test maddelerinin %77'sinin uyumlu olduğu belirlendiğinden dolayı verinin bu modele de uyumlu olduğu kabul edilmiştir. Verinin 2PL ve 3PL modellerde 0,9 değeri olarak her iki modelle de eşit oranda uyumlu olduğu belirlenmiştir. Bu bilgiler ışığında x^2/sd yöntemine göre çalışma evrenine ait veri 1PL, 2PL ve 3PL modellerle uyum sağladığı söylenebilir.

Tablo 3.15'te madde uyum yüzdeleri incelendiğinde 2PL modelde tüm örneklem büyüklüklerinde maddelerin uyum yüzdelerinin %100 olduğu; 1PL modelde uyum yüzdelerinin %77 ile %97 arasında değiştiği ve 3PL modelde ise çalışma evrenine (%97) ait veri dışındaki verilerde maddelerin %100'ünün modelle uyum sağladığı belirlenmiştir.

3.1.2.2. Ortalama Standartlaştırılmış Artık (Standardized Residual) Analizi İle Model Veri Uyumunun Sınanması

Farklı örneklem büyüklüklerinde ve modellerde elde edilen ortalama standartlaştırılmış artık değerler Tablo 3.16’da gösterilmiştir.

Tablo 3.16

Büyük Örneklerde 10 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model	2PL Model	3PL Model
	Ort. Std. Artık	Ort. Std. Artık	Ort. Std. Artık
1000	2,472	0,815	0,751
2000	3,546	1,188	1,048
3000	4,253	1,419	1,272
5000	5,717	1,843	1,649
6288*	6,337	2,051	1,850

*Çalışma evreni

Tablo 3.16 incelendiğinde, 1000 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 0,751 ile 3PL modelde ulaştığı ve bu örneklem verisi için en uyumlu modelin 3PL model olduğu görülmektedir. 2000 kişilik örnekleme ait ortalama standartlaştırılmış artık değerler incelendiğinde, en düşük değer 1,048 ile 3PL modelde elde edildiği ve dolayısıyla 2000 kişilik örneklem verisi için de en uyumlu modelin 3PL model olduğu görülmektedir. 3000 kişilik örneklem verisi için en düşük değer 1,272 ile 3PL modelde elde edildiği ve 3000 kişilik örneklem verisi için en uyumlu modelin 3PL model olduğu görülmektedir. 5000 kişilik örneklem verisi için en düşük değer 1,649 ile yine 3PL modelden elde edildiği ve 5000 kişilik örneklem verisi için de 3PL modelin en uyumlu model olduğu belirlenmiştir. Son olarak çalışma

evreninden elde edilen en düşük standartlaştırılmış artık değerlerinin 1,850 ile 3PL modelde elde edildiği ve bu veriyle de en uyumlu modelin 3PL model olduğu sonucuna varılmıştır.

20 maddelik testten farklı örneklem büyüklüklerinde ve modellerde elde edilen ortalama standartlaştırılmış artık değerler Tablo 3.17’de gösterilmiştir.

Tablo 3.17

Büyük Örneklerde 20 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model	2PL Model	3PL Model
	Ort. Std. Artık	Ort. Std. Artık	Ort. Std. Artık
1000	1,244	0,485	0,464
2000	1,729	0,640	0,509
3000	2,058	0,655	0,511
5000	2,729	0,806	0,602
6288*	1,244	0,485	0,464

*Çalışma evreni

Tablo 3.17 incelendiğinde, 1000 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 0,464 ile 3PL modelde ulaştığı ve bu örneklem verisi için en uyumlu modelin 3PL model olduğu görülmektedir. 2000 kişilik örnekleme ait ortalama standartlaştırılmış artık değerler incelendiğinde, en düşük değer 0,509 ile 3PL modelde elde edildiği ve dolayısıyla 2000 kişilik örneklem verisi için de en uyumlu modelin 3PL model olduğu görülmektedir. 3000 kişilik örneklem verisi için en düşük değer 0,511 ile 3PL modelde elde edildiği ve 3000 kişilik örneklem verisi için en uyumlu modelin 3PL model olduğu görülmektedir. 5000 kişilik örneklem verisi için en

düşük değerin 0,602 ile yine 3PL modelden elde edildiği ve 5000 kişilik örneklem verisi için de 3PL modelin en uyumlu model olduğu belirlenmiştir. Son olarak çalışma evreninden kestirilen parametreler için en düşük standartlaştırılmış artık değerlerin 0,464 ile 3PL modelde elde edildiği ve bu veriyle de en uyumlu modelin 3PL model olduğu sonucuna varılmıştır.

30 maddelik testten farklı örneklem büyüklüklerinde ve modellerde elde edilen ortalama standartlaştırılmış artık değerler Tablo 3.18’de gösterilmiştir.

Tablo 3.18

Büyük Örneklerde 30 Maddelik Teste Ait Ortalama Standartlaştırılmış Artık Değerler

Örneklem	1PL Model		2PL Model		3PL Model	
	Ort.	Std. Artık	Ort.	Std. Artık	Ort.	Std. Artık
1000		1,431		0,539		0,529
2000		1,847		0,656		0,584
3000		2,287		0,701		0,613
5000		2,990		0,773		0,639
6288*		3,249		0,871		0,692

*Çalışma Evreni

Tablo 3.18 incelendiğinde, 1000 kişilik örnekleme ait verinin en düşük ortalama standartlaştırılmış artık değerine 0,529 ile 3PL modelde ulaştığı ve bu örneklem verisi için en uyumlu modelin 3PL model olduğu görülmektedir. 2000 kişilik örnekleme ait ortalama standartlaştırılmış artık değerler incelendiğinde, en düşük değerin 0,584 ile 3PL modelde elde edildiği ve dolayısıyla 2000 kişilik örneklem verisi için de en uyumlu modelin 3PL model olduğu görülmektedir. 3000 kişilik örneklem verisi için en düşük değerin 0,613 ile 3PL modelde elde edildiği ve 3000 kişilik örneklem verisi için en

uyumlu modelin 3PL model olduğu görülmektedir. 5000 kişilik örneklem verisi için ise en düşük değerin 0,639 ile yine 3PL modelden elde edildiği ve 5000 kişilik örneklem verisi için de 3PL modelin en uyumlu model olduğu belirlenmiştir. Son olarak 6288 kişilik çalışma evreninden kestirilen parametrelere ait verinin en düşük standartlaştırılmış artık değerlerinin 0,692 ile yine 3PL modelde elde edildiği ve bu veriyle de en uyumlu modelin 3PL model olduğu sonucuna varılmıştır.

3.1.2.3. G^2 Olabilirlik Oranı (Likelihood Ratio) Testi İle Model Veri Uyumunun Sınanması

Büyük örneklerde 10 maddelik teste ait G^2 değerleri Tablo 3.19’da gösterilmiştir.

Tablo 3.19

Büyük Örneklerde 10 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

Model	1000		2000		3000		5000		6288**	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	7436,82		14966,61		22540,29		37197,89		47028,08	
2PL	7347,44	89,39*	14710,26	256,35*	22205,39	334,90*	36666,17	531,72*	46349,24	678,84*
3PL	7347,33	0,10	14652,77	57,49*	22142,82	62,57*	36523,98	142,19*	46187,67	161,57*

* $p(X^2_{df=10} > 18.30) < 0.05$, **Çalışma evreni

Tablo 3.19’da 1000, 2000, 3000, 5000 kişilik örneklem ve çalışma evrenine ait G^2 değerleri incelendiğinde, bu örneklem türlerinde 2PL modele ait G^2 değerleri ile 1PL modele ait G^2 değerleri arasında manidar bir fark olduğu belirlenmiştir. Bu durum 1000, 2000, 3000, 5000 kişilik örneklem ve çalışma evrenine ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığı şeklinde yorumlanmaktadır. Bu örneklem türlerine ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri karşılaştırıldığında 2000, 3000, 5000 kişilik örneklem ve çalışma evrenine ait verilerde bu iki değer arasındaki farkın manidar olduğu, dolayısıyla verilerin 3PL modele 2PL modelden daha iyi uyum sağladığı sonucuna varılmıştır.

Sadece 1000 kişilik örneklem verisinden 2PL ve 3PL modellerde elde edilen G^2 değerleri arasında manidar bir farkın olmadığı ve verinin her iki modele de uyum sağladığı belirlenmiştir.

Büyük örneklemelerde 20 maddelik teste ait G^2 değerleri Tablo 3.20’de gösterilmiştir.

Tablo 3.20

Büyük Örneklemelerde 20 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

Model	1000		2000		3000		5000		6288**	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	18170,52		36501,09		54632,27		90804,98		114474,95	
2PL	18061,79	108,73*	36240,58	260,51*	54250,97	381,30*	90271,75	533,22*	113793,92	681,03*
3PL	18048,41	13,38	36154,42	86,16*	54156,30	94,67*	90060,30	211,45*	113526,12	267,80*

*P ($X^2_{df=20} > 31.41$) < 0.05, **Çalışma evreni

Tablo 3.20’de 1000, 2000, 3000, 5000 kişilik örneklemelere ve 6288 kişilik çalışma evrenine ait G^2 değerleri incelendiğinde bu örneklemelerin tümünde ve çalışma evreninde 2PL modele ait G^2 değeri ile 1PL modele ait G^2 değeri arasında manidar bir fark olduğu belirlenmiştir. Bu durum 1000, 2000, 3000, 5000 kişilik örneklemelere ve çalışma evrenine ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığı şeklinde yorumlanmaktadır. Bu örneklemelere ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri karşılaştırıldığında 2000, 3000, 5000 kişilik örneklemelere ve çalışma evrenine ait verilerde bu iki değer arasındaki farkın manidar olduğu, dolayısıyla bu verilerin 3PL modele 2PL modelden daha iyi uyum sağladığı sonucuna varılmıştır. Sadece 1000 kişilik örneklem verisinden 2PL ve 3PL modellerde elde edilen G^2 değerleri arasındaki farkın manidar olmadığı ve 2PL ve 3PL modellerin ikisinin de veriye uyum sağladığı belirlenmiştir.

Büyük örneklemelerde 30 maddelik teste ait G^2 değerleri Tablo 3.21’de gösterilmiştir.

Tablo 3.21

Büyük Örneklerde 30 Maddelik Teste Ait G^2 Olabilirlik Oranı Değerleri

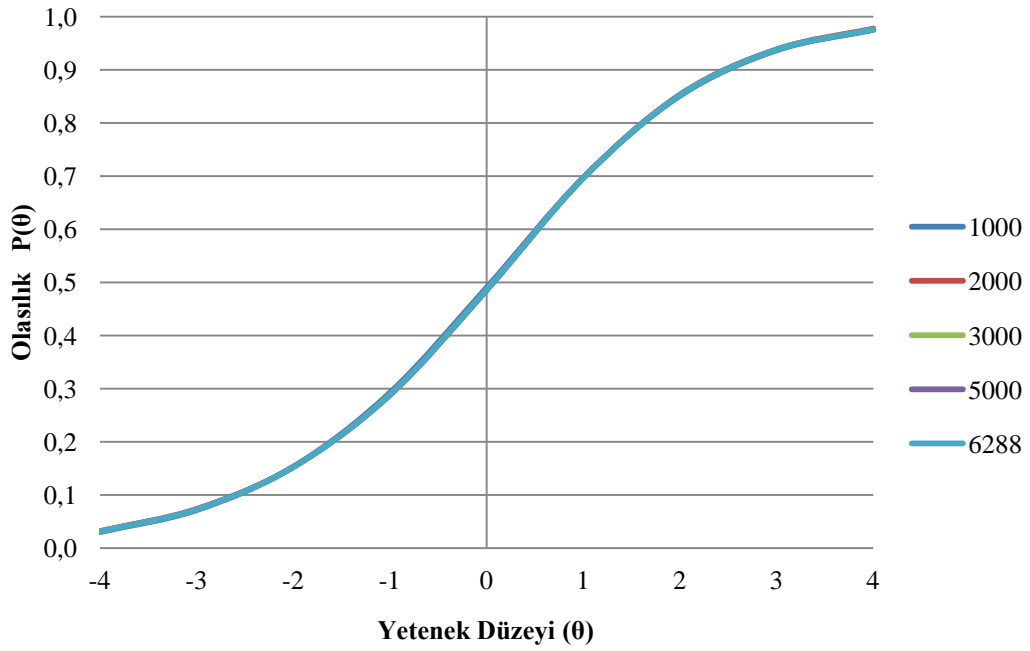
Model	1000		2000		3000		5000		6288**	
	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2	-2 log	G^2
1PL	29652,96		59029,22		88537,11		147311,15		185346,80	
2PL	29217,74	435,22*	58200,06	829,16*	87238,18	1298,93*	145193,31	2117,85*	182857,03	2489,78*
3PL	29217,37	0,37	58117,81	82,25*	87111,90	126,27*	144975,91	217,39*	182556,95	300,07*

*p ($X^2_{df=10} > 43,77$) < 0.05, **Çalışma evreni

Tablo 3.21’de 1000, 2000, 3000, 5000 kişilik örneklemelere ve çalışma evrenine ait G^2 değerleri incelendiğinde bu örneklemelerin tümünde ve çalışma evreninde 2PL modele ait G^2 değeri ile 1PL modele ait G^2 değeri arasında manidar bir fark olduğu belirlenmiştir. Bu durum 1000, 2000, 3000, 5000 kişilik örneklemelere ve çalışma evrenine ait verilerin 2PL modele 1PL modelden daha iyi uyum sağladığı şeklinde yorumlanmaktadır. Bu örneklemelere ait verilerden 3PL modelde elde edilen G^2 değerleri ile 2PL modelde elde edilen G^2 değerleri karşılaştırıldığında 2000, 3000, 5000 kişilik örneklemelere ve çalışma evrenine ait verilerde bu iki değer arasındaki farkın manidar olduğu, dolayısıyla verilerin 3PL modele 2PL modelden daha iyi uyum sağladığı sonucuna varılmıştır. Sadece 1000 kişilik örneklem verisinden 2PL ve 3PL modellerde elde edilen G^2 değerleri arasındaki farkın manidar olmadığı, dolayısıyla modelin her iki veriye de uyum sağladığı belirlenmiştir.

3.1.2.4. Test Karakteristik Eğrilerinin İncelenmesi İle Model Veri Uyumunun Sınanması

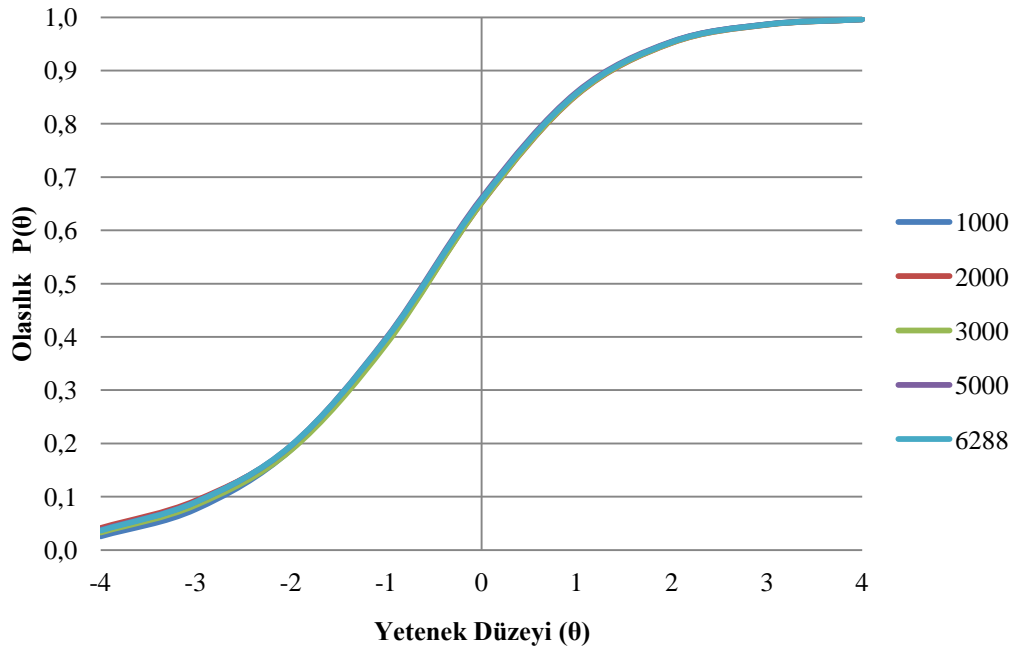
Büyük örneklerde 10 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri Şekil 19’da gösterilmiştir.



Şekil 19. Büyük örneklerde 10 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri.

Şekil 19 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 1PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle de aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklemelere uygulanan 10 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak kabul edilmiştir.

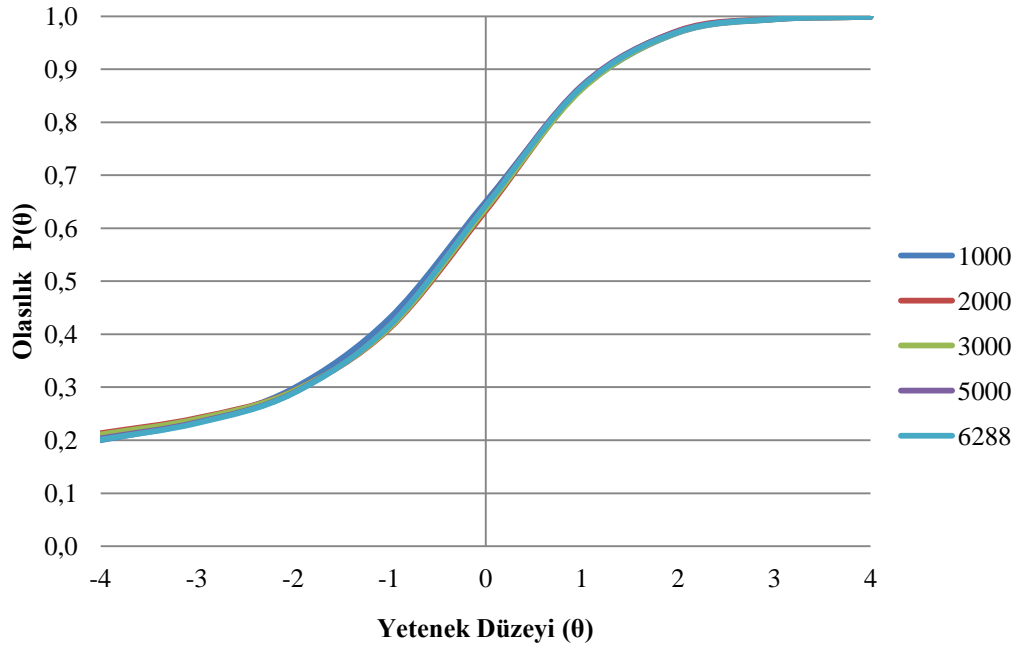
Büyük örneklerden 2PL modelde elde edilen test karakteristik eğrileri Şekil 20’de gösterilmiştir.



Şekil 20. Büyük örneklerde 10 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 20 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerden 2PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum, veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerde 10 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

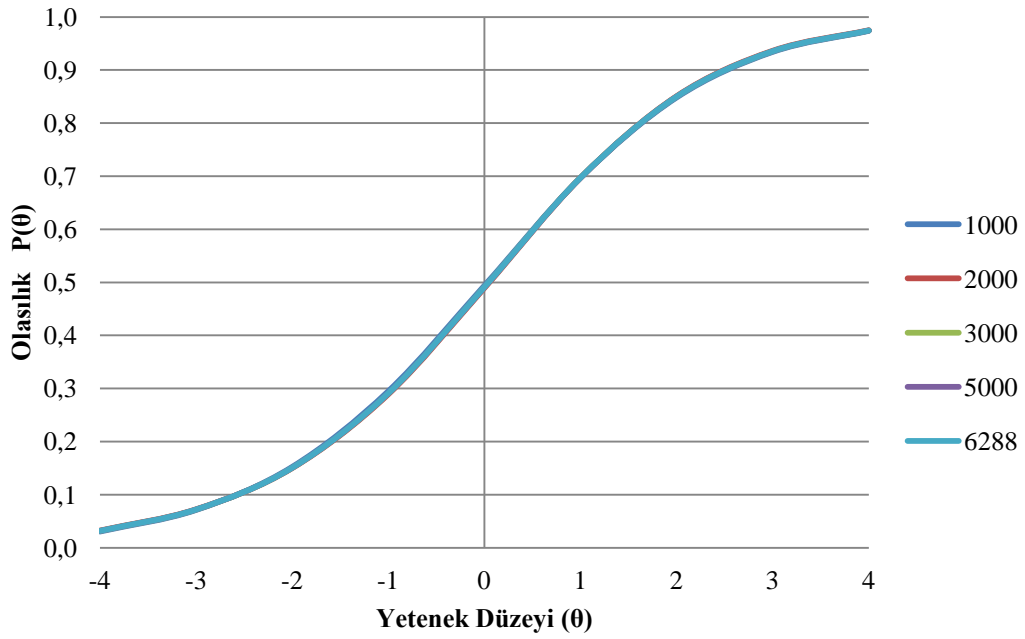
Büyük örneklerde 10 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri Şekil 21’de gösterilmiştir.



Şekil 21. Büyük örneklerde 10 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 21 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 3PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerle uygulanan 10 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

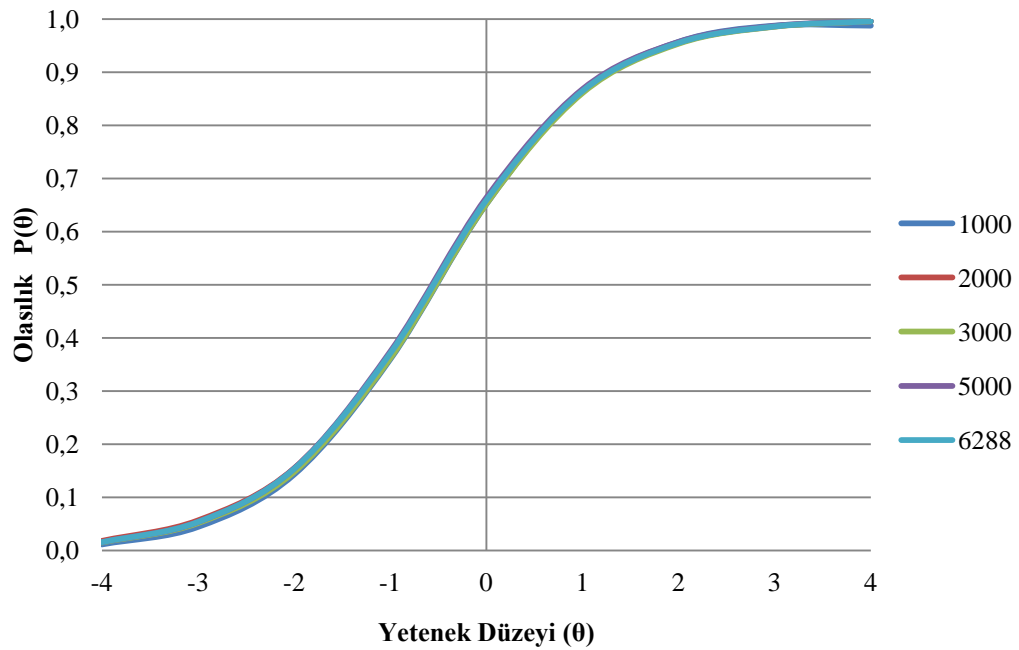
Büyük örneklerde 20 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri Şekil 22'de gösterilmiştir.



Şekil 22. Büyük örneklerde 20 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri.

Şekil 22 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 1PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerle uygulanan 20 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak kabul edilmiştir.

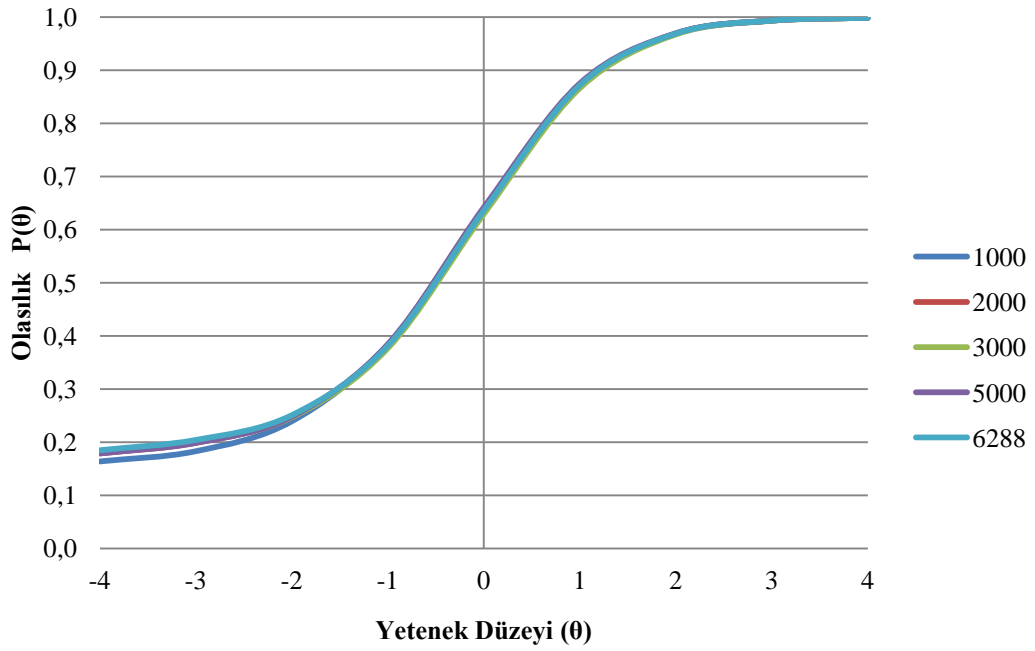
Büyük örneklerde 20 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri Şekil 23'te gösterilmiştir.



Şekil 23. Büyük örneklerde 20 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 23 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerden 2PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden kestirilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerle uygulanan 20 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

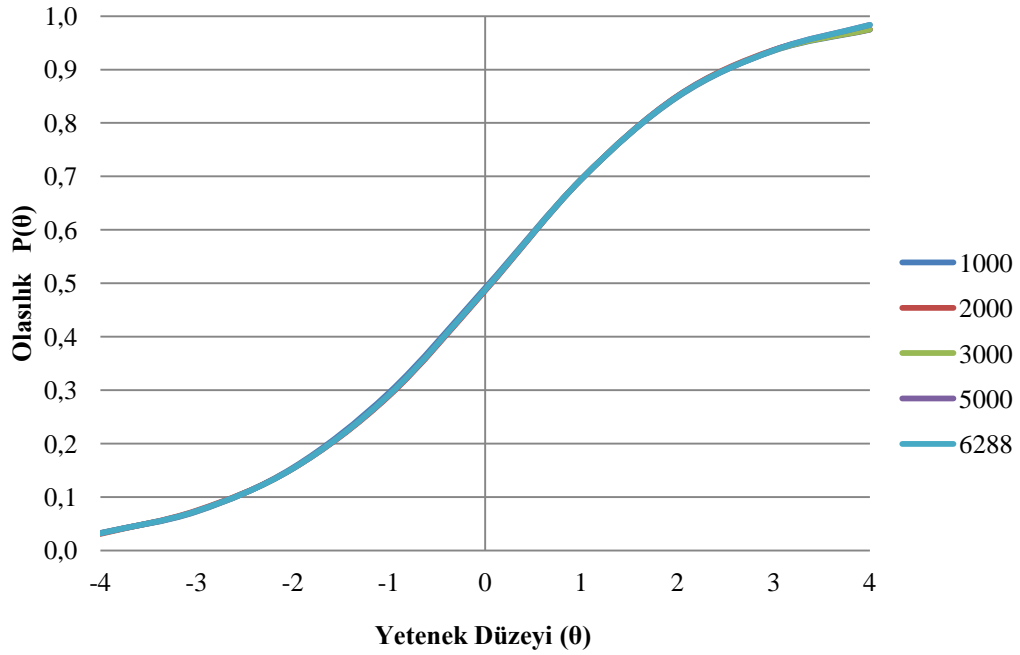
Büyük örneklerde 20 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri Şekil 24'de gösterilmiştir.



Şekil 24. Büyük örneklerde 20 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 24 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 3PL modelde elde edilen test karakteristik eğrilerinin çalışma evrenine ait test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklemeye uygulanan 20 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

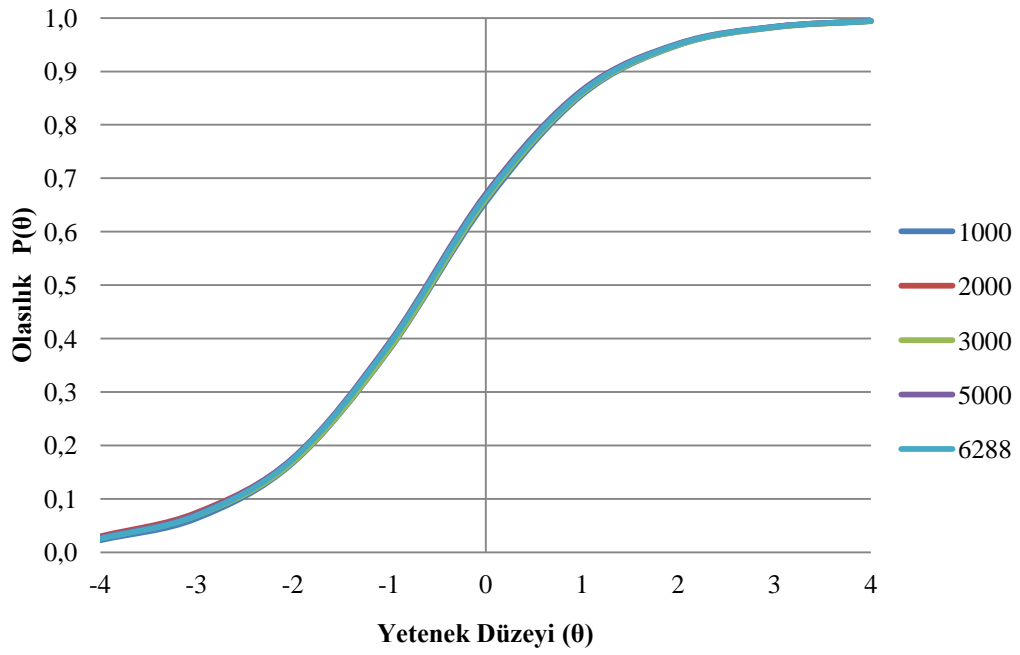
Büyük örneklerde 30 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri Şekil 25'te gösterilmiştir.



Şekil 25. Büyük örneklerde 30 maddelik testten 1PL modelde elde edilen test karakteristik eğrileri.

Şekil 25 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 1PL modelde elde edilen test karakteristik eğrilerinin çalışma evrenine ait test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerle uygulanan 30 maddelik test verisinin 1PL model ile uyum sağladığının kanıtı olarak kabul edilmiştir.

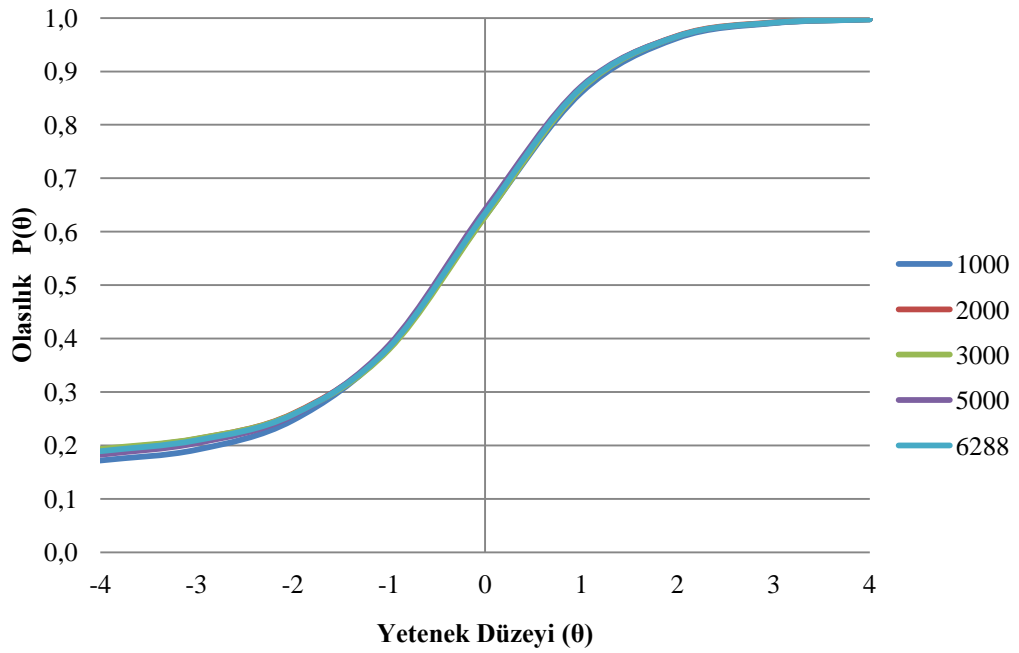
Büyük örneklerde 30 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri Şekil 26'da gösterilmiştir.



Şekil 26. Büyük örneklerde 30 maddelik testten 2PL modelde elde edilen test karakteristik eğrileri.

Şekil 26 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerden 2PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklemelere uygulanan 20 maddelik test verisinin 2PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

Büyük örneklerden 3PL modelde elde edilen test karakteristik eğrileri Şekil 27'de gösterilmiştir.



Şekil 27. Büyük örneklerde 30 maddelik testten 3PL modelde elde edilen test karakteristik eğrileri.

Şekil 27 incelendiğinde, sırasıyla 1000, 2000, 3000, 5000 kişilik örneklerde 3PL modelde elde edilen test karakteristik eğrilerinin çalışma evreninden elde edilen test karakteristik eğrisiyle aynı olduğu görülmektedir. Bu durum veride değişmezliğin sağlandığının ve dolayısıyla büyük örneklerle uygulanan 30 maddelik test verisinin 3PL model ile de uyum sağladığının kanıtı olarak kabul edilmiştir.

Sonuç olarak büyük örneklerle uygulanan 10 maddelik test verisinde model veri uyumunun sınanmasında kullanılan yöntemler ve veri ile uyumlu modeller Tablo 3.22'de gösterilmiştir.

Tablo 3.22

Büyük Örneklerde 10 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G ²	TKE
1000	1,2,3	3	2,3	1,2,3
2000	1,2,3	3	3	1,2,3
3000	2	3	3	1,2,3
5000	2	3	3	1,2,3
6288*	2	3	3	1,2,3

*Çalışma evreni

Tablo 3.22 incelendiğinde, model veri uyumunun sınanmasında kullanılan dört farklı yöntemden 3'nün büyük örneklerde model veri uyumunun tespitinde örneklem büyüklüğünden etkilenmediği belirlenmiştir. χ^2/sd yönteminin ise model veri uyumunun incelenmesinde örneklem büyüklüğünden etkilendiği belirlenmiştir. Bu durumda model veri uyumunun büyük örneklerde ve 10 maddelik testlerde örneklem büyüklüğüne göre değişiklik göstermediği sonucuna varılmıştır.

Büyük örneklerle uygulanan 20 maddelik test verisinde model veri uyumunun sınanmasında kullanılan farklı yöntemler ve uyumlu modeller Tablo 3.23'te gösterilmiştir.

Tablo 3.23

Büyük Örneklerde 20 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G ²	TKE
1000	1,2,3	3	2,3	1,2,3
2000	1,2,3	3	3	1,2,3
3000	1,2,3	3	3	1,2,3
5000	1,2,3	3	3	1,2,3
6288*	1,2,3	3	3	1,2,3

*Çalışma evreni

Tablo 3.23 incelendiğinde, model veri uyumunun sınanmasında kullanılan dört farklı yöntemden her birinin büyük örneklerde örneklem büyüklüğünün değişmesinden etkilenmediği görülmektedir. Bu durumda model veri uyumunun büyük örneklerle ait 20 maddelik test verisinde örneklem büyüklüğüne göre değişiklik göstermediği sonucuna varılmıştır.

Büyük örneklerde 30 maddelik test verisinde model veri uyumunun sınanmasında kullanılan farklı yöntemler ve uyumlu modeller Tablo 3.24'te gösterilmiştir.

Tablo 3.24

Büyük Örneklemelerde 30 Maddelik Teste Ait Model Veri Uyumu

Örneklem	Yöntemlere Göre Uyumlu Modeller			
	χ^2/sd	SAA	G ²	TKE
1000	1,2,3	3	2,3	1,2,3
2000	1,2,3	3	3	1,2,3
3000	1,2,3	3	3	1,2,3
5000	1,2,3	3	3	1,2,3
6288*	2,3	3	3	1,2,3

*Çalışma Evreni

Tablo 3.24 incelendiğinde, model veri uyumunun sınanmasında kullanılan dört farklı yöntemden her birinin örneklem büyüklüğünden etkilenmeden model veri uyumunu sınıdığı görülmektedir. Bu bilgiler ışığında model veri uyumunun büyük örneklemelere uygulanmış 30 maddelik testlerde örneklem büyüklüğüne göre değişiklik göstermediği sonucuna varılmıştır.

Birinci alt probleme ait bulgular toplu olarak değerlendirildiğinde, küçük örneklemelerde model veri uyumunun tespitinde kullanılan Standartlaştırılmış Artık Analizi (SAA) yönteminin örneklem büyüklüğünden ve kısmen de olsa test uzunluğundan etkilendiği; dolayısıyla bu yöntemin kullanılabilmesi için 10 maddelik bir test için en az 500, 20 ve 30 maddelik bir test için en az 750 kişilik örneklemelere ihtiyaç duyulduğu; χ^2/sd yönteminin test uzunluğu ve örneklem büyüklüğünden kısmen etkilendiği; 20 maddelik testlerde örneklem büyüklüğünden etkilenmeyen bu yöntemin, madde sayısı 10 olduğu durumlarda, 3000'den büyük örneklemelerde; madde sayısının 30 olduğu durumlarda ise 5000'den büyük örneklemelerden sonra örneklem büyüklüğünden etkilendiği; G² Olabilirlik Oranı ve test karakteristik eğrilerinin incelenmesi yöntemlerinin test uzunluğu ve örneklem büyüklüğünden etkilenmediği ve dolayısıyla bundan sonraki

arařtırmalarda model veri uyumunun sınanmasında bu alıřmaya dahil edilen farklı test uzunlukları ve rneklem byklklerinde kullanılmalarının uygun olabileceđi sylenbilir.

3.2. İKİNCİ ALT PROBLEME İLİŐKİN BULGULAR VE YORUMLAR

İkinci Alt Problem: Test uzunluđu ve rneklem byklđ farklılařtırıldıđında, MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilmiş madde parametreleri alıřma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilen madde parametrelerine gre nasıl deđiřmektedir?

- 2.a. Kk rneklemlerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilen madde parametreleri, alıřma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilen madde parametrelerine gre nasıl deđiřmektedir?
- 2.b. Byk rneklemlerde (1000, 2000, 3000, 5000) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilen madde parametreleri, alıřma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine gre kestirilen madde parametrelerine gre nasıl deđiřmektedir?

3.2.1. Kk rneklemlerde (150, 250, 350, 500, 750) 10, 20 ve 30 Maddelik Testlerden MTK'nın 1PL, 2PL ve 3PL Modellerine Gre Kestirilen Madde Parametreleri, alıřma Evreninden MTK'nın 1PL, 2PL ve 3PL Modellerine Gre Kestirilen Madde Parametrelerine Gre Nasıl Deđiřmektedir?

Bu alt problemi yanıtlamak amacıyla kk rneklemlere uygulanan 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile alıřma evreninden kestirilen madde parametreleri arasındaki korelasyon katsayıları incelenmiřtir. 10 maddelik test iin elde edilen veriler Tablo 3.25'te gsterilmiřtir.

Tablo 3.25

Küçük Örneklem ve Çalışma Evreninden 10 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklem	1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c
150	,939**	,309	,940**	,606	,944**	,733*
250	,997**	-,311	,997**	,207	,993**	,757*
350	,992**	,083	,992**	,481	,993**	,415
500	,976**	,347	,978**	,345	,974**	,940**
750	,999**	,706*	,999**	,901**	,998**	,922**

*p<0.05 , **p<0.01

Tablo 3.25 incelendiğinde, 1PL modelde kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,939 değeri ile 150 kişilik ve en yüksek 0,999 değeri ile 750 kişilik örneklemelerden elde edildiği görülmektedir. 10 maddelik bir testte “b” parametresinin kestiriminde 150 kişilik örneklemle bile çalışma evreninden kestirilen madde parametreleriyle yüksek korelasyona sahip madde parametreleri elde edildiği belirlenmiştir.

10 maddelik test için küçük örneklemelerde 2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyon katsayısının 0,083 değeri ile en düşük 350 ve 0,706 değeri ile en yüksek 750 kişilik örneklemelerden elde edildiği görülmektedir. Bu durumda 2PL modelde kestirilen “a” parametresinin 750 kişilik örnekleme kadar yeterli düzeyde korelasyon katsayısı olan 0,70 değerine ulaşamadığı gözlenmiştir. 10 maddelik test için 2PL modelde kestirilen bir diğer parametre olan “b” parametreleri incelendiğinde çalışma evreninden kestirilen “b” parametreleriyle en düşük korelasyonun 0,940 değeri ile 150 kişilik örneklemden, en yüksek korelasyonun ise 0,999 değeri ile 750 kişilik örneklemden elde edildiği belirlenmiştir. 2PL modelde

tüm küçük örneklemlerde “a” parametresi kestiriminin aksine çalışma evreninden kestirilen “b” parametresi değerleri ile yüksek korelasyon gösteren verilerin elde edildiği belirlenmiştir. Bu durumda “b” parametresi kestiriminin örneklem büyüklüğünden bağımsız olarak başarı ile kestirilebildiği söylenebilir.

3PL modelde kestirilen madde parametrelerine ait korelasyonlar incelendiğinde küçük örneklemlerden elde edilen 10 maddelik teste ait “a” parametresi değerlerinin çalışma evreninden kestirilen “a” parametresi değerleri ile en düşük korelasyonu 0,207 değeri ile 250 kişilik örneklemden, en yüksek korelasyonu için 0,901 değeri ile 750 kişilik örneklemden kestirilen “a” parametresi ile aldığı görülmektedir. Bu durumda “a” parametresinin 3PL modelde 750 kişilik örneklemlerle de başarılı bir şekilde kestiriminin yapılabileceği ortaya çıkmaktadır. 3PL modelde kestirilen “b” parametresine ait korelasyon katsayıları incelendiğinde en düşük korelasyon katsayısının 0,944 ile 150 kişilik örneklemden, en yüksek korelasyonun ise 0,998 ile 750 kişilik örneklemden elde edildiği belirlenmiştir. Diğer korelasyon katsayılarının ise 0,97 üzerinde değerler almış olması oldukça dikkat çekicidir. 10 maddelik test için küçük örneklemlerden 3PL modelde kestirilen “c” parametrelerinin çalışma evreninden kestirilen “c” parametreleriyle korelasyonundan elde edilen değerler incelendiğinde “c” parametresi için en düşük korelasyon katsayısının 0,415 ile 350 kişilik örneklemden elde edildiği görülmektedir. Ancak 150 ve 250 kişilik örneklemlerden elde edilen korelasyon katsayılarının sırasıyla 0,733 ve 0,757 olması oldukça dikkat çekicidir. En yüksek korelasyon katsayıları ise sırasıyla 0,940 ve 0,922 ile 500 ve 750 kişilik örneklemlerden elde edilmiştir. Bu durum 350 kişilik örneklemden elde edilen 0,350 değerinin istisnai bir durum olduğu ve örneklemden kaynaklanan bir hatadan meydana gelmiş olabileceğini düşündürmektedir. Bu durum göz ardı edildiğinde “c” parametresinin 10 maddelik bir testte, 150 kişilik örneklem büyüklüğünde bile “a” parametresine göre nispeten yüksek korelasyon katsayıları ile kestirilebildiğini göstermektedir.

Küçük örneklemlerde 20 maddelik testten MTK'nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile çalışma evreninden kestirilen madde parametreleri arasındaki korelasyon katsayıları Tablo 3.26'da verilmiştir.

Tablo 3.26

Küçük Örneklem ve Çalışma Evreninden 20 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklem	1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c
150	,962**	,329	,964**	,516*	,955**	,485*
250	,993**	,642**	,993**	,583**	,990**	,671**
350	,993**	,642**	,994**	,581**	,990**	,697**
500	,984**	,795**	,985**	,614**	,972**	,820**
750	,998**	,793**	,998**	,840**	,995**	,829**

*p<0.05, **p<0.01

Tablo 3.26 incelendiğinde, 1PL modelde küçük örneklemde kestirilen “b” parametrelerinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,962 değeri ile 150 kişilik ve en yüksek 0,998 değeri ile 750 kişilik örneklemde elde edildiği görülmektedir. Bu durum “b” parametresinin kestirilmesi için 150 kişilik örneklemin yeterli olabileceği şeklinde yorumlanmıştır.

2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyonunun 0,329 değeri ile en düşük 150 ve 0,793 değeri ile en yüksek 500 kişilik örneklemde elde edildiği görülmektedir. Bu durumda 2PL modelde kestirilen “a” parametresinin 500 kişilik örneklemden sonra yeterli düzeyde korelasyon katsayısı olan 0,70 değerine ulaştığı gözlenmiştir. 2PL modelde kestirilen bir diğer parametre olan “b” parametrelerine ait korelasyonlar incelendiğinde çalışma evreninden kestirilen “b” parametreleriyle en düşük korelasyonun 0,964 değeri ile 150 kişilik örneklemden, en yüksek korelasyon katsayısının ise 0,998 değeri ile 750 kişilik örneklemden elde edildiği belirlenmiştir.

3PL modelde kestirilen madde parametrelerine ait korelasyonlar incelendiğinde 20 maddelik test için küçük örneklemelerden ve çalışma evreninden kestirilen “a” parametresine ait korelasyon katsayıları incelendiğinde en düşük korelasyonun 0,516 değeri ile 150 kişilik örneklemden, en yüksek korelasyonun ise 0,840 değeri ile 750 kişilik örneklemden elde edildiği görülmektedir. 3PL modelde kestirilen “b” parametresine ait korelasyon katsayıları incelendiğinde en düşük korelasyon katsayısının 0,955 ile 150 kişilik örneklemden, en yüksek korelasyonun ise 0,995 ile 750 kişilik örneklemden elde edildiği belirlenmiştir. 3PL modelde kestirilen bir diğer parametre olan “c” parametresinin korelasyon katsayıları incelendiğinde “c” parametresi için en düşük korelasyon katsayısının 0,485 ile 150 kişilik örneklemden, en yüksek 0,829 ile 750 kişilik örneklemden elde edildiği görülmektedir.

Küçük örneklemelerde 30 maddelik testten MTK'nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile çalışma evreninden aynı modellerde kestirilen madde parametreleri arasındaki korelasyon katsayıları Tablo 3.27’te sunulmuştur.

Tablo 3.27

Küçük Örneklemeler ve Çalışma Evreninden 30 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklem	1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c
150	,958**	,658**	,961**	,719**	,948**	,535**
250	,979**	,830**	,978**	,770**	,981**	,548**
350	,987**	,779**	,987**	,749**	,987**	,734**
500	,982**	,895**	,983**	,831**	,975**	,735**
750	,994**	,870**	,995**	,871**	,995**	,756**

**p<0.01

Tablo 3.27 incelendiğinde, küçük örneklemelerden 1PL modelde kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,958 değeri ile 150 kişilik örneklemden ve en yüksek 0,994 değeri ile 750 kişilik örneklemelerden elde edildiği görülmektedir. Bu durum “b” parametresinin kestirilmesi için 150 kişilik örneklemin dahi yeterli olabileceği şeklinde yorumlanmıştır.

Küçük örneklemelerden 2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyon katsayısının 0,658 değeri ile en düşük 150 kişilik örneklemden ve 0,870 değeri ile en yüksek 750 kişilik örneklemden elde edildiği görülmektedir. Bu durumda 2PL modelde kestirilen “a” parametresinin 250 kişilik örneklemden sonra yeterli düzeyde korelasyon katsayısı olan 0,70 ve üzeri değere ulaştığı gözlenmiştir. Bu durum 250 kişilik örneklemin “a” parametresinin kestirimi için yeterli olabileceği şeklinde yorumlanmıştır. 2PL modelde kestirilen bir diğer parametre olan “b” parametreleri incelendiğinde en düşük korelasyonun 0,961 değeri ile 150 kişilik örneklemden, en yüksek korelasyon katsayısının ise 0,995 ile 750 kişilik örneklemden elde edildiği belirlenmiştir. Bu bulgular “b” parametresinin kestiriminin diğer bulgularla tutarlı bir şekilde 150 kişilik örneklem ile yapılabileceğini ortaya koymaktadır. Elde edilen bulgular ışığında 2PL modelde “a” parametresinin başarılı kestirimi için 250 kişilik bir örneklemin yeterli olabileceği belirlenmiştir.

Küçük örneklemelerden 3PL modelde kestirilen “a” parametresine ait korelasyon katsayıları incelendiğinde, en düşük korelasyonun 0,719 değeri ile 150 kişilik örneklemden, en yüksek korelasyonun ise 0,871 değeri ile 750 kişilik örneklemden elde edildiği görülmektedir. 3PL modelde “b” parametresine ait korelasyon katsayıları incelendiğinde, en düşük korelasyon katsayısının 0,948 ile 150 kişilik örneklemden, en yüksek korelasyon katsayısının ise 0,995 ile 750 kişilik örneklemden elde edildiği belirlenmiştir. Aynı şekilde, “c” parametresinin korelasyon katsayıları incelendiğinde ise en düşük korelasyon katsayısının 0,535 ile 150 kişilik örneklemden, en yüksek 0,756 ile 750 kişilik örneklemden elde edildiği görülmektedir. Elde edilen bu değerler 3PL modelde “a” ve “b” parametrelerinin 150 kişilik örneklem büyüklüğünde bu korelasyon değeri göz önünde bulundurulmak kaydıyla kestiriminin yapılabileceği şeklinde

yorumlanmıştır. 350 kişilik örneklemden elde edilen “c” parametresi ile çalışma evreninden elde edilen “c” parametreleri arasında yüksek korelasyon elde edildiğinden dolayı bu korelasyon değeri göz önüne alınarak “c” parametresinin kestiriminde 350 kişilik bir örneklemin bazı durumlar için yeterli olabileceği ve genel olarak tüm korelasyon değerleri gözönüne alındığında 3PL modelde “a”, “b” ve “c” parametrelerinin bazı çalışmalarda 350 kişilik bir örnekleme de kestiriminin yapılabileceği söylenebilir.

3.2.2. Büyük Örneklerde (1000, 2000, 3000, 5000) 10, 20 ve 30 Maddelik Testlerden MTK’nın 1PL, 2PL ve 3PL Modellerine Göre Kestirilen Madde Parametreleri, Çalışma Evreninden MTK’nın 1PL, 2PL ve 3PL Modellerine Göre Kestirilen Madde Parametrelerine Göre Nasıl Değişmektedir?

Bu alt problemi yanıtlamak amacıyla büyük örneklerde 10, 20 ve 30 maddelik testlerden MTK’nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile çalışma evreninden kestirilen madde parametreleri arasındaki korelasyon katsayıları incelenmiştir. 10 maddelik testten elde edilen korelasyon katsayıları Tablo 3.28’de sunulmuştur.

Tablo 3.28

Büyük Örnekler ve Çalışma Evreninden 10 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklem	1PL Model		2PL Model		3PL Model		
	b	a	b	a	b	c	
1000	,994**	,864**	,994**	,929**	,994**	,938**	
2000	,999**	,826**	,999**	,918**	1**	,988**	
3000	,999**	,886**	,999**	,971**	1**	,926**	
5000	1**	,967**	1**	,991**	1**	,968**	

**p<0.01

Tablo 3.28 incelendiğinde, 1PL modelde kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,994 değeri ile 1000 kişilik ve en yüksek 1 tam değeri ile 5000 kişilik örneklemelerden elde edildiği görülmektedir. 1000 kişilik örneklemelerden kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle çok yüksek korelasyona sahip olduğu belirlenmiştir.

10 maddelik test maddeleri için büyük örneklemelerde 2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyonunun 0,826 değeri ile en düşük 2000 ve 0,967 değeri ile en yüksek 5000 kişilik örneklemelerden elde edildiği görülmektedir. Bu durumda tüm örneklem büyüklüklerinde oldukça yüksek korelasyon katsayılarına ulaşıldığı görülmektedir. 2PL modelde kestirilen diğer parametre olan “b” parametrelerine ait korelasyon katsayıları incelendiğinde, en düşük korelasyonun 0,994 değeri ile 1000 kişilik örneklemelerden, en yüksek korelasyon katsayısının 1,00 ile 5000 kişilik örneklemelerden elde edildiği belirlenmiştir. 2PL modelde büyük örneklemelerin tümünde “a” parametresi kestiriminin aksine “b” parametresi kestirimlerine ait korelasyon katsayılarının yüksek değerler aldığı belirlenmiştir.

3PL modelde kestirilen madde parametrelerine ait korelasyon katsayıları incelendiğinde “a” parametresi değerlerinin çalışma evreninden kestirilen “a” parametresi değerleri ile en düşük korelasyonunun 0,918 değeri ile 2000 kişilik örneklemelerden, en yüksek korelasyon katsayısının ise 0,991 değeri ile 5000 kişilik örneklemelerden kestirilen “a” parametresinden elde edildiği görülmektedir. Bu durumda “a” parametresinin 3PL modelde tüm büyük örneklemelerde başarılı bir şekilde kestiriminin yapılabileceği ortaya çıkmaktadır. En düşük değerlerin diğer bulguların aksine 1000 değil 2000 kişilik örneklemelerden elde edilmesi 2000 kişilik veride küçük de olsa bir örnekleme hatası olabileceğinin işareti olarak alınmıştır.

3PL modelde kestirilen “b” parametresine ait korelasyon katsayıları incelendiğinde en düşük korelasyon katsayısının 0,994 ile 1000 kişilik örneklemelerden, en yüksek korelasyon katsayısının ise 1 değeri ile 2000, 3000 ve 5000 kişilik örneklemelerden elde

edildiği görülmektedir. Bu durumda “b” parametresinin de tüm büyük örneklerde başarı ile kestirildiği görülmektedir.

10 maddelik test için büyük örneklerden 3PL modelde kestirilen “c” parametrelerinin çalışma evreninden kestirilen “c” parametreleriyle korelasyonundan elde edilen değerler incelendiğinde “c” parametresi için en düşük korelasyon katsayısının 0,926 ile 3000 kişilik örneklemden elde edildiği, en yüksek korelasyon katsayısının ise 0,988 ile 2000 kişilik örneklemden elde edildiği görülmektedir. Tablodaki diğer değerler incelendiğinde, “c” parametresine ait korelasyon katsayılarının örneklem büyüklüğüne göre doğrusal olarak artmadığı görülmüştür ve korelasyon katsayılarının diğer bulguların aksine örneklem büyüklüğünden bağımsız değerler aldığı görülmektedir. Bu durum “c” parametresi kestiriminin madde sayısının düşük olduğu testlerde küçük de olsa örnekleme hatalarına karşı diğer parametrelere göre daha hassas olduğu şeklinde yorumlanmıştır.

Büyük örneklerde 20 maddelik test maddeleri için MTK’nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile çalışma evreninden kestirilen madde parametreleri arasındaki korelasyon katsayıları Tablo 3.29’da sunulmuştur.

Tablo 3.29

Büyük Örnekler ve Çalışma Evreninden 20 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklem	1PL Model		2PL Model		3PL Model		
	b	a	b	a	b	c	
1000	,994**	,935**	,994**	,874**	,989**	,915**	
2000	,998**	,926**	,999**	,929**	,997**	,906**	
3000	,999**	,977**	,999**	,957**	,998**	,962**	
5000	1**	,997**	1**	,995**	1**	,996**	

**p<0.01

Tablo 3.29 incelendiğinde, 1PL modelde kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,994 değeri ile 1000 kişilik ve en yüksek 1,00 değeri ile 5000 kişilik örneklemelerden elde edildiği görülmektedir.

2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyonunun 0,926 değeri ile en düşük 2000 ve 0,997 değeri ile en yüksek 5000 kişilik örneklemelerden elde edildiği görülmektedir. Bu durumda tüm örneklem büyüklüklerinde oldukça yüksek korelasyon katsayılarına ulaşıldığı görülmektedir. 2PL modelde kestirilen diğer parametre olan “b” parametrelerine ait korelasyon katsayıları incelendiğinde en düşük korelasyon katsayısının 0,994 ile 1000 kişilik örneklemelerden, en yüksek korelasyon katsayısının 1 ile 5000 kişilik örneklemelerden elde edildiği belirlenmiştir. 2PL modelde büyük örneklemelerin tümünde “a” ve “b” parametreleri kestirimlerinin çalışma evreninden kestirilen parametrelerle korelasyonunun oldukça yüksek değerler aldığı görülmektedir.

3PL modelde kestirilen “a” parametresi değerlerinin çalışma evreninden kestirilen “a” parametresi değerleri ile en düşük korelasyonunun 0,874 değeri ile 1000 kişilik örneklemelerden, en yüksek korelasyon katsayısının ise 0,995 değeri ile 5000 kişilik örneklemelerden kestirilen “a” parametresinden elde edildiği görülmektedir. Bu durum “a” parametresinin 3PL modelde tüm büyük örneklemelerde yüksek korelasyonla kestiriminin yapılabileceğini ortaya koymaktadır..

3PL modelde kestirilen “b” parametresine ait korelasyon katsayıları incelendiğinde, en düşük korelasyon katsayısının 0,989 ile 1000 kişilik örneklemelerden, en yüksek korelasyon katsayısının ise 1 değeri ile 5000 kişilik örneklemelerden elde edildiği görülmektedir. Diğer korelasyon katsayılarının da 0,989 ve üzeri olduğu dikkate alındığında “b” parametresinin de tüm büyük örneklemelerde yüksek korelasyon değerleriyle kestirilebildiği görülmektedir.

Tablo 3.34’te 3PL modelde kestirilen “c” parametrelerinin çalışma evreninden kestirilen “c” parametreleriyle korelasyonundan elde edilen korelasyon katsayıları incelendiğinde, “c” parametresi için en düşük korelasyon katsayısının 0,906 ile 2000 kişilik örneklemden elde edildiği, en yüksek korelasyon katsayısının ise 0,996 ile 5000 kişilik örneklemden elde edildiği görülmektedir. Tablodaki diğer değerler incelendiğinde “c” parametresinin büyük örneklemlerde yapılan kestiriminin tümünde yüksek korelasyon katsayılarına ulaşıldığı görülmektedir. 10 maddelik teste ait 3PL modelde kestirilen “c” parametrelerinin korelasyonunun örneklem büyüklüğü arttıkça doğrusal bir şekilde artmadığı gözlenmişti. Bu dağınık görünümün yerini nispeten daha düzenli şekilde artan korelasyon katsayılarına bıraktığı tespit edilmiştir.

Büyük örneklemlerde 30 maddelik testten MTK’nın 1PL, 2PL ve 3PL modellerinde kestirilen madde parametreleri ile çalışma evreninden kestirilen madde parametreleri arasındaki korelasyon katsayıları Tablo 3.30’da sunulmuştur.

Tablo 3.30

Büyük Örneklemler ve Çalışma Evreninden 30 Maddelik Testte Kestirilen Madde Parametrelerinin Korelasyon Katsayıları Matrisi

Örneklemler	1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c
1000	,993**	,972**	,994**	,934**	,989**	,857**
2000	,998**	,959**	,997**	,955**	,998**	,851**
3000	,999**	,988**	,999**	,982**	,999**	,952**
5000	1**	,995**	1**	,995**	,999**	,991**

**p<0.01

Tablo 3.30 incelendiğinde, 1PL modelde kestirilen “b” parametresinin çalışma evreninden kestirilen “b” parametresiyle korelasyon katsayısının en düşük 0,993 değeri

ile 1000 kişilik örneklemden ve en yüksek 1 korelasyon katsayısı ile 5000 kişilik örneklemlerden elde edildiği görülmektedir.

Büyük örneklemlerden 2PL modelde kestirilen “a” parametresinin çalışma evreninden kestirilen “a” parametresi ile korelasyonunun 0,959 değeri ile en düşük 2000 ve 0,995 değeri ile en yüksek 5000 kişilik örneklemlerden elde edildiği görülmektedir. Bu durumda tüm örneklem büyüklüklerinde oldukça yüksek korelasyon katsayılarına ulaşıldığı söylenebilir. 2PL modelde kestirilen diğer parametre olan “b” parametrelerine ait korelasyon katsayıları incelendiğinde, en düşük korelasyonun 0,994 değeri ile 1000 kişilik örneklemden, en yüksek korelasyon katsayısının 1,00 ile 5000 kişilik örneklemden elde edildiği belirlenmiştir. 2PL modelde büyük örneklemlerin tümünde “a” ve “b” parametresi kestirimlerinin çalışma evreninden kestirilen parametrelerle korelasyon katsayısının oldukça yüksek değerler aldığı görülmektedir.

Büyük örneklemlerden 3PL modelde kestirilen “a” parametresi değerlerinin çalışma evreninden kestirilen “a” parametresi değerleri ile en düşük korelasyon katsayısının 0,934 değeri ile 1000 kişilik örneklemden, en yüksek korelasyon katsayısının ise 0,995 değeri ile 5000 kişilik örneklemden “a” parametresi için elde edildiği görülmektedir. Bu durum “a” parametresinin 3PL modelde tüm büyük örneklemlerde başarılı bir şekilde kestiriminin yapılabileceğini ortaya çıkarmaktadır. 3PL modelde kestirilen “b” parametresine ait korelasyon katsayıları incelendiğinde, en düşük korelasyon katsayısının 0,989 ile 1000 kişilik örneklemden, en yüksek korelasyon katsayısının ise 0,999 değeri ile 5000 kişilik örneklemlerden elde edildiği görülmektedir. Diğer korelasyon katsayılarının da 0,998 ve üzeri olduğu dikkate alındığında “b” parametresinin de tüm büyük örneklemlerde başarı ile kestirilebildiği görülmektedir. 3PL modelde kestirilen “c” parametrelerinin korelasyonundan elde edilen değerler incelendiğinde “c” parametresi için en düşük korelasyon katsayısının 0,851 ile 2000 kişilik örneklemden elde edildiği, en yüksek korelasyon katsayısının ise 0,991 ile 5000 kişilik örneklemden elde edildiği görülmektedir. Tablodaki diğer değerler incelendiğinde “c” parametresinin kestiriminin büyük örneklemlerde en düşük 0,851 gibi yüksek korelasyon değerleriyle yapılabildiği görülmektedir.

İkinci alt probleme ait tüm bulgular ışığında; MTK'nın 1PL modeli kullanılarak kestirilen "b" parametresinin örneklem büyüklüğünden ve hatta kullanılan modelden bağımsız olarak tüm modellerde çok yüksek korelasyon değerleriyle kestirilebildiği; 2PL modelde uygun örneklem büyüklüğünün belirlenmesinde "a" parametresi kestirimlerinin belirleyici olduğu; 2PL modelde yapılacak madde parametresi kestirimlerinde, 10 maddelik bir test için en az 750 ($r_a=0,706$, $r_b=0,999$); 20 maddelik bir test için en az 500 ($r_a=0,795$, $r_b=0,998$) ve 30 maddelik bir test için en az 250 ($r_a=0,830$, $r_b=0,978$) kişilik örneklemelerin yeterli olabileceği belirlenmiştir. 3PL modelde parametre kestiriminde örneklem büyüklüğünün belirlenmesinde 10 ve 20 maddelik testlerde "a" parametresi, 30 maddelik testte ise "c" parametresinin belirleyici rol oynadığı söylenebilir. Ayrıca, elde edilen bulgular değerlendirildiğinde, MTK'da madde parametreleri kestiriminde 10 maddelik bir test için en az 750 ($r_a=0,901$, $r_b=0,998$, $r_c=0,922$), 20 maddelik bir test için en az 750 ($r_a=0,840$, $r_b=0,995$, $r_c=0,829$) ve 30 maddelik bir test için en az 350 ($r_a=0,749$, $r_b=0,987$, $r_c=0,734$) kişilik örneklemelerin bazı çalışmalar için yeterli kabul edilebileceği söylenebilir. Bunun yanı sıra 2PL ve 3PL modellerde "a" parametresinin kestiriminin madde sayısının artmasından oldukça olumlu etkilenirken 3PL modelde "c" parametresinin kestiriminin güçleştiği söylenebilir. Bu durum mevcut parametre kestirim programlarının "c" parametresinin kestiriminde yaşadığı güçlüklerin bir yansıması şeklinde yorumlanabilir.

3.3. ÜÇÜNCÜ ALT PROBLEME İLİŞKİN BULGULAR VE YORUMLAR

Üçüncü Alt Problem: Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıl değişmektedir?

- 3.a. Küçük örneklemelerde (150, 250, 350, 500, 750) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıldır?

3.b. Büyük örneklemlerde (1000, 2000, 3000, 5000) 10, 20 ve 30 maddelik testlerden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıldır?

3.3.1. Küçük Örneklerde (150, 250, 350, 500, 750) 10, 20 ve 30 Maddelik Testlerden MTK'nın 1PL, 2PL ve 3PL Modellerine Göre Kestirilmiş Madde Parametrelerine Ait Standart Hata Değerleri Nasıldır?

Araştırmanın bu alt problemine cevap verebilmek amacıyla 10 maddelik testten elde edilen ortalama hata değerleri küçük örneklem için ayrı ayrı elde edilmiş ve bu değerler Tablo 3.31'de gösterilmiştir.

Tablo 3.31

Küçük Örneklerde 10 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri

Örneklem	1PL Model	2PL Model		3PL Model		
	b	a	b	a	b	c
150	0,231	0,181	0,179	0,247	0,195	0,349
250	0,177	0,146	0,144	0,204	0,147	0,284
350	0,152	0,120	0,122	0,154	0,125	0,243
500	0,129	0,101	0,104	0,133	0,105	0,201
750	0,103	0,082	0,085	0,109	0,087	0,157

Tablo 3.31 incelendiğinde, 1PL modelde “b” parametresi için elde edilen ortalama standart hata değerleri en yüksek 0,231 ile 150 kişilik örnekte, en düşük 0,103 değeri ile 750 kişilik örnekte elde edilmiştir. Hata değerlerinin örneklem büyüklüğü arttıkça doğrusal bir şekilde düştüğü gözlenmektedir.

10 maddelik test maddeleri için 2PL modelde kestirilen “a” parametresine ait ortalama hata değerleri incelendiğinde, en yüksek ortalama hata değerinin 0,181 ile 150 kişilik örneklemden, en düşük hata değerinin ise 0,082 değeri ile 750 kişilik örneklemden elde edildiği; 2PL modelde kestirilen bir diğer parametre olan “b” parametresine ait standart hata değerleri incelendiğinde, en yüksek değere 0,179 ile 150 kişilik örnekleme, en düşük değere ise 0,085 değeri ile 750 kişilik örnekleme ulaşıldığı görülmektedir. Bu bilgiler ışığında 1PL modelde olduğu gibi 2PL modelde de ortalama standart hata değerlerinin örneklem büyüklüğü arttıkça düştüğü söylenebilir.

3PL modelde elde edilen ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek değerin 0,247 ile 150 kişilik örneklemden, en düşük değerin ise 0,109 ile 750 kişilik örneklemden elde edildiği görülmektedir. “b” parametresine ait standart hata değerleri incelendiğinde, en yüksek standart hata değerine 0,195 ile 150 kişilik örnekleme, en düşük değere ise 0,087 değeri ile 750 kişilik örnekleme ulaşıldığı görülmektedir. “c” parametresinin kestirilmesinde elde edilen standart hata değerleri incelendiğinde “c” parametresinin diğer tüm parametrelere göre daha yüksek hata değerlerinde kestirildiği görülmektedir. “c” parametresine ait ortalama standart hata değerleri incelendiğinde; 0,349 ile en yüksek 150 kişilik örneklemden; ve en düşük, 0,157 ile 750 kişilik örneklemden elde edildiği görülmektedir. Bu bilgiler ışığında “a”, “b” ve “c” parametrelerine ait hata değerlerinin örnek büyüklüğünden etkilendiği ve örneklemden kişi sayısı arttıkça hata değerinin düzenli olarak düştüğü gözlenmektedir.

20 maddelik test maddeleri için küçük örneklemlerden 1PL, 2PL ve 3PL modelde elde edilen ortalama standart hata değerleri Tablo 3.32’de gösterilmiştir.

Tablo 3.32

Küçük Örneklerde 20 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri

Örneklem	1PL Model	2PL Model		3PL Model		
	b	a	b	a	b	c
150	0,216	0,170	0,217	0,228	0,162	0,305
250	0,165	0,137	0,121	0,180	0,125	0,237
350	0,139	0,116	0,108	0,151	0,111	0,208
500	0,117	0,097	0,088	0,123	0,089	0,179
750	0,096	0,081	0,074	0,103	0,076	0,139

Tablo 3.32 incelendiğinde, 1PL modelde “b” parametresi için elde edilen ortalama standart hata değerlerinin en yüksek 0,216 ile 150 kişilik örneklemden, en düşük 0,096 ile 750 kişilik örneklemden elde edildiği belirlenmiştir.

2PL modelde “a” ve “b” parametrelerine ait ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek ortalama hata değerinin 0,170 ile 150 kişilik örneklemden; en düşük ortalama standart hata değerinin ise 0,081 ile 750 kişilik örneklemden elde edildiği görülmektedir. 2PL modelde kestirilen “b” parametresine ait ortalama hata değerleri incelendiğinde ise en yüksek 0,217 ile 150 kişilik örneklemden, en düşük 0,074 ile 750 kişilik örneklemden elde edildiği görülmektedir.

3PL modelde kestirilen “a”, “b” ve “c” parametrelerine ait ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek ortalama standart hata değerinin 0,228 ile 150 kişilik örneklemden, en düşük ortalama standart hata değerinin ise 0,103 ile 750 kişilik örneklemden elde edildiği; “b” parametresinden elde edilen ortalama standart hata değerleri incelendiğinde ise yine en yüksek standart hata değerinin 0,162 ile 150 kişilik örneklemden, en düşük standart hata değerinin ise 0,076 ile 750 kişilik örneklemden elde edildiği, “c” parametresine ait ortalama standart hata

değerleri incelendiğinde en yüksek değer diğer iki parametrede olduğu gibi 0,305 değeri ile 150 kişilik örneklemden, en düşük ortalama standart hata değerinin ise 0,139 değeri ile 750 kişilik örneklemden elde edildiği görülmektedir.

30 maddelik testten küçük örneklerde 1PL, 2PL ve 3PL modellerde elde edilen ortalama standart hata değerleri Tablo 3.33'te sunulmuştur.

Tablo 3.33

Küçük Örneklerde 30 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri

Örneklem	1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c
150	0,211	0,176	0,163	0,226	0,173	0,312
250	0,161	0,141	0,130	0,185	0,128	0,248
350	0,137	0,118	0,112	0,149	0,110	0,215
500	0,114	0,099	0,093	0,124	0,094	0,177
750	0,094	0,081	0,074	0,101	0,074	0,142

Tablo 3.33 incelendiğinde, 1PL modelde “b” parametresi için elde edilen ortalama standart hata değerlerinin en yüksek 0,211 ile 150 kişilik örneklemden, en düşük 0,094 ile 750 kişilik örneklemden elde edildiği belirlenmiştir.

2PL modelde kestirilen “a” ve “b” parametrelerine ait ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek ortalama hata değerinin 0,176 ile 150 kişilik örneklemden, en düşük değerin 0,081 ile 750 kişilik örneklemden elde edildiği görülmektedir. 2PL modelde kestirilen “b” parametresine ait ortalama hata değerleri incelendiğinde ise en yüksek 0,163 ile 150 kişilik örneklemden, en düşük 0,074 ile 750 kişilik örneklemden elde edildiği görülmektedir.

3PL modelde kestirilen “a”, “b” ve “c” parametrelerine ait ortalama standart hata deęerleri incelendięinde, “a” parametresi iin en yksek ortalama standart hata deęerinin 0,226 ile 150 kiřilik rneklemden, en dřk ortalama standart hata deęerinin ise 0,101 ile 750 kiřilik rneklemden elde edildięi, “b” parametresinden elde edilen ortalama standart hata deęerleri incelendięinde, yine en yksek standart hata deęerinin 0,173 ile 150 kiřilik rneklemden, en dřk ortalama standart hata deęerinin ise 0,074 ile 750 kiřilik rneklemden elde edildięi, “c” parametresine ait ortalama standart hata deęerlerinde en yksek deęerin 0,312 deęeri ile 150 kiřilik rneklemden, en dřk deęerin ise 0,142 deęeri ile 750 kiřilik rneklemden elde edildięi grlmektedir.

3.3.2. Byk rneklemlere (1000, 2000, 3000, 5000) Uygulanan 10 Maddelik Bir testten MTK’nın 1PL, 2PL ve 3PL Modellerine Gre Kestirilmif Madde Parametrelerine Ait Standart Hata Deęerleri Nasıldır?

10 maddelik testten byk rneklemlerde 1PL, 2PL ve 3PL modellerde elde edilen madde parametrelerine ait ortalama standart hata deęerleri Tablo 3.34’te gsterilmiřtir.

Tablo 3.34

Byk rneklemlerde 10 Maddelik Testten Elde Edilen Ortalama Standart Hata Deęerleri

rneklemler	1PL Model	2PL Model		3PL Model		
	b	a	b	a	b	c
1000	0,090	0,072	0,071	0,094	0,077	0,131
2000	0,064	0,052	0,054	0,067	0,054	0,096
3000	0,052	0,042	0,042	0,055	0,045	0,077
5000	0,041	0,032	0,034	0,041	0,035	0,061

Tablo 3.34’te 10 maddelik testten 1PL modelde kestirilen “b” parametresine ait ortalama standart hata deęerleri incelendięinde, en yksek 0,090 deęeriyle 1000 kiřilik rneklemden, en dřk 0,041 deęeriyle 5000 kiřilik rneklemden elde edildięi

görülmektedir. Ayrıca daha önceki bulgularla paralel olarak örneklem büyüklüğü arttıkça ortalama standart hata değerlerinin de 1PL model için düzenli olarak düşüş gösterdiği görülmektedir.

2PL modelde kestirilen “a” ve “b” parametrelerine ait ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek ortalama hata değerinin 0,72 ile 1000 kişilik örneklemden, 0,32 ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir. “b” parametresine ait ortalama standart hata değerleri incelendiğinde “a” parametresine ait ortalama standart hata değerleri ile neredeyse aynı değerler aldığı görülmektedir. “b” parametresi için elde edilen ortalama standart hata değerleri en yüksek 0,071 ile 1000 kişilik örneklemden ve en düşük 0,034 ile 5000 kişilik örneklemden elde edilmiştir.

3PL modelde elde edilen ortalama standart hata değerlerinin “a” parametresi için en yüksek 0,094 ile 1000 kişilik örneklemden, 0,041 ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir. “b” parametresi için elde edilen ortalama standart hata değerleri arasında da yine aynı ilişki tespit edilmiştir. Şöyle ki en yüksek ortalama standart hata değeri yine 1000 kişilik örneklemden 0,77 değeri ile, en düşük değer ise 0,035 ile 5000 kişilik örneklemden elde edilmiştir. “c” parametresinin büyük örneklemelerden elde edilen ortalama hata değerleri incelendiğinde 0,131 değeri ile en yüksek 1000 kişilik örneklemden, 0,061 ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir.

20 maddelik testten büyük örneklemelerde 1PL, 2PL ve 3PL modellerde elde edilen madde parametrelerine ait ortalama standart hata değerleri Tablo 3.35’te gösterilmiştir.

Tablo 3.35

Büyük Örneklerde 20 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri

Örneklem	1PL Model	2PL Model		3PL Model		
	b	a	b	a	b	c
1000	0,083	0,068	0,059	0,083	0,064	0,117
2000	0,059	0,049	0,044	0,060	0,045	0,084
3000	0,042	0,040	0,035	0,050	0,038	0,068
5000	0,038	0,031	0,027	0,038	0,029	0,053

Tablo 3.35'te 20 maddelik testten 1PL modelde kestirilen "b" parametresine ait ortalama standart hata değerleri incelendiğinde, en yüksek değer 0,083 değeriyle 1000 kişilik örneklemden, en düşük değer ise 0,038 ile 5000 kişilik örneklemden elde edildiği görülmektedir.

2PL modelde kestirilen "a" ve "b" parametrelerine ait ortalama standart hata değerleri incelendiğinde, "a" parametresi için en yüksek ortalama hata değerinin 0,068 ile 1000 kişilik örneklemden, 0,031 ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir. 2PL modelde kestirilen "b" parametresine ait ortalama standart hata değerleri en yüksek 0,059 ile 1000 kişilik örneklemden ve en düşük 0,027 ile 5000 kişilik örneklemden elde edildiği görülmektedir.

3PL modelde elde edilen ortalama standart hata değerlerinin "a" parametresi için en yüksek 0,083 ile 1000 kişilik örneklemden, 0,038 ile en düşük 5000 kişilik örneklemden elde edildiği, "b" parametresi için elde edilen ortalama standart hata değerlerinin, en yüksek 1000 kişilik örneklemden 0,064 değeri ile, en düşük 0,029 değeri ile 5000 kişilik örneklemden elde edildiği görülmektedir. "c" parametresi için elde edilen ortalama hata değerleri incelendiğinde, 0,117 değeri ile en yüksek 1000 kişilik örneklemden; 0,053 değeri ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir.

30 maddelik testten büyük örneklerde 1PL, 2PL ve 3PL modellerde elde edilen madde parametrelerine ait ortalama standart hata değerleri Tablo 3.36’da gösterilmiştir.

Tablo 3.36

Büyük Örneklerde 30 Maddelik Testten Elde Edilen Ortalama Standart Hata Değerleri

Örneklem	1PL Model	2PL Model		3PL Model		
	b	a	b	a	b	c
1000	0,081	0,070	0,063	0,084	0,065	0,123
2000	0,058	0,050	0,046	0,064	0,045	0,087
3000	0,037	0,041	0,037	0,050	0,037	0,070
5000	0,037	0,031	0,029	0,038	0,029	0,056

Tablo 3.36’da 30 maddelik testten 1PL modelde kestirilen “b” parametresine ait ortalama standart hata değerleri incelendiğinde, en yüksek değer 0,081 değeriyle 1000 kişilik örneklemden, en düşük değerin ise 0,037 ile 5000 kişilik örneklemden elde edildiği görülmektedir.

2PL modelde kestirilen “a” ve “b” parametrelerine ait ortalama standart hata değerleri incelendiğinde, “a” parametresi için en yüksek ortalama hata değerinin 0,070 ile 1000 kişilik örneklemden, 0,031 ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir. 2PL modelde kestirilen “b” parametresine ait ortalama standart hata değerlerinin ise en yüksek 0,063 ile 1000 kişilik örneklemden ve en düşük 0,029 ile 5000 kişilik örneklemden elde edildiği görülmektedir.

3PL modelde elde edilen ortalama standart hata değerlerinin “a” parametresi için en yüksek 0,084 ile 1000 kişilik örneklemden, 0,038 ile en düşük 5000 kişilik örneklemden, “b” parametresi için elde edilen ortalama standart hata değerlerinin, en yüksek 0,65 değeri ile 1000 kişilik örneklemden, en düşük 0,029 değeri ile 5000 kişilik örneklemden, “c” parametresi için elde edilen ortalama hata değerlerinin ise 0,123

değeri ile en yüksek 1000 kişilik örneklemden, 0,056 değeri ile en düşük 5000 kişilik örneklemden elde edildiği görülmektedir.

Üçüncü alt probleme ait bulgular toplu olarak değerlendirildiğinde, 10, 20 ve 30 maddelik testlerde en yüksek hata değerlerinin “c” parametresinden elde edildiği; “c” parametresinin kestiriminin soru sayısı arttıkça zorlaştığı; “c” parametresinin kestiriminde daha düşük hata değerlerine ulaşmak için soru sayısı arttıkça daha büyük örneklere ihtiyaç duyulduğu; tüm madde parametrelerine ait ortalama standart hata değerlerinin örneklem büyüklüğünün artmasından oldukça olumlu etkilendiği ve örneklem büyüklüğünün artmasından en olumlu etkilenen parametrenin “c” parametresi olduğu söylenebilir. Ayrıca, küçük örneklemlerde madde sayısı 30’a ulaştığında 20 maddelik testten elde edilen ortalama standart hata değerlerinden çok daha yüksek ortalama standart hata değerleri elde edildiği gözlenmiştir. Bu durum madde sayısının artmasının daha fazla sayıda parametre kestirimi gerektirdiğinden dolayı küçük örneklemlerde yeterli bilginin elde edilememesinin yüksek standart hata değerlerine sebep olduğu şeklinde yorumlanmıştır. Ayrıca madde sayısının artmasının küçük örneklemlerde ortalama standart hata değerlerini 20 maddeden sonra olumsuz etkilediği, bu olumsuz etkinin büyük örneklemlerde küçük örneklere oranla sınırlı kaldığı tespit edilmiştir. Büyük örneklemlerde standart hata değerlerinin daha az olmasının sebebi olarak parametre kestiriminde daha fazla örneklemin daha fazla bilgiye ulaşılması anlamını taşıdığından dolayı bu tür bir sonuca ulaşıldığı şeklinde yorumlanmıştır. Bu bilgiler ışığında, küçük örneklemlerde özellikle 2PL ve 3PL modellerde madde parametresi kestirimleri yapmak için testteki madde sayısının en fazla 30 ile sınırlandırılmasının ortalama standart hata değerlerinin kabul edilebilir düzeylerde kalmasını sağlamak açısından önemli olduğu söylenebilir.

BÖLÜM IV

SONUÇLAR VE ÖNERİLER

Bu bölümde her bir alt problem için elde edilen bulgulara ve yorumlara dayalı olarak ulaşılan sonuçlara ve sonraki araştırmalara yönelik önerilere yer verilecektir.

4.1. SONUÇLAR

Araştırmanın problemlerine yönelik olarak elde edilen sonuçlar sırasıyla aşağıda sunulmuştur.

1. Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında 1PL, 2PL ve 3PL modellerde model veri uyumu nasıl değişmektedir?

a. Küçük örneklerde 10 maddelik bir testte model veri uyumu χ^2/sd oranı, Standartlaştırılmış Artık Analizi (SAA), G^2 Olabilirlik Oranı, test karakteristik eğrilerinin incelenmesi yöntemleri olmak üzere toplam dört farklı yöntemle sınanmıştır. SAA dışındaki yöntemlerde model veri uyumunun örneklem büyüklüğünden etkilenmediği ve bu durumda model veri uyumunun 10 maddelik testlerde örneklem büyüklüğüne göre değişmediği sonucuna varılmıştır.

Küçük örneklerde 20 maddelik testte model veri uyumu yine dört farklı yöntemle sınanmıştır. Bu sına sonucunda, madde sayısı artmasına rağmen, SAA yönteminin küçük örneklerde örneklem büyüklüğünden etkilendiği ve bu yöntem dışındaki yöntemlerin örneklem büyüklüğünün artmasından etkilenmediği belirlenmiştir. Ayrıca test uzunluğunun artmasının da model veri uyumu üzerinde etkisinin olmadığı belirlenmiştir. Şöyle ki, 10 maddelik testten elde edilen sonuçlar ile 20 maddelik testten elde edilen sonuçlar karşılaştırıldığında, Standartlaştırılmış Artık Analizi dışındaki yöntemlerde aynı sonuçlara ulaşıldığı belirlenmiştir.

Küçük örneklerde 30 maddelik test verisinde model veri uyumu analizi sonuçları incelendiğinde, yine SAA dışındaki yöntemlerin örneklem büyüklüğünden etkilenmediği belirlenmiştir. Bu bilgiler ışığında daha önceki sonuçlara paralel olarak 30 maddelik bir testin uygulandığı küçük örneklerde model veri uyumunun örneklem büyüklüğünün artmasından etkilenmediği sonucuna varılmıştır. Bunun yanısıra, 20 maddelik ve 30 maddelik testlerde model veri uyumuna ait tablolar birbirleriyle karşılaştırıldığında yöntemlerin tüm örneklem büyüklüklerinde aynı sonuçlara ulaştığı görülmektedir. Bu bilgiler ışığında küçük örneklerde model veri uyumunun 10 maddeden fazla uzunluktaki testlerde test uzunluğundan da etkilenmediği sonucuna varılmıştır.

b. Büyük örneklerde 10 maddelik bir testte model veri uyumu daha önce belirtilen dört farklı yöntemle sınanmıştır. Büyük örneklerde SAA yöntemi küçük örneklerin aksine 10 maddelik test verisinde örneklem büyüklüğünden etkilenmemiştir. SAA'nın, 500 kişilik büyüklükten sonra sonra örneklem büyüklüğünden etkilenmediği belirlenmiştir. SAA'nın aksine x^2/sd oranı örneklem büyüklüğü 3000'e ulaştığında farklı sonuçlar vermeye başlamıştır. Bunun nedeni bu yöntemin büyük örnek gruplarına uygulandığı durumlarda modelden çok küçük empirik sapmaların bile birçok maddenin aslında uygulamada oldukça iyi iş görebilecekleri halde modele uymayan maddeler olarak işaretlenmesine sebep olduğu (Hambleton, Swaminathan ve Rogers, 1991) şeklinde açıklanabilir.

Bunların yanısıra, model veri uyumunun sınanmasında kullanılan diğer üç yöntem kendi içinde tutarlı sonuçlar vermiştir. Kısaca, her ne kadar büyük örneklerde model veri uyumunun sınıandığı x^2/sd oranı örneklem büyüklüğü çok arttığında örneklem büyüklüğünden etkilenmeye başlasa da diğer tüm yöntemler örneklem büyüklüğünden etkilenmemişlerdir. Bu bilgiler ışığında büyük örneklemelere uygulanan 10 maddelik bir testte model veri uyumu örneklem büyüklüğünden etkilenmemektedir sonucuna varılmıştır.

Büyük örneklerde 20 maddelik testte model veri uyumunun sınanmasında daha önce olduğu gibi dört farklı yöntem kullanılmıştır. Bu yöntemlerin tümü neredeyse tüm

örneklem büyüklüklerinde kendi içinde tutarlı sonuçlar vermiştir. Bu bilgiler ışığında model veri uyumunun örneklem büyüklüğünden etkilenmediği sonucuna varılmıştır. Aynı örneklem büyüklüklerinde 10 maddelik testten elde edilen model veri uyumu sonuçları karşılaştırıldığında, χ^2/sd yöntemi dışındaki yöntemlerde aynı sonuçlar elde edilmiştir. Bu bilgiler ışığında model veri uyumunun test uzunluğundan da etkilenmediği sonucuna varılmıştır. Bunun yanısıra, daha önce büyük örneklemelere uygulanan 10 maddelik testte örneklem büyüklüğünden etkilendiği belirlenen χ^2/sd yönteminin madde sayısı 20'ye ulaştığında örneklem büyüklüğünden etkilenmediği, dolayısıyla bu yöntemin çok büyük örneklemelerde en az 20 madde ile kullanılmasının uygun olabileceği sonucuna varılmıştır.

Büyük örneklemelerde 30 maddelik test verisinde model veri uyumu analizi sonuçları incelendiğinde χ^2/sd için çalışma evrenine ait veri dışında tüm yöntemlerin örneklem büyüklüğünden etkilenmediği belirlenmiştir. Ayrıca 20 ve 30 maddelik testlerden elde edilen model veri uyumu analizi sonuçları karşılaştırıldığında bir sonuç dışında tamamen aynı sonuçların elde edildiği görülmüştür. Bu bilgiler ışığında model veri uyumunun büyük örneklemelerde test uzunluğu ve örneklem büyüklüğünden etkilenmediği sonucuna varılmıştır.

2. Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametreleri, çalışma evreninden MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerine göre nasıl değişmektedir?

a. Bu araştırma probleminin cevaplanması amacıyla çalışma evreninden kestirilen madde parametreleri ile küçük örneklemelerde kestirilen madde parametreleri arasındaki korelasyon katsayıları incelenmiştir. Bunun sonucunda Küçük örneklemelerde 10 maddelik bir testte MTK'nın 1PL modelinde "b" parametresinin kestirimi için 150 kişilik bir örneklemin yeterli olabileceği sonucuna varılmıştır. Benzer bir şekilde, Goldman ve Raju (1986) çalışmalarında kullandıkları en küçük örneklem büyüklüğü olan 250 kişilik örneklemin "b" parametresinin kestirimi için yeterli olduğu sonucuna

varmışlardır. Bu çalışmada elde edilen sonuç Goldman ve Raju (1986)'nın bu çalışmasının sonucu ile paraleldir. Bunun yanısıra bu çalışmada “b” parametresinin küçük örneklerde 2PL ve 3PL modellerde de modelden bağımsız olarak yüksek korelasyon değerleriyle kestiriminin yapılabileceği; en başarılı kestirilen parametrenin “b” parametresi olduğu; bunun yanısıra, 10 maddelik bir testte “a” parametresinin 2PL ve 3PL modellerde kestirimi için “b” parametresinin aksine en az 750 kişiye ihtiyaç duyulduğu belirlenmiştir. 3PL modelde “c” parametresinin kestirimi için de 150 kişiye uygulanmış 10 maddelik bir testin yeterli olabileceği sonucuna varılmıştır.

20 maddelik testten küçük örneklerde elde edilen bulgular, 10 maddelik testten elde edilen bulgularla paraleldir. Elde edilen bulgular ışığında, “b” parametresinin yine 20 maddelik teste ait verilerde de tüm örneklem büyüklüklerinde ve tüm modellerde yüksek korelasyon katsayılarına sahip olduğu sonucuna varılmıştır. Madde sayısının artmasının 2PL modelde “a” parametresinin kestiriminin daha başarılı şekilde yapılmasını sağlaması dikkat çekicidir. Madde sayısının artırılmasıyla 2PL modelde “a” parametresine ait korelasyon katsayıları gözle görülür bir şekilde yükselmiştir. Madde sayısının artması “c” parametresinin kestirimini daha güç hale getirmiştir. Daha önce 150 kişilik örnekleme bile yüksek korelasyon katsayıları alan “c” parametreleri madde sayısının artması ile yüksek korelasyon katsayılarına en az 750 kişilik örnekleme ulaşmıştır. Bu nedenle 20 maddelik bir testte “c” parametresinin kestirimi için asgari 750 kişilik bir örneklemin kullanılmasının uygun olabileceği sonucuna varılmıştır.

Küçük örneklerde madde sayısının 20'den 30'a çıkarılması en fazla etkiyi “a” parametresine ait korelasyon katsayısı değerleri üzerinde yapmıştır. 2PL ve 3PL modellerde “a” parametresine ait korelasyon katsayılarında hissedilir derece artış görülmüştür. “b” parametresi korelasyon katsayıları yine modelden bağımsız olarak oldukça yüksek değerler almaya devam etmiştir. Tüm bunlara rağmen madde sayısının 20'den 30'a çıkarılmasının madde parametreleri dışında çok büyük bir etki yaratmadığı sonucuna varılmıştır.

b. Büyük örneklere uygulanan 10 maddelik bir testte MTK'nın 2PL modelinde "a" parametresinin başarılı bir şekilde kestirilmesi için 1000 kişilik bir örneklem büyüklüğünün yeterli olacağı sonucuna ulaşılmıştır. 3PL modelde "a" parametresinin başarılı bir şekilde kestirimi için de 1000 kişilik örneklemin yeterli olacağı belirlenmiştir. 10 maddelik bir testte madde parametresi kestirimi yapılması için 2000 kişilik örneklemden fazlasının gerekli olmayabileceği sonucuna varılmıştır. Bu sonuç Hulin, Lissak ve Drasgow (1982)'un elde ettiği sonuçla paraleldir. Çünkü bu örneklem büyüklüğünden sonra korelasyon katsayıları neredeyse tam korelasyon değerlerine ulaşmaktadır. Son olarak "c" parametresinin kestiriminin madde sayısının düşük olduğu testlerde küçük örnekleme hatalarına karşı oldukça hassas olduğu sonucuna varılmıştır.

Büyük örneklemlerde madde sayısının artması 1PL modelde "b" parametresinin kestiriminde bir değişikliğe yol açmamıştır. Madde sayısının artması 2PL modelde "a" parametresinin daha yüksek korelasyon katsayılarıyla kestirilmesine neden olmuştur. Soru sayısının 10'dan 20'ye çıkarılması 2PL modele ait madde parametresi kestirimlerini daha yüksek korelasyona sahip olacak şekilde etkilemiştir.

Büyük örneklemlerde madde sayısının artırılması "b" parametresine ait korelasyonların küçük miktarda artış göstermesine; aynı zamanda "a" parametresinin hem 2PL hem de 3PL modellerde daha yüksek korelasyon katsayıları ile kestirilmesine neden olmuştur. Madde sayısının artırılması 3PL modelde "c" parametrelerinin korelasyonunun düşmesine neden olmuştur. Bu da madde sayısının artmasının "c" parametresinin sağlıklı bir şekilde kestirimini güçleştirdiği sonucuna varılmıştır. Bu sonuç Thissen ve Weiner (1982)'in çalışmalarında vardığı sonuca paraleldir.

3. *Test uzunluğu ve örneklem büyüklüğü farklılaştırıldığında, MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilmiş madde parametrelerine ait standart hata değerleri nasıl değişmektedir?*

a. 10 maddelik testten küçük örneklemelerde elde edilen standart hata değerleri örneklem büyüklüğü arttıkça azalmaktadır. 10 maddelik test verisinden en yüksek hata değeri ile kestirilen parametre “c” parametresidir.

20 maddelik testten küçük örneklemelerde elde edilen ortalama standartlaştırılmış hata değerleri 10 maddelik testten elde edilen değerlerde olduğu gibi örneklem büyüklüğü arttıkça azalmaktadır. 20 maddelik testte de en yüksek hata ile kestirilen parametre “c” parametresidir. Buna karşın örneklem büyüklüğü artırıldığında standart hata değerlerinde en büyük düşüş de yine daha önce belirtildiği gibi “c” parametresinde görülmektedir. Bunun yanısıra madde sayısının artması tüm parametrelerde standart hata değerlerinin düşmesine neden olmaktadır.

30 maddelik testten küçük örneklemelerde elde edilen standart hata değerleri diğer alt problemlerde olduğu gibi örneklem büyüklüğünün artmasından olumlu etkilenmektedir. 30 maddelik testte de tüm modellerde örneklem büyüklüğü arttıkça standart hata değerleri düşmektedir. 30 maddelik testte de en yüksek hata değeri ile kestirilen parametre “c” parametresidir. Örneklem büyüklüğünün artması sonucu en olumlu etki “c” parametresine ait standart hata değerlerinde görülmektedir. 30 maddelik testten elde edilen standart hata değerlerinde en dikkat çekici sonuç standart hata değerlerinin madde sayısı arttıkça artmasıdır. Özellikle 2PL ve 3PL modellerde 30 maddelik testten elde edilen standart hata değerlerinin 20 maddelik testten elde edilen standart hata değerlerinden daha yüksek olduğu belirlenmiştir. Bu durum özellikle 2PL ve 3PL modeller gibi daha fazla parametrenin kestiriminin yapıldığı modellerde madde sayısının 20'den fazla olmasının madde parametrelerinin kestirimini zorlaştırdığı ve dolayısıyla hata değerlerinde artışa neden olduğunu düşündürmektedir.

b. 10 maddelik testten elde edilen standart hata deęerleri büyük örneklerde de örneklem büyüklüğünün artmasından olumlu etkilenebilmektedir. Örneklem büyüklüğü arttıkça hata deęerleri düşmektedir. “c” parametresi büyük örneklerde de en yüksek hata ile kestirilen parametredir. Ayrıca “c” parametresi örneklem büyüklüğünün artmasından en olumlu etkilenen parametredir. Şöyle ki, örneklem büyüklüğü arttıkça en büyük düşüş “c” parametresinden elde edilen ortalama standart hata deęerlerinde elde edilmiştir.

20 maddelik testten elde edilen standart hata deęerleri büyük örneklerde de örneklem büyüklüğünden olumlu etkilenmektedir. Örneklem büyüklüğünün artması standart hata deęerlerinin düşmesini sağlamaktadır. Büyük örneklerde de en yüksek standart hata deęeri ile kestirilen parametre “c” parametresidir. Yine “c” parametresi örneklem büyüklüğünün artmasından en olumlu etkilenen parametredir.

30 maddelik testten büyük örneklerde elde edilen standart hata deęerleri dięer alt problemlerde olduęu gibi örneklem büyüklüğünün artmasından olumlu etkilenmekte ve örneklem büyüklüğü arttıkça standart hata deęerleri düşmektedir. Ayrıca örneklem büyüklüğünün artmasından en olumlu etkilenen parametre daha önce olduęu gibi “c” parametresidir. Örneklem büyüklüğünün artmasıyla en büyük düşüş “c” parametresine ait standart hata deęerlerinde görülmektedir. 20 ve 30 maddelik testlerden elde edilen standart hata deęerleri karşılaştırıldığında 2PL ve 3PL modellerde elde edilen standart hata deęerlerinin bir miktar arttıęı belirlenmiştir. Bu artış küçük örneklerde 20 maddelik test ile 30 maddelik testten elde edilen standart hata deęerleri arasındaki farka göre düşük kalmıştır. Büyük örneklerden elde edilen standart hata deęerleri madde sayısındaki artışa küçük örneklerle göre daha dayanıklıdır. Bu durumda küçük örneklerde madde sayısının 30 ile sınırlanması hata parametrelerinin kabul edilebilir düzeylerde kalması açısından önemlidir.

4.2. ÖNERİLER

Araştırmanın sonuçları ışığında bundan sonra bu konuda araştırma yapmak isteyen araştırmacılara yönelik öneriler aşağıda sıralanmıştır.

1. Bu çalışmada, Madde Tepki Kuramı'nın tek boyutlu modelleri kullanılarak madde parametreleri kestirimi yapılmıştır. Benzer bir çalışma, MTK'nın çok boyutlu modelleri kullanılarak tekrarlanabilir.
2. Bu çalışmada kullanılan testlerin uzunluğu, MTK'nın kısa testlerde etkinliğinin ortaya konması amacıyla 30 madde ile sınırlı tutulmuştur. Bundan sonraki araştırmalarda benzer bir çalışma daha fazla madde kullanılarak tekrarlanabilir.
3. Bu çalışmada madde parametreleri kestirimi temel alınmıştır. Benzer kapsamda bir çalışma yetenek parametrelerinin kestirimi üzerine yapılabilir.
4. Bu çalışmada, madde parametreleri tek bir bilgisayar programı ile kestirilmiştir. Benzer bir çalışma madde parametrelerinin iki farklı program ile kestirilip elde edilen bulguların birbirleriyle karşılaştırılması şeklinde yapılabilir.
5. Bu çalışma Madde Tepki Kuramı ile sınırlanmıştır. Benzer kapsamda bir çalışma Klasik Test Teorisi ile de yapıp elde edilen bulgular bu araştırmada elde edilen bulgularla karşılaştırılabilir.

KAYNAKÇA

- Adams, R. A., Wilson, M., ve Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1-23.
- Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *The Journal of the Royal Statistical Society*, 34, 42-54.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123- 140.
- Arlı, M., ve Nazik H. (2004). *Bilimsel arařtırmaya giriş*. Ankara : Gazi Kitabevi.
- Baker, F. B. (1998). An investigation of the item parameter recovery of a Gibbs sampling procedure. *Applied Psychological Measurement*, 22, 153-169.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F. B., ve Kim, S. (2004). *Item Response Theory: Parameter Estimation Techniques* (2nd ed.). New York: Marcel Dekker.
- Bock, R. D., ve Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika*, 35, 179-197.
- Bock, R. D., ve Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Büyüköztürk, Ş. (2003). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem A Yayıncılık .

- Chernyshenko, O.S., Stark, S., Chan, K., Drasgow, F. ve Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562.
- Chuah, S. C., Drasgow F., ve Luecht, R. (2010). How big is big enough? Sample size requirements for cast item parameter estimation. *Applied measurement in education*, 19(3), 241-255.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guilford Press.
- De la Torre, J., ve Hong, Y. (2010). Parameter estimation with small sample size: A higher-order IRT model approach. *Applied Psychological Measurement*, 34(4), 267-285.
- DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. London : Oxford Press.
- Doğan, N. (2002). *Klasik test kuramı ve örtük özellikler kuramının örneklem büyüklükleri bağlamında karşılaştırılması*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- Drasgow, F., Levine, M., Tsien, S., Williams, B., ve Mead, A. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-165.
- Edelen, M. O. ve Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(1), 5-18.
- Embretson, S. E., ve Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

- Fidelman, C. (2012). *A retrospective linking of ECLS-K and ECLS-B reading scores*. Paper presented at Federal Committee on Statistical Methodology Policy Seminar, Washington, DC.
- Gao, F., ve Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*(4), 351-380.
- Goldman, S.H., ve Raju, N. S. (1986). Recovery of one- and two-parameter logistic item parameters: An empirical study. *Educational and psychological measurement, 46*(1),11-21.
- Swaminathan, H., ve Gifford, J. A. (1986). Bayesian estimation in three-parameter logistic model. *Psychometrika, 51*,589-601.
- Hambleton, R., ve K., Swaminathan H. (1985). *Item response theory. Principals and applications*. Norwell, MA: Kluwer Academic Publishers.
- Hambleton R.K., Swaminathan H., ve H. J. Rogers (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.
- Hambleton, R. K., Jones R.W., ve Rogers, H. J. (1993). Influence of item parameter estimation errors in test development. *Journal of educational measurement, 30*, 143-155.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. R.L. Linn(ed.), *Educational Measurement*, (s.147-200). Washington, DC: American Council of Education.
- Hanson, B. A., ve Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the

common-item equating design. *Applied Psychological Measurement*, 26(1), 3-24.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, 55, 577-602.

Hulin, C. L., Lissak, R. I., ve Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.

Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel yayın dağıtım.

Kang, T., ve Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, 31, 331-358.

Kirisci, L., Hsu, T., ve Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25(2), 146-162.

Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, 61A, 273-287.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55, 187-193.

Liu, Y., Schulz, E. M., ve Yu, L. (2008). Standart error estimation of 3PL IRT true score equating with an MCMC method. *Journal of Educational and Behavioral Statistics*, 33 (3), 257-278.

Lord, F. M. A. (1952). A theory of test scores. *Psychometric monograph*, No: 7.

- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- Orlando, M. (2004). *Critical issues to address when applying item response theory models*. Paper presented at the conference on improving health outcomes assessment, National Cancer institute, Bethesda, MD, USA.
- Patsula, L. N., ve Gessaroli M. E. (1995). *A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different test lengths and sample sizes*. (Yayınlanmamış Master Tezi). University of Ottawa, Ontario, Canada.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57.
- Rao, C. R. ve Sinharay, S. (2007). *Handbook of Statistics 26: Psychometrics*. The Netherlands : North-holland Publications.
- Reise, S., ve Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, 27(2), 133-44.
- Reeve B. (2006). *An introduction to modern measurement theory*. Rockville, MD: National Cancer Institute. <http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf>
- Russell, D. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in personality and social psychology bulletin. *Personality and Social Psychology Bulletin*, 28,1629-1646.

- Scherbaum, C., Finlinson, S., Barden, K., ve Tamanini, K. (2006). Applications of item response theory to measurement issues in leadership research. *The Leadership Quarterly*, 17(4), 366-386.
- Sireci, S. G. (1991). *An investigation of the stability of IRT item parameters estimated from small data sets*. Paper presented at the annual Conference of Northeastern Educational Research Association, Ellenville, New York.
- Steinberg, L., ve Thissen, D. (1995). Item response theory in personality research. P. E. Shrout & S. T. Fiske (Ed.), *Personality research, methods, and theory: a festschrift honoring Donald W. Fiske* (pp. 161-181). Hilldale, NJ: Erlbaum.
- Thissen, D., ve Wainer, H. (1982). Some standart errors in item response theory. *Psychometrika*, 47, 397-412.
- Toland, M. (2008). *Determining the accuracy of item parameter standart error of estimates in BILOG-MG 3*. Yayınlanmamış Doktora Tezi. University of Nebraska, US.
- Weiss, D. J. (1995). Improving individual differences measurement with item response theory and computerized adaptive testing. D. J. Lubinski, & R. V. Dawis (Ed.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 49-79). Palo Alto, CA: Davies-Black Publishing.
- Wright, B.D. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conferences on Testing Problems*. Princeton, NJ: Educational Testing Service.
- Wright, B. D., ve Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105(2), 276-289.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, 52, 275-291.

Ek 1**Küçük Örneklemelere Ait Betimsel İstatistikler**

Betimsel İstatistikler	Küçük Örneklemeler															Çalışma Evreni		
	150			250			350			500			750			6288		
	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
Ort.	7,05	12,80	19,07	6,90	12,48	18,52	7,09	12,74	18,99	7,23	13,01	19,14	7,05	12,75	18,97	7,08	12,84	19,04
Std.Sap.	2,40	4,86	6,63	2,44	4,71	6,43	2,37	4,57	6,30	2,34	4,50	6,22	2,39	4,61	6,45	2,43	4,69	6,47
Varyans	5,79	23,70	43,96	5,95	22,25	41,35	5,63	20,88	39,88	5,45	20,29	38,66	5,70	21,21	41,61	5,92	21,97	41,97
Çarpıklık	-,44	-,28	-,21	-,42	-,25	-,13	-,68	-,26	-,24	-,71	-,43	-,33	-,55	-,25	-,19	-,56	-,30	-,24
Basıklık	-1,01	-1,10	-1,11	-,92	-1,07	-1,07	-,36	-,94	-,93	-,36	-,73	-,77	-,68	-,93	-,94	-,68	-,95	-,94
Güvenirlilik	,75	,87	,88	,75	,85	,87	,74	,84	,87	,74	,84	,87	,74	,84	,88	,76	,85	,88

Ek 2**Büyük Örneklemelere Ait Betimsel İstatistikler**

Betimsel İstatistikler	Büyük Örneklemeler												Çalışma Evreni		
	1000			2000			3000			5000			6288		
	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
Ort.	7,02	12,7	18,82	7,05	12,82	19,00	7,02	12,74	18,91	7,10	12,89	19,12	7,08	12,84	19,04
Std.Sap.	2,47	4,77	6,57	2,45	4,70	6,48	2,47	4,74	6,54	2,43	4,69	6,46	2,43	4,69	6,47
Varyans	6,1	22,76	43,12	6,01	22,08	41,92	6,08	22,47	42,71	5,91	21,96	41,77	5,92	21,97	41,97
Çarpıklık	-,58	-,31	-,23	-,56	-,30	-,24	-,52	-,27	-,21	-,58	-,32	-,27	-,56	-,30	-,24
Basıklık	-,63	-,95	-,98	-,70	-,97	-,98	-,75	-,98	-,96	-,67	-,94	-,93	-,68	-,95	-,94
Güvenirlilik	,77	,86	,88	,76	,85	,88	,76	,86	,88	,76	,85	,88	,76	,85	,88

Ek 3

Çalışmada Kullanılan Test Maddelerine Ait Faktör Yükleri

30 Maddelik Test		20 Maddelik Test		10 Maddelik Test	
Madde Numarası	Faktör yükü	Madde Numarası	Faktör Yüğü	Madde Numarası	Faktör Yüğü
1	0,66	1	0,66	1	0,70
2	0,68	2	0,68	2	0,71
3	0,45	3	0,63	3	0,70
4	0,63	4	0,62	4	0,68
5	0,54	5	0,61	5	0,71
6	0,61	6	0,71	6	0,75
7	0,60	7	0,60	7	0,69
8	0,71	8	0,59	8	0,75
9	0,60	9	0,66	9	0,77
10	0,58	10	0,70	10	0,70
11	0,66	11	0,65		
12	0,47	12	0,59		
13	0,57	13	0,72		
14	0,69	14	0,67		
15	0,64	15	0,65		
16	0,58	16	0,74		
17	0,57	17	0,75		
18	0,71	18	0,69		
19	0,41	19	0,64		
20	0,67	20	0,63		
21	0,65				
22	0,55				
23	0,73				
24	0,75				
25	0,53				
26	0,53				
27	0,52				
28	0,69				
29	0,63				
30	0,62				

Ek 5

 χ^2 /sd yöntemine göre Madde-Model Uyumu

	1PL Model			2PL Model			3PL Model		
	Uyumlu Madde Sayısı	Uyumsuz Maddeler	Uyum Katsayısı	Uyumlu Madde Sayısı	Uyumsuz Maddeler	Uyum Katsayısı	Uyumlu Madde Sayısı	Uyumsuz Maddeler	Uyum Katsayısı
150-10	10	-	1,00	10	-	1,00	10	-	1,00
150-20	20	-	1,00	20	-	1,00	20	-	1,00
150-30	30	-	1,00	30	-	1,00	30	-	1,00
250-10	10	-	1,00	10	-	1,00	10	-	1,00
250-20	20	-	1,00	20	-	1,00	20	-	1,00
250-30	30	-	1,00	30	-	1,00	30	-	1,00
350-10	10	-	1,00	10	-	1,00	10	-	1,00
350-20	20	-	1,00	20	-	1,00	20	-	1,00
350-30	30	-	1,00	30	-	1,00	30	-	1,00
500-10	10	-	1,00	10	-	1,00	9	M5	0,90
500-20	20	-	1,00	20	-	1,00	20	-	1,00
500-30	30	-	1,00	30	-	1,00	30	-	1,00
750-10	10	-	1,00	10	-	1,00	10	-	1,00
750-20	20	-	1,00	20	-	1,00	20	-	1,00
750-30	29	M19	0,97	30	-	1,00	30	-	1,00
1000-10	10	-	1,0	10	-	1,00	9	M8	0,90
1000-20	20	-	1,0	20	-	1,00	20	-	1,00
1000-30	29	M19	0,97	30	-	1,00	30	-	1,00
2000-10	9	M8	0,9	9	M4	0,90	9	M8	0,90
2000-20	20	-	1,0	20	-	1,00	20	-	1,00
2000-30	27	M12,19,24	0,9	30	-	1,00	30	-	1,00
3000-10	6	M3,6,8,9	0,60	8	M7,8	0,80	5	M1,3,5,8,10	0,50
3000-20	20	-	1,00	20	-	1,00	20	-	1,00
3000-30	26	M3,12,19,24	0,87	30	-	1,00	30	-	1,00
5000-10	6	M3,5,8,9	0,60	7	M4,7,8	0,70	0	TÜMÜ	0,00
5000-20	19	M17	0,95	18	M5,16	0,90	19	M6	0,95
5000-30	25	M3,12,18,19,24	0,83	30	-	1,00	30	-	1,00
6288-10	5	M3,5,6,8,9	0,50	7	M4,7,8	0,70	0	TÜMÜ	0,00
6288-20	18	M5,6	0,90	16	M5,7,8,16	0,80	19	M17	0,95
6288-30	23	M2,3,12,18,19,23,24	0,77	30	-	1,00	29		0,97

EK 6

Tetrakorik Korelasyon Matrisine Dayalı Faktör Analizi için SPSS Syntax'leri

10 Maddelik Test için yazılan syntax

```

FACTOR MATRIX=IN(COR=*)
  /MISSING LISTWISE
  /ANALYSIS Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10
  /PRINT INITIAL EXTRACTION
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(0.05) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /METHOD=CORRELATION.

```

20 Maddelik Test için yazılan syntax

```

FACTOR MATRIX=IN(COR=*)
  /MISSING LISTWISE
  /ANALYSIS Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17
    Q18 Q19 Q20
  /PRINT INITIAL EXTRACTION
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(0.05) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /METHOD=CORRELATION.

```

30 Maddelik Test için yazılan syntax

```

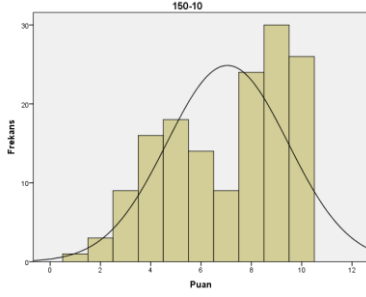
FACTOR MATRIX=IN(COR=*)
  /MISSING LISTWISE
  /ANALYSIS Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17
    Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25 Q26 Q27 Q28 Q29 Q30
  /PRINT INITIAL EXTRACTION
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(0.05) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /METHOD=CORRELATION.

```

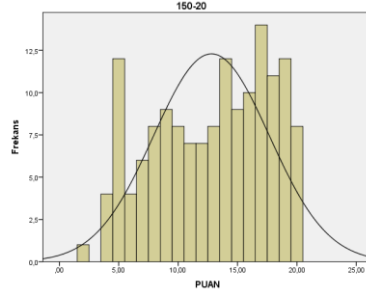
Ek 7

Büyük ve Küçük Örneklemelere Ait Histogram Grafikleri

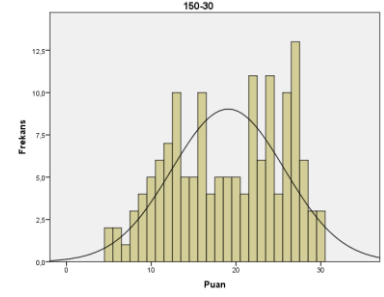
150-10



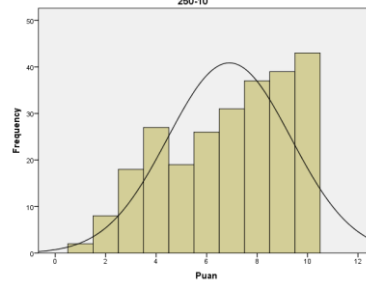
150-20



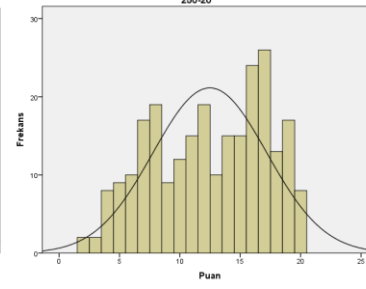
150-30



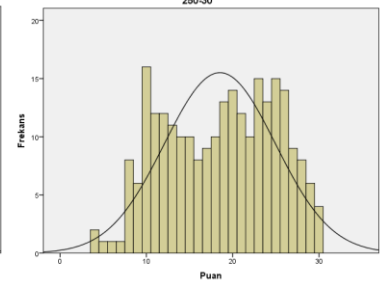
250-10



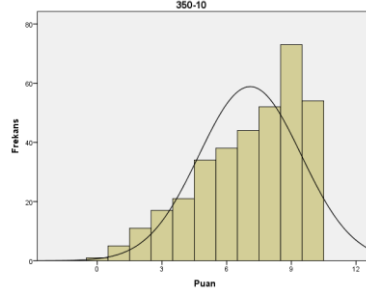
250-20



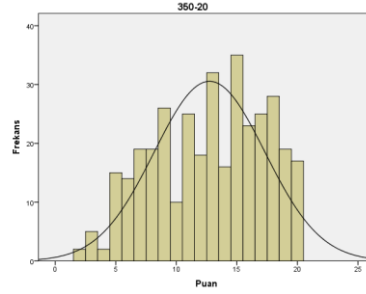
250-30



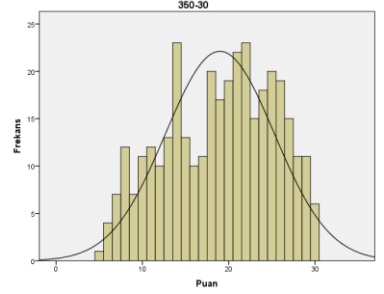
350-10



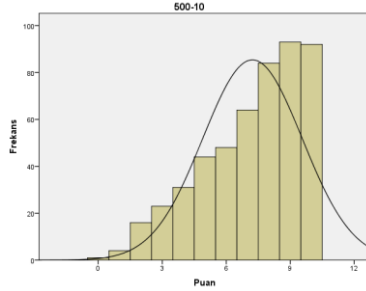
350-20



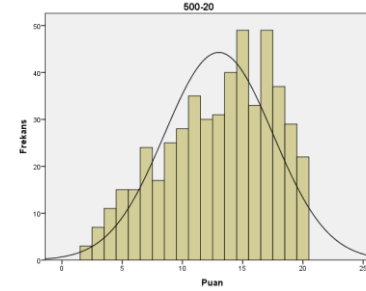
350-30



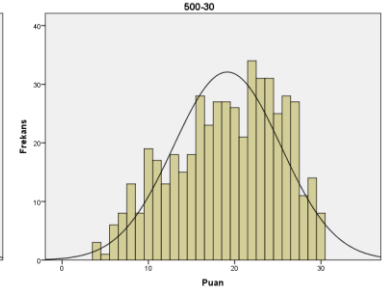
500-10



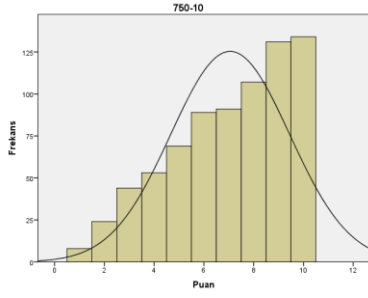
500-20



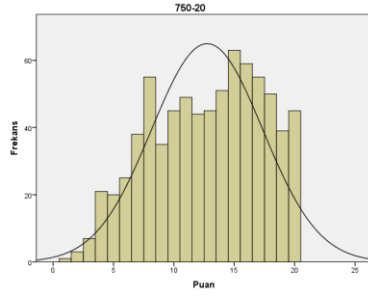
500-30



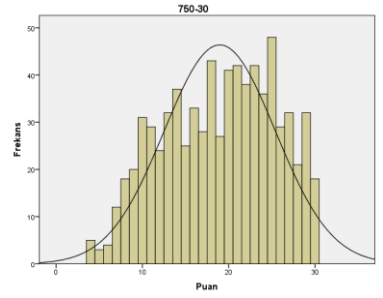
1000-10



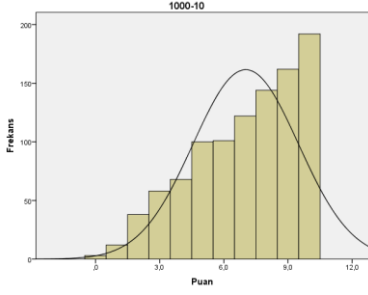
2000-20



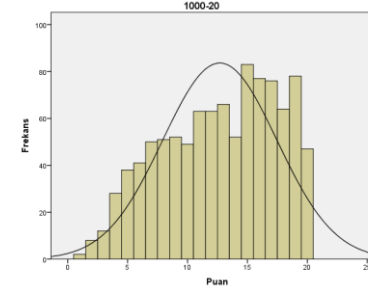
2000-30



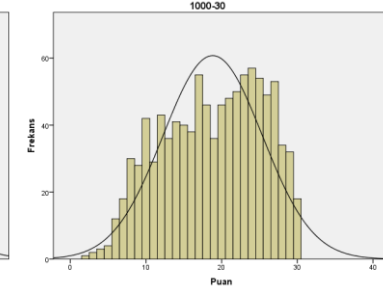
2000-10



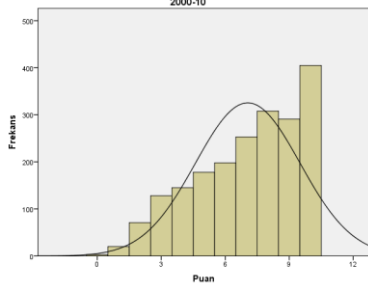
2000-20



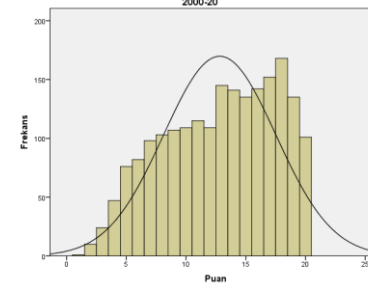
2000-30



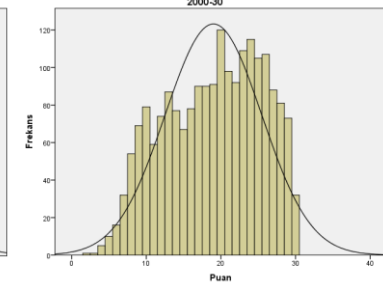
3000-10



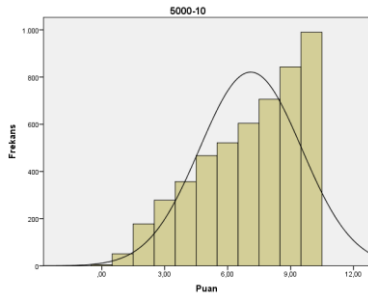
3000-20



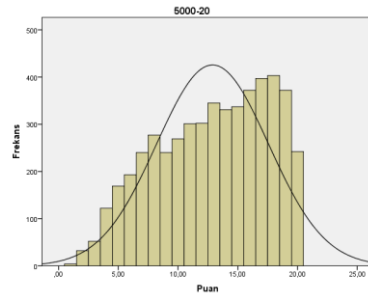
3000-30



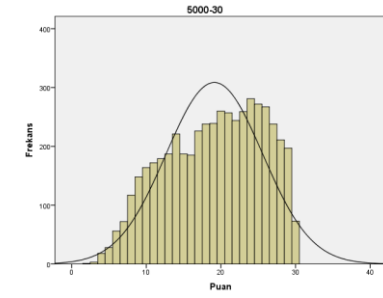
5000-10



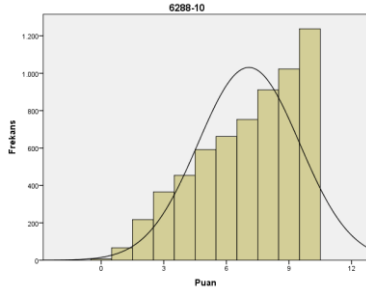
5000-20



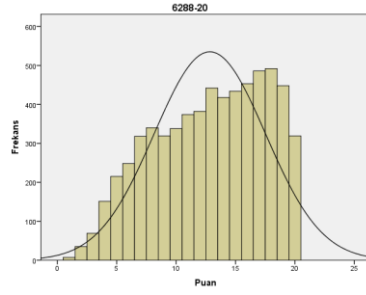
5000-30



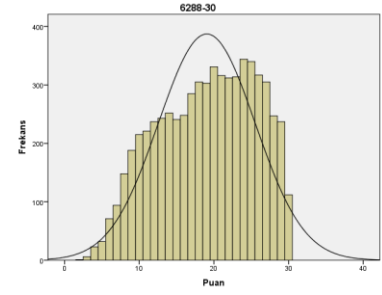
6288-10



6288-20



6288-30



EK 8

Araştırmanın Alt Problemlerine İlişkin Toplu Bulgular

Birinci Alt Probleme İlişkin Toplu Bulgular

Örneklem	10 Maddelik Test				20 Maddelik Test				30 Maddelik Test			
	x^2/sd	SAA	G ²	TKE	x^2/sd	SAA	G ²	TKE	x^2/sd	SAA	G ²	TKE
150	1,2,3	2	2,3	1,2,3	1,2,3	2	2,3	1,2,3	1,2,3	2	2,3	1,2,3
250	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3
350	1,2,3	2	2,3	1,2,3	1,2,3	2	2,3	1,2,3	1,2,3	2	2,3	1,2,3
500	1,2,3	3	2,3	1,2,3	1,2,3	2	2,3	1,2,3	1,2,3	2	2,3	1,2,3
750	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3
1000	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3	1,2,3	3	2,3	1,2,3
2000	1,2,3	3	3	1,2,3	1,2,3	3	3	1,2,3	1,2,3	3	3	1,2,3
3000	2	3	3	1,2,3	1,2,3	3	3	1,2,3	1,2,3	3	3	1,2,3
5000	2	3	3	1,2,3	1,2,3	3	3	1,2,3	1,2,3	3	3	1,2,3
6288*	2	3	3	1,2,3	1,2,3	3	3	1,2,3	2,3	3	3	1,2,3

İkinci Alt Probleme İlişkin Toplu Bulgular

Örneklemler	10 Maddelik Test						20 Maddelik Test						30 Maddelik Test					
	1PL Model		2PL Model		3PL Model		1PL Model		2PL Model		3PL Model		1PL Model		2PL Model		3PL Model	
	b	a	b	a	b	c	b	a	b	a	b	c	b	a	b	a	b	c
150	,939**	0,309	,940**	0,606	,944**	,733*	,962**	0,329	,964**	,516*	,955**	,485*	,958**	,658**	,961**	,719**	,948**	,535**
250	,997**	-0,311	,997**	0,207	,993**	,757*	,993**	,642**	,993**	,583**	,990**	,671**	,979**	,830**	,978**	,770**	,981**	,548**
350	,992**	0,083	,992**	0,481	,993**	,415	,993**	,642**	,994**	,581**	,990**	,697**	,987**	,779**	,987**	,749**	,987**	,734**
500	,976**	0,347	,978**	0,345	,974**	,940**	,984**	,795**	,985**	,614**	,972**	,820**	,982**	,895**	,983**	,831**	,975**	,735**
750	,999**	,706*	,999**	,901**	,998**	,922**	,998**	,793**	,998**	,840**	,995**	,829**	,994**	,870**	,995**	,871**	,995**	,756**
1000	,994**	,864**	,994**	,929**	,994**	,938**	,994**	,935**	,994**	,874**	,989**	,915**	,993**	,972**	,994**	,934**	,989**	,857**
2000	,999**	,826**	,999**	,918**	1**	,988**	,998**	,926**	,999**	,929**	,997**	,906**	,998**	,959**	,997**	,955**	,998**	,851**
3000	,999**	,886**	,999**	,971**	1**	,926**	,999**	,977**	,999**	,957**	,998**	,962**	,999**	,988**	,999**	,982**	,999**	,952**
5000	1**	,967**	1**	,991**	1**	,968**	1**	,997**	1**	,995**	1**	,996**	1**	,995**	1**	,995**	,999**	,991**

Üçüncü Alt Probleme İlişkin Toplu Bulgular

Örneklem	10 Maddelik Test						20 Maddelik Test						30 Maddelik Test											
	1PL Model		2PL Model			3PL Model			1PL Model		2PL Model			3PL Model			1PL Model		2PL Model			3PL Model		
	b	a	b	a	b	c	b	a	b	a	b	c	b	a	b	a	b	a	b	c				
150	0,231	0,181	0,179	0,247	0,195	0,349	0,216	0,170	0,217	0,228	0,162	0,305	0,211	0,176	0,163	0,226	0,173	0,312						
250	0,177	0,146	0,144	0,204	0,147	0,284	0,165	0,137	0,121	0,180	0,125	0,237	0,161	0,141	0,130	0,185	0,128	0,248						
350	0,152	0,120	0,122	0,154	0,125	0,243	0,139	0,116	0,108	0,151	0,111	0,208	0,137	0,118	0,112	0,149	0,110	0,215						
500	0,129	0,101	0,104	0,133	0,105	0,201	0,117	0,097	0,088	0,123	0,089	0,179	0,114	0,099	0,093	0,124	0,094	0,177						
750	0,103	0,082	0,085	0,109	0,087	0,157	0,096	0,081	0,074	0,103	0,076	0,139	0,094	0,081	0,074	0,101	0,074	0,142						
1000	0,090	0,072	0,071	0,094	0,077	0,131	0,083	0,068	0,059	0,083	0,064	0,117	0,081	0,070	0,063	0,084	0,065	0,123						
2000	0,064	0,052	0,054	0,067	0,054	0,096	0,059	0,049	0,044	0,060	0,045	0,084	0,058	0,050	0,046	0,064	0,045	0,087						
3000	0,052	0,042	0,042	0,055	0,045	0,077	0,042	0,040	0,035	0,050	0,038	0,068	0,037	0,041	0,037	0,050	0,037	0,070						
5000	0,041	0,032	0,034	0,041	0,035	0,061	0,038	0,031	0,027	0,038	0,029	0,053	0,037	0,031	0,029	0,038	0,029	0,056						