

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI

ÖLÇME EŞDEĞERLİĞİNİ İNCELEMEDE KULLANILAN YÖNTEMLERİN
FARKLI KOŞULLAR ALTINDA
İSTATİSTİKSEL GÜÇ ORANLARI AÇISINDAN KARŞILAŞTIRILMASI

DOKTORA TEZİ

ALPEREN YANDI

Ankara, Kasım, 2017

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI

ÖLÇME EŞDEĞERLİĞİNİ İNCELEMEDE KULLANILAN YÖNTEMLERİN
FARKLI KOŞULLAR ALTINDA
İSTATİSTİKSEL GÜÇ ORANLARI AÇISINDAN KARŞILAŞTIRILMASI

DOKTORA TEZİ

ALPEREN YANDI

TEZ DANIŞMANI: PROF. DR. NİZAMETTİN KOÇ

Ankara, Kasım, 2017

ONAY SAYFASI

Eğitim Bilimleri Enstitüsü Müdürlüğü'ne

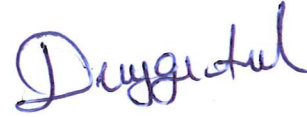
Alperen YANDI'nın hazırladığı “Ölçme Eşdeğerliğini İncelemede Kullanılan Yöntemlerin Farklı Koşullar Altında İstatistiksel Güç Oranları Açısından Karşılaştırılması” başlıklı bu çalışma, jürimiz tarafından Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Programı'nda Doktora Tezi Olarak Kabul Edilmiştir.

İmza

Başkan: Prof. Dr. Nizamettin KOÇ (Danışman)



Üye: Prof. Dr. Duygu ANIL



Üye: Prof. Dr. Zekeriya NARTGÜN



Üye: Doç. Dr. İsmail KARAKAYA



Üye: Yrd. Doç. Dr. Ergül DEMİR



ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 15/11/2017 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca/...../20..... tarihinde kabul edilmiştir.

Prof. Dr. İsmail GÜVEN
Eğitim Bilimleri Enstitüsü Müdürü

T Z B LDİRİMİ

Tezin i indeki b t n bilgilerin etik davranıř ve akademik kurallar  er evesinde elde edilerek sunulduėunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu  alıřmada bana ait olmayan her t rl  ifade ve bilginin kaynaėına eksiksiz atıf yapıldıėını bildiririm.



Alperen YANDI

ÖZET

ÖLÇME EŞDEĞERLİĞİNİ İNCELEMEDE KULLANILAN YÖNTEMLERİN FARKLI KOŞULLAR ALTINDA İSTATİSTİKSEL GÜÇ ORANLARI AÇISINDAN KARŞILAŞTIRILMASI

Yandı, Alperen

Doktora, Ölçme ve Değerlendirme Anabilim Dalı

Tez danışmanı: Prof. Dr. Nizamettin Koç

Kasım 2017, xii + 91 Sayfa

Bu araştırmanın amacı, farklı koşullar altında çoklu puanlanan verilerde kullanılabilen ölçme eşdeğerliği belirleme yöntemlerinin karşılaştırılmasıdır. Bu amaca yönelik olarak gerçek uygulamalarda karşılaşılabilecek farklı koşullar için en uygun ölçme eşdeğerliği belirleme yönteminin hangisi olabileceğine dair kanıt aranmıştır. Çalışma kapsamında çoklu kategoride puanlanan, farklı örneklem büyüklüklerine, farklı test uzunluklarına ve farklı oranlarda değişen madde fonksiyonuna sahip madde içerecek şekilde üretilmiş yapay verilerde, çoklu grup doğrulayıcı faktör analizi, madde parametrelerinin karşılaştırılması ve olabilirlik oran testi yöntemleri kullanılarak ölçme eşdeğerliği ile ilgili elde edilen sonuçlar karşılaştırılmıştır. Elde edilen bulgular incelenerek farklı koşullarda hangi ölçme eşdeğerliği test etme yönteminin daha yüksek güç düzeyine sahip olduğu belirlenmiştir. Araştırma, bu yönleriyle kuramsal çalışmalara katkı sunduğundan temel bir araştırmadır. Örneklem büyüklüğü düzeyleri gerçek veri kullanılan araştırmalarda karşılaşımla sıklığı yüksek olduğu düşünüldüğünden 250/250, 500/500, 1000/1000 olacak şekilde veri üretimi gerçekleştirilmiştir. Değişen madde fonksiyonu içeren madde oranı değişkeni ise her bir koşul için sırasıyla %10, %20, %30'a kadar değişen madde fonksiyonu içeren madde olacak şekilde veri setleri oluşturulmuştur. Tekrarlama sayısı 100 olarak alınmıştır. Tepki kategori sayısı, madde tepki kuramı modeli, değişen madde fonksiyonu türü, değişen madde fonksiyonu büyüklüğü, bireylerin yetenek dağılımları değişkenleri ise tüm veri setleri aynı olacak şekilde belli düzeylerde sabit olarak alınmıştır. Örneklem büyüklüğü koşulunun 1000/1000 olduğu koşullarda ise yöntemlerin istatistiksel güç oranlarının madde sayısı değişkenine göre değişkenlik gösterdiği saptanmıştır. Bu sonuçlara ek olarak üretilen yapay veriler için, tüm yöntemlere ait istatistiksel güç oranlarının değişen madde fonksiyonu içeren madde oranındaki artıştan olumsuz etkilendiği sonucuna ulaşılmıştır. Veri

setlerindeki deęiřen madde fonksiyonu ieren madde sayısı arttıa yntemlerin istatistiksel g oranlarında azalma olmaktadır. Ayrıca elde edilen sonulara gre, 27 farklı kořul altında  farklı yntem iin toplam 81 ikili karřılařtırmanın 38’inde olabilirlik oran testinin; 24’nde oklu grup doęrulayıcı faktr analizi ynteminin; 9’unda ise madde parametrelerini karřılařtırma ynteminin dięer iki yntemden manidar řekilde daha yksek g oranları vermiřtir. 10 karřılařtırmada ise  yntem arasında manidar bir farklılık olmadığı sonucu elde edilmiřtir. Yapılan ikili karřılařtırmalarda olabilirlik oran testi ynteminin odak ve referans grubu rneklem byklęnn 250 olarak alındığı dięer iki ynteme gre daha yksek g oranları verdięi belirlenmiřtir. Bunun yanı sıra oklu grup doęrulayıcı faktr analizi ynteminin daha yksek g oranları verdięi kořulların 40 ve 60 maddelik veri setleri olduęu ortaya konmuřtur. Madde parametrelerini karřılařtırma ynteminin ise dięer iki yntemden daha yksek g oranı verdięi kořulların tamamı 20 maddelik veri setleri iin yapılmıř analizlerin sonularında elde edilmiřtir. 1000/1000 rneklem byklę kořulunda, 20 maddelik veri setlerinde madde parametrelerini karřılařtırma; 40 maddelik veri setlerinde olabilirlik oran testi; 60 maddelik veri setlerinde ise oklu grup doęrulayıcı faktr analizi ynteminin tercih edilmesi gerektięi ıkarımı yapılabilir. Elde edilen sonular doęrultusunda, odak ve referans grubu rneklem byklęlerinin 250 olarak retildięi, 60 madde ve %10 deęiřen madde fonksiyonu ieren madde oranı olan ve odak ve referans grupları rneklem byklęlerinin 500 olduęu, 60 maddelik kořullar dıřında olabilirlik oran testi ynteminin kullanılması nerilebilir.

Anahtar szckler: lme Eřdeęerlięi, Deęiřen Madde Fonksiyonu, oklu Grup Doęrulayıcı Faktr Analizi, Olabilirlik Oran Testi, Madde Parametrelerini Karřılařtırma Yntemi

ABSTRACT**COMPARISON OF THE METHODS OF EXAMINING
MEASUREMENT EQUIVALENCE UNDER DIFFERENT
CONDITIONS IN TERMS OF STATISTICAL POWER RATIOS**

Yandı, Alperen

Ph.D., Department of Measurement and Evaluation

Advisor: Prof. Dr. Nizamettin Koç

November 2017, xii + 91 Pages

The purpose of this study is to compare the measurement equivalence methods available for multiple scoring data under different conditions. For this purpose, evidence has been sought to identify which method is the most appropriate for determining measurement equivalence for different conditions that may be encountered in real applications. In the study, the results of measurement equivalence obtained by using multi group confirmatory factor analysis, comparison of item parameters method and likelihood ratio test from the artificial datasets generated with different sample sizes, different number of items, and different ratio of the items with differential item function were compared. The findings were examined and which measurement equivalence test method had higher power level in different conditions was determined. The research is a basic research since it contributed to the theory by these points. Sample size levels were generated as 250/250, 500/500, 1000/1000 because of considering these levels to be more frequently encountered in researches using real data. Besides, 20, 40 and 60 items were used for test length variable. Datasets containing items with differential item function were set at 10%, 20% and 30%, respectively. The number of replications was taken as 100. The response category number, the item response theory model, the type of differential item function, the differential item function size, and the individual's skill variance variables are fixed at a certain level as all data sets are the same. It has been found that the statistical power ratios of the methods vary according to the test length. In addition to these results, for artificial data generated, statistical power ratios for all methods were adversely affected by the increase in the ratio of item with differential item function. As the number of items with differential item function increases in data sets, the statistical power ratios decrease. Furthermore, according to the results, 38, 24 and 9 of the total 81 binary comparisons for three

different methods under 27 different conditions, the likelihood ratio test, multi group confirmatory factor analysis and comparison of item parameters method yielded higher power ratios significantly than the other two methods respectively. In ten comparisons, there was no significant difference between the three methods. In the binary comparisons, the likelihood ratio test had higher statistical power ratios than the other two methods when sample size of focal and reference group were 250. Moreover, it is noteworthy that multi group confirmatory factor analysis has higher power ratios for data sets had 40 and 60 items. All higher statistical power ratio results for comparison of item parameter method obtained for data sets had 20 items. When the sample size is 1000/1000, it can be said that comparison of item parameters method for test length 20-item data sets; likelihood ratio test for test length 40-item data sets; for multi group confirmatory factor analysis test length 60-item condition were preferred. According to results, it is suggested to use the likelihood ratio test in all conditions except for the conditions that the sample size of the focal and reference group is 250 and the test length 60 items - 10% items with differential item function and the test length 60-item condition that the sample size of the focal and reference groups is 500.

Keywords: Measurement Equivalence, Differential Item Function, Multi Group Confirmatory Factor Analysis, Likelihood Ratio Test, Comparison of Item Parameters.

ÖNSÖZ

Sosyal bilimlerde yapılan araştırmaların sonuçları doğrultusunda önemli kararlar alınmakta ve düzenlemeler gerçekleştirilmektedir. Karar alınması ve düzenlemelerin belirlenmesi sürecinde bireysel farklılıkların incelenmesi gerekmektedir. Alınan kararların ve yapılan düzenlemelerin etkili olması, psikolojik yapılar bakımından farklı düzeydeki bireylerin özelliklerinin göz önünde bulundurulması ile sağlanmaktadır. Bu nedenlere sosyal bilimlerde yapılan çalışmaların birçoğunda bireysel farklılıkların ölçülmesi durumları ile sıklıkla karşılaşilmektedir. Bireysel farklılıkların belirlenmesi için ölçme araçları kullanılmaktadır. Bu ölçme araçlarının elde edilme sürecinde geliştirme ve uyarlama çalışmaları yapılmaktadır.

Ölçme aracı geliştirme ve uyarlama süreçlerinde gerçekleştirilmesi önemli birtakım psikometrik incelemeler mevcuttur. Bu psikometrik incelemelerde ele alınan en önemli özelliklerden ikisi güvenirlik ve geçerliktir. Bir ölçme aracının güvenirlik ve geçerlik incelemelerinde farklı analizler gerçekleştirilebilir. İncelenen ölçme aracının yapısına bağlı olarak farklı türde güvenirlik ve geçerlik başlıklarına ilişkin analizler yapılabilir. Bunun yanı sıra yapılan güvenirlik ve geçerlik analizlerinin çalışmanın hangi aşamasında gerçekleştirildiği de önem arz etmektedir. Bir ölçme aracı için güvenirlik ve geçerlik analizlerinin tümünün nihai form ortaya konmadan önce gerçekleştirilmesi gerekmektedir. Nihai form uygulaması yapılan bir ölçme aracı için yapılan güvenirlik ve geçerlik analizleri sonrasında geriye dönük bir adım atmak ve eksiklikleri gidermek mümkün olmamaktadır. Bu nedenle elde edilen sonuçlar ile ilgili soru işaretleri ortaya çıkabilmektedir.

Bireysel farklılıkların incelendiği çalışmalarda kullanılan bir ölçme aracı özelinde ise yapılması gereken en önemli geçerlik incelemelerinden biri ise ölçme eşdeğerliği analizleridir. Ölçme eşdeğerliği en genel haliyle bir ölçme aracının farklı alt gruplardaki bireyler tarafından aynı şekilde anlaşılması olarak tanımlanabilir. Ölçme eşdeğerliği analizleri farklı kuramlara dayalı yöntemlerle gerçekleştirilebilir. Sosyal bilimlerde yapılan araştırmalarda ölçme eşdeğerliği analizlerinin gerçekleştirilmesindeki eksiklikler ve diğer geçerlik incelemelerinde yapılan zamanlama hatalarına değinmek ve diğer araştırmacılara farklı ölçme eşdeğerliği belirleme yöntemlerinin gerçek uygulama koşullarına yakın koşullarda hangi düzeyde güç oranları verdiğine ilişkin bilgi sağlamak adına bu çalışma gerçekleştirilmiştir.

Araştırma süresince desteğini esirgemeyen danışmanım Sayın Hocam Prof. Dr. Nizamettin Koç'a, bu süreçte önerileri ile bana yol gösteren tez izleme komitesi üyeleri Hocalarım Sayın Prof. Dr. Nükhet Demirtaşlı'ya ve Doç. Dr. İsmail Karakaya'ya;

Öneri ve eleştirileriyle araştırmama katkı sağlayan jüri üyeleri, değerli hocalarım Prof. Dr. Duygu Anıl'a, Prof. Dr. Zekeriya Nartgün'e ve Yrd. Doç. Dr. Ergül Demir'e;

Yapmış olduğum bu araştırma boyunca verilerin elde edilmesi ve analiz sürecinde verdiği bilgilerle beni aydınlatan Yrd. Doç. Dr. Önder Sünbül'e;

Ayrıca araştırma sürecinde yardım ve destekleri ile katkı sunan arkadaşlarım Merve Şahin Kürşad'a, Tuğçe Daşöz'e, Emine Tunç'a, Taşkın Soysal'a, Haluk Güngör'e, kuzenim Necmettin Kaplan'a ve adını sayamadığım diğer tüm arkadaşlarıma;

Son olarak süreçte karşılaştığım her durumda anlayışı ve yardımları ile yanımda olan eşim Betül Yandı'ya; bugünlere gelmememin mimarları anneme, babama ve kardeşime;

teşekkürlerimi sunarım.

Alperen YANDI

İÇİNDEKİLER

ONAY SAYFASI.....	ii
TEZ BİLDİRİMİ	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
ÇİZELGELER DİZİNİ	xi
GRAFİKLER DİZİNİ	xii
ŞEKİLLER DİZİNİ.....	xii
1. GİRİŞ	1
1.1. Problem Durumu	1
1.2. Amaç.....	22
1.3. Önem	23
1.4. Sınırlılıklar.....	24
1.5. Kısaltmalar.....	24
2. YÖNTEM	26
2.1. Araştırma Modeli.....	26
2.2. Veriler	26
2.3. Verilerin Analizi	33
3. BULGULAR VE YORUMLAR.....	43
3.1. Çoklu grup doğrulayıcı faktör analizi yöntemi istatistiksel güç oranlarına ilişkin bulgular.....	43
3.2. Olabilirlik oran testi yöntemi istatistiksel güç oranlarına ilişkin bulgular	51
3.3. Madde parametrelerini karşılaştırma yöntemi istatistiksel güç oranlarına ilişkin bulgular.....	59
3.4. Ortak etkiye ilişkin bulgular	66
3.5. Gerçek veri üzerinde gerçekleştirilen analizlere ilişkin bulgular	71
4. SONUÇ VE ÖNERİLER	73
4.1. Sonuçlar	73
4.2. Öneriler.....	76
5. KAYNAKLAR	80
6. EKLER.....	88

ÇİZELGELER DİZİNİ

Çizelge

1.Yapay Veri Üretim Koşulları.	27
2. Veri Setleri İçin Oluşturulan Kod Numaraları.	31
3. Veri Dosyaları İçin Hesaplanan Betimsel İstatistiklere Ait Sonuçlar.	32
4. Ölçme Eşdeğerliği Test Etmede Aşamalı Test Etme Yolu Adımları.	34
5.ÇGDFA Yöntemi İle Elde Edilen İstatistiksel Güç Oranları.	44
6.OOT Yöntemi İle Elde Edilen İstatistiksel Güç Oranları.....	51
7.MPK Yöntemi İle Elde Edilen İstatistiksel Güç Oranları.	60
8.Örneklem Büyüklüğü, Madde Sayısı ve DMF İçeren Madde Sayısı Koşulları İçin Kullanılan Yöntemlerden Elde Edilen İstatistiksel Güç Oranları.....	67
9.Farklı Koşullar Altında Yöntemler İçin Elde Edilen İstatistiksel Güç Oranlarının Karşılaştırıldığı Anova Testi Sonuçları.....	69
10.ST018 Kodlu - Genel Değer Alt Ölçeği Verileri Kullanılarak Gerçekleştirilen Analiz Sonuçları.	71

GRAFİKLER DİZİNİ

Grafik

1. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (ÇGDFA)	46
2. Madde sayısı değişkenine göre istatistiksel güç oranları (ÇGDFA)	48
3. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (ÇGDFA).....	50
4. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (OOT)	53
5. Madde sayısı değişkeni için istatistiksel güç oranları (OOT)	56
6. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (OOT).....	58
7. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (MPK).....	62
8. Madde sayısı değişkeni için istatistiksel güç oranları (MPK).....	64
9. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (MPK).....	66
10. Değişimlenen koşulların yöntemler üzerindeki ortak etkileri	68

ŞEKİLLER DİZİNİ

Şekil

1. Yapısal Değişmezlik Modeli	35
2. Metrik değişmezlik modeli	36
3. Ölçek değişmezliği modeli	37
4. Katı değişmezlik modeli	38

1. GİRİŞ

Bu bölümde problem, amaç, önem, sınırlılıklar ve konu ile ilgili tanımlara yer verilmiştir.

1.1. Problem Durumu

Bu başlık altında gerçekleştirilen bu araştırmanın ortaya çıkış sürecine ilişkin gerekçeler ele alınmıştır. Bu doğrultuda gerekçeler farklı alt başlıklar şeklinde verilmiştir. Aynı ayrı verilen alt başlıklar sentezlenerek problem durumu ortaya konmaya çalışılmıştır. Bu süreç iki ana başlık altında sunulmuştur. Bu başlıklar, (i) bireysel farklılıkların ölçülmesi ve (ii) geçerlik incelemeleridir.

Bireysel farklılıkların ölçülmesi

Sosyal bilimlerde yapılan araştırmalar bireylerin katılımı ile ulaşılan veriler üzerinden yürütülmektedir. Bireylerin temel veri kaynağı olmasındaki ana etken her bireyin diğerinden farklı özelliklere sahip olması, başka bir ifadeyle her bireyin eşsiz olmasıdır. Bireyler, yaşadıkları çevre, içinde bulundukları kültür yapılarına bağlı olarak uyarıcılara verdikleri tepkiler ve sahip oldukları psikolojik özellikler bakımından farklılık gösterirler. Bu farklılıklar ve bunların ölçülmesi, kuramların ortaya konması, geliştirilmesi, çeşitli alanlarda yapılacak olan düzenleme ve değişikliklere temel oluşturmaktadır. Nitekim bilim tarihinde önemli gelişmelere yol açan çalışmaları gerçekleştiren Darwin, Galton gibi bilim insanları, çalışmalarında bireysel farklılıkların belirlenmesinin önemli olduğunu vurgulamıştır.

Bireysel farklılıkların belirlenme sürecinde elde edilen sözlü veya yazılı veriler üzerinde yapılan analizlerin sonuçlarına dayalı olarak ortaya konan yorumlarla yeni bakış açıları geliştirilmektedir. Yapılan bu yorumların ve geliştirilen bakış açılarının sonucunda, yeni kuramlar ortaya konularak, kavramlar yeniden tanımlanmaktadır. Kavramlara getirilen yeni tanımlamalar ve geliştirilen kuramlar, çeşitli alanlarda düzenleme ve değişiklikleri de zorunlu kılmaktadır. Bu şekilde bir değişim sürecine temel oluşturan bireysel farklılıkların ortaya çıkarılmasında ölçme işlemi çok önemli bir yere sahiptir. Galton (1884) yapmış olduğu ölçme uygulamaları ile çeşitli değişkenler açısından bireysel farklılıkların ortaya koymuş ve ölçme işleminin bu süreçte önemli bir rolü olduğunu belirtmiştir.

Eğitim bilimleri alanında da, ölçme uygulamaları esas alınarak önemli kararlar alınmakta ve köklü düzenlemeler gerçekleştirilmektedir. Bu değişim sürecinde bireylerin psikolojik yapıları sahip oluş düzeyleri ve düzey farklılıkları sıklıkla incelenmektedir. Eğitim sürecinde bireylerde geliştirilmesi planlanan özellikler üzerinde psikolojik yapıların etkisi olduğundan, bu yapılarla ilgili ölçme uygulamaları önem kazanmaktadır. Psikolojik yapılar doğrudan ölçülmesi mümkün olmayan örtük değişkenlerdir. Bireylerin psikolojik yapıları

sahip oluş düzeyleri gözlenen değişkenlere (ölçme aracında bulunan maddeler) verdikleri tepkiler doğrultusunda ortaya konmaktadır. Psikolojik özelliklerin ölçülmesi için farklı sayıda maddelerin bulunduğu tek veya çok boyutlu ölçme araçları geliştirilmiştir. Psikolojik ölçme aracı, bireyleri yalnızca ölçülmek istenen özellik bakımından uyaran ve özelliği ölçülen bireyde sadece söz konusu özelliğin bulunışluluk düzeyini yansıtacak tepki düzeylerinin yer aldığı maddeler takımı olarak tanımlanabilir (Erkuş, 2012).

Araştırmacılar kullandıkları ölçme araçlarını, geliştirme veya uyarlama çalışmaları ile elde etmektedirler. Ölçek geliştirme çalışmalarında, hedef grubu temsil eden daha küçük bir grup üzerinde, ardışık şekilde ilerleyen bir dizi aşama izlenerek uygun ölçme araçları geliştirilmeye çalışılmaktadır. Ölçme aracı uyarlama çalışmalarında ise, belli kültüre ve dile sahip bir grup üzerinde geliştirilen ölçme aracının, farklı bir kültür ve dile uyarlanması amaçlanır. Bu çalışmalar sonucunda geliştirilen veya uyarlanan ölçme aracının psikometrik özellikler bakımından belli bir düzeye ulaşmış olması gerekmektedir. Bu kapsamda bir ölçme aracında bulunması gereken psikometrik özelliklerden en önemlisi olan geçerlik ile ilgili incelemeler ön plana çıkmaktadır.

Geçerlik incelemeleri

Bir ölçme aracının psikometrik özellikler bakımından nitelikli olması, elde edilen sonuçların doğruluğu açısından önem arz etmektedir. Ölçme araçlarında aranan psikometrik özellikler geçerlik, güvenirlik ve kullanılabilirlik olarak sıralanabilir. Geçerlik, bu psikometrik özelliklerden en önemlisidir. Bir ölçme aracının geçerli olarak nitelendirilebilmesi için, ölçülecek özellik veya özellikleri diğer özelliklerle karıştırmadan tam ve doğru olarak ölçebilmesi gerekmektedir (Tekin, 2000). Geçerlik kavramı ile ilgili yapılan çalışmalarda araştırmacılar, ölçme araçlarının kullanımının ve bu araçlarla elde edilen puanlara dayalı olarak verilen kararların etik, sosyal ve deneysel olarak doğrulanabilirliği üzerine yoğunlaşmaktadır (Borsboom, Mellenbergh ve van Heerden, 2004). Messick (1995) geçerlik kavramını uygulamalarla elde edilen puanlara dayalı olarak yapılan yorumların, bu yorumlara göre ortaya çıkan sonuçların deneysel ve kuramsal açıdan desteklenmesi şeklinde ele almaktadır.

Geçerlik kavramı temelde ölçme sonuçlarının doğruluğu ile ilgili kanıt toplama süreci olarak nitelendirilebilir. Geçerlik türleri ise ölçme aracının hangi özelliği ile ilgili nasıl kanıt toplanabileceğine işaret etmektedir. Ölçme aracının yapısına uygun olarak araştırmacılar farklı geçerlik türleri ile ilgili incelemeler yapmaktadır. Araştırma süreçlerinde kullanılan ölçme araçlarının geçerlik inceleme süreçlerinde dikkat çekilmesi gereken önemli noktalar mevcuttur. Bu noktalardan ilki incelemelerin araştırmanın hangi aşamasında yapıldığıdır. Araştırmalarda nihai uygulamalar yapılmadan önce kullanılması planlanan ölçme aracına ilişkin geçerlik

incelemelerinin gerçekleştirilmesi gerekmektedir. Bunun nedeni olarak nihai form uygulaması gerçekleştirilmiş bir ölçme aracı ile ilgili olarak yapılan geçerlik çalışması sonucunda ortaya çıkabilecek problemlerin giderilmesine ilişkin olumsuzluklar gösterilebilir. Nihai form uygulamasından sonra geçerlik konusunda eksiklikleri olduğu belirlenen bir ölçme aracına müdahale etmek mümkün olmamaktadır. Aynı zamanda elde edilmiş olan bulguların doğruluğu ile ilgili kuşklar ortaya çıkmaktadır. Geçerlik incelemeleri ile ilgili üzerinde durulması gereken bir diğer nokta ise inceleme sürecinde gerçekleştirilen analizlerdir. Yapılan birçok çalışmada farklı örneklerde kullanılan ölçme araçları için faktör analizleri gerçekleştirildiği ve geçerlik kanıtı olarak bu analiz sonuçlarının sunulduğu görülmektedir. Ancak bir ölçme aracının geçerliğini yalnızca faktör analizi sonuçlarına dayandırmak doğru bir yaklaşım değildir. Araştırmaların amaçları doğrultusunda kullanılan ölçme araçları ile ilgili olarak farklı geçerlik testleri yapılmalıdır. Bu doğrultuda geçerlik incelemelerinde kullanılacak analizlerin doğru şekilde seçilmesi ve yapılması önemlidir. Geçerlik incelemelerine ilişkin soru işaretleri içeren bir çalışmanın bulgularının işaret edeceği hususlarda da belli şüpheler ortaya çıkabilir. Bu nedenle araştırmacıların bu süreçte hassas davranmaları gerekmektedir.

Geçerlik kavramı ve geçerlik türleri ile ilgili yapılan tanımların ortak noktalarında ölçülmek istenen özelliğin, ölçme aracında ne derece temsil edildiği ve bu araçla farklı özelliklerle karıştırılmadan ne kadar doğru ölçülebildiği soruları bulunmaktadır. Ölçme aracının ölçmek istediği özelliği ölçebilme derecesi, sistematik hata kaynaklarından olumsuz etkilenmektedir. Sistematik hata kaynakları ölçülmek istenen özelliğe başka özelliklerin karışmasına neden olduğundan dolayı geçerliği olumsuz etkilemektedir. Araştırmacılar geçerlik inceleme sürecinde olası sistematik hata kaynaklarını üzerinde durmalıdır. Yapılan çalışma sonuçlarının doğruluğu için bu kaynakların kontrol edildiğine ilişkin kanıtlar sunulmalıdır.

Yapılan bu araştırmanın gerekçeleri bu iki başlık altında özetle açıklanmaya çalışılmıştır. Bu başlıklar altında önemli adımların temelini oluşturan ve birçok çalışmanın temel amacı olan bireysel farklılıkların ölçülme sürecine ve geçerlik incelemelerine ilişkin bilgiler verilmiştir. Bu kısımdan sonraki başlıklarda bu iki konu birlikte ele alınarak problem durumunun ortaya çıkış sürecindeki ayrıntılara değinilmiştir.

Farklı alt gruplarda yer alan bireylerin farklılıkların belirlenmesi ve geçerlik incelemeleri

Sosyal bilimler alanında yapılan çalışmaların önemli bir kısmında bireysel farklılıklar belli bir değişkene göre oluşturulan gruplara üye olma bağlamında ortaya konmaya çalışılmaktadır. Araştırmacılar belli bir değişken bakımından farklı alt gruplarda yer alan bireyleri karşılaştırmakta ve bu bireylerin ölçülen yapılarla sahip oluş düzeylerindeki farklılaşmalarını incelemektedir. Daha önceki başlıklarda da vurgulandığı gibi bu

araştırmaların sonuçları temel alınarak yapılan çıkarımlar doğrultusunda bireylerle ilgili kararlar verilmekte, eğitim öğretim süreçleri ile ilgili değişiklikler yapılmakta ve program değerlendirmeleri gerçekleştirilmektedir. Bireyler ve gruplar arasındaki farklılıklara dayalı olarak önemli kararlar alınması, bu farklılıkların en doğru şekilde ortaya çıkarılmasını gerektirmektedir. Bu durum diğer tüm çalışmalarda olduğu gibi geçerlik incelemelerin eksiksiz ve doğru şekilde yapılmasına işaret etmektedir. Bunun nedeni geçerlik incelemelerinin doğru şekilde yapılmasının incelenen gruplarla ilgili alınan kararları doğrudan etkiliyor olmasıdır.

Geçerlik incelemelerinde araştırmacılar sistematik hata kaynaklarına odaklanmaktadır. Farklı alt gruplarda yer alan bireylerin katılımı ile gerçekleştirilen çalışmalarda karşılaşılabilecek en önemli sistematik hata kaynaklarından birisi yanlılıktır. Messick (1995) geçerliği etkileyen en önemli sistematik hata kaynaklarından birinin “yanlılık” olduğunu belirtmektedir. Yanlılık, ölçme yapılan uygulama durumları, bireylerin toplumsal rolleri, uyarılma çalışmalarındaki çeviri sorunları, psikolojik yapıların kültürler arasında farklılaşması gibi nedenlerle ortaya çıkabilmektedir. Angoff (1993), yanlılığı farklı gruplarda yer alan ve bir örtük özellik bakımından aynı düzeyde olan bireylerin, söz konusu örtük özelliği ölçmeyi amaçlayan ölçekteki bir maddeye aynı tepkiyi verme olasılıklarının farklılaşması durumu olarak tanımlamaktadır. Öte yandan diğer bir tanıma göre yanlılık bir araştırma kapsamında bulunan bireyler için elde edilen puanların bulundukları alt gruba bağlı olarak sistematik hata içermesidir (Camilli ve Shepard, 1994; Zumbo 1999).

Van de Vijver (1998) ölçme uygulamalarında ortaya çıkabilecek yanlılık türlerini yapı yanlılığı, yöntem yanlılığı ve madde yanlılığı olmak üzere üç ana başlık altında toplamıştır. Yapı yanlılığı, yapılan çalışmalar kapsamında konu edinilen örtük özelliğin, ölçme yapılan gruplardaki bireyler için farklı anlamlar ifade etmesidir. Yöntem yanlılığı, örneklem özelliklerinden, uygulama durumlarından ve ölçme aracından kaynaklanan yanlılıklardır (Van de Vijver, 1998). Madde yanlılığı ise çalışmalarda kullanılan ölçme araçlarında madde düzeyinde, bireyler arasında örtük özellik dışındaki farklı bir değişken kaynaklı farklılıkların ortaya çıkması, bir başka ifadeyle ölçme sonuçlarına sistematik hataların karışmasıdır.

Yanlılık durumunun ortaya çıkması ilk olarak geçerliği daha sonrasında yapılan ölçme işlemlerine dayalı olarak verilen kararların doğruluğunu olumsuz etkilemektedir (Holland ve Wainer, 1993). Bu nedenle farklı gruplar üzerinde yapılan araştırmalarda yanlılık incelemelerinin yapılması önem kazanmaktadır. Yapı ve yöntem yanlılığı için araştırmacılar, uygulamalar öncesinde kuramsal olarak çalışmalar yapabilir. Öte yandan madde düzeyindeki yanlılıkların belirlenme sürecinde atılacak ilk adım ele alınan maddenin cevap fonksiyonlarının incelenmesidir. Madde yanlılığının belirleme süreci için yapılan incelemelerde ölçme eş

değerliği kavramı karşımıza çıkmaktadır. Ölçme eş değeriği en genel tanımıyla gözlenen değışkenler ile örtük değışkenler arasındaki ilişkinin inceleme yapılan gruplar arasında aynı olması anlamına gelmektedir (Drasgow ve Kanfer, 1985; Widaman ve Reise, 1997). Bir başka deyişle ölçme eşdeğeriği, bireylerin ölçülen yapı ile ilgili düzeylerinin gruptan bağımsız olarak ortaya çıkması ile ilgili bir durumdur.

Ölçme eşdeğeriği

Alanyazında ölçme eşdeğeriği ile ilgili farklı tanımlamalar bulunmaktadır. Byrne ve Watkins (2003) 'e göre ölçme eşdeğeriği, ölçme aracında bulunan maddelerin uygulanan alt gruplarda aynı şekilde algılanma ve yorumlanmasıdır. Flowers, Raju ve Oshima (2002) ise ölçme eşdeğeriğini, bir örtük özellik bakımından aynı düzeydeki fakat farklı gruplardaki bireylerin, bu özellikle ilgili bir ölçme aracındaki maddelerden ve alt boyutlardan aynı gözlenen puanı almaları olarak açıklamaktadırlar. Bir başka tanıma göre ise ölçme eşdeğeriği, iki ya da daha fazla grup üzerinde uygulanan ölçme aracı ile ölçülmek istenen bir yapının ölçme parametrelerinin matematiksel olarak birbirine denk olmasıdır (Little, 1997). Bir yapının parametrelerinin matematiksel olarak birbirine denk olması, farklı gruplar üzerinde uygulanan bir ölçeğin maddelerine ait faktör yüklerinin ve hata varyanslarının, ölçekte bulunan faktörler arası korelasyonların aynı olması anlamına gelmektedir (Byrne, 2008; Jöreskog ve Sörbom, 1993).

Yapılan çalışmalarda farklı alt grupların karşılaştırılabilmesi için ölçme eşdeğeriğinin mutlaka sağlanmış olması gerektiği belirtilmektedir (Reise, Widaman ve Pugh, 1993; Vandenberg ve Lance, 2000). Ancak grup karşılaştırılmasını amaçlayan çalışmalarda ölçme eşdeğeriği incelemelerine gereken önem ve özen gösterilmemektedir. Bu durum yapılan araştırmalarda elde edilen sonuçların anlamlılığını ve daha da önemlisi bu sonuçlara dayalı alınan kararların doğruluğunu olumsuz etkilemektedir. Steenkamp ve Baumgartner (1998) ölçme eşdeğeriğinin kullanılmamasının sebeplerini şu şekilde sıralamışlardır:

- Ölçme eşdeğeriğinin aşamaları hakkında birçok farklı görüşün bulunması,
- Ölçme eşdeğeriği aşamaları ile ilgili olarak bilim insanları arasında görüş birliğinin sağlanamamış olması,
- Araştırmacıların örtük ve gözlenen değışkenlere yönelik ölçme modellerine ve bunların test edilmesine olan uzaklığı,
- Ölçme eşdeğeriği aşamalarının test edilmesinde kullanılan yöntemler ile ilgili karmaşıklıklar ve bu aşamaların test edilmesi ile ilgili örneklerin az olması,

- K lt rlerarası karşılařtırmaların anlamlı olabilmesi i in hangi  l  mlere y nelik eřitliĐin saĐlanması gerektiĐi ile ilgili bilgi eksikliĐi,
- K lt rlerarası  l me eřdeĐerliĐi i in  l  tleri g steren bir rehberin olmamasıdır.

Yukarıda  zetlenen t m bilgiler ıřıĐında sosyal bilimlerde farklı alt grup karşılařtırmaları yapmayı ama layan  alıřmaların ge erlik inceleme s recinde  l me eř deĐerliĐi analizlerinin  nemli bir par a olduĐu sonucuna ulařılabilir. Ancak bu amaca sahip bir ok  alıřma olmasına karřın  l me eř deĐerliĐi incelemelerinin  eřitli nedenlerle yapılmadıĐı g r lmektedir. Bu t r  alıřmalarda ge erlik incelemelerinde bu řekilde bir eksikliĐin olması bir ok olumsuz durumu beraberinde getirebilir.  l me eř deĐerliĐi analizlerine gereken  nemin verilmemesinin nedenlerinin bařında y ntemlerin iřleyiř s recine iliřkin bilgi eksiklikleri ve karmařıklıklar gelmektedir. Bu nedenle farklı  l me eřdeĐerliĐi inceleme y ntemlerine iliřkin  alıřmalar yapılarak bu hususta arařtırmacılara bilgi saĐlanmalıdır. Farklı  l me eřdeĐerliĐi inceleme y ntemleri kullanılarak yapılan  alıřmalarla arařtırmacılara bu y ntemlerle ilgili olarak bilgi sunulabilir. Y ntemlerin iřleyiř s recine ve farklı kořullar altında kullanıldıĐında ortaya  ıkan durumlarla ilgili sunulacak bilgiler, ge erlik inceleme s recindeki bu eksikliĐin giderilmesi hususunda katkı saĐlayabilir.

 l me eř deĐerliĐi belirleme s reci

 l me eřdeĐerliĐinin test edilmesinde iki ayrı kurama dayanan y ntemlerin  n plana  ıktıĐı g r lmektedir. Bunlardan ilki Madde Tepki Kuramı (MTK)'na, diĐeri ise Yapısal Eřitlik Modellemesi (YEM)'ne dayanan y ntemlerdir. Bunun yanı sıra inceleme s recinde s rekli ya da s reksiz oluřuna g re  rt k deĐiřken modeller (latent variable models) de kullanılabilmektedir (G ng r Culha, 2012). YEM altında  l me eřdeĐerliĐi inceleme s recinde kullanılan y ntem olarak  oklu Grup DoĐrulayıcı Fakt r Analizi ( GDFA) bulunmaktadır. MTK altında bulunan y ntemlerin bazıları ise Madde Karakteristik EĐrilerini Karşılařtırma, Madde Parametrelerini Karşılařtırma (MPK), Raju'nun Alan  l  mleri, Lord'un Ki-Karesi ve Olabilirlik Oran Testi (OOT) y ntemleridir (Embretson, Reise, 2000; Hambleton, Swaminathan ve Rogers, 1991).

Bu arařtırma kapsamında farklı kuramlara dayalı y ntemlerin istatistiksel g   oranları karşılařtırılmıřtır. Bu arařtırma kapsamında YEM temelli y ntemlerden ortalama kovaryans temelli  oklu grup doĐrulayıcı fakt r analizi ve MTK temelli y ntemlerden olabilirlik oran testi ile madde parametrelerin karşılařtırma y ntemi tercih edilmiřtir. Bu y ntemlerin tercih edilmesinin nedeni farklı kuramlara dayalı y ntemlerle ilgili olarak arařtırmacılara bilgi saĐlamaktır. Bununla birlikte farklı y ntemlerle ilgili saĐlanan bu bilgiler ıřıĐında ge erlik inceleme s re leri ile ilgili arařtırmacılara katkı sunmaktır. Y ntemlerden  oklu grup doĐrulayıcı fakt r analizinin tercih edilme nedenlerinde birisi ulusal alanyazında YEM temelli

yöntemlerin yapay veri kullanılarak farklı koşullar altında sınanmamış olmasıdır. MTK altındaki yöntemler kullanılarak yapay veri üzerinde yürütülen araştırmalar mevcut olsa da ulusal alanyazın incelendiğinde YEM temelli yöntemler için bu tür araştırmalara rastlanmamıştır. Ölçme eşdeğerliği sürecinde çoklu grup doğrulayıcı faktör analizi yönteminin kullanımına ilişkin bilgilerin ortaya konması, araştırmacıların doğru adımlar atabilmesi için bu yöntemin kullanılmasına karar verilmiştir. MTK temelli yöntemlerden ikisinin ele alınmasının nedeni ise işleyiş süreci farklı olan yöntemlerle ilgili olarak araştırmacılara ışık tutmaktır. Ayrıca aynı kurama dayalı yöntemlerin hangi durumlarda tercih edilebileceğine ilişkin önerilerle ölçme eş değeriği inceleme süreci ile ilgili olarak araştırmacılara bilgi sağlanması planlanmıştır. Bu kapsamda çalışma sürecinde kullanılan bu yöntemlerle ölçme eşdeğerliği analizlerinin gerçekleştirilme sürecine ilişkin bilgiler alt başlıklar halinde sunulmuştur.

Çoklu Grup Doğrulayıcı Faktör Analizi Yöntemi

Yapısal Eşitlik Modellemesi çatısı altında yer alan doğrulayıcı faktör analizi yönteminin amaçlarından biri, farklı gruplar altında ölçme ve yapısal modeller arasındaki eşdeğerliği eş zamanlı olarak test etmektir (Asil, 2010). Ölçme modeli, gözlenen değişkenlerin örtük değişkenlere bağlandığını gösterirken, yapısal model örtük değişkenlerin kendi aralarındaki ilişkileri betimler.

Ölçme eşdeğerliğinin incelendiği çalışmalarda kovaryans ve ortalama yapılarının birlikte modele dâhil edilmesi önerilmektedir (Little, 1997; Meredith, 1993). Bu nedenle Jöreskog (1971) tarafından ortaya konulan kovaryans modeli ile Sörbom (1974)'un ortalama ve kovaryans (mean and covariance structure) modellerini ayırt etmek gerekmektedir.

Çoklu grup doğrulayıcı faktör analizinde kovaryans temelli olarak her birey için oluşturulan faktör analitik modelin matematiksel gösterimi şu şekildedir:

$$Y_{il} = \lambda_{1i}\eta_1 + \lambda_{2i}\eta_2 + \dots + \lambda_{jm}\eta_{mi} + \varepsilon_{ji}$$

Bu denklemde Y_{il} , i'nci bireyin, l'inci gözlenen değişkendeki gözlenen puanını ifade etmektedir. Regresyon katsayıları, λ_{im} , j maddesinin m faktöründeki faktör yükleri, η_{mi} bir veya daha fazla faktör, ε_{ji} normal dağılım gösteren tesadüfi hata terimlerini gösterir.

Varyans ve kovaryans yapılarının ek olarak diğer modelde gözlenen değişken ortalamaları da modele eklenmektedir (Somer, Korkmaz, Dural ve Can, 2009; Wu, Li ve Zumbo, 2007). Bu şekilde kurgulan modelde her birey için ortaya çıkan faktör analitik modelin matematiksel gösterimi şu şekildedir:

$$Y_{il} = T_j + \lambda_1(\alpha_1 + \eta_1) + \lambda_2(\alpha_2 + \eta_2) + \dots + \lambda_{jm}(\alpha_m + \eta_{mi}) + \varepsilon_{ji}$$

Bu denklemde Y_{il} , i'nci bireyin, l'inci gözlenen değişkendeki gözlenen puanını ifade etmektedir. Regresyon katsayıları, λ_{im} , j maddesinin m faktöründeki faktör yükleri, η_{jm} bir veya daha fazla faktör, ϵ_{ji} normal dağılım gösteren tesadüfi hata terimlerini gösterir. Ancak kovaryans temelli çoklu grup doğrulayıcı faktör analizinden farklı olarak bu modelde α_m ile gösterilen faktör ortalamaları ve bir yanıtın regresyon sabitini gösteren Γ_j terimi modele dâhil edilmiştir. Kovaryans temelli çoklu grup doğrulayıcı faktör analizinde araştırmacılar faktör yükleri, faktör varyans ve kovaryans yapıları ve hata varyanslarının eşdeğerliğini test edebilirken, ortalama kovaryans temelli çoklu grup doğrulayıcı faktör analizi yönteminde faktör ortalamalarının ve regresyon sabitlerinin eş değeriğini test edebilirler.

Olabilirlik Oran Testi

Thissen, Steinberg ve Gerrard (1986) tarafından önerilen olabilirlik oran testi yönteminde çoklu puanlanan verilerde değişen madde fonksiyonu olup olmadığı, sınırlandırılmış (dar) ve genişletilmiş model olarak adlandırılan iki modele ilişkin olabilirlik oranlarını karşılaştırarak analiz yapılır. Olabilirlik oran testi analizlerinde, odak ve referans grubu olarak alınan iki grupta kestirilen madde parametrelerinin eş değeri olup olmadığı incelenir. Madde parametrelerinin eşitliğinin test edilmesi için kurulan sınırlandırılmış modelde, her iki grubun madde parametrelerinin eş değeri olduğu; genişletilmiş modelde ise i. maddenin parametrelerinin odak ve referans grupta farklı diğer maddelere ait parametrelerinin eş değeri olduğu varsayılır. Söz konusu iki model için de olabilirlik oranı hesaplanır. Genişletilmiş modelde parametreleri her iki grupta serbest bırakılan madde(ler), değişen madde fonksiyonu gösterdiği öngörülen maddeler olabilir. Olabilirlik oran testinde amaç, sınırlandırılmış ve dar model için uyum sonuçları gösteren $-2 \log$ olabilirlik değerlerinin ($-2 \log$) arasındaki farkın manidarlığı değerlendirilerek değişen madde fonksiyonu varlığı belirlenmeye çalışılır. Eğer genişletilmiş modelde parametreleri serbest bırakılan madde, değişen madde fonksiyonu gösteriyorsa, genişletilmiş model için elde edilen uyum istatistiği değerinin daha iyi olması beklenmektedir. Çünkü değişen madde fonksiyonu içeren maddeye ilişkin her iki grupta yer alan bireyler farklı yanıtlar vereceğinden, bu maddeye ilişkin parametrelerin serbest bırakılması maddeye ilişkin daha doğru kestirimler yapılmasına olanak tanımaktadır. Olabilirlik oran testi analizlerinde sınırlandırılmış ve genişletilmiş modelden elde edilen $-2 \log$ değerlerinin karşılaştırılması için G^2 istatistiği kullanılır. G^2 istatistiği, bir bağıntı kullanılarak hesaplanmaktadır. Bu bağıntı:

$$G^2 = -2 (\log \text{sınırlandırılmış model} - \log \text{genişletilmiş model}) \text{ şeklindedir.}$$

Bağıntıdaki $\log \text{sınırlandırılmış model}$, her iki grupta da madde parametrelerinin eşit kabul edildiği durumda hesaplanan olabilirlik değeri iken; $\log \text{genişletilmiş model}$ ise belli maddelerin ve

onlara ait parametrelerin gruplarda serbest kestirilmesine izin verilmesiyle elde edilen olabilirlik oran değeridir. Belirtilen bağıntı kullanılarak elde edilen G^2 değeri, iki model arasındaki parametre sayısı farkına eşit olan serbestlik derecesindeki (Thissen, Steinberg ve Wainer, 1993) χ^2 değeri ile karşılaştırılarak model uyum istatistikleri arasındaki farkın manidar olup olmadığı incelenir. G^2 değeri tablodaki kritik χ^2 değerinden büyükse, madde parametrelerinin gruplar arasında manidar farklılık gösterdiği ve genişletilmiş modelde parametreleri serbest bırakılan maddenin değişen madde fonksiyonu gösterdiği sonucuna ulaşılır.

Maddede, değişen madde fonksiyonu bulunduğunda, sırasıyla madde parametrelerinin biri hariç tutularak teste devam edilir. G^2 istatistiğinin değerlendirilmesine ilişkin olarak Greener (2004) tarafından da ölçütler oluşturulmuştur. Bu ölçütlere göre parametreleri serbest bırakılan madde:

$3,84 < G^2 < 9,40$ ise ihmal edilebilir düzeyde,

$9,40 < G^2 < 41,90$ ise orta düzeyde,

$G^2 \geq 41,90$ ise yüksek düzeyde değişen madde fonksiyonu göstermektedir

(Akt: Acar ve Kelecioğlu, 2010).

Bock, Muraki ve Pfeifferberger'in Madde Parametrelerini Karşılaştırma Yöntemi

Madde parametrelerindeki kaymayı (drift) temel alan bu yöntem Bock, Muraki ve Pfeifferberger (1988) tarafından geliştirilmiştir. Ayrıca Muraki ve Engelhard (1989) çalışmaları ile yöntemin geliştirilme sürecine önemli katkılar sağlamıştır (Thissen ve diğerleri, 1993) Bu yöntem, karşılaştırma grupları için kestirilen madde ayırt edicilik ve madde istatistiksel güçlük parametrelerinin farklılıklarını temel almaktadır (Morales, Reise ve Hays, 2000; Reise, Smith ve Furr, 2001). Bu yöntemi diğer DMF belirleme yöntemlerinden ayıran temel nokta madde yerleşim (location) fonksiyonuna odaklanmasıdır.

Çoklu puanlanan madde parametrelerinin kestirilmesi sürecinde kullanılabilecek modellerden biri olan değiştirilmiş aşamalı tepki modelinde (The Modified Graded Response Model; Muraki, 1990) madde yerleşim parametresinin kestirim süreci açıklanmıştır. Bu model aşamalı tepki modelinin (Graded Response Model) sınırlandırılmış hali gibi düşünülebilir. Bu modelde ayırt edicilik parametresinin kestiriminde aşamalı tepki modeli ile aynı süreç gerçekleşmektedir. Bir başka ifadeyle her bir madde için bir adet ayırt edicilik parametresi kestirilmektedir. Ancak bu iki model eşik parametrelerinin kestirilmesinde farklılaşmaktadır. Aşamalı tepki modelinde her bir madde için kategori sayısının bir eksiği kadar kategori eşik (threshold) parametresi (bij) kestirilmektedir. Değiştirilmiş aşamalı tepki modelinde ise tüm ölçek için bir set kategori eşik parametresi (cj) ve her bir madde için ise bir adet yerleşim

parametresi (b_i) kestirilmektedir. Yerleşim parametresi, ait olduğu maddenin zorluk derecesini/kategori eşik değerini gösterir (Embretson ve Reise, 2000). Bu parametreler arasındaki ilişki;

$b_{ij} = b_i c_j$ eşitliği ile gösterilebilir.

Madde parametrelerinin karşılaştırılması yönteminde, farklı gruplarda kestirilen madde parametreleri arasındaki farklar alınarak değişen madde fonksiyonu istatistiği hesaplanır. Ancak, değişen madde fonksiyonu istatistiğinin hesaplanmasında her iki grup için kestirilen parametrelerin metrik eşitliğinin sağlanması gerekir. Değişen madde fonksiyonu istatistiği hesaplanmasında kullanılan bağıntılar şu şekildedir (Bock ve diğerleri, 1988; Morales, Reise ve Hays, 2000; Reise, Smith ve Furr, 2001):

$$\text{Değişen madde fonksiyonu istatistiği} = \hat{a}_{i(\text{referans})} - \hat{a}_{i(\text{odak})}$$

$$\text{Değişen madde fonksiyonu istatistiği} = \hat{b}_{i(\text{referans})} - \hat{b}_{i(\text{odak})}$$

Bu aşamadan sonra hesaplanan değişen madde fonksiyonu istatistiğinin standardize edilmesi gerekmektedir. Kestirilen istatistiklere ilişkin standardize etme işlemi aşağıdaki bağlantılardan yapılmaktadır (Muraki ve Engelhard, 1989; Aktaran Thissen ve diğerleri, 1993: 84).

$$\text{Standardize değişen madde fonksiyonu istatistiği} = \frac{DMF_{ia}}{\sqrt{\text{var}\hat{a}_{i(\text{referans})} + \text{var}\hat{a}_{i(\text{odak})}}}$$

$$\text{Standardize değişen madde fonksiyonu istatistiği} = \frac{DMF_{ib}}{\sqrt{\text{var}\hat{b}_{i(\text{referans})} + \text{var}\hat{b}_{i(\text{odak})}}}$$

Elde edilen standardize değişen madde fonksiyonu istatistiklerinin değerlendirilmesinde Thissen ve diğerleri, (1986), her bir madde için elde edilen standardize değişen madde fonksiyonu istatistiğinin karesinin alınarak bir serbestlik derecesinde χ^2 kritik değeri ile karşılaştırılmasını önermişlerdir. Bir madde için elde edilen standardize değişen madde fonksiyonu istatistiği değerinin karesi, kritik χ^2 değerinden büyükse söz konusu madde için değişen madde fonksiyonu olduğu çıkarımı yapılır. Standardize değişen madde fonksiyonu istatistiği pozitif değer aldığında odak grup; negatif değer aldığında ise referans grup lehine bir avantaj durumu olmaktadır.

Madde tepki kuramı temelli yöntemler ile ilgili olarak yapılan çalışmalarda, bazı yöntemlerin ön plana çıktığı, bazılarının ise fazla tercih edilmediği görülmektedir. Araştırmacıların belli yöntemleri daha fazla tercih etme nedeni olarak, istatistiksel güçleri gösterilebilir. En fazla tercih edilen madde tepki kuramı temelli yöntemlerden biri, olabilirlik oran yöntemidir. Öte yandan madde parametreleri karşılaştırılması yönteminin ise çok fazla tercih edilmediği görülmektedir. Korkmaz (2005) farklı örneklem büyüklüklerinde erkek ve kadın gruplarının karşılaştırılmasını amaçlayan çalışmada bu iki yöntemin benzer

istatistiksel güçlere sahip olduğunu belirlemiştir. Aynı zamanda bu çalışmada madde parametrelerini karşılaştırma yönteminin, değişen madde fonksiyonu belirleme sürecinde uygulaması daha kolay bir yöntem olduğu belirtilmiştir. Ancak bu iki yöntemin, istatistiksel güçlerini etkileyen diğer faktörler değişimlendiğinde ne gibi sonuçlar vereceğindeki belirsizliğin giderilmesi gerekmektedir.

YEM ve MTK temelli bu yöntemlerin ölçme eşdeğerliği analiz sürecine ilişkin benzerlik ve farklılıkları mevcuttur. Yöntemlere ilişkin bu bilgiler sıralanarak ele alınan bu yöntemlerin karşılaştırma gereklilikleri ortaya konmaya çalışılmıştır. Araştırma kapsamında yapılan yorumların doğru şekilde gerekçelendirilmesi için yapısal eşitlik modellemesi ve madde tepki kuramı temelli yöntemler arasındaki benzerlik ve farklılıklara değinmek önem arz etmektedir. Yapılan farklı çalışmalar temel alınarak YEM ve MTK temelli yöntemler arasındaki benzerlik ve farklılıklar sırasıyla açıklanmıştır (Raju, Laffitte ve Byrne, 2002; Reise ve diğerleri, 1993):

Yöntemler arası benzerlikler

1. Her iki yöntem örtük değişkenle, gözlemlenen değişkenler arasındaki ilişkileri değerlendirmeyi amaçlar.
2. Ölçme eşdeğerliğinin test edilmesi sürecinde her iki yaklaşımda da farklı gruplarda bulunan ve ölçülen örtük değişken açısından aynı seviyede bulunan bireylerin madde veya ölçek bazında gerçek puanının eş değerliliğini araştırır. Bu durum klasik test teorisi ile benzerlik gösterir. Klasik test teorisinde iki farklı testten alınan puanların eşitliği sınanırken, bu kuramda örtük değişken sabit olarak alındığından iki alt grubun gerçek puanlarının eş değer olup olmadığı kontrol edilir.
3. İki farklı kuramı temel alan yöntemler farklı alt gruplarda örtük değişken puanlarının dağılımlarının aynı olduğunu iddia etmez. Örtük değişkene ilişkin puan dağılımları gruplar arasında tipik farklılık gösterir. Örtük değişken ortalamalarındaki bu farklılık genellikle çalışmalarda “etki” (impact) olarak adlandırılmaktadır. Bu iki başlık altındaki yöntemlerde gizil değişkene ilişkin dağılımın eş değeri değil, örtük değişken bakımından aynı düzeyde olan bireylerin grup üyeliklerinden bağımsız olarak aynı gerçek puana sahip olup olmadığı incelenmektedir.
4. YEM ve MTK temelli yöntemler ölçme eşdeğerliğinin olmadığı durumlarda sorunun kaynağını belirlemek için kullanılabilir. MTK temelli yöntemlerde ölçme eşdeğerliği olmayan maddeler için değişen madde fonksiyonu varlığı saptanmaya çalışılır. YEM temelli yöntemlerde ortaya konan modelin veri setine uygun olup olmadığı test edilir.

Eşdeğerliğin olmadığı belirlendiğinde madde içeriği incelenerek bunun kaynağı belirlenmeye çalışılır.

5. Her iki kuramı temel alan yöntemlerde, ölçme eşdeğerliğinin sağlanamadığı durumlar için madde cevap fonksiyonları kullanılır. Çoklu puanlanan verilerden eşdeğerliğin bozulduğu noktanın orta veya uç noktalarda olmasına uygun şekilde çözümler getirilir.

Yöntemler arası farklılıklar

- 1- Yapısal eşitlik modellemesi ve madde tepki kuramı temelli yöntemlerde örtük değişken ve gözlenen değişkenler arasındaki ilişkinin tanımı birbirinden farklıdır. YEM temelli yöntemlerde bu ilişki doğrusal iken, MTK temelli yöntemlerde bu ilişki doğrusal değildir.
- 2- Bu iki kurama dayanan yöntemlerin örtük ve gözlenen değişkenler arasındaki ilişkiyi tanımlamak için kullandıkları modeller farklılık göstermektedir. YEM temelli yöntemler doğrusal regresyon modeli; MTK temelli yöntemler ise lojistik regresyon modellerini kullanmaktadır. İkili puanlanan verilerde sürekli yapıdaki örtük değişken ile gözlenen değişkenler arasındaki ilişkinin belirlenmesi için lojistik regresyon modeli, doğrusal regresyon modeline göre daha uygundur. Ancak çoklu puanlanan verilerden doğrusal regresyon modelleri, lojistik regresyon modeli kadar verimlidir.
- 3- Yapısal eşitlik modellemesi temelli yöntemler çok boyutlu örtük değişken yapısının ve çok sayıda alt grubun olduğu durumlarda yapılacak ölçme eşdeğerliği analizleri için madde tepki kuramı temelli yöntemlere göre daha verimlidir. Bu durumlara ilişkin çalışmalar (Kim, Cohen ve Park, 1995; Oshima, Raju ve Flowers, 1997) mevcut olsa da bu noktada MTK temelli yöntemlerin geliştirilmesi gerekmektedir.
- 4- İki kuram temelli yöntemlerin farklılaştığı bir diğer nokta, varyanslarının eşitliğinin test edilme durumudur. Yapısal eşitlik modellemesi hesaplamalarında her bir madde için ortaya çıkan hata oranı standarttır; kişiden kişiye değişiklik göstermez. Ancak, madde tepki kuramı için hata oranının bireyden bireye değişiklik gösterebilme durumu söz konusudur. YEM temelli yöntemlerde ölçme eşdeğerliğinin sınırları modellerde eklenen en katı sınırlamalardan biri hata varyanslarının gruplar arası eşdeğerliği incelemelerinde konulmaktadır. Ancak, madde tepki kuramı temelli yöntemlerde hata varyanslarının eşdeğerliğine ilişkin araştırma yapılmamaktadır.

- 5- Çoklu puanlanan verilerde iki yöntem arasında ortaya çıkan diğer bir farklılık ise, kategori cevaplama olasılıkları ile ilgilidir. MTK temelli yöntemlerde yapılacak yetenek kestirimleri ile bir bireyin puanlama kategorilerini işaretleme olasılıkları belirlenebilirken, YEM temelli yöntemlerde bu mümkün değildir. Bu bağlamda MTK ölçme eşdeğerliğinin daha ayrıntılı şekilde belirlenebilmesine olanak tanımaktadır.
- 6- Son olarak iki kuram temelli yöntemler ölçek düzeyinde telafi edici değişen madde fonksiyonu bağlamında farklılaşmaktadır. Telafi edici değişen madde fonksiyonu, zıt yönlü değişen madde fonksiyonuna sahip maddelerin ölçek düzeyinde yapılacak incelemelerde değişen madde fonksiyonu olmadığına yönelik bulgulara yol açması durumudur. Telafi edici değişen madde fonksiyonu ile ilgili durumlar YEM temelli yöntemlerde fazla önemsenmezken, MTK temelli yöntemlerde bu duruma özel belirleme yöntemleri mevcuttur.

Ölçme eş değeri analizi analizleri sürecinde yukarıda özetlenen açılardan benzerlik ve farklılık gösteren yöntemlerin farklı koşullar altında birlikte ele alınarak karşılaştırılmasının araştırmacılara katkı sağlayabilir. Ulusal alanyazın incelendiğinde çoklu grup doğrulayıcı faktör analizi yöntemi ve madde tepki kuramı altındaki yöntemlerle ilgili araştırmaların olduğu görülmektedir (Akalın, 2014; Arıkan Akın, 2015; Asil ve Gelbal, 2012; Atalay, Gök, Kelecioğlu ve Arsan, 2012; Ateşok Deveci, 2008; Başusta ve Gelbal, 2015; Çepni, 2011; Çetin, 2010; Güzeller, 2011; Kan, Sünbül ve Ömür, 2013; Öğretmen, 2006; Öğretmen, 2009; Somer, 2004; Somer ve diğerleri, 2009; Şekercioğlu, 2009; Uzun, 2008). Yapılan bu araştırmalarda, çoklu grup doğrulayıcı faktör analizi yöntemi tek başına ya da madde tepki kuramı altındaki yöntemlerden birkaçı bir arada kullanılmıştır. Yapısal eşitlik modellemesi ve madde tepki kuramı temelli yöntemlerin bir arada kullanıldığı az sayıda araştırmaya rastlanmaktadır (Öğretmen, 2006; Somer ve diğerleri, 2009). Yöntemlerin birlikte ele alınarak değerlendirilmesi analiz süreçlerine ilişkin bilgi eksikliklerinin giderilmesine yardımcı olabilir. Böylece araştırmacılar yapmış oldukları uygulamalarda ortaya çıkan koşullara göre hangi yöntemi kullanarak analiz yapacağına daha kolay karar verebilir. Bu nedenlerden dolayı sosyal bilim alanında önemli ölçüde yer tutan farklı grup karşılaştırmalarına odaklanan araştırmaların geçerlik incelemeleri ile ilgili soru işaretleri giderilebilir.

Yöntem karşılaştırılmasını amaçlayan çalışmalarda, araştırmacılara doğru bilginin sağlanabilmesi için karşılaştırmaların hangi açıdan yapıldığı ortaya konulmalıdır. Bu araştırma kapsamında ele alınan yöntemler değişimlenen koşullar altındaki istatistiksel güç açısından karşılaştırılacaktır. Her bir yöntemin istatistiksel güç durumu değişimlenen koşullar için oran

olarak hesaplanmıştır. Bu çalışma kapsamında istatistiksel güç oranı, üretilen yapay verilerde bulunan değişen madde fonksiyonu içeren maddelerin doğru olarak tespit edilme oranı olarak tanımlanmıştır. Nitekim YEM ve MTK temelli ölçme eş değeri belirlene yöntemlerinin karşılaştırıldığı çalışmalarda istatistiksel güç oranı, “true positive” olarak isimlendirilmektedir (Kim ve Yoon, 2011; Meade ve Lautenschlager, 2004; Stark, Chernyshenko ve Drasgow, 2006). Bunun yanı sıra çalışma kapsamında yapay veriler için elde edilen sonuçlar, istatistiksel güç oranlarını etkilediği belirlenen değişkenlerden gerçek veriler üzerinde değişimlenmesi mümkün olanlara müdahale edilerek gerçek verilerden elde edilen sonuçlarla karşılaştırma yapılmıştır. Böylece yapay verilerin analizinde elde edilen sonuçların gerçek veri analizi sonuçları ile nispeten de olsa desteklenmesi amaçlanmıştır. Farklı kuramsal temelli yöntemlerin istatistiksel güç oranları karşılaştırılarak, araştırmacılara yöntemlerin farklı koşullar altındaki performanslarına ilişkin bilgi sunulmuştur. Bir başka ifadeyle hangi yöntemin hangi koşullar altında daha iyi bir performans sergilediğinin ortaya konulması planlanmıştır.

Araştırma sürecinde elde edilen bulguların yorumlanması ve daha önce ortaya konulmuş bilgilerle kıyaslanması için alanyazında yer alan çalışmaların incelenmesi önemlidir. Daha önce yapılmış olan çalışmaların incelenmesi, bu çalışmada var olan bilgilerden farklı olarak elde edilen sonuçların nedenlerine ilişkin çıkarımda bulunma sürecine ışık tutabilir. Bu araştırma kapsamında alanyazındaki çalışmalar incelenerek YEM ve MTK temelli yöntemleri karşılaştıran çalışmalar özetlenmiştir. İncelenen çalışmaların YEM ve MTK temelli yöntemlerin yapay veriler üzerindeki performanslarını araştıranlarla sınırlı tutulmasının nedeni elde edilen bulguların karşılaştırılabilir olmasını sağlamaktır. Bu doğrultuda alanyazında bu duruma uygun olarak yapılmış olan çalışmalar sunulmuştur.

Reise, Wideman ve Pugh (1993) tarafından yapılan çalışmada katılımcıların ruh halini belirlemek için 5’li likert tipinde NA5 ölçeği (bireylerin ölçme anındaki duygu durumlarını ölçme amacına sahiptir), 540 ve 598 kişiden oluşan iki farklı gruba uygulanmıştır. Elde edilen veriler üzerinde, doğrulayıcı faktör analizi (DFA) ve madde tepki kuramı temelli değişen madde fonksiyonu belirleme yöntemlerinin kullanılarak ölçme eşdeğerliği incelemeleri yapılmıştır. Yapılan bu çalışmada veriler için uygun modeli belirleme ve elde edilen sonuçları değerlendirme başlıkları altında YEM ve MTK yöntemleri karşılaştırılmıştır. Veri seti için uygun model belirleme hususunda YEM temelli yöntemlerin daha gelişmiş ve kullanıcı dostu olduğu belirtilmiştir. Diğer taraftan elde edilen çıktıların değerlendirilmesi noktasında kullanılan uyum indeksleri açısından YEM modellerinin daha alternatifli olduğu (uyumun iyiliği indeksi, karşılaştırmalı uyum indeksi vb.) MTK modellerinde yalnızca olabilirlik oranı, Ki-Kare değerinin değerlendirildiği belirtilmiştir. Ayrıca bu çalışmada YEM ve MTK temelli

yöntemler arasında ortaya konan en önemli farklılık, YEM temelli modellerin istatistiksel güçlük parametrelerini (bjx) önemsemiyor olmasıdır. Nitekim ölçme eşdeğerliği analizlerine ilişkin elde edilen sonuçlarda ölçekte yer alan iki madde için yöntemler arasında farklılıklar olduğu saptanmıştır. Sözü edilen iki madde DFA modellerine göre gruplar arasında değişmezlik sergilerken, MTK modellerine göre ise değişmez olmadığı sonucuna varılmıştır. Yöntemler arasında ortaya çıkan bu fark DFA modellerinde istatistiksel güçlük parametrelerinin dikkate alınmıyor olması ile ilişkilendirilmiştir. Bu çalışma, DFA temelli modellere ortalama yapılarının dâhil edilmemesinin, analizlerde değişen madde fonksiyonu içeren maddelerin belirlenmesinde sorunları beraberinde getirdiğini göstermektedir.

Ankenmann, Witt ve Dunbar (1999) odak ve referans grup örneklem büyüklükleri, yetenek dağılımlarını, DMF içeren maddelerin parametre aralıklarını ve DMF örüntüsünü değişimleyerek elde ettikleri 72 farklı koşul için 100 tekrarlama yaparak ürettikleri veriler üzerinde Olabilirlik Oran Testi ve Mantel Haenszel yöntemlerinin istatistiksel güç ve I. tip hata oranlarını incelemişlerdir. Araştırmada madde sayısı, DMF içeren madde sayısı ve DMF büyüklüğü sabit olarak alınmıştır. Veriler 26 madde (20 ikili, 6 çoklu puanlanan) ve bir çoklu puanlanan maddenin 0.25 büyüklüğünde DMF'ye sahip olacağı şekilde üretilmiştir. İkili puanlanan maddeler için, üç parametrelili lojistik model; çoklu puanlanan maddeler için ise aşamalı tepki modeli temel alınmıştır. Değişimleme yapılan koşullardan örneklem büyüklüğü (O/R) ise 2000/2000, 2000/500 ve 500//500 seviyeleri olacak şekilde ele alınmıştır. Araştırma bulguları, olabilirlik oran testi yöntemi istatistiksel güç oranlarının toplam örneklem büyüklüğü artışına paralel olarak arttığını göstermektedir. Bu çalışmada istatistiksel güç oranlarına ilişkin elde edilen en çarpıcı bulgu ise, olabilirlik oran testi yönteminin 500/500 örneklem büyüklüğünde istatistiksel güç oranlarının, %60'ın üstüne çıkmamış olmasıdır. Bu bulgular doğrultusunda, bu yöntemin küçük örneklemelerde DMF belirleme sürecinde yetersiz kaldığı sonucuna ulaşılmıştır.

Flowers, Raju ve Oshima (2002), MTK ve DFA temelli yöntemlerin, ölçme eşdeğerliği belirlemeye ilişkin istatistiksel güçlerini yapmış oldukları simülasyon çalışmasında ortaya koymaya çalışmışlardır. Bu çalışmada aşamalı tepki modeli temel alınarak 20 maddeden oluşan 5'li likert şeklinde puanlanan tek boyutlu veriler üretilmiştir. Yapılan maddelere tek biçimli ve tek biçimli olmayan değişen madde fonksiyonu yüklenmiştir. Ayrıca referans ve odak grupları için iki farklı yetenek dağılımı tasarlanmıştır. Bunlardan ilkinde iki grup aynı yetenek dağılımına sahipken, diğerinde ise grupların yetenek dağılımları arasında bir standart sapma olacak şekilde veri üretilmiştir. Çalışmada gruplar 1000 kişiden oluşturulmuş ve her bir koşul için beş tekrarlama yapılmıştır. Yapılan çalışmada DFA için yapılan analizlerde yalnızca

yerleşim (location) parametresi ve hem yerleşim hem kesişim parametrelerinin kullanıldığı ortalama kovaryans yapılarını inceleyen modeller, MTK için ise telafisel olmayan değişen madde fonksiyonu belirleme modelleri tercih edilmiştir. Elde edilen sonuçlara göre yalnızca yerleşim parametrelerinin incelendiği modellerin a parametresinde bulunan değişen madde fonksiyonlarını başarıyla belirlediği, ancak b parametresi için başarılı olamadığı saptanmıştır. MTK temelli modellerin ise yerleşim ve kesişim parametrelerinin incelendiği koşullarda b parametresinde değişen madde fonksiyonu olan koşullar için başarılı olduğu belirlenmiştir. Ancak bu iki modelin yalnızca a parametresinde değişen madde fonksiyonu olan koşullarda başarısız olduğu görülmüştür. Bir başka deyişle DFA ve MTK temelli yöntemlerin tek biçimli değişen madde fonksiyonu olan koşullara, tek biçimli olmayanlara göre daha duyarlı olduğu ifade edilmiştir. Öte yandan MTK temelli modellerin, DFA temelli modellere göre daha kabul edilebilir I. ve II. tip hata oranına sahip olduğu rapor edilmiştir.

Meade ve Lautenschlager (2004) yapmış oldukları simülasyon çalışmasında MTK temelli OOT ile ÇGDFA yöntemlerinin ölçme eşdeğerliği sürecine ilişkin istatistiksel güç ve I. tip hata oranlarını karşılaştırmışlardır. Çalışma kapsamında sınanan hipotezler, YEM ve MTK temelli yöntemlerin a ve b parametreleri kaynaklı DMF durumlarının belirlenmesi sürecindeki performansları üzerine kurgulanmıştır. Çalışma kapsamında belirli maddelerin a ve b parametrelerine müdahalede bulunulmuştur. Maddelerin b parametreleri ile ilgili yapılan değişimleme sürecinde üç farklı şekilde DMF oluşturulmuştur. Bu süreçte DMF içeren madde oluşturmak için en büyük cevap kategorisi, en büyük iki cevap kategorisi ve en uç iki cevap kategorisinin b parametreleri değişimlenmiştir. Çalışma kapsamında DMF büyüklüğü 0.40 olarak belirlenmiş; bu büyüklüğün belirgin bir farklılaşmaya yol açacağı için uygun olduğu belirtilmiştir. Çalışma kapsamında 5'li likert şeklinde puanlanan, altı maddeli, tek boyutlu bir yapıda üretilen veriler için örnekleme oluşturan her bir gruba yönelik 150, 500 ve 1000 değerleri kullanılmıştır. Ayrıca her bir veri seti için 100 tekrarlama gerçekleştirilmiştir. Veri setinde bulunan altı maddeden, iki veya dört madde DMF içerecek şekilde üretilmiştir. Çalışma kapsamında elde edilen sonuçlara göre ÇGDFA yöntemine ilişkin sonuçlar incelendiğinde 150/150 kişilik veri setinde bu yöntemin diğer örneklem büyüklüğü değişkenlerine nazaran daha yüksek istatistiksel güç değerleri verdiği saptanmıştır. Bir başka ifadeyle beklentinin aksine ÇGDFA yöntemi istatistiksel güç oranları örneklem büyüklüğündeki artışa ters yönde hareket etmiştir. Bunun yanında değişimleme yapılan koşullardan bağımsız şekilde ÇGDFA yöntemi istatistiksel güç oranlarının genel olarak yetersiz kaldığı açıklanmıştır. Bu oranların üretilen 100 veri seti için %0.00 ile %20.00 arasında değiştiği belirtilmiştir. OOT yöntemine ilişkin istatistiksel güç oranlarında ise örneklem büyüklüğü değişkenindeki değişimle aynı

yönde bir değişim olduğu gözlenmiştir. OOT yöntemi ile gerçekleştirilen analizlerde en düşük istatistiksel güç oranı 150/150 kişilik veri setlerinde elde edilmiştir. Ayrıca her iki yöntemle yapılan analizlerde DMF içeren madde sayısının istatistiksel güç oranlarında belirgin bir farklılaşmaya yol açmadığından önemli bir değişken olmadığı ortaya konmuştur. Elde edilen bulgular OOT yönteminin ÇGDFA yöntemine göre yüksek istatistiksel güç oranlarına sahip olduğunu göstermiştir. OOT yöntemine ait istatistiksel güç oranları, 15 farklı analiz koşulu için %15.00 ile % 100 arasında değişiklik göstermektedir. Bütün bu sonuçlara göre çalışma kapsamında sınanan ilk hipotezi destekler nitelikte bulgular elde edilmiş; ÇGDFA yöntemi ölçme eşdeğerliği belirleme sürecinde yetersiz kalırken, OOT yöntemi istenen düzeyde sonuçlar vermiştir.

Stark, Chernyshenko ve Drasgow (2006), farklı koşullar altında ölçme eşdeğerliğini analiz etme yöntemlerini karşılaştırdıkları çalışmalarında, ÇGDFA ve OOT yöntemlerini kullanmışlardır. Yapılan bu çalışmada tek boyutlu ve on beş maddelik veri setleri için, DMF büyüklüğü, DMF türü, madde cevap kategori sayısı, örneklem büyüklüğü ve ortalama etkisi koşullarında değişimleme yapılmıştır. DMF büyüklüğü değişkeni, DMF yok; yük parametresi (ayrıt edicilik) için -.15 (düşük) ve -.40 (yüksek); kesişim parametreleri için +.25 (küçük) ve +.50 (büyük) olarak alınmıştır. DMF türü değişkeni ise maddelerin yalnızca yük, kesişim ve her iki parametresine de DMF yüklendiği koşullar olarak oluşturulmuştur. Bir başka ifadeyle tek biçimli ve tek biçimli olmayan DMF türleri ele alınmıştır. Üretilen veri setleri ikili ve çoklu (5'li likert) puanlanan şekilde tasarlanmıştır. Ayrıca örneklem büyüklükleri her bir grup için 500 ve 1000 kişi olarak alınmıştır. Bu koşullar altında yapılan analizlerde, serbest temel ve sınırlandırılmış temel model yaklaşımlarının her ikisi ile analizler gerçekleştirilmiştir. Elde edilen bulguların yorumlanma sürecinde iki farklı yol izlenmiştir. Bunlardan ilki bulguların manidarlık düzeyi .05 esas alınarak, ikinci ise Bonferroni düzeltmesi yapılarak yorumlanmasıdır. Bu çalışmada her bir koşul için 50 tekrarlamaya yapılmıştır. Çoklu puanlanan veriler üzerinden elde edilen bulgulara göre örneklem büyüklüğü değişkeni sabit tutularak yapılan incelemelerde DMF büyüklüğünün, düşük ve yüksek olduğu koşullarda ÇGDFA yöntemi için %98 ile %100 arasında istatistiksel güç oranı elde edilmiştir. Ancak OOT yöntemi için DMF büyüklüğü yüksek olduğunda %100; ancak düşük olduğunda ise %58 - %96 aralığında değişen istatistiksel güç oranlarına ulaşılmıştır. Buna göre DMF büyüklüğü yüksek olan koşullarda ÇGDFA ve OOT yöntemleri arasında istatistiksel güç oranı bakımından bir fark olmadığı çıkarımı yapılabilir. Bunun yanı sıra yöntemlere ilişkin belirlenen istatistiksel güç oranlarının, örneklem büyüklüğüne bağlı olarak artış gösterdiği, çalışma sonucunda elde edilen bulgular doğrultusunda yapılan diğer bir yorumdur. Örneklem büyüklüğü değişkeni göz önüne

alınarak yapılan değerlendirmede çoklu puanlanan veriler için ÇGDFA yönteminin daha yüksek istatistiksel güç oranları verdiği belirtilmiştir. Bunun nedeni olarak her iki yöntem düşünüldüğünde, ÇGDFA yönteminin daha basit bir model üzerine kurulu olması gösterilmiştir. OOT yönteminin analiz sürecinde beş, ÇGDFA yöntemi için iki parametre kestirimi yapılmıştır.

González Roma, Hernandez ve Gomez Benito (2006), aşamalı tepki modelini temel alarak farklı koşullar altında ürettikleri veri üzerinde ÇGDFA yönteminin değişen madde fonksiyonu belirleme sürecine ilişkin I. tip hata ve istatistiksel güç oranlarını belirlemeyi amaçlamışlardır. Bu çalışmada, örneklem büyüklüğü, yetenek dağılımları, değişen madde fonksiyonu türü ve değişen madde fonksiyonu büyüklüğü değişkenlerini değiştirmişlerdir. Örneklem büyüklüğü değişkeni için 100, 200, 400 ve 800 değerleri on farklı koşul oluşturacak şekilde düzenlenmiştir. Bu on farklı koşuldan dördünde referans ve odak grup örneklem büyüklükleri eşit, altısında ise referans grup örneklem büyüklüğü odak gruptan daha büyük olacak şekilde veri üretimi gerçekleştirilmiştir. Yetenek dağılımı değişkeninde ise iki farklı değişimleme yapılmış, referans grup ve odak grubun yetenek dağılımı eşit (RG $N(0,1)$ OG $N(0,1)$) ve eşit olmayacak (RG $N(0,1)$ OG $N(-1,1)$) şekilde düzenleme yapılmıştır. Değişen madde fonksiyonu türü için ise verilerin, değişen madde fonksiyonu göstermeyen, tek biçimli olan ve olmayan değişen madde fonksiyonu içeren maddeleri içerecek şekilde üretilmesi tercih edilmiştir. Son olarak değişen madde fonksiyonu büyüklüğü değişkeni için düşük (0.10), orta (0.25) ve yüksek (0.50) olacak biçimde değerler kullanılmıştır. Ortaya çıkan 140 farklı koşul için 100 tekrarlama yapılmıştır. Üretilen on maddelik, beş kategorili puanlanan veri setlerinde yalnızca madde 1 değişen madde fonksiyonu özelliği göstermektedir. Bu çalışmada referans ve odak grup örneklem büyüklüklerinin ve yetenek dağılımlarının eşit olduğu koşullarda ÇGDFA yöntemi istatistiksel güç oranları %1 ile %100 arasında değişkenlik göstermiştir. İstatistiksel güç oranları içinde en düşük oranın, referans ve odak grup örneklem büyüklüğünün 100; değişen madde fonksiyonu büyüklüğünün ise düşük olduğu koşulda hesaplandığı belirlenmiştir. En yüksek oranlar ise diğer örneklem büyüklüğü dağılımlarında, orta – yüksek değişen madde fonksiyonu büyüklüğü olan koşullarda hesaplanmıştır. Elde edilen bulgular örneklem ve değişen madde fonksiyonu büyüklüğü değişkenindeki artışa paralel olarak istatistiksel güç oranlarının da arttığını göstermektedir. Sonuç olarak ÇGDFA yönteminin kabul edilebilir sonuçlar vermesi için DMF büyüklüğünün en az orta düzey; referans ve odak grup örneklem büyüklüklerinden her birinin ise en az 200 olarak alınması gerektiği belirtilmiştir.

Atar ve Kamata (2011) yapmış oldukları çalışmada ordinal lojistik regresyon ve olabilirlik oran testi yöntemlerinin I. tip hata ve istatistiksel güç oranlarını ele almışlardır.

Yapılan çalışmada 200 tekrarlar yapılarak, örneklem büyüklüğü, alt grup örneklem büyüklüğü oranları, DMF büyüklüğü ve DMF durumu değişkenleri değiştirilerek 54 farklı koşul altında üretilen veriler üzerinde çalışılmıştır. Ayrıca üretilen veri setlerinde madde sayısı, DMF içeren madde sayısı, puanlama kategorisi sayısı ve yetenek dağılımı değişkenleri sabit tutulmuştur. Değişimleme yapılan değişkenlerden örneklem büyüklüğü için toplam 600, 1200 ve 2400 sayısına sahip veriler üretilmiştir. Odak referans grubu örneklem büyüklüğü oranları olarak 1:1 ve 2:1 alınmıştır. DMF büyüklükleri ise 0.32 (düşük), 0.43 (orta) ve 0.53 (yüksek) olarak değiştirilmiştir. DMF durumu olarak ise düşük, yüksek ve dengeli değiştirme koşulları oluşturulmuştur. Bu koşullar altında olabilirlik oran testi ile gerçekleştirilen analizlerde 0.43 birim DMF büyüklüğü koşulu için örneklem büyüklüğü değişkeni ele alındığında istatistiksel güç oranlarının, düşük değiştirme koşulunda % 65 - %100; yüksek değiştirme koşulunda %55 - %99; son olarak dengeli değiştirme koşulunda ise %95 - %100 arasında değişiklik gösterdiği belirlenmiştir. Elde edilen bu bulgular orta ve yüksek örneklem büyüklüğünde olabilirlik oran testi yönteminin yüksek istatistiksel güç oranları verdiğini göstermektedir. Buna ek olarak istatistiksel güç oranlarının örneklem büyüklüğündeki artışa bağlı olarak yükseldiği yorumunu yapmışlardır. Bu çalışmada 0.43 lojit birimlik DMF büyüklüğü koşulunda, olabilirlik oran testi yöntemi için en yüksek istatistiksel güç oranları toplam örneklem sayısının 2400 olduğu koşullarda elde edilmiştir.

Kim ve Yoon (2011) yapmış oldukları Monte Carlo çalışmasında çoklu grup doğrulayıcı faktör analizi ve olabilirlik oran testi temelli yöntemlerin istatistiksel güçlerini karşılaştırmışlardır. Çalışmada altı maddelik tek boyutlu veriler üretilmiştir. Örneklem büyüklüğü değişimleme yapılan değişkenlerden biridir ve her bir grup için 100-200-500-1000 olacak şekilde belirlenmiştir. Oluşturulan alt gruplardaki kişi sayıları ise eşit olarak alınmıştır. Ayrıca puanlama tipi olarak ikili ve çok kategorili (5'li likert) olacak şekilde planlama yapılmıştır. Değişen madde fonksiyonu sayısı içeren madde sayısı bir olarak alınmıştır. Değişen madde fonksiyonu büyüklüğü değişkeninin de değiştiği çalışmada düşük ve yüksek olmak üzere, ayırt edicilik parametresi için gruplar arasında 0.20 ve 0.40; kesişim parametreleri için ise 0.30 ve 0.60 olacak şekilde koşullar oluşturulmuştur. Ele alınan bu çalışmada eş değerlik, farklı parametrelere değişen madde fonksiyonu yüklenerek bozulmuştur. Bu koşullar yalnızca faktör yükü, yalnızca eşik parametresi ve bu parametrelerin her ikisinde değişen madde fonksiyonu yüklenmiş olan koşullardır. Bu koşullara ek olarak değişen madde fonksiyonu miktarlarında da değişimleme yapılarak toplam 48 farklı koşul elde edilmiş ve 500 tekrarlar gerçekleştirilmiştir. Elde edilen sonuçlara göre Bonferroni düzeltilmesinin kullanıldığı koşullarda eşdeğerliği bozan maddeleri belirlerken DFA ve MTK temelli yöntemler benzer

sonular verse de, deęiřen madde fonksiyonu gstermeyen maddeleri gsteriyor gibi belirleme (yanlıř pozitif – I. tip hata) oranları DFA temelli yntemlerde daha yksektir. Ayrıca deęiřen madde fonksiyonu byklę ve rneklem byklęnn yanlıř pozitif oranlarını etkileyebileceęi aıklanmıřtır. rneklem byklę ile ilgili elde edilen sonulara gre ise ikili puanlanan verilerde deęiřen madde fonksiyonunun doęru bir řekilde belirlenmesi iin en az 500 kiřilik rnekleme ihtiya olduęu, ancak oklu puanlanan verilerde rneklem byklęnn 100 olmasının hem DFA hem de MTK temelli yntemler iin yeterli olduęu grlmřtr. Ayrıca oklu puanlanan verilerde kesiřim parametresi kaynaklı deęiřen madde fonksiyonu ieren madde olan kořulların tmnde, DFA ve MTK temelli yntemler iin rneklem byklę deęiřkenindeki artıřa paralel olarak istatistiksel g oranlarında da bir artıř olduęu belirlenmiřtir. Belirtilen bu kořulda DFA temelli yntemlerde istatistiksel g oranları %64 ile %100; MTK temelli yntemlerde ise %34-%100 arasında deęiřkenlik gstermektedir.

Kankaras, Vermunt ve Moors (2011), yaptıkları simlasyon alıřmasında oklu grup doęrulayıcı faktr analizi, rtk sınıf analizi ve madde tepki kuramı temelli deęiřen madde fonksiyonu belirleme yntemlerini kullanmıřlardır. alıřma kapsamında bu yntemlerin lme eřdeęerlięini belirlemeye ynelik istatistiksel g oranlarının karřılařtırılması amalanmıřtır. İncelenen bu alıřmada deęiřen madde fonksiyonu kaynakları (yerleřim ve kesiřim parametrelerinden kaynaklanan), rtk deęiřkenin yapısı (kategorik ve srekli), eřdeęer olmayan madde sayısı (1 ve 3 madde), rneklem byklę (200/200 ve 1000/1000), deęiřmezlik dzeyinin belirlenme stratejisi (lek ve madde dzeyinde) ve model seiminde kullanılan kriter (Ki-Kare farkının ve Akaike Bilgi Kriteri) seimleri deęiřimlenerek farklı kořullar oluřturulmuřtur. rneklem byklęnn deęiřimlendięi kořullar karřılařtırıldıęında, oklu grup doęrulayıcı faktr analizi ve madde tepki kuramı temelli deęiřen madde fonksiyonu belirleme yntemlerinin kk rneklemelerde istatistiksel gnn azaldıęı; bu yntemlerin daha doęru sonular verebilmesi iin daha byk rneklemelerin kullanılması gerektięi sonucuna varılmıřtır. Madde tepki kuramı temelli yntemler iin eř deęer olmayan madde sayısı kořulunun da rneklem byklę ile benzer etkiye sahip olduęu ortaya konulmuřtur. MTK temelli yntemle gerekleřtirilen analizlerde eř deęer olmayan madde sayısı arttıka yntemlerin bu tr maddeleri yakalama oranları artarken, II. tip hata oranlarının azaldıęı rapor edilmiřtir. DFA temelli yntemlerle gerekleřtirilen analizlerde ise eř deęer olmayan madde sayısındaki artıřın, istatistiksel g oranında dřk dzeyde bir azalmaya neden olduęu saptanmıřtır. Ancak eř deęer olmayan madde sayısı deęiřkeni, rneklem byklę deęiřkeninin etkisinden farklı olarak daha ok lek dzeyinde yapılan analizleri etkiledięi belirlenmiřtir.

Elosua ve Wells (2013), çoklu grup doğrulayıcı faktör analizi, ordinal lojistik regresyon ve olabilirlik oran testini simülasyon çalışması ile farklı koşullar altında incelemişlerdir. Yapılan bu çalışmada örneklem büyüklüğü (her bir grup için 300–500–1000), değişen madde fonksiyonu türü (yalnızca b parametresinde, yalnızca a parametresinde ve hem a hem b parametresinde) ve değişen madde fonksiyonu büyüklüğü (düşük –0.25 ve yüksek –0.40) değişkenlerinde değişimleme yapılmıştır. Ayrıca Cattell’in kişilik envanterinden elde edilen madde parametreleri kullanılarak 15 maddelik 5’li likert yapıda veriler üretilmiştir. Çalışmada toplam 21 koşul mevcuttur [(3 örneklem büyüklüğü x3 değişen madde fonksiyonu türü x2 değişen madde fonksiyonu büyüklüğü) +3 değişen madde fonksiyonu içermeyen koşul]. Çalışmada her bir koşul için 100 tekrarlamaya gerçekleştirilmiştir. Elde edilen sonuçlara göre değişen madde fonksiyonu içeren maddelerin olduğu koşullarda olabilirlik oran testi için I. tip hata oranının arttığı saptanmıştır. Tek biçimli değişen madde fonksiyonu koşullarında olabilirlik oran testi ve çoklu grup doğrulayıcı faktör analizi yöntemi benzer performanslar sergilerken, tek biçimli olmayan değişen madde fonksiyonu olan koşullarda ise yalnızca olabilirlik oran testi yöntemi kabul edilebilir düzeyde performans sergilemiştir. Çoklu grup doğrulayıcı faktör analizi yönteminde istatistiksel güç oranlarının düşük değişen madde fonksiyonu büyüklüğü için 300/300 – 500/500 – 1000/1000 örneklem büyüklükleri için sırasıyla; %20 – %28 – %51; yüksek düzeyde değişen madde fonksiyonu için ise %59 – %72 – %80 arasında değişkenlik gösterdiği sonucu elde edilmiştir. Olabilirlik oran testi yöntemi için hesaplanan istatistiksel güç oranlarının ise aynı örneklem büyüklüklerinde düşük değişen madde fonksiyonu büyüklüğü koşulunda sırasıyla %22 – %31– %52; yüksek düzeyde değişen madde fonksiyonu için ise %62 – %72 – %74 arasında değişkenlik gösterdiği belirlenmiştir. Bu sonuçlara göre ortalama ve kovaryans eşdeğerliği ve olabilirlik oran testi yöntemlerinin örneklem büyüklüğü ve değişen madde fonksiyonu büyüklüğündeki artışa bağlı olarak istatistiksel güçlerinin arttığı rapor edilmiştir.

Yukarıda üzerinde durulan araştırmalar bir bütün olarak değerlendirildiğinde özetle ölçme eşdeğerliğinin belirlenme sürecine ilişkin olarak farklı bulgu ve görüşler bulunduğu görülmektedir. Ölçme eşdeğerliği aşamaları ile ilgili farklı görüşlerin bulunması nedeniyle, farklı grup karşılaştırmalarının yapıldığı çalışmalarda ölçme eşdeğerliğinin test edilme sürecinin göz ardı edilmesi ile karşılaşılabilmektedir. Ayrıca, gerçek uygulamalarda ortaya çıkabilecek farklı koşullarda farklı ölçme eşdeğerliği belirleme yöntemleri ile elde edilecek sonuçlara yönelik olarak araştırmacılara bilgi sağlayan kaynaklar sınırlıdır. Bu kaynakların sınırlı olması, ölçme eşdeğerliğinin test edilmesinin göz ardı edilmesine yol açabilmektedir. Bu nedenle farklı kuramsal temeller doğrultusunda geliştirilen yöntemlerin gerçek uygulamalarda

ortaya çıkabilecek koşullara benzer şekilde üretilen yapay veriler üzerinde denenmesi ve bu yöntemlerin hangi koşullara daha uygun olduğunun ortaya konulması gerekmektedir. Gerçekleştirilen bu araştırma örneklem büyüklüğü değişkeni düzeyleri, madde sayısı ve değişen madde fonksiyonu içeren madde oranı değişkenleri için değişimleme yapılmıştır. Alanyazında incelenen çalışmalardan farklı olarak örneklem büyüklüğü düzeyleri 250/250, 500/500 ve 1000/1000 olarak belirlenmiştir. Madde sayısı değişkeni için ise 20, 40, 60 olarak değişimlenmiştir. Değişen madde fonksiyonu içeren madde oranı değişkeni için %10, %20 ve %30 düzeyleri ele alınmıştır. Farklı koşullara sahip şekilde üretilen veriler kullanılarak bir çalışmanın yapılması ile farklı yöntemlerin farklı koşullar altında verdiği sonuçları, araştırmacılar tek bir çerçevede altında görebilecektir. Böylece araştırmacılar, yapmış oldukları uygulamalar için en uygun ölçme eşdeğerliği belirleme yönteminin hangisi olduğuna daha kolay karar verebileceklerdir. Çalışmada yapılan analizlerin diğer bölümünde ise gerçek veriler üzerinde analizler gerçekleştirilmiştir. Gerçek veriler üzerinde analiz sonuçları ile yapay verilerden elde edilen sonuçlar birlikte ele alınmıştır. Böylece yapay verilerden elde edilen sonuçların gerçek uygulama verileri üzerinde yapılan analiz sonuçları ile ne düzeyde örtüştüğü incelenmiştir.

1.2. Amaç

Bu araştırmanın genel amacı, farklı koşullar altında çoklu puanlanan verilerde kullanılabilen ölçme eşdeğerliği belirleme yöntemlerinin karşılaştırılmasıdır. Bu amaca yönelik olarak gerçek uygulamalarda karşılaşılabilecek farklı koşullar için en uygun ölçme eşdeğerliği belirleme yönteminin hangisi olabileceğine dair kanıt aranmıştır. Çoklu kategoride puanlanan, farklı örneklem büyüklüğü, farklı test uzunluklarına ve farklı oranlarda değişen madde fonksiyonuna sahip madde içerecek şekilde üretilmiş yapay verilerde, çoklu grup doğrulayıcı faktör analizi, madde parametrelerinin karşılaştırılması ve olabilirlik oran testi yöntemleri kullanılarak ölçme eşdeğerliği ile ilgili elde edilen sonuçlar karşılaştırılmıştır. Ayrıca yapay veri sonuçlarının gerçek uygulamalarda ortaya çıkan sonuçlarla desteklenebilmesi için kullanılan yöntemlerle gerçek uygulamadan elde edilmiş bir veri seti üzerinde de analizler yapılmış ve sonuçlar yapay veri sonuçları ile karşılaştırılmıştır. Bu genel amaç doğrultusunda araştırmada aşağıda belirtilen sorulara cevap aranmıştır:

- 1- Farklı örneklem büyüklükleri, farklı test uzunlukları ve farklı oranda değişen madde fonksiyonu içeren koşulların bulunduğu yapay veriler üzerinde;
 - a. Çoklu grup doğrulayıcı faktör analizi,
 - b. Olabilirlik oran testi yöntemi,

- c. Madde parametrelerinin karşılaştırılması yöntemi ile gerçekleştirilen ölçme eşdeğerliği analizlerinde elde edilen istatistiksel güç oranları nasıl değişmektedir?
- 2- Kullanılan üç yöntemin istatistiksel güç oranları karşılaştırıldığında nasıl bir farklılaşma ortaya çıkmaktadır?
- 3- Yapay veriler üzerinde gerçekleştirilen analizlere ek olarak gerçek veriler üzerinde gerçekleştirilen analiz sonuçlarında yöntemlerin istatistiksel güç oranları nasıl değişmektedir?

1.3. Önem

Gerçekleştirilen test uygulamalarının sonuçlarına göre eğitim ortamlarında, öğretim etkinliklerinde, eğitim politikalarında değişikliklere ve düzenlemelere gidilmekte, bireyler hakkında önemli kararlar verilmektedir. Bu uygulamalarda ölçme araçlarının uygulama yapılan alt gruplarda aynı özelliği ölçüyor olması (ölçme eşdeğerliğinin sağlanması), alınan kararların doğruluk ve isabetlilik düzeyi açısından önemlidir. Bu nedenle elde edilen sonuçlarda ortaya çıkan farklılıkların ölçme aracından kaynaklanmadığının kesinleştirilmesi gerekmektedir. Bu bağlamda, ölçme eşdeğerliğinin test edilmesine ilişkin kullanılan farklı yöntemlerin gerçek uygulamalarda değişkenlik gösterebilecek koşullar altında vermiş oldukları sonuçlar hakkında araştırmacılara bilgi sağlanması önem kazanmaktadır. Bu araştırmada farklı koşullar altında araştırmacılara, ölçme eş değeri yöntemlere ilişkin bilgiler sunulması, gerçek uygulamalarda elde edilen verilerin analizleri için uygun yöntemin belirlenmesi konusunda yardımcı olabilecektir. Buna ek olarak kullanılan yöntemlerin uygulama aşamalarına ilişkin verilen bilgiler, araştırmacıların ölçme eşdeğerliği analiz süreci ile ilgili doğru karar verebilmelerine katkı sağlayabilir. Bu araştırma sonuçları, gerçek veriler üzerinde toplanan verilerin özelliklerine göre ele alınan yöntemlerin kullanılabilirliklerine dair bilgi sağlayabilecektir.

Ölçme eşdeğerliğinin belirlendiği ve yapısal eşitlik modellemesini esas alan yöntemlerin kullanıldığı çalışmalarda, çoklu grup doğrulayıcı faktör analizi yönteminin kullanımı ile ilgili bilgi karmaşıklığı bulunmaktadır. Bu yöntem kullanılarak yapılan analizlerde, yalnızca kovaryans yapıları ve ortalama – kovaryans yapılarını birlikte incelenebilmektedir. Bu durum, ölçme eşdeğerliği analizlerinde araştırmacıların karışıklık yaşamalarına neden olmaktadır. Bu araştırmada sunulan sonuçlarla daha sonraki çalışmalar için bu yöntemle ilgili olan karmaşık noktalar giderilebilecek ve oluşabilecek koşullara uygun olan yöntem hakkında araştırmacılara bilgi sunulabilecektir.

Madde tepki kuramı altındaki çoklu puanlanan verilerde değişen madde fonksiyonu analizleri için alanyazın incelendiğinde, olabilirlik oran testi yönteminin en sık kullanılan yöntemlerden biri olduğu; madde parametrelerinin karşılaştırılması yönteminin ise araştırmacılar tarafından çok fazla tercih edilmediği görülmektedir. Ayrıca madde parametresi karşılaştırma yönteminin farklı koşullar altında kullanıldığı çalışmaların sınırlı sayıda olduğu belirlenmiştir. Madde parametrelerini karşılaştırma yönteminin kullanıcılar için daha kolay olmasına karşın (Korkmaz, 2005), az tercih edilmesinin nedenin istatistiksel güç oranları ile ilişkili olabileceğinin araştırılması gerekmektedir. Bu çalışma ile madde parametrelerini karşılaştırma yöntemine ilişkin istatistiksel güç oranları sunularak araştırmacılara bu yöntemle ilgili ayrıntılı bilgiler sağlanmaya çalışılmıştır. Araştırma bu yönüyle önemli bulunabilir.

Yapılan bu çalışmayla beraber madde tepki kuramı altında kullanılan iki yöntemin hem farklı koşullar altında vermiş olduğu sonuçlara hem de yapısal eşitlik modellemesi altındaki yöntemlerle benzerlik ve farklılıklara ilişkin olarak araştırmacılara katkı sağlanması beklenmektedir. Bütün bu koşullar göz önünde bulundurulduğunda yapılan bu çalışmanın, farklı koşulların elde edilebileceği diğer çalışmalara ölçme eşdeğerliği belirleme süreci konusunda ışık tutabilecek nitelikte olduğu için önemli bulunabilir.

1.4. Sınırlılıklar

- 1- Araştırmada üretilen yapay veriler 5'li likert tipinde çoklu puanlanan veriler ile sınırlıdır.
- 2- Araştırmada üretilen yapay veriler 0.43 (orta düzey) değişen madde fonksiyonu büyüklüğü ile sınırlıdır.
- 3- Araştırmada üretilen yapay veriler tek boyutlu yapıdadır.
- 4- Araştırmada üretilen yapay verilerde maddelere yüklenen değişen madde fonksiyonu tek biçimlidir.
- 5- Araştırma kapsamında her bir veri seti bir odak ve bir referans grup için üretilmiştir.
- 6- Araştırma kapsamında kullanılan yöntemler çoklu grup doğrulayıcı faktör analizi, olabilirlik oran testi ve madde parametrelerini karşılaştırma yöntemi ile sınırlıdır.
- 7- Araştırmada üretilen verilere ilişkin betimsel istatistiklerin belirlenme sürecinde, değişen madde fonksiyonu içeren maddeler toplam puan dışında bırakılmamıştır.

1.5. Kısaltmalar

DFA: Doğrulayıcı Faktör Analizi

DMF: Değişen Madde Fonksiyonu

DMFİ: Değişen Madde Fonksiyonu İstatistiği

KUI: Karşılaştırmalı Uyum İndeksi

UIİ: Uyumun İyiliği İndeksi

MPK: Madde Parametrelerinin Karşılaştırılması

MTK: Madde Tepki Kuramı

ÇGDFA: Çoklu Grup Doğrulayıcı Faktör Analizi

OOT: Olabilirlik Oran Testi

YEM: Yapısal Eşitlik Modellemesi



2. YÖNTEM

Bu bölümde araştırmanın modeli, verilerin üretilmesi, verilerin analizi ve yorumlanması açıklanmaktadır.

2.1. Araştırma Modeli

Bu araştırmada, belirlenen koşullara göre üretilen yapay veriler üzerinde ölçme eşdeğerliği test etme yöntemlerinin istatistiksel güç oranlarını karşılaştırmak ve bu doğrultuda ölçme eş değeri süreci ile ilgili olarak araştırmacılara bilgi sağlamak amaçlanmaktadır. Araştırma, bu yönüyle kuramsal çalışmalara katkı sunduğundan temel bir araştırmadır. Karasar (2005), temel araştırmaların amacının “var olan bilgiye yenilerini katmak” olduğunu belirtmiştir.

2.2. Veriler

Bu başlık altında araştırma kapsamında kullanılan verilerin özellikleri açıklanmıştır. Araştırma kapsamında yöntemlerin karşılaştırılması yapay ve gerçek veriler üzerinde gerçekleştirilmiştir. Yapay veriler araştırma kapsamında belirlenen değişkenlerin değişimlenmesi ile farklı koşullar içerecek şekilde üretilmiştir. Gerçek veri seti ise daha önce yapılan bir çalışma kapsamında maddeleri için değişen madde fonksiyonu incelemesi gerçekleştirilmiş olan bir veri seti olacak şekilde belirlenmiştir. Yapay ve gerçek veri setleri ile ilgili bilgiler ayrı ayrı sunulmuştur.

2.2.1. Yapay veri

Araştırmanın amacına ilişkin gerçekleştirilen analizlere yönelik belirlenen ölçme eşdeğerliği test yöntemleri için iki farklı gruba ait çok kategorili veriler üretilmiştir. Buna ek olarak ele alınan yöntemlerin istatistiksel güç oranı bakımından farklılık gösterdiği yapay veri koşullarına benzer özelliklere sahip gerçek veri setleri üzerinde analizler tekrarlanmıştır. Gerçek veriler üzerinde gerçekleştirilen analizlerden elde edilen sonuçlar yapay veriden elde edilmiş olan sonuçlarla karşılaştırılmaya çalışılmıştır. Üretilen verilerin bazı koşullarında değişimleme yapılmış, bazı koşullar ise alt gruplar için sabit olarak alınmıştır. Değişimleme yapılan ve sabit tutulan koşullar Çizelge 1’de görülmektedir.

Çizelge 1.Yapay veri üretim koşulları.

Değişimleme yapılan koşullar	Sabit tutulan koşullar
Örneklem büyüklüğü	Tepki kategori sayısı
Madde sayısı	Madde tepki kuramı modeli
Değişen madde fonksiyonu içeren madde oranı	Değişen madde fonksiyonu türü
	Değişen madde fonksiyonu büyüklüğü
	Bireylerin yetenek dağılımları

Değişimleme yapılan koşullarda örneklem büyüklüğü düzeyleri için, gerçek veri kullanılan araştırmalarda karşılaşımla sıklığı fazla olduğu düşünüldüğünden 250/250, 500/500, 1000/1000 olacak şekilde veri üretimi gerçekleştirilmiştir. Bunun yanı sıra madde sayısı değişkeni için 20, 40 ve 60 madde kullanılmıştır. Değişen madde fonksiyonu içeren madde oranı ise her %10, %20, %30 olacak şekilde veri setleri oluşturulmuştur. Üretilen veriler için tekrarlama sayısı 100 olarak alınmıştır. Harwell, Stone, Hsu ve Kirişçi (1996) yapay veri çalışmalarında en az 25 tekrar yapılması gerektiğini belirtmiştir. Bunun yanında yapay veri kullanılarak yapılan çalışmalarda tekrarlama sayısının 25 ile 1000 arasında değiştiği görülmektedir (Bilican Demir, 2014; Bolt, 2002; Chang, Mazzeo ve Roussos, 1996; Fortmann-Johnson, 2007; Hidalgo Montesinos ve Lopez Pina, 2002; Koh ve Zumbo, 2008;). Buna göre 3 (örneklem büyüklüğü) x3 (madde sayısı) x3 (değişen madde fonksiyonu içeren madde sayısı) olmak üzere 27 farklı benzetim koşulu oluşmuştur. Her bir koşul 100 tekrar yapıldığında toplam 2700 veri dosyası elde edilmiştir. Yapay veri üretimi için R programı kullanılmıştır. Yapay veri koşullarını gösteren tablo, Ek-1’de sunulmuştur. Değişimleme yapılan ve sabit tutulan koşullar aşağıda açıklanmıştır.

2.2.1.1. Değişimleme yapılan koşullar

Örneklem Büyüklüğü

Alanyazın incelendiğinde, yapılan çalışmalar, yapısal eşitlik modellemesi ve madde tepki kuramı temelli ölçme eşdeğerliği belirleme yöntemlerinin, örneklem büyüklüğü değişkeninden etkilendiğini ortaya koymaktadır. Ölçme eşdeğerliği yöntemleri kullanılarak gerçekleştirilen çalışmalarda, örneklem büyüklüklerinin 400–2000 arasında değişiklik gösterdiği belirlenmiştir (Atalay, ve diğerleri, 2012; Bolt, 2002; Holmes Finch ve French, 2007; Huang, Church ve Katigbak, 1997; Kazelskis, Thames ve Reeves, 2004; Kim ve Cohen, 1998; Korkmaz, 2005; Nachtigall, Kroehne, Funke ve Steyer, 2003; Narayanan ve Swaminathan, 1994; Reise, Wideman ve Pugh, 1993; Somer ve diğerleri, 2009; Somer, 2004).

Alanyazındaki bu sonuçlar doğrultusunda, bu çalışmada örneklem büyüklüğü değişkeni için 500 (250/250), 1000 (500/500) ve 2000 (1000/1000) kişilik veriler üretilmiştir.

Madde sayısı

Ölçme eşdeğerliği belirleme çalışmalarında kullanılan yöntemlerin etkilendiği diğer bir faktör olarak madde sayısı değişkeni ele alınmaktadır. Alanyazında ulaşılabilen gerçek veya yapay veri kullanılarak yapılan çalışmalarda, madde sayılarının 6 ile 180 madde arasında değiştiği görülmektedir (Clauser, Mazor, 1998; Dodeen, 2004; Kazelskis ve diğerleri, 2004; Korkmaz, 2005; Atalay ve diğerleri, 2012). Yapılan araştırmalarda, madde sayısı değişkenlerinin geniş bir aralıkta değiştiği gözlenmektedir. Madde sayısı değişkeninin ölçme eşdeğerliği belirleme yöntemlerine olan etkisi ve gerçek uygulamalarda farklı sayıda maddeden oluşan ölçme araçlarının kullanılması durumu göz önüne alındığında, bu çalışmada bu değişken için değişimleme yapılması gerekli görülmüştür. Bu çalışmada, gerçek uygulama durumlarında karşılaşılabilecek koşullara uygun olması açısından 20, 40 ve 60 madde içeren yapay veriler üretilmiştir.

Değişen madde fonksiyonu içeren madde oranı

Bir testte bulunan değişen madde fonksiyonunu gösteren maddeler, testin geçerliğini olumsuz yönde etkilemektedir. Ayrıca değişen madde fonksiyonu içeren maddeler, değişen madde fonksiyonu belirleme yöntemlerinin gücünü de olumsuz etkilemektedir (Fidalgo, Mellenbergh ve Muniz, 2000; Wang ve Yeh, 2003; Holmes Finch ve French, 2007; Atalay ve diğerleri, 2012). Yapay veri kullanılarak gerçekleştirilen ve farklı ölçme eşdeğerliği belirleme yöntemlerinin kullanıldığı çalışmalarda, değişen madde fonksiyonu içeren madde sayısının birden fazla olması önerilmektedir (Su ve Wang, 2005). Yapılan çalışmalar incelendiğinde yapay veri ile yapılan çalışmalarda, değişen madde fonksiyonu oranının % 0 ile % 40 oranları arasında değiştiği görülmektedir (Rogers ve Swaminathan, 1993; Narayanan ve Swaminathan, 1994; Fidalgo ve diğerleri, 2000; Holmes Finch ve French, 2007). Elde edilen bu sonuçlar doğrultusunda üretilen yapay veriler için, %10, %20 ve %30 oranında değişen madde fonksiyonu içeren madde içeren veriler üretilmiştir. Bu oranlamalar doğrultusunda madde sayısı koşuluna göre her bir veri setinde farklı sayıda DMF içeren madde yer almaktadır. DMF içeren maddeler veri setinin ilk bölümünde yer almaktadır. %10, %20 ve %30 oranları için, 20 maddelik veri setlerinde sırasıyla ilk 2, ilk 4 ve ilk 6 madde; 40 maddelik veri setlerinde sırasıyla ilk 4, ilk 8 ve ilk 12 madde; 60 maddelik veri setlerinde sırasıyla ilk 6, ilk 12 ve ilk 18 madde DMF içermektedir.

2.2.1.2. Sabit tutulan koşullar

Tepki kategori ve boyut sayısı

Üretilen yapay veriler için maddelere ait tepki kategori sayısı, gerçek uygulamalarda psikolojik özelliklerin belirlenmesinde kullanılan ölçme araçları düşünülerek “beş” olarak alınmıştır. Yapay verilerle gerçekleştirilen çalışmalarda tepki kategori sayılarının en az dört alındığı belirlenmiştir (Dodeen, 2004; Garrett, 2009; Bilican Demir, 2014). Ayrıca madde tepki kuramı varsayımlarından tek boyutluluğun karşılanabilmesi için veri setleri tek boyutlu olarak üretilmiştir.

Madde tepki kuramı modeli

Üretilen yapay verilerde cevap örüntülerinin elde edilmesi için Samejima'nın (1968) aşamalı tepki modeli (Graded Response Model) kullanılmıştır. Sıralı ve çoklu puanların matematiksel modellerini içeren bu model performans ve tutum gibi çoklu puanlamalarla ölçülebilen özelliklerin ölçülmesine uygundur (Samejima, 1997). Aşamalı tepki modeli, matematiksel olarak 2 parametrelili lojistik modelin, $K - 1$ sayıda 2 parametrelili lojistik modellerin genelleştirilmiş halidir. Burada “K” tepki kategori sayısını vermektedir (Embretson ve Reise, 2000). Gerçek ve yapay veri üzerinden gerçekleştirilen çalışmalarda aşamalı tepki modelinin kullanıldığı görülmektedir (Cohen, Kim ve Baker, 1993; Kim ve Cohen, 1998; Sireci ve Berberoğlu, 2000; Bolt, 2002; Başokçu ve Öğretmen, 2013; Arıkan Akın, 2015; Köse, 2015)

Aşamalı tepki modelinde bir madde için bir ayırt edicilik parametresi ve tepki kategori sayısının bir eksiği kadar eşik parametresi kestirilir. Bu doğrultuda bir madde için değişen madde fonksiyonu üç farklı koşulda ortaya çıkabilir. Bu koşullar:

- Madde ayırt edicilik parametrelerinin farklı, kategori eşik parametrelerinin eşit olduğu koşul,
- Madde ayırt edicilik parametrelerinin eşit, kategori eşik parametrelerinin farklı olduğu koşul,
- Madde ayırt edicilik parametrelerinin ve kategori eşik parametrelerinin farklı olduğu koşuldur.

Değişen madde fonksiyonu türü

Aşamalı tepki modelinde maddeler, alt gruplarda eşik ve ayırt edicilik parametreleri bakımından farklılaşabilir. Bu durumda maddeler için tek biçimli ve tek biçimli olmayan değişen madde fonksiyonu ortaya çıkabilmektedir. Bu çalışmada maddelerin sadece eşik parametreleri bakımından farklılaştığı koşullar için analizler gerçekleştirileceğinden üretilen verilerde tek biçimli değişen madde fonksiyonlarının olduğu koşullar kullanılmıştır.

Değişen madde fonksiyonu büyüklüğü

Alanyazın incelendiğinde yapılan çalışmalarda DMF büyüklüğü olarak 0.10, 0.15, 0.25, 0.30, 0.32, 0.43, 0.50, 0.53 ve 0.60 gibi değerlerin kullanıldığı görülmektedir (Ankeman ve diğerleri, 1999; Fidalgo ve diğerleri, 2000; Su ve Wang, 2005; González Roma ve diğerleri, 2006; Atar, 2007; Erdem Keklik, 2014; Kim ve Yoon, 2011; Bilican Demir, 2014). Bu çalışmada, yapılan çalışmalarda değişimlenen büyüklükler ve Educational Testing Service tarafından sınıflama göz önüne alınarak değişen madde fonksiyonu büyüklüğü belirlenmiştir. Bu sınıflamada;

- A düzeyi (ihmal edilebilir) DMF için 0.43 lojit birimden daha düşük düzeyde;
- B düzeyi (orta düzey) için 0.43 lojit birimden manidar şekilde farklılaşmayacak düzeyde;
- C düzeyi (yüksek düzey) için ise 0.64 lojit birime eşit veya daha büyük düzeyde bir farklılaşma olmalıdır (Golia, 2012; Zwick, 2012).

Bu çalışmada değişen madde fonksiyonu için belirlenen 0.43 lojit birimlik parametreler arası değişim, değişen madde fonksiyonu büyüklüğü olarak alınmıştır. Değişen madde fonksiyonu için belirlenen bu büyüklük yapılan çalışmalarda orta ve yüksek düzey olarak ele alınmaktadır (Stark ve diğerleri, 2006; Atar ve Kamata, 2011; Elosua ve Wells, 2013). Ayrıca Meade ve Lautenschlager (2004). 0.40 birimlik değişen madde fonksiyonu büyüklüğünün, madde parametrelerinde alt gruplar arasında yeterli düzeyde farklılaşma yaratacağını belirtmişlerdir.

Bireylerin yetenek dağılımları

Roussos ve Stout (1996) ve Tian (1999), gerçek test uygulamaları ve uzman görüşüne dayanarak grupların yetenek dağılım ortalamaları arasında 0.50 ve 1.00 standart sapmalılık farkların gerçek uygulamalara yakın olduğunu vurgulamıştır. Bu doğrultuda gerçekleştirilen çalışmada alt grupların yetenek dağılımları ortalaması 0.00 ve standart sapması ise 1.00 olarak birim normal dağılım gösterecek şekilde sınırlandırılmıştır [N (0.1)].

27 farklı koşulda üretilen 100 tekrarlama dosyası tek veri dosyası içerisinde toplanmıştır. Bu sayede elde edilen bulgular için yapılan yorumların anlaşılabilirlik düzeyinin artırılması hedeflenmiştir. Üretilen veri setleri analiz sürecinde yaşanabilecek karışıklıklar düşünülerek ayrı ayrı kod numaraları verilmiştir. Her bir koşul için üç haneli kod numaraları oluşturulmuştur. Bu kodlardaki rakamların anlamları şu şekildedir:

XYZ = ÖRNEKLEM BÜYÜKLÜĞÜ/MADDE SAYISI/DMFLİ MADDE ORANI(%).

Oluşturulan veri setleri için üretilen kodlar Çizelge 2’de verilmiştir.

Çizelge 2. Veri setleri için oluşturulan kod numaraları.

KOŞUL NUMARASI (XYZ)	TANIM [ÖRNEKLEM BÜYÜKLÜĞÜ/MADDE SAYISI/DMFLİ MADDE ORANI(%)]
111	500/20/10
112	500/20/20
113	500/20/30
121	500/40/10
122	500/40/20
123	500/40/30
131	500/60/10
132	500/60/20
133	500/60/30
211	1000/20/10
212	1000/20/20
213	1000/20/30
221	1000/40/10
222	1000/40/20
223	1000/40/30
231	1000/60/10
232	1000/60/20
233	1000/60/30
311	2000/20/10
312	2000/20/20
313	2000/20/30
321	2000/40/10
322	2000/40/20
323	2000/40/30
331	2000/60/10
332	2000/60/20
333	2000/60/30

Veri üretiminden sonra yapılan bir diğer işlem ise veri dosyaları için betimsel istatistiklerden aritmetik ortalama, standart sapma, çarpıklık ve basıklık katsayılarının hesaplanmasıdır. 27 koşul için üretilen verilere ait betimsel istatistikler Çizelge 3’te verilmiştir.

Çizelge 3. Veri dosyaları için hesaplanan betimsel istatistiklere ait sonuçlar.

KOŞUL NUMARASI	Aritmetik ortalama	Standart Sapma	Çarpıklık Katsayısı	Basıklık katsayısı
111	53.363	15.672	0.250	-0.573
112	53.003	15.654	0.238	-0.593
113	52.710	15.722	0.267	-0.595
121	105.670	29.903	0.256	-0.587
122	105.594	30.218	0.275	-0.595
123	107.887	30.103	0.277	-0.562
131	160.600	44.284	0.231	-0.577
132	161.064	44.661	0.236	-0.623
133	161.364	45.058	0.232	-0.593
211	53.661	15.826	0.223	-0.605
212	53.607	15.771	0.241	-0.585
213	53.332	15.616	0.241	-0.571
221	106.604	30.150	0.259	-0.576
222	107.535	30.351	0.237	-0.607
223	106.807	30.188	0.250	-0.601
231	159.903	44.660	0.263	-0.587
232	160.027	44.989	0.250	-0.620
233	160.315	45.052	0.256	-0.614
311	53.716	15.813	0.226	-0.612
312	53.127	15.753	0.262	-0.583
313	53.003	15.761	0.262	-0.592
321	107.652	30.365	0.226	-0.606
322	107.195	30.221	0.242	-0.607
323	107.478	30.388	0.227	-0.615
331	160.801	44.629	0.248	-0.590
332	159.419	44.712	0.267	-0.586
333	159.863	44.956	0.271	-0.595

Çizelge 3'te yer alan sonuçlara göre üretilen veri setlerinin aritmetik ortalama değerleri 52.710 ile 161.364 değerleri arasında değişkenlik göstermektedir. 20 madde içerecek şekilde üretilen verilerin aritmetik ortalamaları 52.710 ile 53.716, 40 madde içerecek şekilde üretilen verilerin aritmetik ortalamaları 105.594 ile 107.887; son olarak 60 madde içerecek şekilde üretilen verilerin aritmetik ortalamaları ise 159.419 ile 161.364 arasında değişmektedir. Tüm veri setleri için standart sapma değerleri incelendiğinde ise değerlerin 15.616 ile 45.058 arasında olduğu görülmektedir. Ayrıca tüm koşul dosyalarının çarpıklık ve basıklık katsayılarının ± 1 aralığında yer aldığı görülmektedir.

2.2.2. Gerçek veri

Bu araştırma kapsamında üretilen veri setlerinde elde edilen sonuçlarla karşılaştırmak için Atalay Kabasakal ve Kelecioğlu (2012) tarafından yapılan çalışmada analiz edilen veri seti kullanılmıştır. İncelenen bu çalışmada, PISA 2006 öğrenci anketi verileri üzerinde ordinal lojistik regresyon ve poly-SIBTEST yöntemleri ile değişen madde fonksiyonu analizleri

gerçekleştirilmiştir. Analiz sürecinde tutum yapısını ölçmek için kullanılan beş ölçeğin oluşturduğu 32 madde kullanılmıştır.

Yapılan incelemelerde kullanılmak üzere ST18 kodlu, “Fen bilimlerine verilen değer” alt ölçeğinin “Genel değer” alt boyutunda yer alan beş maddenin Türkiye ve Amerika Birleşik Devletleri verileri tercih edilmiştir. Bu ölçeğin tercih edilmesinin nedeni, ölçek kapsamında yer alan maddelerin tümünün ihmal edilebilir, orta ve yüksek düzeyde değişen madde fonksiyonu içeriyor olmasıdır. Beş maddeden 1, 4 ve 5. maddelerin analiz yapılan yöntemlerin her ikisi tarafından yüksek düzeyde; 2. maddenin ordinal lojistik regresyon tarafından yüksek, poly-SIBTEST tarafından ihmal edilebilir düzeyde; 3. maddenin ise ordinal lojistik regresyon tarafından orta, poly-SIBTEST tarafından ihmal edilebilir düzeyde değişen madde fonksiyonu gösterdiği saptanmıştır. Ölçekteki maddelerin tüm düzeylerde değişen madde fonksiyonu içerme durumu karşılaştırmanın anlamlılığı açısından önemlidir. Bunun nedeni ÇGDFA, OOT ve MPK yöntemlerinin değişen madde fonksiyonu düzeyine olan duyarlılığı ile ordinal lojistik regresyon ve poly-SIBTEST yöntemlerinin duyarlılığının farklı olmasıdır. Bunun yanı sıra, bu çalışma kapsamında ele alınan değişen madde fonksiyonu büyüklüğünün (0,43 lojit birim) ordinal lojistik regresyon ve poly-SIBTEST yöntemleri ile gerçekleştirilen analizlerde ne düzeyde bir sapmaya yol açacağının bilinmiyor olmasıdır. Ayrıca tüm maddelerin tek biçimli değişen madde fonksiyonu içerdiği sonucu elde edilmiştir. Bu çalışmanın kapsamında incelenen verilerin tek biçimli değişen madde fonksiyonu içerecek şekilde üretilmiş olması karşılaştırmanın anlamlı hale gelmesini sağlayan bir diğer nedendir.

Bu araştırma kapsamında gerçek veriler için, yapay veriler üzerinde değişimlenen koşullardan sadece örneklem büyüklüğü değişkenine ilişkin değişimleme yapılmıştır. Bunun nedeni yapay veriler için değişimlenen diğer koşullar (madde sayısı ve DMF içeren madde oranı) için aynı veri seti üzerinde müdahale edebilmenin mümkün olmamasıdır. Gerçek veriler üzerinde araştırma sürecinde ele alınan örneklem büyüklüğü düzeylerine uygun olacak şekilde belirlenen odak ve referans gruplarından SPSS 20.0 paket programı aracılığı ile rastgele veri seçimi yapılmıştır.

2.3. Verilerin Analizi

Araştırmada üretilen yapay veriler üzerinde ölçme eşdeğerliğinin belirlenmesi için aşamalı test yöntemi kullanılmıştır. Farklı özelliklere sahip koşullar içeren yapay veriler üzerinde ise yapısal eşitlik modellemesi altındaki yöntemlerden çoklu grup doğrulayıcı faktör analizi yöntemi ile madde tepki kuramı altındaki yöntemlerden olabilirlik oran testi, madde parametrelerinin karşılaştırılması yöntemleri ile analizler gerçekleştirilmiştir. Kullanılan bu

yöntemlerle gerçekleştirilen analiz sürecinde izlenen aşamalara ilişkin bilgiler ayrı ayrı sunulmuştur.

2.3.1. Çoklu grup doğrulayıcı faktör analizi ve ölçme eşdeğerliği

Alanyazında YEM temelli yöntemler ile ölçme eşdeğerliği aşamaları iki farklı şekilde görüş mevcuttur. Bu görüşlerden ilkinde ölçme eşdeğerliği testleri, en az sınırlandırılmış modelden en fazla sınırlandırılmış modele doğru (aşamalı test etme); ikincisinde ise en fazla sınırlandırılmış modelden en az sınırlandırılmış modele doğru (geriye doğru aşamalı test etme) gerçekleştirilmektedir. Geriye doğru test etme yönteminin en fazla sınırlandırılmış modelden analizine başlandığından, bu yöntemin dezavantajlı olduğu belirtilmektedir (Meredith, 1993). Bunun nedeni fazla sınırlandırılmış model kullanıldığında ölçme eşdeğerliğinin sağlanamaması durumunda, eşdeğerliği bozan varyans kaynağının bulunabilmesi için diğer tüm aşamaların gerçekleştirilmesinin gerekli olmasıdır. Ayrıca aşamalı test etme yöntemi kullanılarak yapılan çalışmalarda, ölçme eşdeğerliğinin sağlanamadığı durumlarda kısmi eşdeğerliğin test edilme sürecine geçilerek daha ileri adımlara devam edilebilmektedir (Brown, 2015). Ölçme eşdeğerliği aşamaları konusunda farklı görüşler bulunsa da belirtilen nedenlerden dolayı aşamalı test etme yönteminin kullanılması önerilmektedir. Alanyazında önerilen aşamalı test etme yönteminde izlenen adımlar arasında benzerlik olduğu görülmektedir. Aşamalı test etme yöntemi olarak önerilen iki yaklaşımda hangi adımların izlendiği Çizelge 4'te gösterilmektedir. Ölçme eşdeğerliği test etmede aşamalı test etme yolu adım

Çizelge 4. Ölçme eşdeğerliği test etmede aşamalı test etme yolu adımları.

Meredith (1993)'e göre	Vandenberg ve Lance (2000)'e göre	
• Yapısal değişmezlik • Metrik değişmezlik • Ölçek değişmezliği • Katı değişmezlik	• Yapısal değişmezlik • Metrik değişmezlik • Ölçek değişmezliği • Özgül varyansların eşdeğerliği • Faktör varyanslarının eşdeğerliği • Faktör kovaryanslarının eşdeğerliği • Faktör ortalamalarının eşdeğerliği	} Ölçme modeline ilişkin Yapısal modele ilişkin

Geriye doğru aşamalı test etme yöntemine olan üstünlüklerinden dolayı ölçme eşdeğerliği belirleme sürecinde kullanımı önerilen aşamalı test etme yöntemine ilişkin alanyazında farklı öneriler bulunmaktadır. Ancak yapılan araştırmalar incelendiğinde Meredith (1993) tarafından önerilen aşamaların, araştırmacılar tarafından tercih edildiği görülmektedir (Byrne ve Stewart, 2006; Van de Vijver, 1998; Wu ve diğerleri, 2007). Ayrıca çoklu grup

doğrulamalı faktör analizi yöntemi kullanılarak gerçekleştirilen ölçme eşdeğerliği analiz sürecinde, aşamalı test etme yolunun izlenmesi önerilmektedir (Meredith, 1993; Vandenberg ve Lance, 2000). Bu araştırmada da, Meredith (1993) tarafından önerilen aşamalı test etme yolu izlenerek ölçme eşdeğerliği analizleri gerçekleştirilmiştir. Ölçme eşdeğerliğinin test edilmesine ilişkin önerilen aşamalar şunlardır:

2.3.1.1. Yapısal/Şekilsel değişmezlik (Configural invariance)

2.3.1.2. Metrik değişmezlik (Metric invariance)

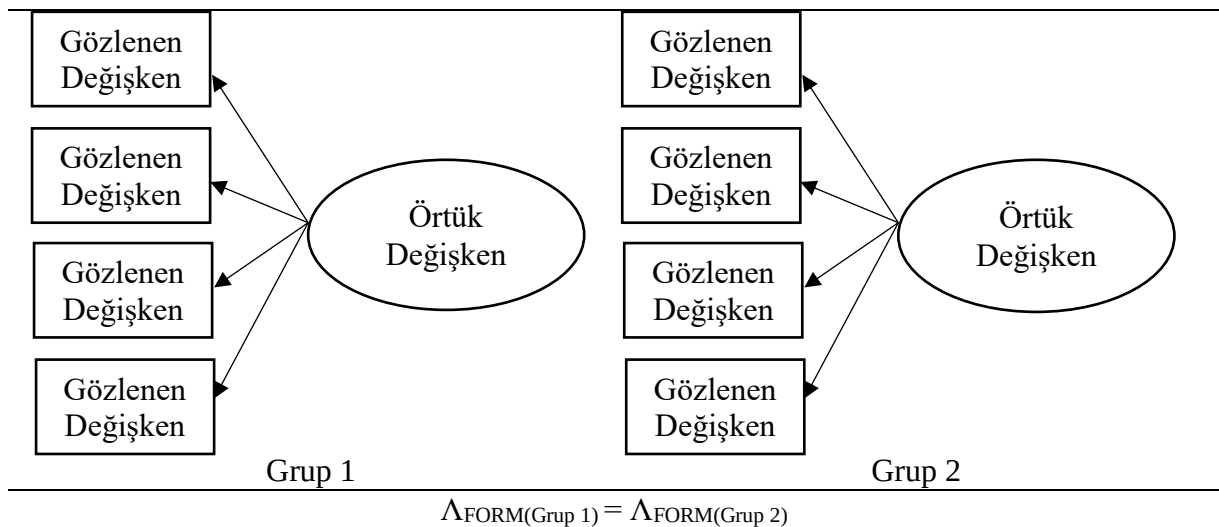
2.3.1.3. Ölçek değişmezliği (Scalar invariance)

2.3.1.4. Katı değişmezlik (Strict invariance).

2.3.1.1 Yapısal/Şekilsel değişmezlik (Configural invariance)

Ölçme eşdeğerliğinin test edilmesindeki temel aşamadır. Bu aşamada farklı gruplara uygulanan ölçme araçlarının, her grupta aynı faktör yapısına sahip olup olmadığı incelenir. Başka bir ifadeyle, yapısal değişmezlik aşamasında farklı gruplara uygulanan ölçme araçlarının faktör sayısı ve örüntülerinin gruplar için eş değer olup olmadığı test edilir. Yapısal eşdeğerliğin test edilmesinde parametreler üzerinde gruplar arası herhangi bir sınırlandırma yapılmaz. Faktör yükleri, hata ve faktör varyansları serbest bırakılır. Yani parametrelerin her bir grupta farklı değerler alabilmelerine izin verilir.

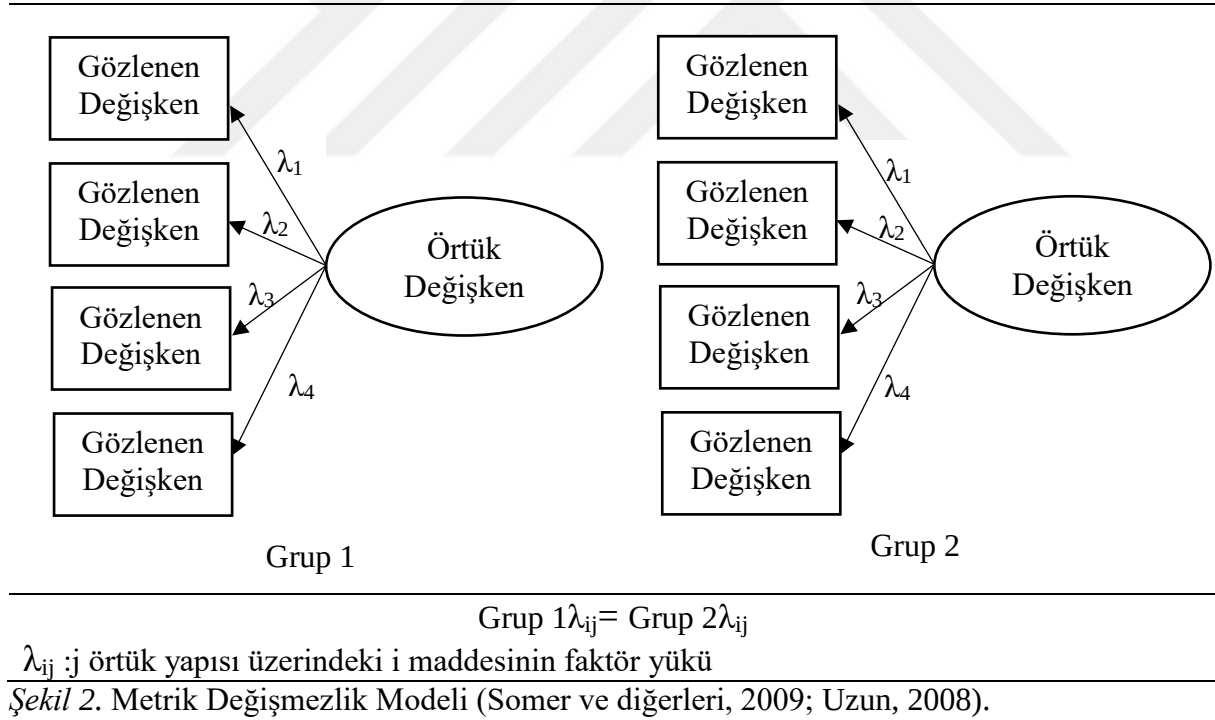
Diğer ölçme eşdeğerliği aşamalarının test edilebilmesi için yapısal eşdeğerliğin sağlanması önemlidir. Çünkü herhangi bir sınırlamanın olmadığı modelin desteklenmemesi, sınırlayıcı modellerin de desteklenmeyeceği anlamına gelmektedir. Nitekim yapısal eşitliğin sağlanamadığı durumlarda diğer değişmezlik aşamalarının kontrolüne geçmek mümkün değildir (Bollen, 1991). Yapısal değişmezlik modeline ilişkin iki farklı şematik gösterim Şekil 1’de sunulmuştur:



Şekil 1. Yapısal Değişmezlik Modeli (Somer ve diğerleri, 2009; Uzun, 2008).

2.3.1.2. Metrik değişmezlik (Metric invariance)

Yapısal değişmezlik, farklı gruplarda bulunan bireylerin maddelere verdikleri cevapların veya maddenin farklı gruplardaki bireyler için benzerliğini test etmez. Bu durumun metrik değişmezlikle test edilmesi gerekir (Steenkamp ve Baumgartner, 1998). Metrik değişmezlik aşamasında ölçme aracında yer alan maddelere ait faktör yüklerinin (λ) ölçme yapılan gruplardaki eşdeğerliği test edilir (Vandenberg ve Lance, 2000). Metrik değişmezlik sınırlayıcı model içermesi nedeniyle yapısal değişmezlikten istatistiksel açıdan daha güçlü bir testtir. Metrik eşdeğerliğin test edilmesinde söz konusu ölçme aracında yer alan maddelere ilişkin faktör yüklerinin (λ) gruplar arası eşit/değişmez olduğu şeklindeki bir hipotez test edilmektedir: $\lambda^{g1} = \lambda^{g2}$ (Önen, 2009). Bu aşamada faktör yükleri gruplar arasında eşitlenir (Hong, Malik ve Lee, 2003). Metrik eşdeğerliğin sağlanmasından bahsedilebilmesi için gözlenen değişkenleri örtük değişkenlere bağlayan regresyon eğimleri paralel olmalıdır (Ekermans, Saklofske, Austin ve Stough, 2011). Metrik eşdeğerliğin sağlanması durumunda maddelerin farklı gruplarda aynı şeyi ifade ettiği belirtilebilir. Metrik değişmezlik modeline ilişkin iki farklı şematik gösterim Şekil 2’de sunulmuştur:



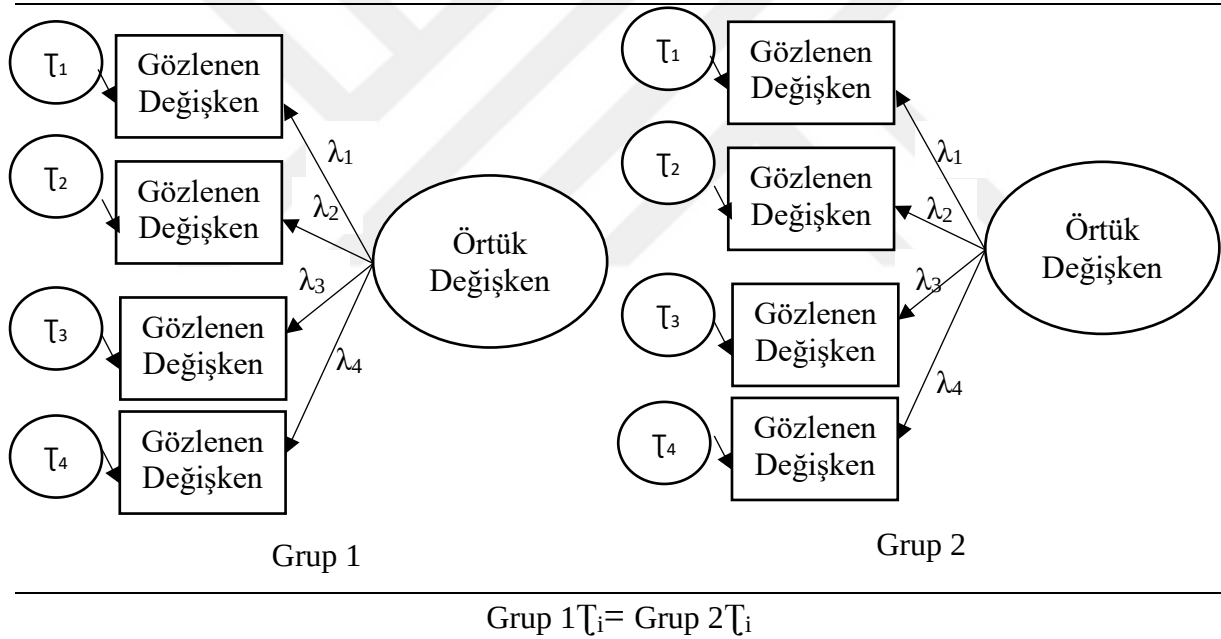
Şekil 2. Metrik Değişmezlik Modeli (Somer ve diğerleri, 2009; Uzun, 2008).

2.3.1.3. Ölçek değişmezliği (Scalar invariance)

Ölçek değişmezliği aşamasında ise yapısal değişmezlikteki faktör örüntüsü, metrik değişmezlik aşamasındaki faktör yüklerine ek olarak regresyon, farklı gruplar için eşitlenmektedir (Brown, 2015; Hong ve diğerleri, 2003). Ölçek değişmezliği, eşit orijinler gerektirmektedir. Başka bir ifadeyle, ölçülmek istenen örtük özellik bakımından aynı düzeyde

olan farklı gruptaki bireylerin gözlenen değişkenlerden aynı puanı alması gerekmektedir. Bu durum ancak maddelere ilişkin regresyon sabitlerinin (τ), gruplar arası eşdeğerliği sağlandığında mümkündür. Son yıllarda ölçme eşdeğerliği araştırmalarında farklı gruplar arasında regresyon sabitinin de eş değer olup olmadığının incelenmesi gerektiği vurgulanmaktadır (Somer ve diğerleri, 2009). Bu bağlamda ölçek değişmezliği, madde regresyon sabitleri vektörünün gruplar arası eş değer olduğu şeklindeki bir hipotezin test edilmesi yoluyla incelenmektedir: $\tau_{g1} = \tau_{g2}$. Regresyon sabiti, örtük değişken sıfır değeri aldığı anda, gözlenen değişkenin alacağı değere eşittir (Salzberger, Sinkovics ve Schlglmich, 1999).

Madde regresyon sabitleriyle ilgili incelemeler, karşılaştırma grupları için yeterli sayıda eş değer regresyon sabitine sahip madde olduğuna işaret ediyorsa araştırmacılar, gruplar arasında gözlenen puanlardaki gerçek farklılıklara ilişkin istenen bilgilere ulaşabilir. Ölçek değişmezliği modeline ilişkin iki farklı şematik gösterim Şekil 3'te sunulmuştur:



τ_i : i maddesinin regresyon sabiti

Şekil 3. Ölçek değişmezliği Modeli (Somer ve diğerleri, 2009; Uzun, 2008).

2.3.1.4. Katı değişmezlik (Strict invariance)

Yapısal, metrik ve ölçek değişmezliği sağlandıktan sonra son aşamada maddelere ilişkin hata terimleri (ϵ) ile ilgili analiz yapılmaktadır. Hata terimi, gözlenen değişkendeki değişimin, bağlı olduğu örtük değişkendeki değişim tarafından açıklanamayan kısmına karşılık gelmektedir (Deshon, 2004). Katı değişmezlik aşamasında gözlenen değişkenin, ölçülmek istenen gruplarda eş değer hata miktarına sahip olup olmadığı incelenir. Yani bu aşamada ölçme

değer arasındaki fark; iki modelin serbestlik derecesi farkı için hesaplanan kritik Ki-Kare değeri ile karşılaştırılır. İki oran arasındaki farkın ($\Delta\chi^2$) manidar şekilde farklılaştığının belirlenmesi halinde ölçek değişmezliğinin sağlanmadığı sonucuna ulaşılmış olmaktadır.

Ki-Kare değerinin manidarlığının örneklem büyüklüğünden etkilenmesi ve model uyumunun daha doğru gerçekleştirilebilmesi için en az iki uyum indeksinin kullanımının gerekli olması nedeniyle (Vandenberg ve Lance, 2000), yapılan bu araştırmada $\Delta\chi^2$ değerine ek olarak karşılaştırmalı uyum indeksi (KUI) ve uyumun iyiliği indeksi (UIİ) değerleri arasındaki farklar da incelenmiştir. ΔKUI ve $\Delta UIİ$ değerlerinin incelenmesinde herhangi bir manidarlık testi yapılamazken, karşılaştırılan iki sınırlandırılmış modele ilişkin uyum indeksleri farkının $(-0.01) - (0.01)$ aralığında olması gerekmektedir (Cheung ve Rensvold, 2002; Wu ve diğerleri, 2007).

Çoklu grup doğrulayıcı faktör analizi yönteminde, aşamalara ilişkin olarak uyum indekslerinden elde edilen kanıtlara ek olarak maddelerin değişen madde fonksiyonu gösterip göstermediğine işaret eden kanıtlara modifikasyon indeksleri üzerinden ulaşılmaktadır. Elde edilen analiz çıktılarında M_i CONST şeklinde bir modifikasyon indeksi elde edilmesi, o maddenin değişen madde fonksiyonuna sahip olduğunu gösterir.

Bu araştırmada çoklu grup doğrulayıcı faktör analizi yönteminin uygulama sürecinde komut dizin dosyaları, Lisrel 8.72 programı kullanılarak hazırlanmış ve analizler gerçekleştirilmiştir. Analizlerde her bir veri seti için yapılan işlemler aşama aşama şu şekildedir:

- 1- Her bir koşulda üretilen veri setleri G1 ve G2 isimli .sav uzantılı olarak iki grup için düzenlenmiştir.
- 2- Düzenlenen veri setleri ayrı ayrı klasörlenmiştir. Bu işlemten sonra 2700 klasör oluşturulmuştur.
- 3- Oluşturulan her bir klasördeki veri özelliklerine uygun yapısal, metrik, ölçek ve katı değişmezlik aşamaları için dört farklı komut dizin dosyası (syntax) oluşturulmuştur. Komut dizin dosyaları örneği Ek-2’de sunulmuştur.
- 4- Komut dizin dosyalarının oluşturulmasından sonra her bir klasördeki veri setleri için kovaryans matrisleri ve ortalamaları içeren dosyalar program üzerinden elde edilmiştir.
- 5- Son olarak her bir klasör için komut dizin dosyaları kullanılarak sonuçlar elde edilmiştir.
- 6- Elde edilen analiz çıktıları excel programı yardımı ile düzenlenerek aşamalara ve maddelere ilişkin incelemelerden sonra sonuçlar sunulmuştur.

2.3.2. Madde tepki kuramı ve ölçme eşdeğerliği

Madde tepki kuramı ile yapılan ölçme eşdeğerliği analizleri değişen madde fonksiyonu analizlerinin incelenmesine dayanmaktadır. Madde tepki kuramı temel alınarak yapılan analizlerde, değişen madde fonksiyonu olmama durumu ölçme eşdeğerliği olarak tanımlanmaktadır. Değişen madde fonksiyonu, ölçme aracında ölçülmek istenen özelliğe sahip oluş düzeyleri aynı alt gruplarda bir maddenin doğru cevaplanma olasılığının farklılaşp farklılaşmadığını gösteren bir fonksiyondur (Zumbo, 1999). Camilli ve Shepard (1994) bir maddenin değişen madde fonksiyonu göstermesinin iki nedeni olabileceğini belirtmiştir. Bunlar, madde yanlılığı ya da gruplar arasındaki gerçek farklılıklardır. Gruplar arasındaki gerçek farklılıklar madde etkisi olarak tanımlanır (Zumbo, 1999). Madde etkisinin söz konusu olması elde edilen sonuçların sistematik hata içermediğinin kanıtıdır. Ancak madde etkisi söz konusu değilse, madde yanlılığına dair kuramsal incelemeler yapılmalı ve uzman görüşlerine başvurulması gerekmektedir.

Değişen madde fonksiyonu iki şekilde ortaya çıkabilir: tek biçimli ve tek biçimli olmayan değişen madde fonksiyonu. İnceleme yapılan madde, özelliğe sahip olma düzeylerinin tümünde aynı grup lehine işliyorsa tek biçimli; farklı düzeylerinde farklı gruplar lehine işliyorsa tek biçimli olmayan değişen madde fonksiyonundan söz edilebilir (Zumbo, 1999). Değişen madde fonksiyonu analizlerinde, uygulama yapılan bireyler belli bir değişkene göre (cinsiyet, sosyo-ekonomik düzey vb.) iki alt gruba ayrılır. Bu alt gruplardan yanıtlama açısından avantajlı olduğu düşünülen grup referans, dezavantajlı olduğu düşünülen ise odak grup olarak tanımlanır. Değişen madde fonksiyonunun olduğu ve olmadığı durumların matematiksel gösterimleri aşağıdaki gibidir:

Değişen madde fonksiyonu olmayan durum	Değişen madde fonksiyonu olan durum
$f(Y \mid \Theta, G = R) = f(Y \mid \Theta, G = O)$	$P(Y \mid \Theta, G = g) \neq P(Y)$

Denklemlerde bulunan Θ , ölçülmek istenen örtük değişkeni; Y , gözlenen değişkene verilen tepkileri; G , grup değişkenini; R , referans grubu ve O odak grubunu göstermektedir. Değişen madde fonksiyonu olma durumuna ilişkin verilen gösterimde, referans ve odak grupların madde yanıtlama olasılıklarının aynı olduğu; değişen madde fonksiyonu olan durumda ise alt gruplarda yanıtlama olasılıklarının farklılaştığı gösterilmektedir.

Yapılan analizlerde elde edilen sonuçların doğru bir biçimde yorumlanabilmesi için analiz yöntemlerine dair ayrıntıların bilinmesi gerekmektedir. Bu çalışmada madde tepki kuramı temelli olan iki yöntem kullanılmıştır.

Bunlar:

2.3.2.1. Olabilirlik oran testi

2.3.2.2. Madde parametrelerinin karşılaştırılması şeklindedir.

Bu iki yöntemin ölçme eşdeğerliği sürecinde kullanımına yönelik bilgiler, aşağıda sırasıyla sunulmuştur.

2.3.2.1. Olabilirlik Oran Tesi

Araştırmada olabilirlik oran testi analizlerinde MULTILOG 7.03 (Thissen, 1991) programı kullanılmıştır. Analizlerde her bir veri seti için yapılan işlemler aşama aşama şu şekildedir:

- 1- Her bir koşul için üretilen veri seti format olarak analiz ve programa uygun hale getirilmiştir. Veri seti formatına ilişkin örnek Ek-3'te sunulmuştur.
- 2- Düzenlenen veri setleri her bir koşul için oluşturulan 27 farklı klasör altında toplanmıştır.
- 3- Oluşturulan her bir klasördeki veri seti için ilk olarak sınırlandırılmış model komut dizin dosyası oluşturulmuştur. Daha sonra her bir madde için serbest model komut dizin dosyaları hazırlanmıştır. Tüm koşul ve tekrarlamalarda bulunan değişen madde fonksiyonu içeren maddeler için toplam 24300 komut dizin dosyası elde edilmiştir.
- 4- Her bir klasördeki veri setleri için komut dizin dosyaları çalıştırılarak sonuçlar elde edilmiştir.
- 5- Elde edilen analiz çıktıları düzenlenerek, her bir çıktıdan -2loglikelihood değerleri çekilmiş ve ardından çekilen bu değerlerin excel programı farkı alınmıştır. Alınan farklar, belirlenen serbestlik derecesindeki kritik Ki-Kare değeri ile karşılaştırılması yapılarak değişen madde fonksiyonu var olma durumu belirlenmiştir.

2.3.2.2. Madde Parametrelerinin Karşılaştırılması Yöntemi

Araştırmada madde parametrelerinin karşılaştırılması yöntemine ilişkin analizler, PARSCALE 4.1 programı (Muraki ve Bock, 1991) kullanılarak gerçekleştirilmiştir. Analizlerde her bir veri seti için yapılan işlemler aşama aşama şu şekildedir:

- 1- Her bir koşul için üretilen veri seti format olarak analiz ve programa uygun hale getirilmiştir. Veri seti formatına ilişkin örnek, Ek-4'te sunulmuştur. Veri dosyaları düzenlenirken, PARSCALE programı belli sayıda sütundan sonraki puanları analize dâhil edemediğinden, 40 ve 60 maddelik veri setleri sırasıyla iki ve üç satıra bölünmüştür. Başka bir ifadeyle 40 madde içeren veri setlerinde her bir birey için iki satır veri; 60 madde içeren veri setlerinde ise üç satır veri bulunmaktadır.

- 2- Düzenlenen veri setleri her bir koşul için oluşturulan tek bir klasör altında toplanmıştır.
- 3- Oluşturulan veri setlerinin her biri için bir adet komut dizin dosyası oluşturulmuştur. Tüm koşul ve tekrarlamalar için toplam 2700 komut dizin dosyası hazırlanmıştır. Komut dizin dosyaları oluşturulurken analiz edilecek veri setinin formatına uygun şekilde düzenlemeler yapılmıştır.
- 4- Her veri seti için komut dizin dosyaları R programına aktarılmıştır. Ardından analizler gerçekleştirilerek sonuçlar elde edilmiştir.
- 5- Elde edilen analiz çıktıları düzenlenerek, her bir çıktıdan madde parametreleri farkının manidarlığını göstermek için oluşturulan tablolar alınmıştır. Excel programı kullanılarak bu tablolarda parametre farklarının manidar olduğu maddeler belirlenerek, değişen madde fonksiyonu var olma durumu incelenmiştir.

İstatistiksel güç oranlarının belirlenmesi

Araştırma kapsamında kullanılan yöntemlerle analizler gerçekleştirildikten sonra her bir yöntem için ayrı ayrı istatistiksel güç oranları hesaplanmıştır. İstatistiksel güç hesaplaması sürecinde veri setinde DMF içeren maddelerin yöntemlerce belirlenmesi ilişkin oranlar yüzde olarak tespit edilmiştir. Oranların belirlenmesinde, uygulanan yöntem tarafından DMF içerecek şekilde üretilmiş maddelerden DMF içeren olarak belirlenenlerin sayısı, veri setindeki DMF içerecek şekilde üretilmiş toplam madde sayısına bölünmüş ve ardından yüzdelik olarak ortaya konmuştur.

3. BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın alt amaçlarına uygun olarak elde edilen bulgulara ve bu bulgularla ilgili yorumlara yer verilmiştir. Çalışma kapsamında ele alınan yöntemlerin her biri için örneklem büyüklüğü, madde sayısı ve değişen madde fonksiyonu içeren madde oranı değişkenleri değişimlenerek oluşturulan koşullardaki istatistiksel güç oranları belirlenmiştir. Oluşturulan 27 farklı koşulun her biri için veri setindeki değişen madde fonksiyonu içeren maddelerin doğru olarak belirlenme oranları hesaplanmış ve ortaya çıkan istatistiksel güç oranları çizelgeler yardımıyla açıklanmıştır. Değişimlenen değişkenlerin istatistiksel güç oranlarına etki yönlerinin net bir biçimde görülebilmesi amacıyla grafikler kullanılmıştır. Hesaplanan istatistiksel güç oranları her bir yöntem ve değişken için ayrı başlıklar halinde sunulmuştur. İstatistiksel güç oranlarına her bir yöntem için tek bir çizelgede verilmiştir. Ancak değişkenlerdeki değişimin etkisinin açıklanabilmesi için her değişken için ayrı bir grafik kullanılmıştır.

3.1. Çoklu Grup Doğrulayıcı Faktör Analizi yöntemi istatistiksel güç oranlarına ilişkin bulgular

Bu başlık altında *“Farklı örneklem büyüklükleri, test uzunlukları ve farklı oranda değişen madde fonksiyonu içeren maddeye sahip yapay veriler üzerinde çoklu grup doğrulayıcı faktör analizi yöntemi ile yapılan analizlerde elde edilen istatistiksel güç oranları nasıl değişmektedir?”* sorusu ile ilgili olarak elde edilen bulgu ve yorumlara yer verilmiştir. Çoklu grup doğrulayıcı faktör analizi yöntemi istatistiksel güç oranları çizelge 5’te verilmiştir. Değişimleme yapılan koşullar için elde edilen grafikler ise sırayla sunulmuştur.

Çizelge 5.ÇGDFA yöntemi ile elde edilen istatistiksel güç oranları.

Örneklem büyüklüğü		Madde sayısı	DMF içeren madde oranı	İstatistiksel güç oranı
Odak	Referans			
250	250	20	10	45.00
500	500	20	10	78.00
1000	1000	20	10	92.00
250	250	20	20	35.00
500	500	20	20	72.50
1000	1000	20	20	88.50
250	250	20	30	29.33
500	500	20	30	58.00
1000	1000	20	30	82.00
250	250	40	10	47.00
500	500	40	10	77.00
1000	1000	40	10	93.50
250	250	40	20	38.00
500	500	40	20	63.75
1000	1000	40	20	86.50
250	250	40	30	27.33
500	500	40	30	55.33
1000	1000	40	30	81.00
250	250	60	10	44.67
500	500	60	10	67.33
1000	1000	60	10	91.00
250	250	60	20	35.17
500	500	60	20	65.67
1000	1000	60	20	86.67
250	250	60	30	26.78
500	500	60	30	54.11
1000	1000	60	30	81.89

Çizelge 5' göre çoklu grup doğrulayıcı faktör analizi istatistiksel güç oranları %26.78 ile %93.50 arasında değişiklik göstermektedir. Bu yöntem için en düşük güç oranı odak ve referans grup örneklem büyüklüğü 250, madde sayısı 60 ve DMF içeren madde oranı %30 olduğunda elde edilmiştir. En yüksek istatistiksel güç oranı ise odak ve referans grup örneklem büyüklüğü 1000, madde sayısı 40 ve DMF içeren madde oranı %10 olan koşul için hesaplanmıştır. Değişimlenen değişkenlerin istatistiksel güç oranlarına etkisi her bir değişken için ayrı ayrı özetlenmiştir.

Örneklem büyüklüğü değişkeni için elde edilen istatistiksel güç oranları

Örneklem büyüklüğü değişkeni için çoklu grup doğrulayıcı faktör analizi yöntemi kullanılarak yapılan analizler sonucunda elde edilen istatistiksel güç oranları Çizelge 5'te verilmiştir.

Çizelge 5 incelendiğinde 20 maddelik veri setleri için elde edilen istatistiksel güç oranlarının, örneklem büyüklüğü değişkenine göre, %29.33 ve %92.00 arasında değişiklik gösterdiği görülmektedir. 20 maddelik veri setlerinde, DMF içeren madde oranı değişkeni sabit tutulduğunda, 250/250, 500/500, 1000/1000 örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %45.00 - %78.00 - %92.00;

%20 DMF içeren madde oranı koşulu için: %35.00 - %72.50 - %88.50;

%30 DMF içeren madde oranı koşulu için: %29.33 - %58.00 - %82.00.

Çizelge 5'te açıklanan bir diğer istatistiksel güç oranları, 40 maddelik veri setleri için hesaplanmıştır. 40 maddelik veri setleri için örneklem büyüklüğü değişkenine göre istatistiksel güç oranları, %27.33 ve %93.50 arasında değişiklik göstermektedir. 40 maddelik veri setlerinde DMF içeren madde oranı değişkeni sabit tutulduğunda, farklı örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %47.00 - %77.00 - %93.50;

%20 DMF içeren madde oranı koşulu için: %38.00 - %63.75 - %86.50;

%30 DMF içeren madde oranı koşulu için: %27.33 - %55.33 - %81.00.

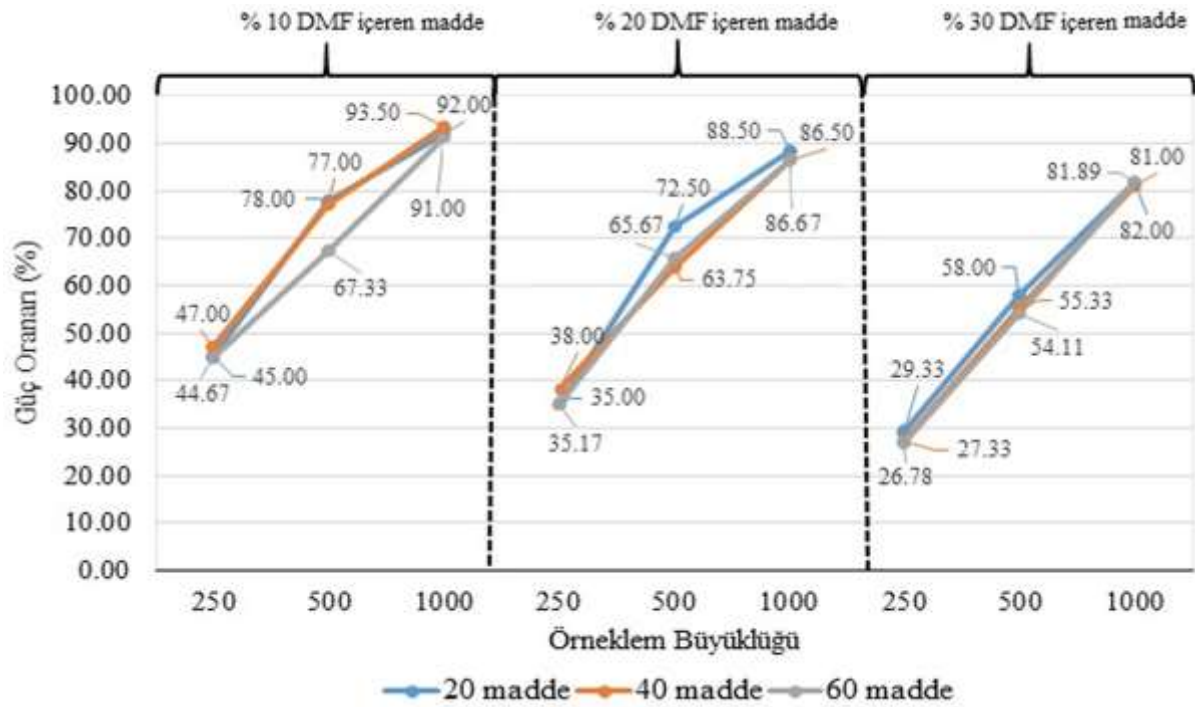
Son olarak Çizelge 5'te örneklem büyüklüğü değişkenine göre 60 maddelik veri setleri için hesaplanmış istatistiksel güç oranları, %26.78 ve %91.00 arasında değişiklik göstermektedir. 60 maddelik veri setlerinde DMF içeren madde oranı değişkeninin sabit olarak alındığı, farklı örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %44.67 - %67.33 - %91.00;

%20 DMF içeren madde oranı koşulu için: %35.17 - %65.67 - %86.67;

%30 DMF içeren madde oranı koşulu için: %26.78 - %54.11 - %81.89.

Örneklem büyüklüğü değişkenine göre istatistiksel güç oranlarındaki değişime ilişkin sonuçlar Grafik 1'de gösterilmiştir.



Grafik 1. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (ÇGDFA).

Grafik 1’de sunulan bulgulara göre madde sayısı ve değişen madde fonksiyonu içeren madde oranı değişkenlerinin tüm düzeylerinde, örneklem büyüklüğü arttıkça ÇGDFA yöntemi için elde edilen istatistiksel güç oranlarının arttığına işaret etmektedir. ÇGDFA yöntemi için en yüksek oranlar örneklem büyüklüğü 1000 olduğunda elde edilmiştir. En yüksek istatistiksel güç oranı, 1000 örneklem büyüklüğü, 40 madde ve %10 değişen madde fonksiyonu içeren madde oranı koşulunda %93.50; en düşük oran ise 250 örneklem büyüklüğü, 60 madde ve %30 değişen madde fonksiyonu içeren madde oranı koşulunda %26.78 olarak hesaplanmıştır.

Bu araştırmada, çoklu grup doğrulayıcı faktör analizi yöntemi için örneklem büyüklüğü değişkeni ile ilgili olarak elde edilen bulgular; Stark ve diğerleri (2006), González Roma ve diğerleri (2010), Kim ve Yoon (2011), Kankaras ve diğerleri (2011) ve Elosua ve Wells (2013) çalışmalarında elde edilen bulgular ile benzerlik göstermektedir. Meade ve Lautenschlager (2004)’ın yapmış olduğu çalışmada ise, elde edilen bulgulardan farklı olarak örneklem büyüklüğündeki artışın ÇGDFA yöntemi için istatistiksel güç oranlarında düşüşe neden olduğu rapor edilmiştir. Bu farklılık, çalışmalar kapsamında ele alınan koşullar arasındaki farklılıklardan kaynaklandığı ileri sürülebilir.

Yapılan bu çalışmalarda örneklem büyüklüğü değişkeni için 100 – 1000 arasında değerler tercih edilmiştir. Stark ve diğerleri (2006), 500 – 1000; González Roma ve diğerleri (2010), 100 – 200 – 400 ve 800; Kim ve Yoon (2011), her bir grup için 100 – 200 – 500 – 1000; Kankaras ve diğerleri (2011), her bir grup için 200 – 1000 ve Elosua ve Wells (2013), her bir

grup için 300 – 500 – 1000 olarak değişimleme gerçekleştirmiştir. İncelenen bu çalışmaların tümünde bu araştırmada belirlendiği gibi örneklem büyüklüğündeki artışla beraber ÇGDFA yönteminin istatistiksel güç oranlarının da arttığı saptanmıştır. Meade ve Lautenschlager (2004) tarafından yapılan çalışmada ise, her bir grup için 150, 500 – 1000 değerlerini alan örneklem büyüklüklerinden ÇGDFA yöntemi için istatistiksel güç oranının en yüksek düzeye 150 değerine ulaştığı belirtilmiştir.

Madde sayısı değişkeni için elde edilen istatistiksel güç oranları

Araştırma kapsamında değişimleme yapılan koşullardan bir diğeri madde sayısıdır. Madde sayısı değişkeni için elde edilen istatistiksel güç oranları örneklem büyüklüğü ve değişen madde fonksiyonu içeren madde sayısı değişkenleri sabit tutularak açıklanmıştır. Çizelge 5'e göre odak ve referans grubu örneklem büyüklükleri 250 olduğunda, madde sayısı değişkenine göre hesaplanan istatistiksel güç oranlarının %26.78 ve %47.00 arasında değiştiği görülmektedir. DMF içeren madde oranı değişkeni sabit tutulduğunda 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %45.00 - %47.00 - %44.67;

%20 DMF içeren madde oranı koşulu için: %35.00 - %38.00 - %35.17;

%30 DMF içeren madde oranı koşulu için: %29.33 - %27.33 - %26.78.

Çizelge 5'de yer alan diğer sonuçlar olan referans ve odak grupları büyüklükleri 500/500 örneklem büyüklüğüne sahip veri setleri için istatistiksel güç oranları %54.11 ve %78.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve DMF içeren madde oranı değişkeni sabit tutulduğunda farklı örneklem büyüklüklerinde, madde sayılarına göre belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %78.00 - %77.00 - %67.33;

%20 DMF içeren madde oranı koşulu için: %72.50 - %63.75 - %65.67;

%30 DMF içeren madde oranı koşulu için: %58.00 - %55.33 - %54.11.

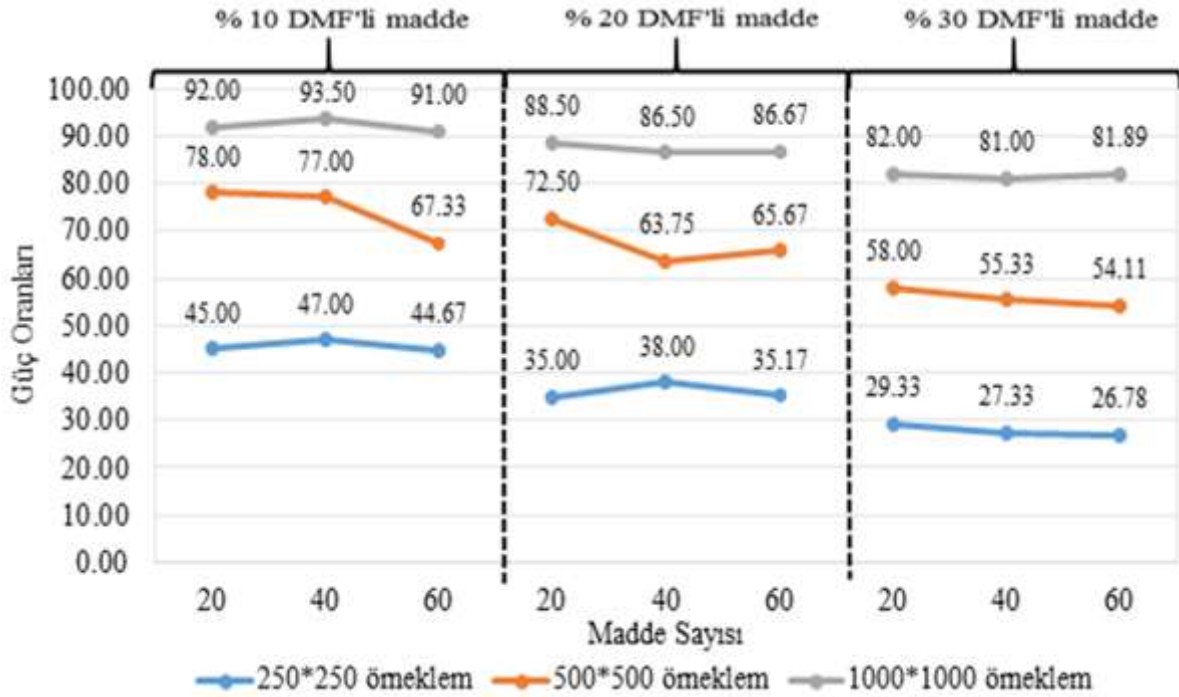
Son olarak Çizelge 5'de örneklem büyüklüğü dağılımının 1000/1000 olduğu koşulda elde edilen istatistiksel güç oranları ise %81.00 ve %93.50 arasında yer almaktadır. Örneklem büyüklüğü ve DMF içeren madde oranı değişkeni sabit tutulduğunda 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %94.50 - %50.75 - %35.00;

%20 DMF içeren madde oranı koşulu için: %94.50 - %53.50 - %39.75;

%30 DMF içeren madde oranı koşulu için: %89.00 - %60.00 - %47.00.

Madde sayısı değişkenine göre istatistiksel güç oranlarındaki değişimin ne yönde olduğunun görülmesi için elde edilen sonuçlar Grafik üzerinde gösterilmiştir. Madde sayısı değişkenine ilişkin istatistiksel güç oranları Grafik 2’de verilmiştir.



Grafik 2. Madde sayısı değişkenine göre istatistiksel güç oranları (ÇGDFA).

Grafik 2’de sunulan bulgulara göre çoklu grup doğrulayıcı faktör analizi yöntemi istatistiksel güç oranlarının madde sayısına bağlı olarak düşük miktarlarda azalma gösterdiği ortaya çıkmaktadır. Elde edilen sonuçlara göre 20 ve 40 maddelik veri setlerinde en yüksek istatistiksel güç oranları elde edilmiştir. Ancak 60 maddelik veri setleri için diğer madde sayısı değişkenlerine göre daha düşük istatistiksel güç oranları hesaplanmıştır. 60 maddelik veri setleri için ortaya çıkan bu azalma analizlerde incelenen modelin karmaşık hale gelmesi ile ilişkilendirilebilir. Üretilen veri setlerinden 20 maddelik olanlar için istatistiksel güç oranları 29.33 – 92; 40 maddelik olanlar için 27.33 – 93.50 ve 60 maddelik olanlar için 26.78 – 91.00 değerleri arasında değişiklik göstermektedir. Diğer değişkenler sabit olarak alındığında 20, 40, 60 maddelik veri setleri için dokuz farklı durum ortaya çıkmıştır. Bu dokuz farklı durumdan altısında 20 maddelik; üçünde ise 40 maddelik veri setlerinde en yüksek istatistiksel güç oranları elde edilmiştir. Bu bulgular doğrultusunda çalışma kapsamında üretilen veri setlerinde ÇGDFA yöntemi ideal istatistiksel güç oranlarının 20 maddelik veri setlerinde gerçekleştirilen analizler sonucunda hesaplandığı ortaya çıkmaktadır.

Değişen madde fonksiyonu içeren madde oranı değişkeni için elde edilen istatistiksel güç oranları

Çizelge 5'te verilen bulgulara göre örneklem büyüklüğü 250/250 olarak alındığında, DMF içeren madde oranı değişkenine göre belirlenen istatistiksel güç oranları, %27.33 ve %47.00 arasında değişmektedir. Örneklem büyüklüğü 250/250 olarak alındığında, %10, %20 ve %30 DMF içeren madde olan veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %45.00 - %35.00 - %29.33;

40 madde koşulu için: %47.00 - %38.00 - %27.33;

60 madde koşulu için: %44.67 - %35.17 - %26.78.

Çizelge 7'de yer alan diğer sonuçlar ise, odak ve referans grubu örneklem büyüklükleri 500 olduğunda hesaplanan istatistiksel güç oranlarıdır. Bu örneklem büyüklüğüne sahip veri setlerinde istatistiksel güç oranları %54.11 ve %78.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve madde sayısı değişkeni sabit tutulduğunda hesaplanan istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %78.00 - %72.50 - %58.00;

40 madde koşulu için: %77.00 - %63.75 - %55.33;

60 madde koşulu için: %67.33 - %65.67 - %54.11.

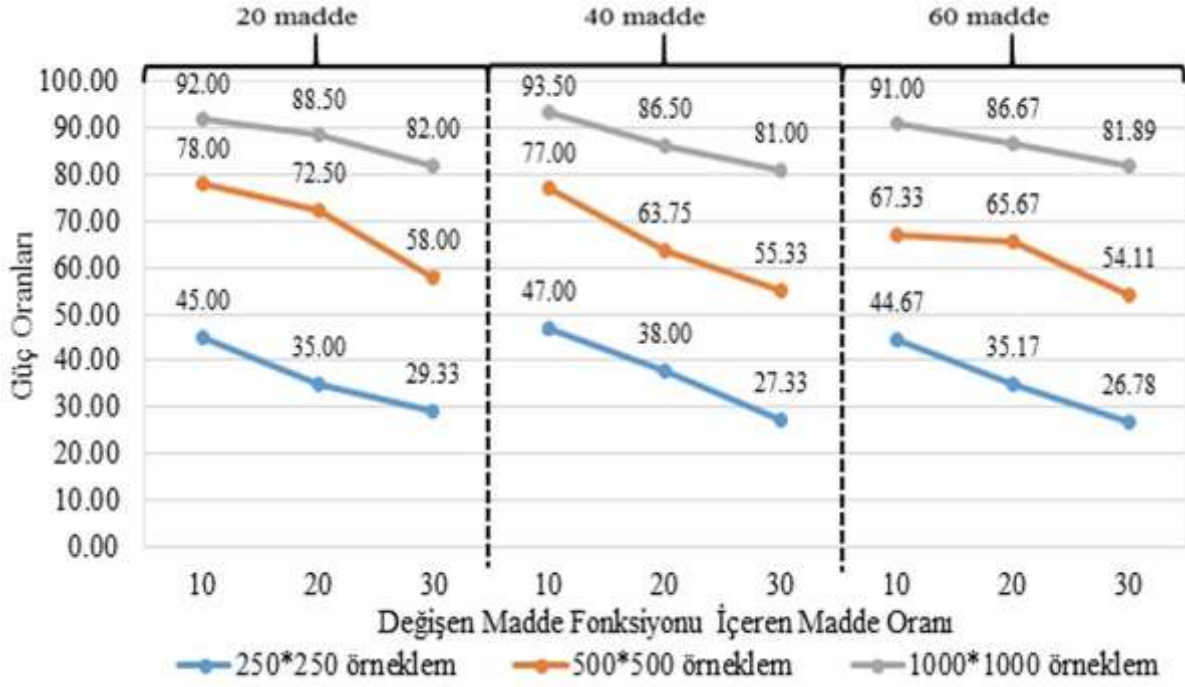
Çizelge 5'te değişen madde fonksiyonu içeren madde oranı değişkeni için sunulan son sonuçlar, odak ve referans grubu örneklem büyüklüklerinin 1000 olduğu koşulda hesaplanan istatistiksel güç oranlarıdır. 1000/1000 örneklem büyüklüğünde istatistiksel güç oranları %93.50 ve %81.00 arasında yer almaktadır. Örneklem büyüklüğü ve madde sayısı değişkeni sabit tutulduğunda DMF içeren madde oranı değişkeni için hesaplanan istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %92.00 - %88.50 - %82.00;

40 madde koşulu için: %93.50 - %86.50 - %81.00;

60 madde koşulu için: %91.00 - %86.67 - %81.89.

Çizelge 5’te verilen istatistiksel güç oranlarının, DMF içeren madde oranı değişkeni için yönünü göstermek için sonuçlar Grafik 3’te sunulmuştur.



Grafik 3. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (ÇGDFA).

Grafik 3’te sunulan değerlere göre DMF içeren madde oranı değişkeninin artmasıyla, çoklu grup doğrulayıcı faktör analizi yönteminin istatistiksel güç oranlarının azaldığı görülmektedir. ÇGDFA yöntemi için en yüksek istatistiksel güç oranlarının değişen madde fonksiyonu içeren madde oranı % 10; en düşük oranlar ise % 30 oranı, değişen madde fonksiyonu içeren madde koşulu için elde edilmiştir. Üretilen veri setlerinde istatistiksel güç oranları, değişen madde fonksiyonu içeren madde oranı % 10 olan koşullar için 44.67 – 93.50; % 20 olan koşullar için 35.00 – 88.50, %30 olan koşullar için 26.78 – 82.00 değerleri arasında değişmektedir. Bu bulgular doğrultusunda çalışma kapsamında üretilen veri setlerinde ÇGDFA yöntemi için ideal istatistiksel güç oranlarının değişen madde fonksiyonu içeren madde oranının %10 olduğu koşul için hesaplandığı göstermektedir. Elde edilen bu bulgular Kankaras ve diğerleri (2011) tarafından yapılan çalışmada elde edilen bulgular ile benzerlik gösterirken, Meade ve Lautenschlager (2004) tarafından yapılan çalışmada elde edilen bulgular ile farklılık göstermektedir.

Kankaras ve diğerleri (2011) tarafından yapılan çalışmada beş maddelik üretilen verilerden bir ve üç madde eşdeğerliği bozacak şekilde değişimlenmiştir. ÇGDFA yöntemi kullanılarak yapılan analizlerde eşdeğerliği bozan madde sayısının bir olduğu koşulda istatistiksel güç oranı %56.50 iken, üç olduğu koşulda ise % 55.90 olarak belirlenmiştir. Bu çalışmada ÇGDFA yöntemi istatistiksel güç

oranları üzerinde eşdeğerliği bozan madde sayısının düşük düzeyde bir azalmaya neden olduğu saptanmıştır. Meade ve Lautenschlager (2004) tarafından yapılan çalışmada ise, analiz edilen veri setleri, veri setini oluşturan altı maddeden iki veya dört maddeyi değişen madde fonksiyonu içerecek şekilde üretilmiştir. Bu veri setleri üzerinde ÇGDFA yöntemi ile gerçekleştirilen analizler, değişen madde fonksiyonu içeren madde sayısı değişkeninin hesaplanan istatistiksel güç oranlarında önemli bir değişikliğe yol açmadığı rapor edilmiştir. Yapılan bu çalışmada veri setinde bulunan madde sayısı altı olarak sınırlandırıldığından, ortaya çıkan modelin basit bir yapıda olması nedeniyle bu farklılığın ortaya çıktığını söylemek mümkündür.

3.2. Olabilirlik oran testi yöntemi istatistiksel güç oranlarına ilişkin bulgular

Bu başlık altında; *“Farklı örneklem büyüklükleri, test uzunlukları ve farklı oranda değişen madde fonksiyonu içeren maddeye sahip yapay veriler üzerinde olabilirlik oran yöntemi ile yapılan analizlerde elde edilen istatistiksel güç oranları nasıl değişmektedir?”* sorusuna yönelik elde edilen bulgu ve yorumlara yer verilmiştir. Olabilirlik oran testi yöntemi istatistiksel güç oranları çizelge 6’da gösterilmiştir.

Çizelge 6.OOT yöntemi ile elde edilen istatistiksel güç oranları.

Örneklem büyüklüğü		Madde sayısı	DMF içeren madde oranı	İstatistiksel güç oranı
Odak	Referans			
250	250	20	10	59.75
500	500	20	10	79.75
1000	1000	20	10	91.25
250	250	20	20	61.00
500	500	20	20	76.33
1000	1000	20	20	88.83
250	250	20	30	60.33
500	500	20	30	75.25
1000	1000	20	30	87.00
250	250	40	10	81.00
500	500	40	10	93.00
1000	1000	40	10	96.75
250	250	40	20	78.25
500	500	40	20	90.75
1000	1000	40	20	95.15
250	250	40	30	75.50
500	500	40	30	88.25
1000	1000	40	30	93.33
250	250	60	10	39.00
500	500	60	10	56.25
1000	1000	60	10	69.50
250	250	60	20	37.00
500	500	60	20	55.75
1000	1000	60	20	66.25
250	250	60	30	36.25
500	500	60	30	52.50
1000	1000	60	30	62.75

Çizelge 6’da verilen sonuçlar olabilirlik oran testi istatistiksel güç oranlarının %36.25 ile %96.75 arasında değiştiğini göstermektedir. Bu yöntem için en düşük güç oranı ÇGDFA yönteminde olduğu odak ve referans grup örneklem büyüklüğü 250, madde sayısı 60 ve DMF içeren madde oranı %30 olduğunda elde edilmiştir. En yüksek istatistiksel güç oranının hesaplandığı koşulda ÇGDFA ile aynı olup odak ve referans grup örneklem büyüklüğü 1000, madde sayısı 40 ve DMF içeren madde oranı %10 olan koşuldur. Değişimlenen değişkenlerin istatistiksel güç oranlarına etkisi her bir değişken için ayrı ayrı özetlenmiştir.

Örneklem büyüklüğü değişkeni için elde edilen istatistiksel güç oranları

Çizelge 6’da sunulan bulgular 20 maddelik veri setlerinde elde edilen istatistiksel güç oranlarının örneklem büyüklüğü değişkenine göre %59.75 ve %91.25 arasında değiştiğini göstermektedir. 20 maddelik veri setlerinde, 250/250, 500/500, 1000/1000 örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %59.75 - %79.75 - %91.25;

%20 DMF içeren madde oranı koşulu için: %61.00 - %76.33 - %88.83;

%30 DMF içeren madde oranı koşulu için: %60.33 - %75.25 - %87.00.

Çizelge 6’da açıklanan istatistiksel güç oranlarından bir diğeri, 40 maddelik veri setlerine ilişkindir. 40 maddelik veri setleri için örneklem büyüklüğü değişkenine göre istatistiksel güç oranları, %75.50 ve %96.75 arasında değişiklik göstermektedir. 40 maddelik veri setlerinde DMF içeren madde oranı değişkeni sabit tutulduğunda, örneklem büyüklüğü değişkeninin her düzeyi için ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %81.00 - %93.00 - %96.75;

%20 DMF içeren madde oranı koşulu için: %78.25 - %90.75 - %95.15;

%30 DMF içeren madde oranı koşulu için: %75.50 - %88.25 - %93.33.

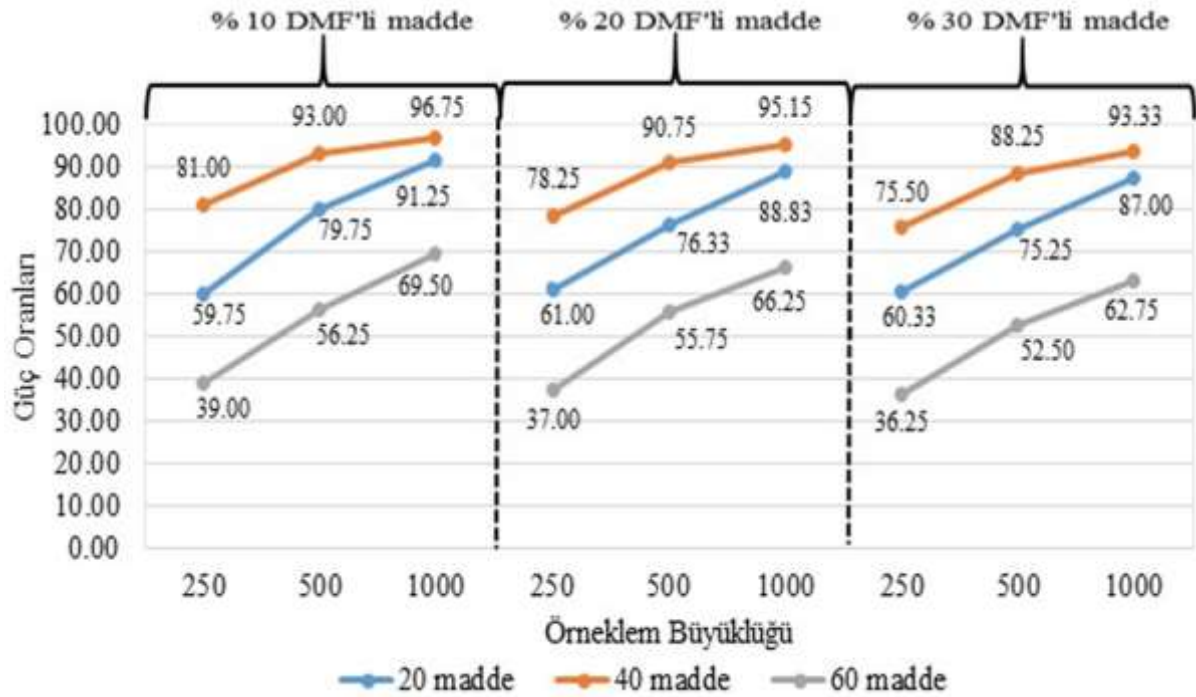
Çizelge 6’da örneklem büyüklüğü değişkenine göre sunulan diğer istatistiksel güç oranları, 60 maddelik veri setleri için hesaplanmıştır. 60 maddelik veri setleri için istatistiksel güç oranları örneklem büyüklüğü değişkenine göre, %36.25 ve %69.50 arasında değişiklik göstermektedir. 60 maddelik veri setlerinde, DMF içeren madde oranı değişkeni sabit tutulduğunda üç örneklem büyüklüğü düzeyi için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %39.00 - %56.25 - %69.50;

%20 DMF içeren madde oranı koşulu için: %37.00 - %55.75 - %66.25;

%30 DMF içeren madde oranı koşulu için: %36.25 - %52.50 - %62.75.

Örneklem büyüklüğü değişkenine göre istatistiksel güç oranlarındaki değişimin daha net bir biçimde görülebilmesi için elde edilen sonuçlar Grafik 4'te gösterilmiştir.



Grafik 4. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (OOT).

Grafik 4'te bulunan eğri incelendiğinde olabilirlik oran testi yöntemi için elde edilen istatistiksel güç oranlarının örneklem büyüklüğündeki artışa bağlı olarak arttığı görülmektedir. Bu sonuç ÇGDFA yöntemi için elde edilen bulgularla benzerlik göstermektedir. OOT yöntemi için en yüksek istatistiksel güç oranları, odak ve referans grupları büyüklüklerinin 1000 olduğu koşullarda elde edilmiştir. Bu yöntemle gerçekleştirilen analizlerde hesaplanan en yüksek istatistiksel güç oranı 1000 örneklem büyüklüğü, 40 madde ve %10 değişen madde fonksiyonu içeren madde oranı koşulunda %96.75; en düşük oran ise 250 örneklem büyüklüğü, 60 madde ve %30 değişen madde fonksiyonu içeren madde oranı koşulunda %36.25 şeklindedir.

Bu araştırmada elde edilen bulgular ile Ankeman ve diğerleri (1999), Meade ve Lautenschlager (2004), Stark ve diğerleri (2006), Atar ve Kamata (2011), Kankaras ve diğerleri (2011), Kim ve Yoon (2011) ve Elosua ve Wells (2013) tarafından yapılan çalışmalarda OOT yöntemi kullanılarak yapılan analizlerde örneklem büyüklüğü değişkeni için elde edilen bulgularla benzerlik göstermektedir.

Yapılan bu çalışmalarda örneklem büyüklüğü değişkeni için 100 – 2000 arasında değerler tercih edilmiştir. Ankeman ve diğerleri (1999), 500 – 2000; Meade ve Lautenschlager (2004), her bir grup için 150 – 500 – 1000; Stark ve diğerleri (2006), 500 – 1000; González Roma, Hernandez ve Gomez Benito (2010), 100 – 200 – 400 ve 800; Kim ve Yoon (2011), her

bir grup için 100 – 200 – 500 – 1000; Kankaras ve diğerleri (2011), her bir grup için 200 – 1000 ve Elosua ve Wells (2013), her bir grup için 300 – 500 – 1000 olarak değerler belirleyerek veri üretimi gerçekleştirmiştir. İncelenen bu çalışmaların tümünde, bu araştırmada belirlendiği gibi örneklem büyüklüğündeki değişimin yönüne paralel olarak OOT yönteminin istatistiksel güç oranlarının da değiştiği saptanmıştır.

Ankeman ve diğerleri (1999), 26 maddelik veri setleri için yapılan analizlerde 500/500 örneklem büyüklüğünde istatistiksel güç oranlarının %60 oranının üstüne çıkmadığını belirlemişlerdir. Her bir grup örneklem büyüklüğünün 500 olarak alındığı koşulda yapılan bu araştırmada, 60 maddelik veri setleri için istatistiksel güç oranları %60'ın üzerine çıkmamıştır. Bu iki bulgu madde sayısının yüksek olduğu koşullarda OOT ile gerçekleştirilen analizler için geniş örneklem kullanılması gerektiğine işaret etmektedir. Kim ve Yoon (2011) tarafından yapılan çalışmada ise çoklu puanlanan verilerde örneklem büyüklüğünün 100 olarak alınmasının bile yeterli düzeyde olduğu belirtilmiştir. Yine bu çalışmada elde edilen bulgularda MTK temelli yöntemler için %34 – %100 aralığında yer almaktadır. Kim ve Yoon tarafından yapılan çalışmada toplam örneklem büyüklüğünün 200 – 2000 olarak değiştiği ve odak ve referans grup büyüklüklerinin eşit olduğu düşünüldüğünde, OOT yöntemi için bu çalışmada da benzer istatistiksel güç oranlarının elde edilmiştir. Elosua ve Wells (2013), her bir grup için 300 – 500 – 1000 örneklem büyüklüğü değerleri alınarak, 0.40 DMF büyüklüğü kullanılarak tek biçimli DMF belirlenmesi için OOT ile gerçekleştirilen analizlerde ise %62 – %72 – %74 istatistiksel güç oranları elde edildiğini saptamışlardır. OOT ile gerçekleştirilen analizlerde belirlenen bu istatistiksel güç oranları, taban değer olarak bu araştırmada elde edilen değerden yüksek iken, tavan değer olarak düşüktür. Ortaya çıkan bu farklılığının incelenen bu çalışmadaki madde sayısının 15 ile sınırlandırılmış olmasına ve ele alınan koşullardan birinin veri setinde değişen madde fonksiyonu içeren maddenin olmadığı bir koşul olmasına bağlı olabilir.

Madde sayısı değişkeni için elde edilen istatistiksel güç oranları

Çizelge 6’da sunulan bulgulara göre odak ve referans grubu örneklem büyüklükleri 250 olarak alındığında 20, 40 ve 60 maddelik veri setleri için elde edilen istatistiksel güç oranlarının %36.25 ve %81.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü 250/250 olarak alındığında ve DMF içeren madde oranı değişkeni sabit tutulduğunda, 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %59.75 - %81.00 - %41.00;

%20 DMF içeren madde oranı koşulu için: %61.00 - %78.25 - %39.00;

%30 DMF içeren madde oranı koşulu için: %60.33 - %75.50 - %36.25.

Çizelge 6’da verilen referans ve odak grupları büyüklükleri 500/500 olduğunda elde edilen istatistiksel güç oranları ise %52.50 ve %93.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve DMF içeren madde oranı değişkeni sabit tutulduğunda farklı örneklem büyüklüklerinde madde sayısı değişkeni düzeyleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %79.75 - %93.00 - %56.25;

%20 DMF içeren madde oranı koşulu için: %76.33 - %90.75 - %55.75;

%30 DMF içeren madde oranı koşulu için: %75.25 - %88.25 - %52.50.

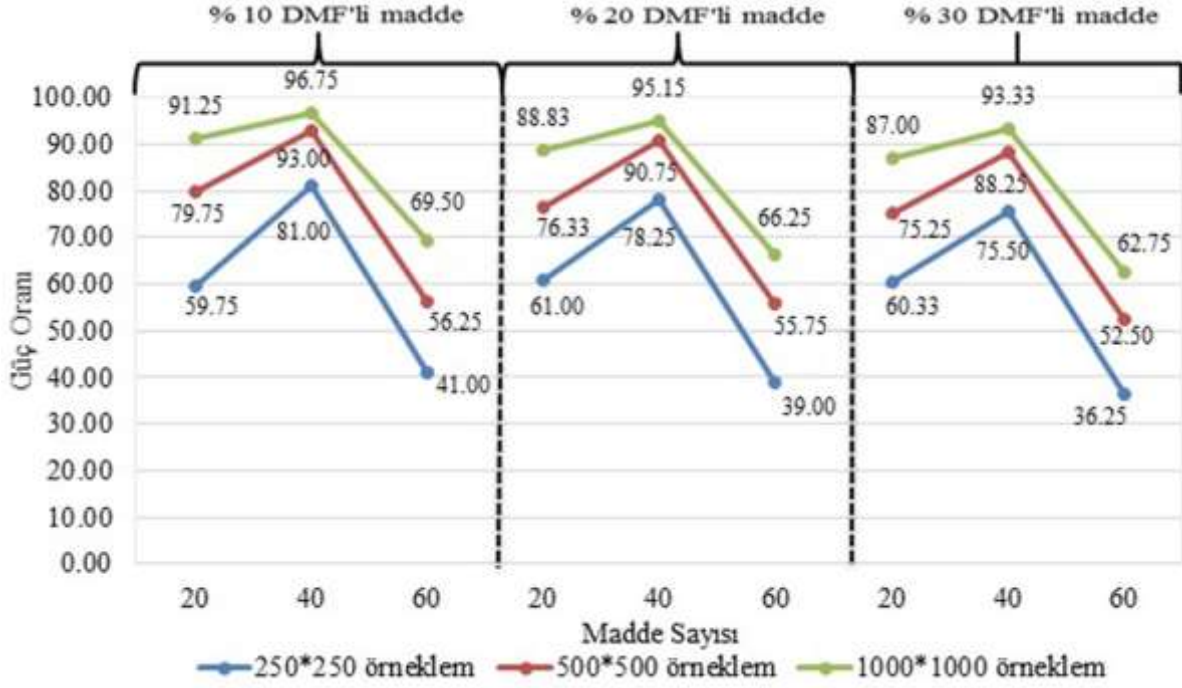
Son olarak Çizelge 6’da odak ve referans grubu örneklem büyüklüklerinin 1000 olduğu koşulda elde edilen istatistiksel güç oranları verilmiştir. Bu örneklem büyüklüğü için 20, 40 ve 60 maddelik veri setlerinde hesaplanan istatistiksel güç oranları, %62.75 ve %96.75 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve DMF içeren madde oranı değişkeni sabit tutulduğunda 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %91.25 - %96.75 - %69.50;

%20 DMF içeren madde oranı koşulu için: %88.83 - %95.15 - %66.25;

%30 DMF içeren madde oranı koşulu için: %87.00 - %93.33 - %62.75.

Madde sayısı değişkenine göre istatistiksel güç oranlarının örneklem büyüklüğü değişkenine göre değişim yönünün görülebilmesi için elde edilen sonuçlar Grafik üzerinde gösterilmiştir. Olabilirlik oran testi kullanılarak yapılan analizlerde madde sayısı değişkenine ilişkin elde edilen istatistiksel güç oranları Grafik 5’te verilmiştir.



Grafik 5. Madde sayısı değişkeni için istatistiksel güç oranları (OOT).

Grafik 5'te sunulan bulgular olabilirlik oran testi yöntemi ile gerçekleştirilen analizlerde yüksek istatistiksel güç oranlarının 40 maddelik veri setleri için olduğuna işaret etmektedir. Başka bir ifadeyle, bu çalışma kapsamında üretilen yapay verilerde olabilirlik oran testi yöntemi için en ideal madde sayısının 40 olduğu ortaya çıkmaktadır. Madde sayısı değişkeni için 20 maddelik veri setlerinden 40 maddelik veri setlerine geçildiğinde bir artış söz konusu iken; 40 maddelik veri setlerinden 60 maddelik veri setlerine geçilirken ciddi bir azalma söz konusudur. Nitekim en düşük istatistiksel güç oranları 60 maddelik veri setleri için elde edilmiştir. Üretilen veri setlerinden 20 maddelik olanlar için istatistiksel güç oranları 59.75 – 91.25; 40 maddelik olanlar için 75.50 – 96.75 ve 60 maddelik olanlar için 36.25 – 69.50 değerleri arasında değişiklik göstermektedir.

Değişen madde fonksiyonu içeren madde oranı değişkeni için elde edilen istatistiksel güç oranları

Çizelge 6'da sunulan verilere göre odak ve referans grubu örneklem büyüklükleri 250 olarak alınan koşul için DMF içeren madde oranı değişkenine göre belirlenen istatistiksel güç oranları %36.25 ve %81.00 arasında değişmektedir. Örneklem büyüklüğü 250/250 olarak alındığında %10, %20 ve %30 DMF içeren madde olan veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %59.75 - %61.00 - %60.33;

40 madde koşulu için: %81.00 - %78.25 - %75.50;

60 madde koşulu için: %39.00 - %37.00 - %36.25.

Çizelge 6'da verilen diğer sonuçlara göre odak ve referans grubu örneklem büyüklükleri 500 olan koşul için hesaplanan istatistiksel güç oranları %52.50 ve %93.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve madde sayısı değişkeni sabit tutulduğunda DMF içeren madde oranı değişkeni düzeyleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %79.75 - %76.33 - %75.25;

40 madde koşulu için: %93.00 - %90.75 - %88.25;

60 madde koşulu için: %56.25 - %55.75 - %52.50.

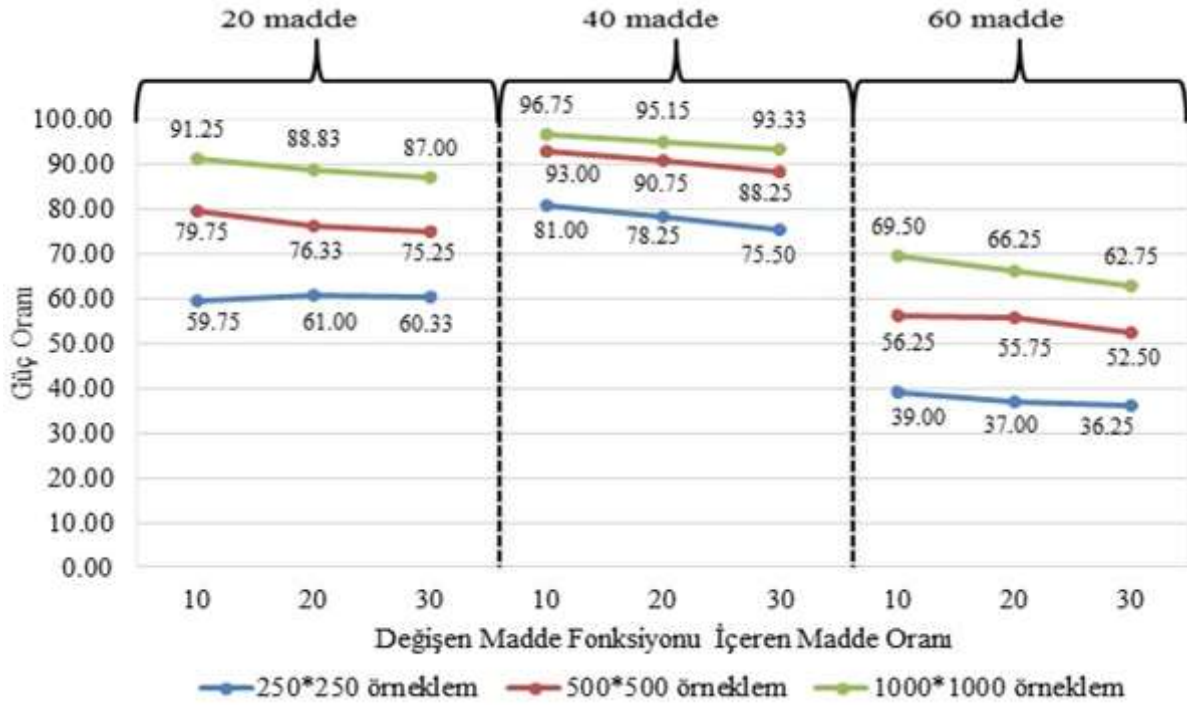
Çizelge 6'da sunulan diğer sonuçlarda ise örneklem büyüklüğü dağılımının odak ve referans için 1000 olduğu koşulda istatistiksel güç oranları, %62.75 ve %96.75 arasında değişiklik göstermektedir. 1000/1000 örneklem büyüklüğü koşulunda, DMF içeren madde oranı değişkeni düzeyleri için 20, 40 ve 60 maddelik veri setlerinde hesaplanan istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %91.25 - %88.83 - %87.00;

40 madde koşulu için: %96.75 - %95.15 - %93.33;

60 madde koşulu için: %69.50 - %66.25 - %62.75.

Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranlarındaki değişim yönünün belirlenebilmesi amacıyla veriler grafik üzerinde gösterilmiştir. DMF içeren madde oranı değişkenine ilişkin istatistiksel güç oranları Grafik 6'da verilmiştir.



Grafik 6. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (OOT).

Grafik 6 incelendiğinde olabilirlik oran testi yöntemi kullanılarak yapılan analizlerde DMF içeren madde oranı değişkeninin, istatistiksel güç oranlarında düşük miktarda bir azalmaya neden olduğu görülmektedir. Ancak bunun, tüm koşullar için geçerli olmadığı da yine Grafik 6’da görülmektedir. Örneğin, örneklem büyüklüğünün 250/250 olduğu 20 maddelik veri setlerinde istatistiksel güç oranları, DMF içeren madde oranı değişkeni %10 olan koşuldan %20 olan koşulla karşılaştırıldığında artarken; bu değişkene ilişkin oranın %20 olduğu koşulla %30 olan koşul karşılaştırıldığında azaldığı belirlenmiştir. Ancak diğer tüm koşullarda istatistiksel güç oranları DMF içeren madde oranı değişkenindeki artışla birlikte düşük düzeylerde azalma göstermektedir.

Üretilen veri setlerinde istatistiksel güç oranları, değişen madde fonksiyonu içeren madde oranı % 10 olan koşullar için 39.00 – 96.75; % 20 olan koşullar için 37.00 – 95.15, son olarak %30 olan koşullar için 36.25 – 93.33 değerleri arasında yer almaktadır. Bu bulgular çalışma kapsamında üretilen veri setleri için, ÇGDFA yönteminde olduğu gibi OOT yöntemi için de ideal istatistiksel güç oranlarının değişen madde fonksiyonu içeren madde oranının %10 olduğu koşullarda elde edildiğini göstermektedir. Elde edilen bu bulgular Kankaras ve diğerleri (2011) tarafından yapılan çalışmada elde edilen bulgular ile benzerlik göstermektedir. Beş maddelik olarak üretilen verilerden bir ve üç maddenin eşdeğerliği bozacak şekilde değişimlendiği bu çalışmada, değişen madde fonksiyonu içeren madde sayısındaki artışın OOT yönteminin istatistiksel gücünü olumsuz etkilediği belirlenmiştir. Meade ve Lautenschlager

(2004) tarafından yapılan çalışmada veri setlerini oluşturan altı maddeden iki veya dört madde değişen madde fonksiyonu içerecek şekilde üretilmiştir. Bu veri setleri üzerinde OOT yöntemi ile gerçekleştirilen analizlerde değişen madde fonksiyonu içeren madde sayısı değişkeninin, istatistiksel güç oranlarında önemli bir değişikliğe yol açmadığı rapor edilmiştir. Değişen madde fonksiyonu içeren madde oranı değişkeni ile ilgili olarak ortaya çıkan bu farklılığın çalışmada ele alınan koşullara ilişkin farklılıklardan dolayı ortaya çıktığı ileri sürülebilir.

3.3. Madde parametrelerini karşılaştırma yöntemi istatistiksel güç oranlarına ilişkin bulgular

Bu kısımda “*Farklı örneklem büyüklükleri, test uzunlukları ve farklı oranda değişen madde fonksiyonu içeren maddeye sahip yapay veriler üzerinde madde parametrelerini karşılaştırma yöntemi ile gerçekleştirilen analizlerde elde edilen istatistiksel güç oranları nasıl değişmektedir?*” sorusuna ilişkin bulgu ve yorumlar açıklanmıştır. Madde parametrelerini karşılaştırma yöntemi ile ilgili yapılan istatistiksel güç oranları incelemelerinde diğer yöntemlerde olduğu gibi farklı koşullar altında DMF içeren maddelerin doğru olarak belirlenme oranları belirlenmiştir. İstatistiksel güç oranları değişimlenen değişkenler için ayrı ayrı rapor edilmiştir. Madde parametrelerini karşılaştırma yöntemi istatistiksel güç oranları çizelge 7’de gösterilmiştir.

Çizelge 7.MPK yöntemi ile elde edilen istatistiksel güç oranları.

Örneklem büyüklüğü		Madde sayısı	DMF içeren madde oranı	İstatistiksel güç oranı
Odak	Referans			
250	250	20	10	54.50
500	500	20	10	78.00
1000	1000	20	10	94.50
250	250	20	20	49.00
500	500	20	20	79.00
1000	1000	20	20	94.50
250	250	20	30	36.00
500	500	20	30	67.17
1000	1000	20	30	89.00
250	250	40	10	28.00
500	500	40	10	41.25
1000	1000	40	10	50.75
250	250	40	20	29.00
500	500	40	20	41.38
1000	1000	40	20	53.50
250	250	40	30	24.33
500	500	40	30	41.58
1000	1000	40	30	60.00
250	250	60	10	18.17
500	500	60	10	30.00
1000	1000	60	10	35.00
250	250	60	20	15.25
500	500	60	20	30.00
1000	1000	60	20	39.75
250	250	60	30	18.44
500	500	60	30	30.44
1000	1000	60	30	47.00

Çizelge 7’de verilen sonuçlar olabirlik oran testi istatistiksel güç oranlarının %15.25 ile %94.50 arasında değiştiğini göstermektedir. Bu yöntem için en düşük güç oranı odak ve referans grup örneklem büyüklüğü 250, madde sayısı 60 ve DMF içeren madde oranı %20 olduğunda elde edilmiştir. En yüksek istatistiksel güç oranının hesaplandığı koşulda odak ve referans grup örneklem büyüklüğü 1000, madde sayısı 20 ve DMF içeren madde oranı %10 - %20 olan koşuldur. Farklı koşullar altında istatistiksel güç oranlarındaki değişim her bir değişken için ayrı ayrı açıklanmıştır.

Örneklem büyüklüğü değişkeni için elde edilen istatistiksel güç oranları

Çizelge 7’de sunulan bulgulara göre 20 maddelik veri setlerinde elde edilen istatistiksel güç oranları örneklem büyüklüğü değişkenine göre %36.00 ve %94.50 arasında değişiklik göstermektedir. 20 maddelik veri setlerinde, DMF içeren madde oranı değişkeni sabit tutulduğunda 250/250, 500/500, 1000/1000 örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %54.50 - %78.00 - %94.50;

%20 DMF içeren madde oranı koşulu için: %49.00 - %79.00 - %94.50;

%30 DMF içeren madde oranı koşulu için: %36.00 - %67.17 - %89.00.

Çizelge 7’de örneklem büyüklüğü değişkenine göre sunulan diğer istatistiksel güç oranları 40 maddelik veri setlerine ilişkindir. 40 maddelik veri setleri için örneklem büyüklüğü değişkenine göre istatistiksel güç oranları %24.33 ve %60.00 arasında değişiklik göstermektedir. 40 maddelik veri setlerinde DMF içeren madde oranı değişkeni sabit tutulduğunda farklı örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %28.00 - %41.25 - %50.75;

% 20 DMF içeren madde oranı koşulu için: %29.00 - %41.38 - %53.50;

%30 DMF içeren madde oranı koşulu için: %24.33 - %41.58 - %60.00.

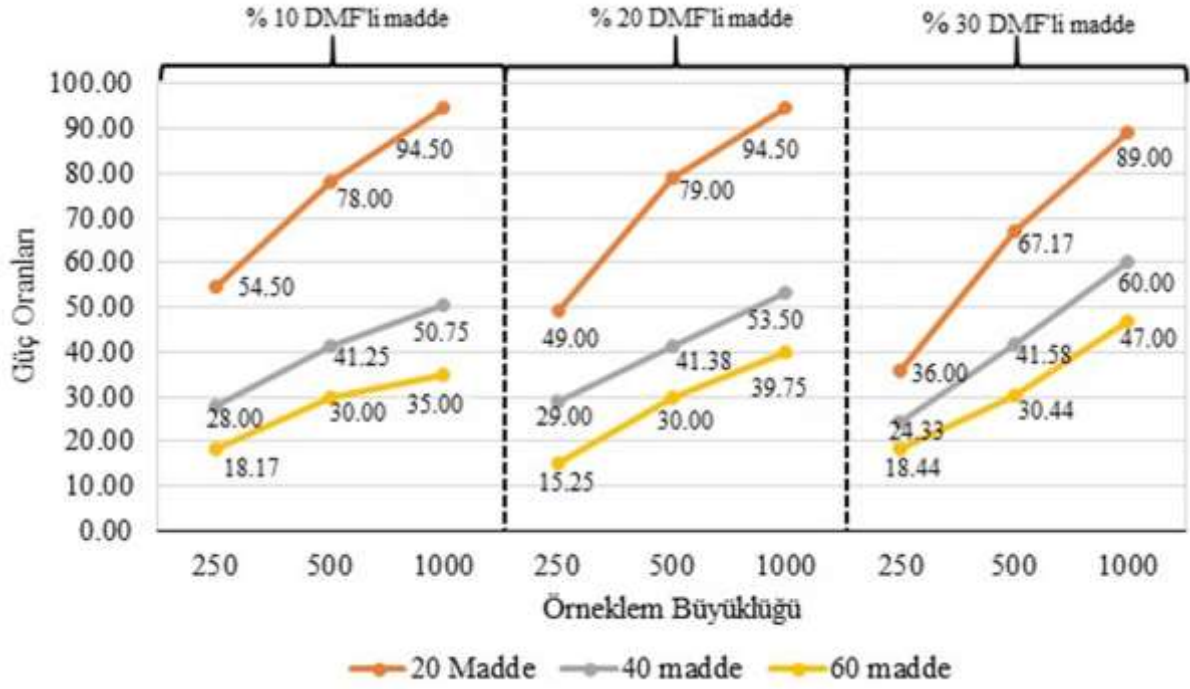
Son olarak Çizelge 7’de örneklem büyüklüğü değişkenine göre sunulan diğer istatistiksel güç oranları 60 maddelik veri setleri için elde edilmiş olanlardır. 60 maddelik veri setleri için istatistiksel güç oranları örneklem büyüklüğü değişkenine göre %15.25 ve %47.00 arasında değişiklik göstermektedir. 60 maddelik veri setlerinde DMF içeren madde oranı değişkeni sabit tutulduğunda farklı örneklem büyüklüklerinde ayrı ayrı belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %18.17 - %30.00 - %35.00;

% 20 DMF içeren madde oranı koşulu için: %15.25 - %30.00 - %39.75;

%30 DMF içeren madde oranı koşulu için: %18.44 - %30.44 - %47.00.

Örneklem büyüklüğü değişkenine göre istatistiksel güç oranlarındaki değişimin daha net bir biçimde görülebilmesi için elde edilen sonuçlar grafik üzerinde gösterilmiştir. Örneklem büyüklüğü değişkenine ilişkin istatistiksel güç oranları Grafik 7’de verilmiştir.



Grafik 7. Örneklem büyüklüğü değişkeni için istatistiksel güç oranları (MPK).

Grafik 7 incelendiğinde örneklem büyüklüğü arttıkça madde parametrelerinin karşılaştırılması yöntemi ile elde edilen istatistiksel güç oranlarının arttığı görülmektedir. Elde edilen istatistiksel güç oranları madde sayısı ve DMF içeren madde oranı koşullarının tümünde referans ve odak grubunun örneklem büyüklükleri 1000/1000 şeklinde olduğunda en yüksek değerlerine ulaşmıştır. Buna ek olarak elde edilen sonuçlar DMF içeren madde oranı değişkeni sabit tutulduğunda madde sayısının istatistiksel güç oranını olumsuz yönde etkilediğini göstermektedir. Madde sayısı ve DMF içeren madde oranı değişkenlerine göre istatistiksel güç oranlarının değişiminin görülebilmesi için sonraki başlıklar altında açıklanmıştır.

Madde sayısı değişkeni için elde edilen istatistiksel güç oranları

Çizelge 7'de sunulan bulgular incelendiğinde odak ve referans grubu örneklem büyüklükleri 250 olarak alındığında, madde sayısı değişkenine göre hesaplanan istatistiksel güç oranlarının %15.25 ve %54.50 arasında değişiklik gösterdiği sonucuna ulaşılmıştır. Örneklem büyüklüğü 250/250 olarak alındığında ve DMF içeren madde oranı değişkeni sabit tutulduğunda 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %54.50 - %28.00 - %18.17;

%20 DMF içeren madde oranı koşulu için: %49.00 - %29.00 - %15.25;

%30 DMF içeren madde oranı koşulu için: %36.00 - %24.33 - %18.44.

Çizelge 7’de yer alan diğer sonuçlar ise referans ve odak grupları büyüklükleri 500/500 olduğunda elde edilen sonuçlardır. 500/500 örneklem büyüklüğüne sahip veri setleri için istatistiksel güç oranları %30.00 ve %79.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve DMF içeren madde oranı değişkeni sabit tutulduğunda 20, 40 ve 60 maddelik veri setleri için hesaplanan istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %78.00 - %41.25 - %30.00;

%20 DMF içeren madde oranı koşulu için: %79.00 - %41.38 - %30.00;

%30 DMF içeren madde oranı koşulu için: %67.17 - %41.58 - %30.44.

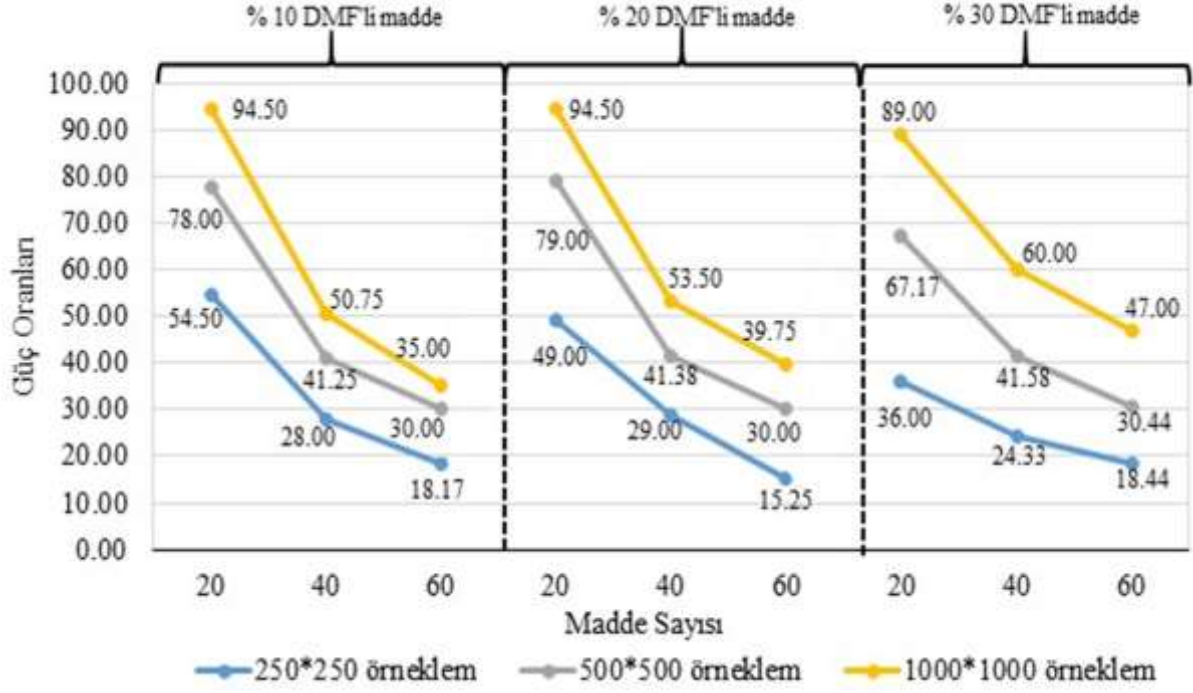
Çizelge 12’nin son kısmında yer alan sonuçlar ise örneklem büyüklüğü dağılımının 1000/1000 olduğu koşulda elde edilen istatistiksel güç oranlarıdır. Bu örneklem büyüklüğü dağılımında 20, 40 ve 60 maddelik veri setleri için istatistiksel güç oranları %94.50 ve %35.00 arasında değişiklik göstermektedir. Referans ve odak grubu örneklem büyüklüğünün 1000 olduğu, DMF içeren madde oranı değişkeninin sabit olarak alındığı 20, 40 ve 60 maddelik veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

%10 DMF içeren madde oranı koşulu için: %94.50 - %50.75 - %35.00;

%20 DMF içeren madde oranı koşulu için: %94.50 - %53.50 - %39.75;

%30 DMF içeren madde oranı koşulu için: %89.00 - %60.00 - %47.00.

Madde sayısı değişkenine göre istatistiksel güç oranlarındaki değişimin daha net bir biçimde görülebilmesi için elde edilen sonuçlar örneklem büyüklüğü değişkeninde olduğu gibi Grafik üzerinde gösterilmiştir. Madde sayısı değişkenine ilişkin istatistiksel güç oranları Grafik 8’de verilmiştir.



Grafik 8. Madde sayısı değişkeni için istatistiksel güç oranları (MPK).

Grafik 8 incelendiğinde madde sayısı arttıkça madde parametrelerinin karşılaştırılması yöntemi ile elde edilen istatistiksel güç oranlarının azaldığı görülmektedir. Örneklem büyüklüğü ve DMF içeren madde oranı değişkenleri sabit alınarak inceleme yapıldığında en yüksek istatistiksel güç oranları 20 maddelik veri setleri için elde edilmiştir. Örneklem büyüklüğü değişkeni için bir önceki başlıkta yapılan yorumlar Grafik 8 üzerinde de görülebilir. Örneklem büyüklüğü değişkeni istatistiksel güç oranlarına olumlu etki yapmaktadır. DMF içeren madde oranı değişkeni için Grafik 8 incelendiğinde net bir yön olmadığı görülmektedir.

Değişen madde fonksiyonu içeren madde oranı değişkeni için elde edilen istatistiksel güç oranları

Çizelge 7’de sunulan bulgular örneklem büyüklüğü 250/250 olarak alındığında DMF içeren madde oranı değişkenine göre belirlenen istatistiksel güç oranlarının %15.25 ve %54.50 arasında yer aldığını göstermektedir. Örneklem büyüklüğü 250/250 olarak alındığında ve madde sayısı değişkeni sabit tutulduğunda %10, %20 ve %30 DMF içeren madde olan veri setleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %54.50 - %49.00 - %36.00;

40 madde koşulu için: %28.00 - %29.00 - %24.33;

60 madde koşulu için: %18.17 - %15.25 - %18.44.

Çizelge 7’de yer alan diğer sonuçlar ise odak ve referans grubu örneklem büyüklükleri 500 olduğunda elde edilen sonuçlardır. 500/500 örneklem büyüklüğüne sahip veri setleri için istatistiksel güç oranları %30.00 ve %79.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve madde sayısı değişkeni sabit tutulduğunda DMF içeren madde oranı değişkeni düzeyleri için belirlenen istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %78.00 - %79.00 - %67.17;

40 madde koşulu için: %41.25 - %41.38 - %41.58;

60 madde koşulu için: %30.00 - %30.00 - %30.44.

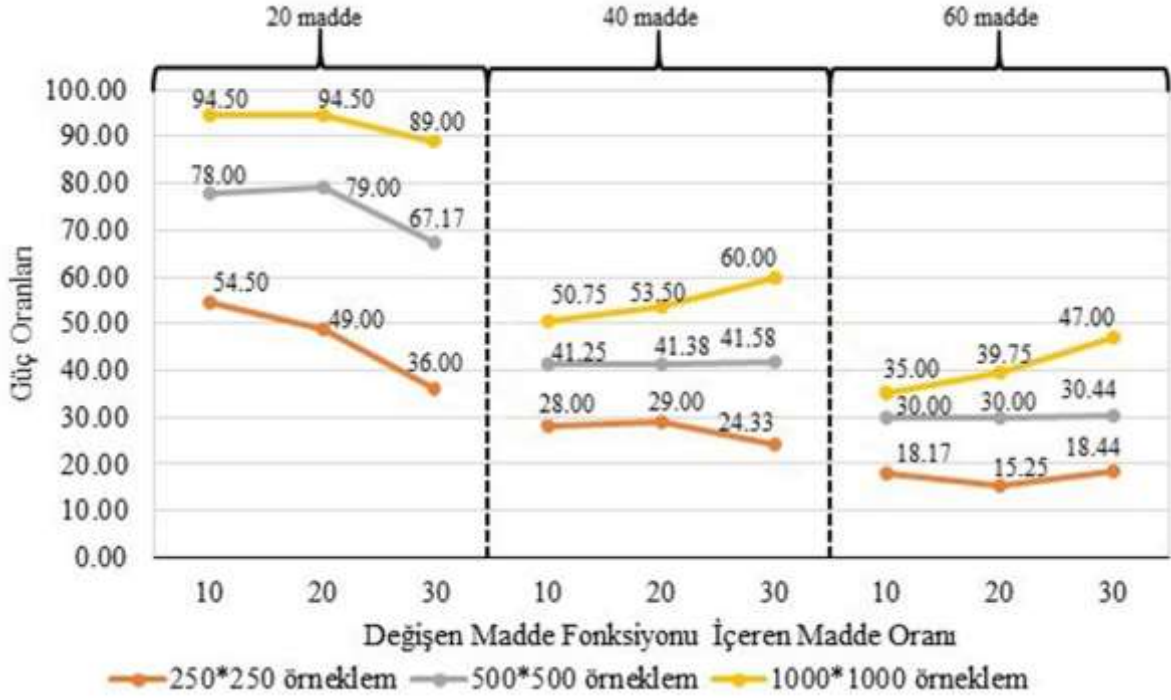
Çizelge 7’de’te sunulan diğer sonuçlar ise örneklem büyüklüğü dağılımının odak ve referans için 1000 olduğu koşulda hesaplanmış olan istatistiksel güç oranlarıdır. 1000/1000 örneklem büyüklüğünde istatistiksel güç oranları %94.50 ve %35.00 arasında değişiklik göstermektedir. Örneklem büyüklüğü ve madde sayısı değişkeni sabit tutulduğunda DMF içeren madde oranı değişkeni için hesaplanan istatistiksel güç oranları sırasıyla şu şekildedir:

20 madde koşulu için: %94.50 - %94.50 - %89.00;

40 madde koşulu için: %50.75 - %53.50 - %60.00;

60 madde koşulu için: %35.00 - %39.75 - %47.00.

Çizelge 7 incelendiğinde DMF içeren madde oranı değişkeni için net bir örüntü olmadığı sonucuna ulaşılmıştır. Nitekim diğer değişkenler için çizilen grafiklerde bu durum ortaya konmuştur. Ancak net bir örüntünün olmadığından emin olunabilmesi için diğer iki değişkende olduğu gibi veriler yine grafik üzerinde gösterilmiştir. DMF içeren madde oranı değişkeni için belirlenen istatistiksel güç oranları Grafik 9’da gösterilmiştir.



Grafik 9. Değişen madde fonksiyonu içeren madde oranı değişkeni için istatistiksel güç oranları (MPK).

Grafik 9 incelendiğinde DMF içeren madde oranı değişkeninin istatistiksel güç oranlarını net bir yönde etkilemediğini anlaşılmaktadır. İstatistiksel güç oranları için DMF içeren madde oranı arttıkça 20 maddelik veri setleri için azalma olduğu; 40 ve 60 maddelik veri setleri için artma ve azalma koşullarının her ikisinin de olduğu görülmektedir.

3.4. Ortak etkiye ilişkin bulgular

Örneklem büyüklüğü, madde sayısı ve değişen madde fonksiyonu içeren madde oranı değişkenleri için üç yöntemle gerçekleştirilen analiz sonuçları Çizelge 14'te bir arada sunulmuştur. Çizelge 14'teki sonuçlar incelenerek kullanılan yöntemlerin hangi koşullar altında daha yüksek istatistiksel güç oranları gösterdiğine ilişkin yorumlar yapılmıştır.

Çizelge 8.Örneklem büyüklüğü, madde sayısı ve DMF içeren madde sayısı koşulları için kullanılan yöntemlerden elde edilen istatistiksel güç oranları.

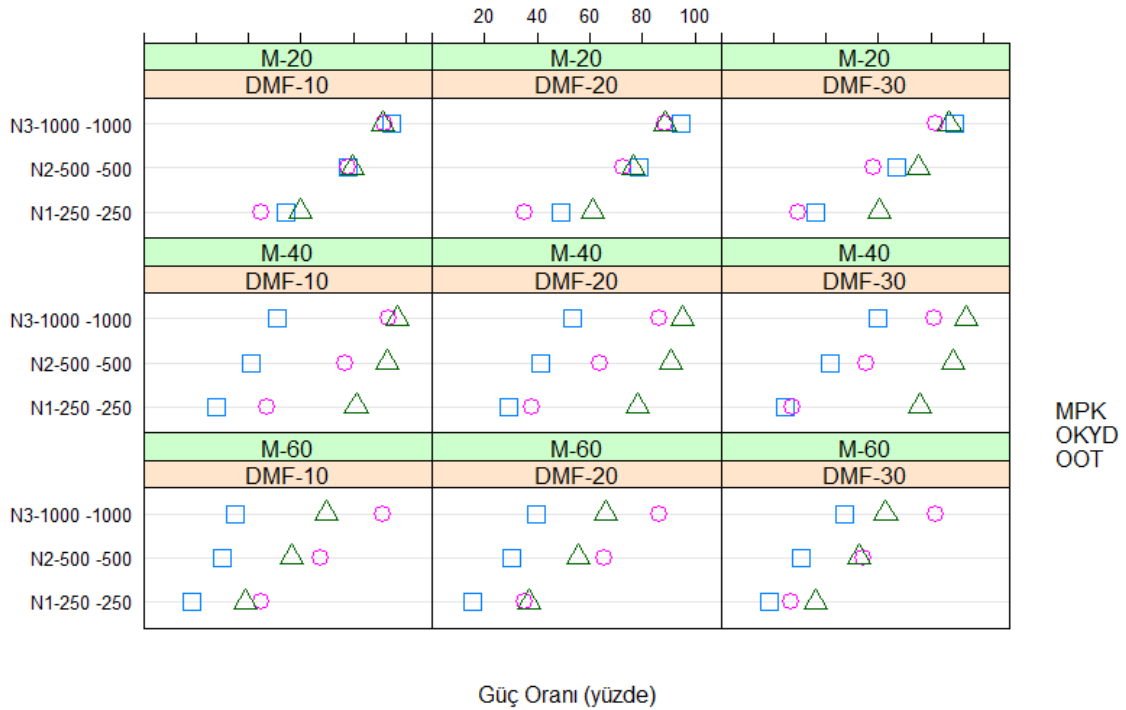
Örneklem Büyüklüğü (O/R)	Madde sayısı	DMF içeren madde oranı	Çoklu Grup Doğrulamalı Faktör Analizi	Olabilirlik oran testi yöntemi	Madde parametrelerini karşılaştırma yöntemi
250/250	20	10	45.00	59.75	54.50
500/500	20	10	78.00	79.75	78.00
1000/1000	20	10	92.00	91.25	94.50
250/250	20	20	35.00	61.00	49.00
500/500	20	20	72.50	76.33	79.00
1000/1000	20	20	88.50	88.83	94.50
250/250	20	30	29.33	60.33	36.00
500/500	20	30	58.00	75.25	67.17
1000/1000	20	30	82.00	87.00	89.00
250/250	40	10	47.00	81.00	28.00
500/500	40	10	77.00	93.00	41.25
1000/1000	40	10	93.50	96.75	50.75
250/250	40	20	38.00	78.25	29.00
500/500	40	20	63.75	90.75	41.38
1000/1000	40	20	86.50	95.15	53.50
250/250	40	30	27.33	75.50	24.33
500/500	40	30	55.33	88.25	41.58
1000/1000	40	30	81.00	93.33	60.00
250/250	60	10	44.67	39.00	18.17
500/500	60	10	67.33	56.25	30.00
1000/1000	60	10	91.00	69.50	35.00
250/250	60	20	35.17	37.00	15.25
500/500	60	20	65.67	55.75	30.00
1000/1000	60	20	86.67	66.25	39.75
250/250	60	30	26.78	36.25	18.44
500/500	60	30	54.11	52.50	30.44
1000/1000	60	30	81.89	62.75	47.00

Çizelge 8 incelendiğinde farklı koşullar altında üretilen verilerden elde edilen sonuçlar doğrultusunda, odak ve referans grubu örneklem büyüklüklerinin 250 olduğu, 60 madde - %10 değişen madde fonksiyonu içeren madde olan ve odak ve referans grupları örneklem büyüklüklerinin 500 olduğu, 60 maddelik koşullar dışındaki tüm koşullarda olabilirlik oran testi yönteminin daha yüksek sonuçlar verdiği ortaya çıkmaktadır. Örneklem büyüklüğü koşulunun 1000/1000 olduğu koşullarda ise yöntemlerin istatistiksel güç oranlarının madde sayısı değişkenine göre değişkenlik gösterdiği saptanmıştır. 1000/1000 örneklem büyüklüğü

koşulunda 20 maddelik veri setleri için madde parametrelerini karşılaştırma, 40 maddelik veri setleri için olabilirlik oran testi, 60 maddelik koşullar için ise çoklu grup doğrulayıcı faktör analizi yönteminin tercih edilmesi gerektiğine yönelik ipuçları vermektedir. bu sonuçlara ek olarak üretilen yapay veriler için, tüm yöntemlere ait istatistiksel güç oranlarının değişen madde fonksiyonu içeren madde oranındaki artıştan olumsuz etkilendiği sonucuna ulaşılmıştır. Veri setlerindeki değişen madde fonksiyonu içeren madde sayısı arttıkça yöntemlerin istatistiksel güç oranlarında azalma olmaktadır. Çalışma kapsamında üretilen yapay veriler için yöntemlerin istatistiksel güç oranlarının en yüksek düzeye ulaştığı koşullar;

- Çoklu grup doğrulayıcı faktör analizi yöntemi 1000/1000, 40 maddelik, %10 değişen madde fonksiyonu içeren madde oranı olan koşul (%93,50),
- Olabilirlik oran testi yöntemi de yine 1000/1000, 40 maddelik, %10 değişen madde fonksiyonu içeren madde oranı olan koşul (%96,75),
- Madde parametrelerini karşılaştırma yöntemi ise 1000/1000, 20 maddelik ve %10-20 değişen madde fonksiyonu içeren madde oranı koşul (%94,50) şeklinde sıralanmaktadır.

Çalışma kapsamında ele alınan yöntemlerin istatistiksel güç oranlarına, değişimleme yapılan koşulların ortak etkileri Grafik 10'da gösterilmiştir.



Grafik 10. Değişimlenen koşulların yöntemler üzerindeki ortak etkileri.

Grafik 10 incelendiğinde değişimleme yapılarak elde edilen 27 farklı koşulun 16'sında olabilirlik oran testi yönteminin diğer iki yöntemle kıyasla daha yüksek istatistiksel güç oranı

verdiği görülmektedir. Çoklu grup doğrulayıcı faktör analizi yöntemi yedi, madde parametreleri karşılaştırma yöntemi ise dört koşulda diğer iki yöntemle göre daha yüksek sonuçlar vermiştir. Güç oranlarının manidar farklılık gösterip göstermediğinin belirlenmesi ve kullanılan yöntemlerin ikili olarak karşılaştırılması için normallik ve varyans homojenliği varsayımlarının sağlandığının belirlenmesinden sonra gerçekleştirilen ANOVA testinden elde edilen sonuçlar ise Çizelge 9’da sunulmuştur.

Çizelge 9.Farklı koşullar altında yöntemler için elde edilen istatistiksel güç oranlarının karşılaştırıldığı ANOVA testi sonuçları.

Koşul	Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	Anlamlı Fark
111	Gruplarasası	11179.167	2	5589.583	132.768	.000	OOT*-MPK
	Gruplariçi	12503.807	297	42.100			OOT*-ÇGDFA
	Toplam	23682.973	299				MPK*-ÇGDFA
211	Gruplarasası	204.167	2	102.083	.508	.602	
	Gruplariçi	59727.034	297	201.101			
	Toplam	59931.201	299				
311	Gruplarasası	579.167	2	289.583	3.163	.044	
	Gruplariçi	27190.956	297	91.552			
	Toplam	27770.123	299				
112	Gruplarasası	33866.667	2	16933.333	506.962	.000	OOT*-MPK
	Gruplariçi	9920.273	297	33.402			OOT*-ÇGDFA
	Toplam	43786.940	299				MPK*-ÇGDFA
212	Gruplarasası	2134.927	2	1067.463	6.373	.002	MPK*-ÇGDFA
	Gruplariçi	49749.010	297	167.505			
	Toplam	51883.937	299				
312	Gruplarasası	2275.260	2	1137.630	9.841	.000	MPK*-ÇGDFA
	Gruplariçi	34332.328	297	115.597			MPK*-OOT
	Toplam	36607.588	299				
113	Gruplarasası	53247.927	2	26623.963	919.131	.000	OOT*-MPK
	Gruplariçi	8603.036	297	28.966			OOT*-ÇGDFA
	Toplam	61850.962	299				MPK*-ÇGDFA
213	Gruplarasası	14897.927	2	7448.963	59.648	.000	OOT*-MPK
	Gruplariçi	37089.678	297	124.881			OOT*-ÇGDFA
	Toplam	51987.604	299				MPK*-ÇGDFA
313	Gruplarasası	2600.000	2	1300.000	8.604	.000	OOT*-ÇGDFA
	Gruplariçi	44875.213	297	151.095			MPK*-OOT
	Toplam	47475.213	299				MPK*-ÇGDFA
121	Gruplarasası	144200.000	2	72100.000	1164.037	.000	OOT*-ÇGDFA
	Gruplariçi	18396.070	297	61.940			OOT*-MPK
	Toplam	162596.070	299				ÇGDFA*-MPK
221	Gruplarasası	140404.167	2	70202.083	713.105	.000	OOT*-ÇGDFA
	Gruplariçi	29238.356	297	98.446			OOT*-MPK
	Toplam	169642.523	299				ÇGDFA*-MPK
321	Gruplarasası	131804.167	2	65902.083	1228.910	.000	OOT*-ÇGDFA
	Gruplariçi	15927.050	297	53.626			OOT*-MPK
	Toplam	147731.217	299				ÇGDFA*-MPK

Çizelge 9. devamı

Koşul	Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	Anlamlı Fark
122	Gruplararası	137554.167	2	68777.083	1017.434	.000	OOT*-ÇGDFA
	Gruplariçi	20076.771	297	67.599			OOT*-MPK
	Toplam	157630.937	299				ÇGDFA*-MPK
222	Gruplararası	122227.127	2	61113.563	751.513	.000	OOT*-ÇGDFA
	Gruplariçi	24152.242	297	81.321			OOT*-MPK
	Toplam	146379.368	299				ÇGDFA*-MPK
322	Gruplararası	96618.167	2	48309.083	532.286	.000	OOT*-ÇGDFA
	Gruplariçi	26955.049	297	90.758			OOT*-MPK
	Toplam	123573.216	299				ÇGDFA*-MPK
123	Gruplararası	164923.927	2	82461.963	1279.027	.000	OOT*-ÇGDFA
	Gruplariçi	19148.312	297	64.472			OOT*-MPK
	Toplam	184072.239	299				ÇGDFA*-MPK
223	Gruplararası	115029.260	2	57514.630	922.558	.000	OOT*-ÇGDFA
	Gruplariçi	18515.745	297	62.343			OOT*-MPK
	Toplam	133545.005	299				ÇGDFA*-MPK
323	Gruplararası	56797.260	2	28398.630	224.111	.000	OOT*-ÇGDFA
	Gruplariçi	37634.939	297	126.717			OOT*-MPK
	Toplam	94432.199	299				ÇGDFA*-MPK
131	Gruplararası	38942.927	2	19471.463	1850.067	.000	OOT*-MPK
	Gruplariçi	3125.846	297	10.525			ÇGDFA*-OOT
	Toplam	42068.773	299				ÇGDFA*-MPK
231	Gruplararası	73511.927	2	36755.963	605.913	.000	OOT*-MPK
	Gruplariçi	18016.636	297	60.662			ÇGDFA*-OOT
	Toplam	91528.563	299				ÇGDFA*-MPK
331	Gruplararası	159616.667	2	79808.333	1062.682	.000	OOT*-MPK
	Gruplariçi	22304.959	297	75.101			ÇGDFA*-OOT
	Toplam	181921.626	299				ÇGDFA*-MPK
132	Gruplararası	29107.260	2	14553.630	5006.544	.000	OOT*-ÇGDFA
	Gruplariçi	863.356	297	2.907			OOT*-MPK
	Toplam	29970.616	299				ÇGDFA*-MPK
232	Gruplararası	67793.927	2	33896.963	625.251	.000	OOT*-MPK
	Gruplariçi	16101.376	297	54.213			ÇGDFA*-OOT
	Toplam	83895.303	299				ÇGDFA*-MPK
332	Gruplararası	110690.427	2	55345.213	618.100	.000	OOT*-MPK
	Gruplariçi	26593.635	297	89.541			ÇGDFA*-OOT
	Toplam	137284.062	299				ÇGDFA*-MPK
133	Gruplararası	15881.087	2	7940.543	1441.022	.000	OOT*-ÇGDFA
	Gruplariçi	1636.575	297	5.510			OOT*-MPK
	Toplam	17517.662	299				ÇGDFA*-MPK
233	Gruplararası	34983.487	2	17491.743	566.971	.000	OOT*-MPK
	Gruplariçi	9162.805	297	30.851			ÇGDFA*-MPK
	Toplam	44146.292	299				
333	Gruplararası	61057.140	2	30528.570	328.802	.000	OOT*-MPK
	Gruplariçi	27575.857	297	92.848			ÇGDFA*-OOT
	Toplam	88632.997	299				ÇGDFA*-MPK

*Lehine fark olan grup

Tablo 9’da sunulan sonuçlara göre; 27 farklı koşul altında üç farklı yöntem için toplam 81 ikili karşılaştırmanın 38’inde olabilirlik oran testinin, 24’ünde çoklu grup doğrulayıcı faktör analizi yönteminin, 9’unda ise madde parametrelerini karşılaştırma yönteminin diğer iki yöntemden manidar olarak daha yüksek güç oranları verdiği görülmektedir. 10 karşılaştırmada ise üç yöntem arasında manidar bir farklılık olmadığı saptanmıştır. Yapılan ikili

karşılaştırmalarda olabilirlik oran testi yönteminin odak ve referans grubu örneklem büyüklüğünün 250 olduğu durumlarda diğer iki yöntemle göre daha yüksek güç oranları verdiği sonucuna ulaşılmıştır. Bunun yanı sıra çoklu grup doğrulayıcı faktör analizi yönteminin daha yüksek güç oranları verdiği koşulların 40 ve 60 maddelik veri setleri olduğu sonucu elde edilmiştir. Madde parametrelerini karşılaştırma yönteminin ise diğer iki yöntemden daha yüksek güç oranı verdiği koşulların tamamı 20 maddelik veri setleri için yapılmış analizlerde elde edilmiştir.

3.5. Gerçek veri üzerinde gerçekleştirilen analizlere ilişkin bulgular

Bu kısımda, “*Yapay veriler üzerinde gerçekleştirilen analizlere ek olarak gerçek veriler üzerinde gerçekleştirilen analiz sonuçlarında yöntemlerin istatistiksel güç oranları nasıl değişmektedir?*” sorusuna ilişkin bulgu ve yorumlara yer verilmiştir.

Gerçek veriler üzerinde ölçme eşdeğerliği analizlerine geçilmeden önce Türkiye ve Amerika Birleşik Devletleri örneklemelerine ait veriler içerisinde geçersiz ve kayıp veriler temizlenmiştir. Daha sonra ise kalan veriler içerisinde 250/250, 500/500 ve 1000/1000 olacak şekilde rastgele veri setleri oluşturulmuştur. Veri setlerinde madde sayısı değişkeni ve değişen madde fonksiyonu içeren madde oranı değişkenleri için değişimleme yapılması mümkün olmadığından sadece örneklem büyüklüğü değişkeni için ele alınan düzeylere göre veri seçimi yapılmıştır. Ölçme eşdeğerliği analizlerine geçilmeden önce oluşturulan bu veri setleri için basıklık çarpıklık katsayıları belirlenmiş ve doğrulayıcı faktör analizi yapılmıştır. Böylece veri setlerinin normallik ve tek boyutluluk açısından kontrolleri gerçekleştirilmiştir.

Gerçek veri setlerine ait basıklık ve çarpıklık katsayıları Çizelge 10’da sunulmuştur.

Çizelge 10.ST018 kodlu - Genel Değer alt ölçeği verileri kullanılarak gerçekleştirilen analiz sonuçları.

Örneklem büyüklüğü (O/R)	DMF içeren madde sayısı	Çoklu Grup Doğrulayıcı Faktör Analizi yöntemi		Olabilirlik Oran Testi yöntemi		Madde Parametrelerinin Karşılaştırılması yöntemi	
		Belirlenen madde sayısı	Belirlenen maddeler	Belirlenen madde sayısı	Belirlenen maddeler	Belirlenen madde sayısı	Belirlenen maddeler
250/250	5	0	-	2	m2 – m3	0	-
500/500	5	5	Tüm maddeler	4	m2 – m3 m4 – m5	3	m2 – m3 m5
1000/1000	5	5	Tüm maddeler	5	Tüm maddeler	4	m2 – m3 m4 – m5

Çizelge 10 incelendiğinde 250/250 örneklem büyüklüğünde ÇGDFA ve MPK yöntemleri değişen madde fonksiyonu içeren beş maddeden hiçbirini saptayamamıştır. Olabilirlik oran testi ise iki maddenin değişen madde fonksiyonu içerdiğini belirlemiştir. Bu iki maddenin orta düzey değişen madde fonksiyonuna sahip olduğu belirlenmiştir. Örneklem büyüklüğü 500/500 düzeyine arttırıldığında ÇGDFA tüm maddelerin, OOT dört maddenin, MPK üç maddenin değişen madde fonksiyonu içerdiğine yönelik sonuçlar vermiştir. 1000/1000 örneklem büyüklüğünde ise ÇGDFA ve OOT yöntemleri tüm maddelerin MPK yöntemi ise dört maddenin değişen madde fonksiyonuna sahip olduğunu ortaya konmuştur. Bu sonuçların yanı sıra OOT yöntemi ile değişen madde fonksiyonu içerdiği belirlenen maddeler için ortaya konan farklılaşma düzeyinin örneklem büyüklüğündeki artışa bağlı olduğu sonucuna ulaşılmıştır. OOT yöntemi ile 250/250 örneklem büyüklüğünde belirlenen maddelerin orta düzeyde değişen madde fonksiyonu içerdiği görülürken, örneklem büyüklüğü 1000/1000 olduğunda bu koşul yüksek düzeye çıkmıştır. 500/500 örneklem büyüklüğünde ihmal edilebilir ve orta düzeyde değişen madde fonksiyonu içerdiği belirlenen maddelere ilişkin düzeylerin ise sırasıyla orta ve yüksek düzeye çıktığı sonucuna ulaşılmıştır.

4. SONUÇ VE ÖNERİLER

Bu bölümde gerçekleştirilen analizler sonucunda elde edilen bulgular ışığında sonuçlar ve öneriler sunulmuştur.

4.1. Sonuçlar

Bu araştırmada kapsamında, 5’li likert yapıda üretilen yapay verilerin örneklem büyüklüğü, madde sayısı ve DMF içeren madde oranı değişkenleri açısından farklı yöntemlere ilişkin istatistiksel güç oranlarındaki değişim incelenmiştir. Veri üretimi sürecinde örneklem büyüklüğü (250/250, 500/500, 1000/1000), madde sayısı (20, 40, 60) ve DMF içeren madde oranı (%10, %20, %30) olmak üzere toplam 27 koşul oluşacak şekilde değişimleme yapılmıştır. DMF büyüklüğü 0,43 lojit birim olarak belirlenmiştir. Ayrıca yöntemlerin istatistiksel güç oranları bu değişkenler açısından karşılaştırılarak diğer araştırmacılara ölçme eşdeğerliği analiz sürecine ilişkin bilgi sağlanması amaçlanmıştır. Bu amaç doğrultusunda, aşamalı tepki modeli esas alınarak üretilen tek boyutlu veriler üzerinde çoklu grup doğrulayıcı faktör analizi, olabilirlik oran testi ve madde parametrelerinin karşılaştırılması yöntemleri ile analizler gerçekleştirilmiştir.

Elde edilen bulgular, kullanılan üç yöntem içinde örneklem büyüklüğü değişkenindeki artışın istatistiksel güç oranlarını olumlu etkilediğini göstermiştir. Başka bir ifadeyle örneklem büyüklüğü 250/250 koşulundan 1000/1000 koşuluna doğru arttıkça tüm yöntemlerin istatistiksel güç oranlarının arttığı belirlenmiştir. Ancak madde sayısındaki artış ve örneklem büyüklüğünün 250/250 olarak alındığı koşullarda yöntemlerin istatistiksel güç oranlarında azalmalar olduğu görülmüştür. Nitekim 60 maddelik koşullarda üç yöntemin istatistiksel güç oranları %45’i aşamamıştır. Bunun yanı sıra madde parametreleri karşılaştırma yönteminin örneklem büyüklüğünün 250/250 olduğu koşullarda yetersiz olduğu saptanmıştır. Bu koşulların tümünde MPK yöntemi istatistiksel güç oranları % 18.17-%54.44 aralığında değişiklik göstermiştir.

Madde sayısı değişkeni için ele alınan koşullar incelendiğinde yöntemlerin istatistiksel güç oranlarının bu değişkene ait düzeylerden farklı şekilde etkilendiği ortaya konmuştur. Üretilen veri setlerindeki madde sayılarındaki artış ÇGDFA ve MPK yöntemlerinin istatistiksel güç oranlarında azalmaya yol açtığı belirlenmiştir. Bu azalma MPK yönteminde yüksek miktarlarda iken ÇGDFA yöntemi için daha düşük miktarlardadır. ÇGDFA yöntemi ile yapılan analizlerde madde sayısının 20’den 40’a doğru arttığı koşulların birkaçında düşük düzeylerde artışlar gözlenmiştir. OOT yöntemi için ise madde sayısı değişkeninin etkisi artışa bağlı olarak

değişkenlik göstermektedir. Madde sayısının 20'den 40'a doğru artış gösterdiği koşullar için bu yöntemin istatistiksel güç oranı artarken, 40 maddeden 60 maddeye doğru artışın olduğu koşullarda oranlar azalmıştır. Çalışma kapsamında üretilen veriler için OOT yöntemi ile gerçekleştirilen analizlerde en yüksek istatistiksel güç oranları 40, en düşük istatistiksel güç oranları 60 maddelik veri setleri için elde edilmiştir. Yöntemlerin istatistiksel güç oranları yeterliği için en az %70 ve üzeri olması (Cohen ve Cohen, 1983) durumu göz önünde bulundurulmuştur. OOT yönteminin performansının 20 ve 40 maddelik veri setlerinde yeterli düzeyde olduğu belirlenmiştir. 60 maddelik veri setlerinde ise büyük örneklemelerde ve DMF içeren madde oranı değişkeninin %10 ve %20 olduğu koşullarda yeterli olduğu saptanmıştır. ÇGDFA yöntemi için ise büyük örneklemelerde istatistiksel güç oranlarının madde sayısı değişkeninin tüm düzeylerinde yeterli olduğu görülmüştür. MPK yöntemi için ortaya konulan bulgularda ise bu yöntemin büyük örneklemelerde ve 20 maddelik veri setlerinde yeterli olduğu sonucuna ulaşılmıştır.

Değişen madde fonksiyonu içeren madde oranı değişkenindeki artışın ÇGDFA ve OOT yöntemlerinin istatistiksel güç oranlarında azalmaya neden olduğu belirlenmiştir. MPK yönteminde ise istatistiksel güç oranlarındaki değişimin madde sayısı değişkenine bağlı olarak değiştiği görülmüştür. DMF içeren madde oranı değişkeni %10'dan %30'a doğru arttığında, 20 maddelik veri setleri için istatistiksel güç oranları genel olarak azalırken, 40 ve 60 maddelik veri setlerinde genel olarak artış olduğu gözlenmiştir. ÇGDFA ve OOT yöntemleri için en yüksek istatistiksel güç oranları diğer değişkenlerde esas alındığında DMF içeren madde oranının %10 olduğu koşullar için elde edilmiştir. MPK yöntemi için en yüksek istatistiksel güç oranları, 20 maddelik veri setlerinde DMF içeren madde oranı %10 ve %20; 40 ve 60 maddelik veri setlerinde ise genel olarak DMF içeren madde oranı %20 olduğunda hesaplanmıştır.

Elde edilen bulgular ışığında karşılaştırılan yöntemlere ilişkin istatistiksel güç oranlarının değişimleme yapılan değişkenlere göre farklılık gösterdiği ortaya konmuştur. Yapılan incelemelerde 20 maddelik ve 1000/1000 örneklem büyüklüğüne sahip veri setleri için elde edilen bulgularda, değişen madde fonksiyonu içeren madde oranı değişkeninin her düzeyinde MPK yönteminin diğer yöntemlere göre ön planda olduğu belirlenmiştir. 20 maddelik ve 500/500 örneklem büyüklüğünde, değişen madde fonksiyonu içeren madde oranı %10 ve %30 olan veri setleri için elde edilen bulgular OOT, bu oranın %20 olduğu veri setleri için ise MPK yönteminin daha yüksek düzeyde istatistiksel güç oranları verdiği saptanmıştır. Örneklem büyüklüğünün 250/250, madde sayısının 20 olduğu veri setlerinde ise OOT yönteminin diğer iki yöntemle göre daha yüksek düzeyde istatistiksel güç oranlarına sahip

olduğu ortaya konmuştur. Ayrıca OOT yönteminin diğer iki yöntemle göre örneklem büyüklüğündeki azalmadan daha az etkilendiğini görülmektedir. 40 maddelik veri setleri üzerinde yapılan analizlerde ise örneklem büyüklüğü düzeylerinin tümünde OOT yöntemi en yüksek istatistiksel güç değerlerini, ÇGDFA yöntemi ikinci en büyük, MPK yöntemi ise en düşük istatistiksel güç değerlerini vermiştir. 20 maddelik veri setlerinde özellikle büyük örneklemelerde yüksek istatistiksel güç değerleri veren (%89.00 – % 94.50 arasında) MPK yönteminin madde sayısındaki artıştan etkilendiği belirlenmiştir. MPK yöntemi ile gerçekleştirilen analizlerde madde sayısı değişkeninin önemli olduğu ortaya konmuştur. 60 maddelik veri setlerinde gerçekleştirilen analizlerde, 1000/1000 ve 500/500 örneklem büyüklüğü koşullarında değişen madde fonksiyonu içeren madde oranı değişkeninin tüm düzeylerinde ÇGDFA yönteminin en yüksek istatistiksel güç oranları verdiği sonucuna ulaşılmıştır. Madde sayısı 60 olan 250/250 örneklem büyüklüğünde ise %10 değişen madde fonksiyonu içeren madde oranı düzeyinde ÇGDFA; diğer düzeylerde ise OOT yöntemi için en yüksek istatistiksel güç oranlarının elde edildiği görülmektedir. ÇGDFA yönteminin madde sayısı 60 olan koşullarda diğer yöntemlere göre daha yüksek oranlara sahip olmasının gerekçesi olarak analizlerin daha basit bir model üzerine kurulu olarak gerçekleştirilmesi gösterilebilir. 250/250 örneklem büyüklüğünde ise OOT yönteminin istatistiksel güç oranlarının diğer iki yöntemle ve diğer değişken düzeylerindeki oranlara göre daha yüksek olmasını, örneklem büyüklüğü değişkenindeki azalmadan daha az etkileniyor olması ile ilişkilendirmek mümkündür.

Gerçek verilerle yapılan analizlerde örneklem büyüklüğü değişkenindeki değişimin yönü ile karşılaştırılan yöntemlerin değişen madde fonksiyonu içeren maddeleri belirlemedeki gücünün arttığı sonucu elde edilmiştir. Bu durum çalışma kapsamında da üretilen yapay veriler üzerinde gerçekleştirilen analizlerle benzer bir nitelik taşımaktadır. İncelenen gerçek verilerdeki ölçme modeli, üretilen verilerdekine göre daha basit yapılı olsa da ÇGDFA yönteminin 250/250 örneklemde yetersiz kaldığını göstermektedir. Ancak ÇGDFA yöntemi ile 500/500 örneklem büyüklüğü için gerçekleştirilen analizlerde diğer yöntemlerden farklı olarak değişen madde fonksiyonu içeren tüm maddeler saptanmıştır. Üretilen verilerde 60 maddelik ve 500/500 ve 1000/1000 örneklem büyüklüklerinde ÇGDFA yönteminin istatistiksel güç oranlarının yüksek olması durumu, test edilen ölçme modelinin basitliği ile ilişkilendirilmiştir. Gerçek verilerde ele alınan basit model ve büyük örneklemelerde ÇGDFA yöntemi ile ortaya konan bu durum göz önüne alındığında, basit modeller için gerçekleştirilen analizlerde bu yönteminin daha uygun bir yöntem olduğu ortaya konmuştur. 250/250 örneklem

büyükliğünde ise OOT yönteminin diğer yöntemlere kıyasla değişen madde fonksiyonu içeren maddeleri tespit edebiliyor olması, yapay veriler üzerinde gerçekleştirilen analiz sonuçları ile uyumludur. Yapay veriler ile gerçekleştirilen analizlerde örneklem büyüklüğündeki azalmadan en az etkilenen yöntemin OOT olduğu belirlenmiştir. Aynı duruma gerçek verilerle gerçekleştirilen analiz bulgularında da ulaşılmıştır.

Yapılan analizlerde ayrıca analiz yöntemlerinin gerçekleştirilme süreci göz önüne alındığında MPK yönteminin diğer yöntemlere göre daha kullanışlı olduğu ileri sürülebilir. MPK yönteminde veri seti kullanılan programın tanıyabileceği şekilde düzenlendikten sonra her bir veri seti için tek bir komut dosyası ile analizler gerçekleştirilebilmektedir. Ayrıca elde edilen sonuçlar tek bir çıktı dosyası üzerinden kolaylıkla görülebilmektedir. ÇGDFA yöntemi için gerçekleştirilen analizlerde ise her bir veri seti için önce alt grup veri dosyaları oluşturulmakta, sonraki adımda ise bu dosyalar için ortalama ve kovaryans çıktıları elde edilmektedir. Analiz için gerekli dosyalar elde edildikten sonra, test edilen dört aşama için ayrı komut dosyası kullanılarak analizler gerçekleştirilmektedir. Sonuçlar için tercih edilen uyum indeksleri ve modifikasyon indeksleri incelenmektedir. OOT yöntemi ise uygulama açısından diğer yöntemlere nazaran daha fazla zaman ve dikkat gerektirmektedir. İlk olarak kullanılan programın tanıyabileceği şekilde düzenlenen veri setindeki madde sayısının bir fazlası kadar komut dosyası oluşturulur. Oluşturulan komut dosyalarının her biri için çıktı dosyasına ulaşılmaktadır. Bütün çıktı dosyalarından ilk olarak sınırlandırılmış modellere ait olanlar içerisinde parametre sayısı ve $-2\log\text{likelihood}$ değerleri saptanmaktadır. Bu işlem her bir madde için elde edilen çıktı dosyaları için tekrarlanmakta ve genişletilmiş modellere ilişkin $-2\log\text{likelihood}$ değerleri belirlenmektedir. Son olarak, belirlenen serbestlik derecesindeki kritik Ki-Kare değeri ile sınırlandırılmış ve genişletilmiş modellere ait $-2\log\text{likelihood}$ değeri farkları alınarak hesaplan G^2 değerleri karşılaştırılmakta ve ele alınan madde ile ilgili karar verilmektedir.

4.2. Öneriler

Yapılan bu araştırmadan elde edilen bulgular ve sonuçlar doğrultusunda, öneriler diğer araştırmalara ve yöntemlerin kullanımına yönelik olmak üzere iki ayrı başlık altında toplanmıştır:

Yöntemlerin kullanımına yönelik öneriler

- 1- Yapılan araştırmalarda odak ve referans grubu örneklem büyüklüğünün 250 ve 500 aralığında olduğu durumlarda ölçme eş değeri için olabilirlik oran testi kullanılması önerilebilir.

- 2- 40 maddelik tek boyutlu bir ölçme aracının uygulandığı araştırmalarda, araştırmacıların ölçme eş değeriği analizleri için olabilirlik oran testini tercih etmeleri önerilebilir.
- 3- Madde sayısının 60 olduđu uygulamalarda ölçme eş değeriği analizlerinin çoklu grup doğrulayıcı faktör analizi yöntemi ile yapılması önerilebilir.
- 4- Yapılan uygulamalarda odak ve referans grubu örneklem büyüklüğünün 1000 seviyelerinde, madde sayısının ise 20 olduđu durumlarda madde parametrelerini karşılaştırma yöntemi ile ölçme eş değeriği incelemelerinin yapılması önerilebilir.
- 5- Olabilirlik oran testi yönteminin 20 ve 40 madde uygulanan durumlarda istatistiksel güç oranı bakımından yeterli ($>70\%$) düzeye ulaşabilmesi için odak ve referans grubu örneklem büyüklüğünün her birinin en az 500 olacak şekilde uygulama yapılması gerektiği önerilebilir.
- 6- Çoklu grup doğrulayıcı faktör analizi yönteminin ise madde sayısından bağımsız olarak yeterli düzeyde istatistiksel güç oranlarına ulaşabilmesi için örneklem büyüklüklerinin 1000 seviyelerine çıkartılması önerilebilir.
- 7- Madde parametrelerini karşılaştırma yönteminin 20 maddelik veri setlerinde yeterli güç oranına ulaşabilmesi için odak ve referans grubu örneklem büyüklüğünün 1000 düzeyine ulaşacak şekilde planlama yapılması önerilebilir.

Diğer araştırmalara yönelik öneriler

- 1- Yapılan bu araştırmanın amacı, genel olarak ölçme araçlarında aranan en önemli psikometrik özellik olan geçerliğe kanıt sağlayan ölçme eşdeğerliği analizlerinde kullanılan farklı kuramları temel alan yöntemleri karşılaştırmaktır. Bu karşılaştırmalar sonucunda da gerçek uygulamalardaki koşullara göre uygun ölçme eşdeğerliği yönteminin belirlenmesine ilişkin araştırmacılara bilgi sağlamaktır. Ölçme eşdeğerliği ile geçerlik arasındaki bu ilişki nedeniyle bu konuda gerçekleştirilecek analizler ölçme aracı uyarlama ve geliştirme sürecinde önem kazanmaktadır. Ancak, yapılan birçok çalışmada ölçme eşdeğerliği çalışmalarının nihai form üzerinden elde edilen veriler doğrultusunda gerçekleştirildiği görülmektedir. Nihai formu uygulanmış bir ölçme aracına ait veriler üzerinden geçerliğe ilişkin kanıt toplamak, geriye dönük olarak ölçme aracının geçerliği ile ilgili olumsuzlukların ve yapılan değerlendirmelere ilişkin hataların önüne geçilmesi noktasında araştırmacılara ve uygulayıcılara herhangi bir yarar sağlamayabilecektir.

Bu nedenle ölçme eşdeğerliği çalışmalarının ölçme aracı geliştirme ve uyarlama sürecinde yapılması önerilmektedir.

- 2- Bu araştırmada değişen madde fonksiyonu büyüklüğü 0.43 lojit birim olarak alınmıştır. Yapılacak olan farklı çalışmalarda değişen madde fonksiyonu büyüklüğü değişkeninde değişimleme yapılarak analizler gerçekleştirilebilir.
- 3- Bu araştırmada analizler çok kategorili puanlama şekli olarak 5-4-3-2-1 olmak üzere beş kategoriye sahip veriler üzerinde gerçekleştirilmiştir. Diğer araştırmalarda kategori sayıları arttırılarak ya da azaltılarak yine çoklu puanlanan verilerle çalışma tekrarlanabilir. Ayrıca verilerin üretiminde aşamalı tepki modeli temel alınmıştır. Yapılacak farklı çalışmalarda diğer çok kategorili puanlama modellerine dayalı veriler üretilerek karşılaştırma yapılabilir. Çoklu puanlanan verilerin yanı sıra ikili puanlanan veriler üretilerek uygun yöntemlerle, aynı değişkenler için değişimleme yapılarak çalışma tekrarlanabilir.
- 4- Gerçekleştirilen bu araştırmada tekrarlama sayısı 100 olarak alınmış ve yöntemlerin istatistiksel güç oranları üzerinde durulmuştur. Yapılacak olan çalışmalarda tekrarlama sayısı artırılarak, yöntemlerin hata düzeyleri karşılaştırılabilir.
- 5- Bu araştırma kapsamında yetenek dağılımı değişkeni üzerinde herhangi bir değişimleme yapılmamıştır. Gelecek araştırmalarda yetenek dağılımı değişkeni ele alınarak araştırma yapılabilir.
- 6- Bu araştırma kapsamında veriler tek boyutlu olarak üretilmiştir. Yapılacak olan çalışmalarda çok boyutlu olarak üretilen verilerle analizler gerçekleştirilerek istatistiksel güç oranları karşılaştırılabilir.
- 7- Bu araştırma kapsamında veri setleri tek biçimli değişen madde fonksiyonu içeren maddelere sahip olacak şekilde üretilmiştir. Yeni çalışmalarda veri setlerinde tek biçimli olan veya olmayan değişen madde fonksiyonu içeren maddeler bir arada veya yalnızca tek biçimli olmayan maddeler yer alacak şekilde üretim gerçekleştirilerek analiz tekrarlanabilir.
- 8- Bu araştırmada örneklem büyüklüğü düzeyi 250/250, 500/500 ve 1000/1000 değerleri kullanılmıştır. Yapılacak olan araştırmalarda örneklem büyüklüğü değişkeni için daha küçük ve daha büyük düzeyler ele alınabilir. Bunun yanı sıra bu çalışmada, referans ve odak grup örneklem büyüklüğü olarak 1:1 oranı tercih edilmiştir. Örneklem büyüklüğü oranı farklı miktarlarda alınarak çalışma yeniden tekrarlanabilir.

- 9- Bu arařtırmada madde sayısı deęiřkeni olarak 20, 40 ve 60 deęerleri kullanılmıřtır. Madde sayısı deęiřkeni daha kk olacak řekilde belirlenerek yeni arařtırmalar gerekleřtirilebilir.
- 10- Bu arařtırmada deęiřen madde fonksiyonu ieren madde oranı deęiřkeni iin yzdelik olarak 10, 20 ve 30 deęerleri tercih edilmiřtir. Yapılabilecek olan yeni arařtırmalarda bu deęiřken iin daha kk ve byk miktarlarda deęerler ve daha fazla sayıda dzey belirlenerek arařtırma tekrarlanabilir.



5. KAYNAKLAR

- Acar, T., & Kelecioğlu, H. (2010). Comparison of differential item functioning determination techniques: HGLM, LR and IRT-LR. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(2), 639 – 649.
- Akalın, Ş. (2014). *Kamu Personeli Seçme Sınavı genel yetenek testinin madde yanlılığı açısından incelenmesi*. (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Angoff, W.H. (1993). Perspectives On Differential Item Functioning Methodology. Holland and Wainer (Ed.), *Differential Item Functioning* içinde (s. 3-23). Lawrence Erlbaum Associates, Publishers, New Jersey.
- Ankenmann, R. D., Witt, E. A., & Dunbar, S. B. (1999). An Investigation of the Power of the Likelihood Ratio Goodness-of-Fit Statistic in Detecting Differential Item Functioning. *Journal of Educational Measurement*, 36(4), 277-300.
- Arıkan Akın, Ç. (2015). Değişen madde fonksiyonu belirlemede MTK-olabilirlik oranı, ordinal lojistik regresyon ve poly-sıbtest yöntemlerinin karşılaştırılması. *E-Uluslararası Eğitim Araştırmaları Dergisi*, 6(1), 1- 16.
- Asil, M. (2010). *Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi*. (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Asil, M., & Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim*, 37(166), 236-249.
- Atalay, K., Gök, B., Kelecioğlu, H., & Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon araştırması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43, 270-281.
- Atalay Kabasakal, K. & Kelecioğlu, H. (2012). Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning. *Eğitim Bilimleri Fakültesi Dergisi*, 45(2), 77.
- Atar, B. (2007). *Differential Item Functioning Analyses For Mixed Response Data Using IRT Likelihood-Ratio Test, Logistic Regression and Gllamm Procedures*. (Unpublished doctoral dissertation). Florida State University.
- Atar, B., & Kamata, A. (2011). Comparison of IRT likelihood ratio test and logistic regression DIF detection procedures. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 41(41).
- Ateşok Deveci, N. (2008). *Üniversitelerarası Kurul Yabancı Dil Sınavı'nın madde yanlılığını bakımından incelenmesi*. (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Başokçu, O. T., & Öğretmen, T. (2013). Öğretmen Öz-yeterlilik Ölçeğinde Değişen Madde Fonksiyonlarının Ağırlıklandırılmış Cevap Modeli ile Belirlenmesi. *Ege Eğitim Dergisi*, 14(2), 67-81.

- Başusta, N.B., & Gelbal, S. (2015). Gruplararası karşılaştırmada ölçme eşdeğerliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: PegemA yayıncılık.
- Bilican Demir, S. (2014). *Çoklu puanlanan maddelerde madde işlev farklılığının mantel test ve olabilirlik oran testi ile karşılaştırılması*. (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Bock, R. D., Murakl, E., & Pfeifferberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement*, 25(4), 275-285.
- Bollen, K.A. (1991). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Bolt, D. M. (2002). A Monte Carlo Comparison of Parametric and Nonparametric Polytomous DIF detection methods. *Applied Measurement in Education*, 15, 113-141.
- Borsboom, D., Mellenbergh, G.J & van Heerden, J. (200). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Byrne, B.M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872-882.
- Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural equation modeling*, 13(2), 287-321.
- Bryne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
Web: <http://jcc.sagepub.com/cgi/reprint/34/2/155.pdf> adresinden alınmıştır.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items* (4. Baskı). Thousand Oaks, CA: Sage.
- Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for Polytomously Scored Items: An Adaptation of the SIBTEST Procedure. *Journal of Educational Measurement*, 33(3), D15-353.
- Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: issues and practice*, 17(1), 31-44.
- Cohen, A. S., Kim, S.-H., & Baker, F. B. (1993). Detection of Differential Item Functioning in the Graded Response Model. *Applied Psychological Measurement*, 17, 335-350.

- Çepni, Z. (2011). *Değişen madde fonksiyonlarının sıbtest, mantel haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Çetin, B. (2010). Cross-cultural structural parameter invariance on PISA 2006 student questionnaires. *Eğitim Araştırmaları-Eurasian Journal of Educational Research*, 38, 71-89.
- DeShon, R.P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, 46(1), 137-149.
- Dodeen, H. (2004). Stability of differential item functioning over a single population in survey data. *The Journal of experimental education*, 72(3), 181-193.
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662.
- Ekermans, G., Saklofske, D.H., Austin, E., & Stough, C. (2011). Measurement invariance and differential item functioning of the Bar-On EQ-i: S measure over Canadian, Scottish, South African and Australian samples. *Personality and Individual Differences* 50, 286–290.
- Elosua, P., & Wells, C. S. (2013). Detecting DIF in polytomous items using MACS, IRT and ordinal logistic regression. *Psicológica*, 34(2)
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Erdem Keklik, D. (2014). Değişen Madde Fonksiyonunu Belirlemede Mantel-Haenszel ve Lojistik Regresyon Tekniklerinin Karşılaştırılma. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(2), 12 - 25.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme – I*. Ankara: PegemA Yayıncılık
- Fidalgo, A. M., Mellenbergh G. J., & Muniz, J. (2000). Effects of Amount of DIF, Test Length, and Purification Type on Robustness and Power of Mantel-Haenszel Procedures. *Methods of Psychological Research Online*. 5(3), 43-53.
- Flowers, C.P., Raju, N.S., & Oshima, T. C. (2002). A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory. *Paper presented at the Annual Meeting of the National Council on Measurement in Education*, New Orleans.
- Fortmann-Johnson, K., A. (2007). *The evaluation of new criteria for polytomous DIF in the DFIT framework*. (Unpublished doctoral dissertation). Illinois Institute of Technology, Chicago.
- Galton, F. (1884). Measurement of character. *Fortnightly Review*, 36, 179-185.

- Garrett, P. (2009). *A Monte Carlo Study Investigating Missing Data, Differential Item Functioning and Effect Size*. (Unpublished doctoral dissertation). Georgia State University.
- González-Romá, V., Hernandez, A., & Gomez-Benito, J. (2006). Power and Type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research*, 41(1), 29-53.
- Güngör Culha, D. (2012). *Örtük sınıf analizlerinde ölçme eşdeğerliğinin incelenmesi*. (Yayımlanmamış doktora tezi). Ege Üniversitesi, İzmir.
- Güzeller, C. O. (2011). A study of cross-cultural equivalence of computer attitude in PISA 2009 student questionnaire. *Eğitim ve Bilim*, 36(162), 320 - 327.
- Golia, S. (2012). Differential Item Functioning classification for polytomously scored items. *Electronic Journal of Applied Statistical Analysis*, 5(3), 367-373.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. CA: Sage.
- Harwell, M., Stone, C. A., Hsu, T. -C., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20, 101-125.
- Hidalgo - Montesinos M. D., & Lopez - Pina J. A. (2002). Two-stage equating in differential item functioning detection under the graded response model with the raju area measures and the lord statistic. *Educational and Psychological Measurement*. 62(1), 32-44.
- Holmes Finch, W., & French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and psychological Measurement*, 67(4), 565-582.
- Holland, P.W., & Wainer, H. (Editörler). (1993). *Differential Item Functioning*. Lawrence Erlbaum Associates, Publishers, New Jersey.
- Hong, S., Malik, M.L., & Lee, M. (2003). Testing configural, metric, scalar and latent mean invariance across genders in sociotropy and autonomy using a non- western sample. *Educational and Psychological Measurement*, 63(4), 636-654.
- Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 28, 192-218.
- Jöreskog, K. G. (1970). Simultaneous factor analysis in several populations. *ETS Research Report Series*, 1970(2).
- Jöreskog, K., & Sörbom, D. (1993), *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc.

- Kan, A., Sünbül, Ö., & Ömür, S. (2013). 6. - 8. sınıf seviye belirleme sınavları alt testlerinin çeşitli yöntemlere göre değişen madde fonksiyonlarının incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 207- 222.
- Kankaras, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, 40(2), 279-310.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel yayın dağıtım
- Kazelskis, R., Thames, D., & Reeves, C. (2004). The elementary reading attitude survey: factor invariance across gender and race. *Reading Psychology*, 25, 111-120.
- Kim, S. H., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement*, 22(4), 345-355.
- Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32(3), 261-276.
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, 18(2), 212-228.
- Koh, K. H., & Zumbo, B. D. (2008). Multi-group confirmatory factor analysis for testing measurement invariance in mixed item format data. *Journal of Modern Applied Statistical Methods*, 7(2), 471-477.
- Korkmaz, M. (2005). *Madde cevap kuramına dayalı olarak çok kategorili maddelerde madde ve test yanlılığının (işlevsel farklılığın) incelenmesi* (Doktora tezi) Ege Üniversitesi, İzmir.
- Köse, İ. A. (2015). PISA 2009 öğrenci anketi alt ölçeklerinde (Q32-Q33) bulunan maddelerin değişen madde fonksiyonu açısından incelenmesi. *Kastamonu Eğitim Dergisi*, 23(1), 227-240
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53-76.
- Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7(4), 361-388.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.

- Morales, L. S., Reise, S. P., & Hays, R. D. (2000). Evaluating the equivalence of health care ratings by whites and Hispanics. *Medical care*, 38(5), 517.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14, 5971.
- Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data*. Chicago: Scientific Software, Inc.
- Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). (Why) should we use SEM? Pros and Cons of Structural Equation Modeling. *Methods of Psychological Research Online*, 8(2), 1- 22.
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315-328.
- Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement*, 34(3), 253-272.
- Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği*. (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Öğretmen, T. (2009). Değişen madde ve test fonksiyonunun belirlenmesinde madde tepki kuramına dayalı parametrik yöntemlerin karşılaştırılması. *Eğitim ve Bilim*, 34(152), 113-125.
- Önen, E. (2009). *Ölçme eşdeğerliğinin yapısal eşitlik modelleme teknikleri ile incelenmesi*. (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517.
- Reise, S. P., Smith, L., & Furr, R. M. (2001). Invariance on the NEO PI-R neuroticism scale. *Multivariate Behavioral Research*, 36(1), 83-110.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552-566.
- Rogers, J. H., & Swaminathan, H. (1993). A comparison of logistic regression and mantel-haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105-116.
- Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on sibtest and mantel-haenszel type I error performance. *Journal of Educational Measurement*, 33, 215-230

- Salzberger, T., Sinkovics, R.R., & Schlegelmilch, B.B. (1999). Data equivalence in cross cultural research: A comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal*, 7(2), 23-38.
- Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Report Series*, 1968(1).
- Samejima, F. (1997). Graded response model. Wim J. van der Linden, Ronald K. Hambleton (Ed.), *Handbook of modern item response theory* içinde (s. 85-100). Springer New York.
- Smith, L. L. (2002). On the usefulness of item bias analysis to personality psychology. *Personality and Social Psychology Bulletin*, 28(6), 754-763.
- Sireci, S.G., & Berberoğlu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, 13(3), 229 – 248.
- Somer, O. (2004). Gruplararası karşılaştırmalarda ölçek eşdeğerliğinin incelenmesi: Madde ve test fonksiyonlarının farklılaşması. *Türk Psikoloji Dergisi*, 19(53), 69 - 82.
- Somer, O., Korkmaz, M., Dural, S., & Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde tepki kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi*, 24(64), 61-75.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27(2), 229-239.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292.
- Steenkamp E. M., & Baumgartner H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1). Oxford University Press.
- Su, Y.-H., & Wang, W.-C. (2005). Efficiency of the mantel, generalized mantel-haenszel, and logistic discriminant function analysis methods in detecting differential item functioning for polytomous items. *Applied Measurement in Education*, 18, 313-350.
- Şekercioğlu, G. (2009). *Çocuklar için benlik algısı profilinin uyarlanması ve faktör yapısının farklı değişkenlere göre eşitliğinin test edilmesi*. (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: the concept of item bias. *Psychological Bulletin*, 99(1), 118-128.

- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. P.W. Holland and H. Wainer (Eds.), *Differential item functioning* içinde (s. 67-113). Hillsdale, N J: Lawrence Erlbaum.
- Thissen, D. (1991). *MULTILOG: Multiple category item analysis and test scoring using item response theory*. Chicago, IL: Scientific Software International.
- Tian, F. (1999). *Detecting DIF in polytomous item responses*. (Yayımlanmamış doktora tezi). University of Ottawa, Ottawa.
- Uzun, B. (2008). *TIMSS - R Türkiye örnekleminde fen başarısını etkileyen değişkenlerin cinsiyetler arası eşdeğerliğinin incelenmesi*. (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Van de Vijver, F. J. (1998). Towards a theory bias and equivalence. *Zuma Nachrichten: Cross-Cultural Survey Equivalence*, 3, 41-65.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Wang, W. C., & Yeh, L. Y. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, 27, 479-498.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007) Decoding the meaning of factorial invariance and updating the practice of multiple-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, 12, 1-26.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores*. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, 2012(1).

6. EKLER

EK 1. Yapay veri koşulları

Koşullar			
Koşul no	Örneklem (O/R)	Madde sayısı	DMF içeren madde yüzdesi
1	250/250	20	10
2			20
3			30
4		40	10
5			20
6			30
7		60	10
8			20
9			30
10	500/500	20	10
11			20
12			30
13		40	10
14			20
15			30
16		60	10
17			20
18			30
19	1000/1000	20	10
20			20
21			30
22		40	10
23			20
24			30
25		60	10
26			20
27			30

EK 2. Çoklu grup doğrulayıcı faktör analizi Yöntemi Komut Dizin Dosyaları Örneği

YAPISAL DEĞİSMEZLİK

Group 1: GRUP1

Observed Variables:

M1 M2 M3 M4 M5

Covariance Matrix from File XXX.COV

Means from File XXX.DAT

Sample Size: 500

Latent Variables:

GD

Equations:

M1 = GD

M2 = GD

M3 = GD

M4 = GD

M5 = GD

Group 2: GRUP2

Covariance Matrix from File YYY.COV

Means from File YYY.DAT

Sample Size: 500

Latent Variables:

GD

Equations:

M1 = GD

M2 = GD

M3 = GD

M4 = GD

M5 = GD

Set the error variance of GD free

Path Diagram

End of Problem

ÖLÇEK DEĞİSMEZLİĞİ

Group 1: GRUP1

Observed Variables:

M1 M2 M3 M4 M5

Covariance Matrix from File ZZZ.COV

Means from File ZZZ.DAT

Sample Size: 1000

Latent Variables:

GD

Equations:

M1 = CONST + GD

M2 = CONST + GD

M3 = CONST + GD

M4 = CONST + GD

M5 = CONST + GD

Group 2: GRUP2

Covariance Matrix from File TTT.COV

Means from File TTT.DAT

Sample Size: 1000

Equations:

Set the error variance of GD free

Path Diagram

End of Problem

METRIK DEĞİSMEZLİK

Group 1: GRUP1

Observed Variables:

M1 M2 M3 M4 M5

Covariance Matrix from File XXX.COV

Means from File XXX.DAT

Sample Size: 250

Latent Variables:

GD

Equations:

M1 = CONST + GD

M2 = CONST + GD

M3 = CONST + GD

M4 = CONST + GD

M5 = CONST + GD

Group 2: GRUP2

Covariance Matrix from File YYY.COV

Means from File YYY.DAT

Sample Size: 250

Latent Variables:

GD

Equations:

M1 = CONST

M2 = CONST

M3 = CONST

M4 = CONST

M5 = CONST

Set the error variance of GD free

Path Diagram

End of Problem

KATI DEĞİSMEZLİK

Group 1: GRUP1

Observed Variables:

M1 M2 M3 M4 M5

Covariance Matrix from File ZZZ.COV

Means from File ZZZ.DAT

Sample Size: 1000

Latent Variables:

GD

Equations:

M1 = CONST + GD

M2 = CONST + GD

M3 = CONST + GD

M4 = CONST + GD

M5 = CONST + GD

Group 2: GRUP2

Covariance Matrix from File TTT.COV

Means from File TTT.DAT

Sample Size: 1000

Equations:

Path Diagram

End of Problem

EK 3. Olabilirlik Oran Testi Komut Dizin Dosyası Örneği

Alperen YANDI

DIF model

>PROBLEM RANDOM, INDIVIDUAL, DATA='STMG.dat',

NITEMS=10, NGROUPS=2, NEXAMINEES=10341, NCHARS=5;

>TEST ALL, GRADED, NC=(4(0)10), HIGH=(4(0)10);

>EQUAL AJ,

ITEMS=(6,7,8,9,10),

WITH= (1,2,3,4,5);

>EQUAL BK=(1,2,3),

ITEMS=(6,7,8,9,10),

WITH= (1,2,3,4,5);

>ESTIMATE NCYCLES=100;

>SAVE;

>END;

4

1234

1111111111

2222222222

3333333333

4444444444

(5A1,1X,I1,1X,10A1)

EK 4. Madde Parametreleri Karşılaştırma Yöntemi Komut Dizin Dosyası Örneği

ALPEREN- DIF ANALYSIS USING PARTIAL CREDIT MODEL
SIMULATED GRM DATA, TWO SAMPLES

```
>FILES  DFNAME='500MPK.dat' ;
>INPUT  NIDCHAR=4, MGROUP=2, NTOTAL=5 ;
(4A1,1X,1A1,1X,5A1)
>TEST   TNAME=SON, ITEM=(1,2,3,4,5),
        INAME=(M1,M2,M3,M4,M5),NBLOCK=5 ;
>BLOCK1 NIT=1, NCAT=4, ORIGINAL=(1,2,3,4) ;
>BLOCK2 NIT=1, NCAT=4, ORIGINAL=(1,2,3,4) ;
>BLOCK3 NIT=1, NCAT=4, ORIGINAL=(1,2,3,4) ;
>BLOCK4 NIT=1, NCAT=4, ORIGINAL=(1,2,3,4) ;
>BLOCK5 NIT=1, NCAT=4 ORIGINAL=(1,2,3,4) ;
>MGROUP DIF=(0,1,0,0), GNAME=(TR,USA), GCODE=('1','2');
>CALIB  LOGISTIC, GRADED, NQPT=30, CYCLES=(100,1,1,1,1), SCALE=1.0, NEWTON=25,
        CRIT=(0.005), POSTERIOR, SKIPC ;
>PRIOR  SMU=(1.0(0)5), SSIGMA=(0.0001(0)5);
>SCORE  ;
```