



# HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

## BAYESCİ MADDE TEPKİ KURAMI MODELLERİNDE ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

Merve AYVALLI KARAGÖZ

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye... En İyiyeye...*



Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

BAYESCİ MADDE TEPKİ KURAMI MODELLERİNDE ÖLÇME DEĞİŞMEZLİĞİNİN  
İNCELENMESİ

EXAMINING MEASUREMENT INVARIANCE IN BAYESIAN ITEM RESPONSE THEORY  
MODELS

Merve AYVALLI KARAGÖZ

Doktora Tezi

Ankara, 2023

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Merve AYVALLI KARAG¼Z'¼n hazırladıđı “Bayesci Madde Tepki Kuramı Modellerinde ¼lçme Deđiřmezliđinin İncelenmesi” bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde ¼lçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı	Prof. Dr. Nuri DOđAN	İmza
J¼ri Üyesi (Danıřman)	Prof. Dr. H¼lya KELECİOđLU	İmza
J¼ri Üyesi	Prof. Dr. Burcu ATAR	İmza
J¼ri Üyesi	Doç. Dr. C. Deha DOđAN	İmza
J¼ri Üyesi	Doç. Dr. Z¼lfikar DENİZ	İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, ¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 17 / 01 / 2023 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. Selahattin GELBAL  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bu araştırmanın amacı farklı koşullarda üretilen simülasyon veride Bayesci MTK modellerinde ölçme değişmezliğinin Bayes faktör ile incelenmesi ve alt gruplar içeren gerçek veride yapıya ait ölçme değişmezliğinin ÇGDFA ile test edilerek maddelerin değişmezliğinin Bayes faktör ile incelenmesidir. Bu amaç doğrultusunda Bayesci madde tepki kuramı modelleri kapsamında öncelikle simülasyon verilerde farklı örneklem büyüklüklerinde ve madde güçlük parametreleri arasındaki farkın ( $d_k$ ) 0.00, 0.10, 0.30, 0.50 ve 0.70 olduğu durumlarda Bayes faktör ile ölçme değişmezliği incelenmiştir. Gerçek verilerde ikili grup karşılaştırmaları hem ÇGDFA ile hem de Bayes faktör ile gerçekleştirilmiştir. Gerçek veri uygulaması TIMSS 2019 matematik başarı testi ile yapılmıştır. Bayesci analizler R programı ile WinBUGS üzerinden, ÇGDFA Mplus programı kullanılarak gerçekleştirilmiştir. Simülasyon veriden elde edilen sonuçlarda  $d_k = 0.3$  ve grup büyüklüğü 1500 ve daha büyük olduğunda Bayes faktörün ölçme değişmezliğinin sağlanmadığı yönünde kanıt ürettiği belirlenmiştir.  $d_k = 0.5$  olduğunda ise Bayes faktör ile ölçme değişmezliği sonuçlarının grup büyüklüğüne bağlı olarak değiştiği, büyük örneklerde Bayes faktörün ölçme değişmezliğinin sağlanmadığına dair kanıt üretmesinin daha olası olduğu belirlenmiştir.  $d_k = 0.7$  olduğunda ise grup büyüklüğünden bağımsız olarak Bayes faktör sonuçlarının ölçme değişmezliğinin sağlanmadığı yönünde kanıt ürettiği görülmüştür. Gerçek verilerden elde edilen ÇGDFA sonuçlarında Şili-Türkiye ve Türkiye-Singapur örneklerinde ölçme değişmezliği metrik değişmezlik düzeyinde sağlanırken Şili-Singapur örneklerinde yapısal değişmezlik düzeyinde sağlanmıştır. Gerçek verilerden elde edilen Bayes faktör sonuçlarında ise Şili-Türkiye ve Türkiye-Singapur örneklerinde Bayes faktör sonuçlarının genel olarak simülasyon veriye ait sonuçlarla tutarlılık gösterdiği, ancak Şili-Singapur örneğinden elde edilen sonuçların simülasyon veriden elde edilen sonuçlarla tutarlı olmadığı belirlenmiştir.

**Anahtar sözcükler:** Ölçme değişmezliği, Bayesci madde tepki kuramı modelleri, Bayes faktör, Çoklu-grup doğrulayıcı faktör analizi, TIMSS

## Abstract

The aim of this study is to examine measurement invariance in Bayesian IRT models with Bayes factor in simulated data generated under different conditions and to examine measurement invariance in real data with MGCFA and Bayes factor. For this purpose, within the scope of Bayesian IRT models, measurement invariance was examined with Bayes factor in simulated data at different sample sizes ( $n$ ) and when the difference between item difficulty parameters ( $d_k$ ) was 0.00, 0.10, 0.30, 0.50 and 0.70. In real data, pairwise group comparisons were performed with both MGCFA and Bayes factor. Real data application was conducted with TIMSS 2019 mathematics achievement test. Bayesian analyses were conducted with the R software, and MGCFA was conducted with Mplus. According to simulation data results, in the case of  $d_k=0.3$  and  $n=1500$  or more, the Bayes factor evidenced that measurement invariance was not achieved. In the case of  $d_k=0.5$ , the Bayes factor results were found to vary based on group size. In large samples, the Bayes factor was found to be more likely to evidence that measurement invariance was not achieved. In the case of  $d_k=0.7$ , the Bayes factor results evidenced that measurement invariance was not achieved regardless of group size. According to the MGCFA results, measurement invariance was achieved at the metric invariance level in Chile-Turkey and Turkey-Singapore samples, and at the configural invariance level in Chile-Singapore samples. The Bayes factor results of Chile-Turkey and Turkey-Singapore samples were generally consistent with those of the simulated data, but the Chile-Singapore results were not.

**Keywords:** Measurement invariance, Bayesian item response theory models, Bayes factor, Multi-group confirmatory factor analysis, TIMSS

## Teşekkür

Lisansüstü eğitimim boyunca bilgisi ve deneyimleriyle yoluma ışık tutarak desteğini hiç esirgemeyen, akademik duruşu ve çalışma disipliniyle beni kendine hayran bırakan kıymetli hocam ve danışmanım Prof. Dr. Hülya KELECİOĞLU'na,

Tez izleme komitelerimde yer alan ve tez sürecimin en başından sonuna kadar pozitif yaklaşımları, yapıcı eleştirileri ve fikirleriyle tezimin gelişmesine katkı sağlayan değerli hocalarım Prof. Dr. Burcu ATAR ve Doç. Dr. C. Deha DOĞAN'a,

Önerileriyle tezime katkı sağlayan değerli jüri üyesi hocalarım Prof. Dr. Nuri DOĞAN ve Doç. Dr. Kaan Zülfikar DENİZ'e,

Doktora süreci boyunca birbirimizi motive ettiğimiz, desteklerini, yardımlarını ve güler yüzlerini hiç esirgemeyen arkadaşlarım, yoldaşlarım Dr. Sinem DEMİRKOL'a ve Öğr. Gör. Mehmet Ali IŞIKOĞLU'na,

Tezimi yazabilmem için iş yerinde bana her türlü kolaylığı sağlayarak beni motive eden, destekleyen çalışma arkadaşlarım Öğr. Gör. Dr. Ebru KAYA BAŞAR ve Öğr. Dr. Deniz ÖZEL'e,

Doktora ders sürecinde bana evlerinin kapılarını açan, hayattaki şansım, canım dostum Kübra ZERAY'a ve biricik annesi canım teyzem Nafize ZERAY'a,

Bana her zaman güvenen ve arkamda duran, beni bugünlere getiren, en büyük destekçilerim canım annem Nebahat AYVALLI ve canım babam Mustafa AYVALLI'ya, canım kardeşlerim Melike AYVALLI ve Mehmet Ali AYVALLI'ya,

Bu zorlu süreçte gösterdiği özveri ve anlayışla beni hep destekleyen, motive eden, her yorulduğumda beni toparlayan canım eşim Tunçberk KARAGÖZ'e,

Doktora eğitimim boyunca yurt içi burs desteği sağlayan TÜBİTAK'a,

sonsuz teşekkürlerimi sunarım.

## İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	viii
Simgeler ve Kısaltmalar Dizini.....	ix
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	6
Araştırma Problemi.....	7
Sınırlılıklar.....	8
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	9
Ölçme Değişmezliği.....	9
Doğrulayıcı Faktör Analizi Temelli Yöntemler.....	10
Madde Tepki Kuramı Temelli Modeller.....	13
Bayesci Madde Tepki Kuramı Modellerinde Ölçme Değişmezliği.....	15
İlgili Araştırmalar.....	19
Bölüm 3 Yöntem.....	23
Araştırmanın Türü.....	23
Çalışma Grubu.....	23
Araştırma Verileri.....	24
Verilerin Analizi.....	28
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	35
Araştırmanın Birinci Alt Problemine İlişkin Bulgular.....	35
Araştırmanın İkinci Alt Problemine İlişkin Bulgular.....	42
Araştırmanın Üçüncü Alt Problemine İlişkin Bulgular.....	46



Tartışma.....	52
Bölüm 5 Sonuç ve Öneriler.....	55
Sonuçlar.....	55
Öneriler.....	58
Kaynaklar.....	61
EK-A: Simülasyon veriye ait madde güçlük parametre kestirimleri.....	lxxiv
EK-B: Örnek İz Grafikleri.....	lxxv
EK-C: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu Onay Bildirimi.....	lxxviii
EK-D: Etik Beyanı.....	lxxix
EK-E: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu.....	lxxx
EK-F: Thesis/Dissertation Originality Report.....	lxxxii
EK-G: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	lxxxiii

## Tablolar Dizini

Tablo 1 <i>Çalışma grubunda yer alan katılımcı sayıları ve ortalama başarı puanları</i>	24
Tablo 2 <i>TIMSS 2019 8. Sınıf matematik çerçevesi</i>	26
Tablo 3 <i>Analize dâhil edilen çoktan seçmeli matematik maddeleri</i>	28
Tablo 4 <i>Madde güçlük parametreleri arasında farkın olmadığı durumlarda (<math>dk = 0.0</math>) ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları</i>	36
Tablo 5 <i>Madde güçlük parametreleri arasındaki farkın <math>dk = 0.1</math> olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları</i>	37
Tablo 6 <i>Madde güçlük parametreleri arasındaki farkın <math>dk = 0.3</math> olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları</i>	38
Tablo 7 <i>Madde güçlük parametreleri arasındaki farkın <math>dk = 0.5</math> olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları</i>	40
Tablo 8 <i>Madde güçlük parametreleri arasındaki farkın <math>dk = 0.7</math> olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları</i>	41
Tablo 9 <i>Ükelere ait doğrulayıcı faktör analizi sonuçları</i>	42
Tablo 10 <i>Şili ve Türkiye örneklemi için çoklu grup doğrulayıcı faktör analizi sonuçları</i>	43
Tablo 11 <i>Şili ve Singapur örneklemi için çoklu grup doğrulayıcı faktör analizi sonuçları</i>	44
Tablo 12 <i>Türkiye ve Singapur örneklemi için çoklu grup doğrulayıcı faktör analizi sonuçları</i>	45
Tablo 13 <i>Ükelere ait madde güçlük parametre kestirimleri</i>	47
Tablo 14 <i>Şili-Türkiye Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark</i>	48
Tablo 15 <i>Türkiye-Singapur Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark</i>	49
Tablo 16 <i>Şili-Singapur Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark</i>	51

## Simgeler ve Kısaltmalar Dizini

**BF:** Bayes faktör

**ÇGDFA:** Çoklu-grup doğrulayıcı faktör analizi

**DFA:** Doğrulayıcı faktör analizi

**DMF:** Değişen madde fonksiyonu

**MCMC:** Markov chain Monte Carlo

**MTK:** Madde tepki kuramı

**OECD:** Organisation for Economic Co-operation and Development

**TIMSS:** Trends In International Mathematics and Science Study

## Bölüm 1

### Giriş

Bu bölümde öncelikle problem durumu açıklanmış; araştırmanın amacı, önemi, araştırma problemi ve alt problemlere yer verilmiş ve araştırmanın sınırlılıkları belirtilmiştir.

#### Problem Durumu

Eğitimde ve psikolojide bilişsel ve duyuşsal özellikler gibi gizil değişkenleri, farklı özelliklere sahip gruplarda ölçmek için uygulanan testlerin sıklığı gitgide artmaktadır. Bu tür çalışmalar, genellikle belirli gruplar arasında bir karşılaştırma içerir. Özellikle uluslararası uygulanan geniş ölçekli testlerde farklı gruplar arasında ya da zamana karşı karşılaştırma yapma amacı olan matematik başarısı, fen başarısı ve okuma becerileri gibi akademik konuların yanı sıra tutum, motivasyon, kaygı gibi psikolojik yapılar da ölçülmektedir (Davidov vd., 2014). Yapılan araştırmalarda ölçülen gizil yapılarda gruplar arası anlamlı karşılaştırmaların yapılabilmesi için ölçülen yapının tüm alt gruplarda aynı olması gerekir. Ölçme değişmezliği farklı gruplara ait sonuçları veya aynı grubun zamana karşı ölçümlerini karşılaştırmak için önemli bir ön koşuldur. Böylece ölçme aracının uygulandığı farklı kültür, cinsiyet ya da bölge gibi değişkenlik gösteren demografik özelliklere sahip kişiler ya da gruplar arasında bu farklılıklar göz önünde bulundurularak anlamlı karşılaştırmaların yapılması mümkün olmaktadır. Bu durum ölçülen yapının farklı gruplardaki genellenebilirliğini sağlayabilmek adına önemlidir (Brown, 2006).

Ölçme değişmezliğinin incelenmesi için farklı yöntemler kullanılmaktadır (Millsap, 2011). Bunlar doğrulayıcı faktör analizi temelli yöntemler ve madde tepki kuramı temelli yöntemler olmak üzere iki farklı grupta incelenebilir. Kavramsal olarak MTK'deki madde ayırt edicilik indeksi DFA'da faktör yüklerine benzerdir ve MTK'deki madde güçlük parametresi kategorik değişkenler için uygulanan DFA'da eşik parametresine karşılık geldiğinden (Takane ve de Leeuw, 1987) ampirik çalışmalar DFA ve MTK temelli yöntemlerin birbirinin alternatifi olduğunu ortaya koymuştur (McDonald, 2013).

Doğrulayıcı faktör analizi temelli yöntemlerde ölçme modellerinin gruplar arasındaki benzerliği çoklu-grup doğrulayıcı faktör analizi ile test edilerek ölçme değişmezliği incelenir. Bu yöntemin en önemli avantajlarından birisi, ölçme değişmezliğinin faktör yükleri, kesişimler, artık varyans, faktör varyansı, faktör kovaryansı, gizli ortalamalar gibi tüm yönleriyle incelenebilmesidir. Bu yöntemde yuvalanmış (hiyerarşik) modeller arası karşılaştırmalar yapılarak ölçme değişmezliği testi yapılır (Meredith, 1993; Steenkamp ve Baumgartner, 1998). Daha yüksek değişmezlik seviyeleri, gruplar arasında daha katı parametre eşitliği ve kısıtlamaları gerektirir ve elde edilmesi daha zordur. Hesaplama zorluklarının yanı sıra geleneksel yaklaşımları kullanan ölçme değişmezliği çalışmaları, genellikle gizil yapıların geleneksel yöntem altındaki katı varsayımlar nedeniyle ölçme değişmezliğini sağlayamadığını göstermişlerdir (Schmitt ve Kuljanin, 2008). Üç seviyeli çoklu-grup DFA (yapısal, metrik ve skaler değişmezlik) fikri uygulanabilir olsa da pratikte geleneksel çoklu-grup DFA'ya dayalı tam skaler değişmezliği elde etmek zordur. Skaler değişmezlik, özellikle birçok grup ve birçok faktör içeren büyük ölçekli çalışmalarda genellikle zayıf model uyumuna sahiptir. Bu nedenle metrik değişmezliğin son yıllarda çoklu grup analizlerinde değişmezliğin sağlanmasına ilişkin yeterli bir seviye olduğu kabul edilmesine rağmen, gruplar arasında faktör ortalamalarının karşılaştırılmasını engellediği raporlanmıştır (Muthén ve Asparouhov, 2018). Ayrıca metrik değişmezliğin sağlandığı ancak skaler değişmezliğin başarısız olduğu durumlar, maddenin farklı gruplar tarafından aynı şekilde algılandığı sonucuna varmak için yetersiz kalmaktadır. Değişmezlik modeli başarısız olduğunda, araştırmacılar, kısmi değişmezlik olarak da adlandırılan faktör ortalamalarını karşılaştırmak için genellikle bazı madde parametrelerini serbest kestirirler (Marsh vd., 2018). Kısmi değişmezlik yaklaşımı, kabul edilebilir bir model uyumu elde etmek için modifikasyon indekslerine dayalı aşamalı seçim modelini kullanır. Bu post-hoc analizlerdeki çoklu model spesifikasyonları Tip I veya Tip II hata olasılığını ayarlamayı imkânsız hale getirir (Byrne vd., 1989). MacCallum ve arkadaşları (1992) model modifikasyonlarının sayısının az sayıda olması gerektiğini ve yalnızca model uyumundaki ciddi sorunları gideren modifikasyonların yapılması gerektiğini önermişlerdir.

Doğrulayıcı faktör analizi ile değişmezlik testlerinde, gizil yapıdaki bir madde referans madde olarak seçilir ve faktör yükü 1'e sabitlenir ya da gizil yapının varyansı 1'e sabitlenerek metrik standart hale getirilir (Brown, 2006; Dimitrov, 2012). Bu yöntemle referans olarak belirlenen değişkenin değişmez olup olmadığının test edilmesine izin verilmediğinden, eğer değişken gruplar arasında değişmez değil ise elde edilen sonuçlar yanıltıcı olabilir (Cheung ve Rensvold, 1999; Yoon ve Millsap, 2007). Varyansı standartlaştırarak oluşturulan bu alternatif metrik tanımlama yöntemi, gizil yapının gruplar arası varyansının eşit olduğunu varsayar. Ancak kültürler arası karşılaştırmalar için bu durum geçerli olmayabilir (Cheung ve Rensvold, 2000; Millsap, 2011).

MTK temelli yöntemlerde ölçme değişmezliği ise alt gruplarda, madde düzeyi gözlemleri değerlendirmek için kullanılan madde yanlılığı yöntemleri (değişen madde fonksiyonu) ile belirlenebilir. MTK temelli yöntemlerde, madde düzeyinde doğrusal bir ilişki olduğunu varsayan DFA temelli yöntemlerden farklı olarak gizil yapı ile madde düzeyinde gözlenen puan arasında doğrusal olmayan bir ilişki ortaya koyulmaktadır. MTK temelli yöntemler arasında Olabilirlik oran testi, Lord'un  $\chi^2$  testi, Raju alan ölçüleri yöntemi, Wald istatistiği gibi yöntemler bulunmaktadır (Millsap, 2011).

MTK kapsamında yapılan değişmezlik testleri, referans ve odak gruplar için ayrı ayrı elde edilen madde parametrelerinin ya da madde karakteristik eğrilerinin karşılaştırılması yoluyla yapılır. Bunun için maddelerin ortak bir ölçekte sunulması gerekir. Madde parametrelerinin ortak bir ölçeye yerleştirilebilmesi için de bir ortak maddeye (anchor item) ihtiyaç vardır. Eğer ortak madde değişmez değil ise, oluşturulan bu ortak ölçeğin geçerliği ile ilgili sorun olacağından, değişmezlik testi sonuçlarının hatalı olmasına yol açar (Finch, 2005; Stark vd., 2006; Woods, 2009).

Bir maddenin değişmez olması, aynı yetenek düzeyindeki gruplarda, ölçülen özelliğe ait bir maddenin tüm gruplardaki kişiler tarafından aynı cevaplanma olasılığına sahip olması ile ilgilidir. Bir maddenin iki grup için ölçme değişmezliği testinde kurulan sıfır hipotezi, her iki gruptaki madde parametresinin eşitliğine, alternatif hipotez ise gruplardaki madde

parametrelerinin farklı olduğuna karşılık gelir (Verhagen vd., 2016). MTK temelli modellerden olabilirlik oran testi madde parametrelerinin değişmezliğini, parametrelerin uyum indeksleri ile değerlendiren bir modeldir (Thissen vd., 1993). Bununla birlikte Wald teste (Lord, 1980) dayanan ve madde parametrelerinin eşitliğini iki grup arasında değerlendiren ölçme değişmezliği testleri de bulunmaktadır (Woods vd., 2013). Ölçme değişmezliği testi için geleneksel yöntemlerin dışında Muthen ve Asparouhov (2013) tarafından Alignment yöntemi önerilmiştir. Alignment yönteminde tüm madde parametrelerinin gruplar arasında değişmesine izin verilir ve daha sonra gruplar arasındaki küçük ya da büyük farklılıklara yönelik döndürme yöntemi uygulanır. Bu yöntemde değişmezlik testi madde parametreleri arasındaki önemli farklılıkları inceleyerek ilerler (Muthen ve Asparouhov, 2013).

Mevcut ölçme değişmezliği yöntemlerinde, ölçme değişmezliği hipotezinin lehine kanıt toplayamama ve analiz öncesinde ortak madde belirleme gerekliliği gibi temel iki sınırlılık bulunmaktadır. (Verhagen, vd., 2016).

Verhagen ve Fox (2013); ülkeler, okullar gibi yuvalanmış ve büyük gruplarda ölçme değişmezliği hipotezinin karşılaştırılabilmesi için gruplar arasında madde parametrelerinin varyansına dayalı Bayesci MTK kapsamında değişmezliğin değerlendirilebilmesi için Bayes faktör önermişlerdir. Bayesci MTK kapsamındaki yöntemler geleneksel yöntemlerden daha esneklerdir ve frekansçı yöntemlerin aksine ölçme değişmezliği hipotezinin hem lehine hem de aleyhine kanıt toplanmasını sağlayarak ölçme değişmezliği testinin daha geniş kapsamlı biçimde değerlendirilmesine imkân tanır (Verhagen, 2012). Ayrıca bu yöntemle ortak maddeye de ihtiyaç duyulmamaktadır (Verhagen vd., 2016).

Araştırmada Bayes faktör hesaplanarak yapılan hipotez testi ile ölçme değişmezliği testi diğer yöntemlere kıyasla çeşitli avantajlara sahiptir. Bu avantajlardan ilki iki farklı kültürün yer aldığı bir ölçme uygulamasında, maddelerin değişmezliği ile ilgili bilgi mevcut olmadığı durumda ölçme değişmezliğinin Bayes faktör ile direkt olarak değerlendirilebilmesi mümkündür. Frekansçı yaklaşımlarda uygulanan ölçme değişmezliği testlerinde yalnızca sıfır hipotezi test edilir ve alternatif hipotez ile ilgili kanıtlar değerlendirilmeden elde edilen sonuçlar

özellikle düşük güç düzeyine sahip çalışmalarda  $H_0$  lehine yanlı sonuçlara yol açabilir (Verhagen ve Wagenmakers, 2014). Hipotez testinde Bayes yaklaşımını kullanmak değişmezliğin direkt olarak test edilebilmesine olanak sağlar ve değişmezlik lehine kanıt düzeyi Bayes faktör ile ölçülür. Böylece bir hipotezi reddetmeye odaklanmak yerine hem  $H_0$  hem de  $H_1$  hipotezi için kanıtlar değerlendirilebilir. Test sonucunda Bayes faktör  $H_0$  ya da  $H_1$  lehine kanıt üretebileceği gibi hangi hipotezin tercih edilmesi gerektiğine dair kararsız da kalabilir (Kass ve Raftery, 1995).

Diğer avantajlarından birisi ise değişmezlik testinde ortak maddeye ihtiyaç duyulmamasıdır. Diğer ölçme değişmezliği testleri en az bir ortak maddeye ihtiyaç duyar. Özellikle çoklu grup karşılaştırmalarında önceden değişmezliği sağlanmış bir maddenin testte yer almaması durumunda değişen madde fonksiyonu içeren bir ortak madde seçimi değişmezlik testinin sonuçlarını etkileyeceğinden yanlı kestirimler yapılmasına neden olabilir (Woods ve Grimm, 2011). Ayrıca madde parametrelerinde herhangi bir kısıtlama yapılmasına gerek olmaması, Bayes faktör ile ölçme değişmezliği testinin avantajları arasındadır ve yöntemin keşif amaçlı analizler için kullanılabilmesini sağlar (Verhagen vd., 2016).

Psikoloji ve eğitimde Bayes yaklaşımı ile yapılan ölçme değişmezliği testleri artmaktadır. Mevcut Bayes yaklaşımını temel alan ölçme değişmezliği testleri, değişmezliği belirleyebilmek için genellikle Fox (2010) tarafından önerilen rastgele madde etkileri modelini kullanılır. Bu modelde genel ortalamadan sapmalar her gruba özgü madde parametresi için belirlenir. Bu sapmalara göre gruplar arasındaki varyans, ölçme değişmezliğini belirlemek için kullanılır. Verhagen ve Fox (2013) bir evrenden rastgele seçilen bir örnekleme Bayes faktörün çoklu değişmezlik hipotezlerinde kullanılabilmesini, ölçme değişmezliği testlerinde düşük Tip 1 hata oranıyla beraber iyi bir güç düzeyine sahip olduğunu belirlemişlerdir. Ancak rastgele madde etkileri modeli yalnızca çoklu grup karşılaştırmalarında kullanıma uygundur. Verhagen vd. (2016), ikili grup karşılaştırmaları için madde güçlük parametre farklılıklarını doğrudan değerlendirme imkânı veren Bayes faktör testini önermişlerdir. Çoklu grupların Bayesci yöntemlerle karşılaştırılmasına yönelik çalışmalar olsa da az sayıda grubun karşılaştırılmasına



ilişkin Bayes yaklaşımı ile ölçme değişmezliği çalışmaları literatürde sınırlı sayıdadır. Kùltürler arası test uygulamalarının artması, ölçme değişmezliğinin önemini de artırmaktadır. Ölçme değişmezliğinin belirlenmesinde sıklıkla kullanılan doğrulayıcı faktör analizine dayalı geleneksel yöntemler, farklı model uyumlarının karşılaştırılmasını gerektirmektedir ve bu modellerin her birinin uyum göstermesi beklendiđi için bu yöntemler zaman alıcıdır (White, 2000). Bu bağlamda ölçme değişmezliğinin değerlendirilebilmesi için daha pratik yöntemlere ihtiyaç duyulmaktadır.

### **Araştırmanın Amacı ve Önemi**

PISA, TIMSS, PIRLS, TALIS gibi uluslararası geniş ölçekli test uygulamalarının artması, birden fazla ülkenin ya da farklı kültürel özelliklere sahip alt gruplara ait test sonuçlarının karşılaştırılmasını da beraberinde getirmektedir. Bununla birlikte özellikle uygulama sonuçlarında ön sıralarda olan ülkelerdeki eğitim sistemlerine ilişkin bilgiler diğer katılımcı ülkelerdeki eğitim politikalarını da etkilemektedir (OECD, 2019). Farklı gruplara bir test uygularken test sonuçlarının grup üyeleri arasında karşılaştırılabilir olması da önemlidir. Gruplar arasında anlamlı karşılaştırmalar yapabilmek için ölçülen gizil değişken ( $\theta$ ) ortak bir ölçekte ölçülmelidir. Ortak ölçek kullanılarak yapılan analizlerde ölçme değişmezliği varsayımının ihlal edilip edilmediđi göz önünde bulundurulmalıdır (Fox, 2010).

OECD (2019) geniş ölçekli testlerde değişmezlik analizleri ile ilgili yayınladıđı raporunda yeni ölçme değişmezliği yaklaşımlarının esas olarak bir “kavram kanıtı” olduđunu ve bunların uygulanmasının geleneksel ÇGDFA ve MTK yaklaşımlarıyla ilgili mevcut sorunları azaltıp azaltmayacađı konusunda henüz kesin bir cevap vermediđini vurgulamıştır. Ayrıca bu yaklaşımların beklentileri karşılayabileceđine dair karar vermeden önce daha fazla tecrübeye ve araştırmaya ihtiyaç olduđu belirtilmiştir.

Özellikle ortak madde ihtiyacı olmaksızın doğrudan madde parametreleri arasındaki farka dayalı olarak Bayes faktör ile belirlenen ölçme değişmezliği testlerinin geniş ölçekli testlerde uygulanması ölçme değişmezliği ile ilgili alanyazına katkı sağlayacađından teorik

olarak oldukça önem taşımaktadır. Ayrıca özellikle ortak madde ihtiyacının olduğu durumlarda Bayes faktör ile ölçme değişmezliği analizleri, hangi maddelerin değişmez olduğunu keşfetme amacıyla temel bir başlangıç analizi olarak kullanılabilir. Simülasyon çalışmasından elde edilen bulgular ışığında da ölçme değişmezliği hipotezinin lehine ve aleyhine toplanan kanıtlarla, direkt olarak madde güçlük parametreleri arasındaki farka dayalı bir değerlendirme yapılabilecektir.

Bu kapsamda araştırmanın amacı, farklı koşullarda üretilen simülasyon veride Bayesci MTK modellerinde ölçme değişmezliğinin Bayes faktör ile incelenmesi ve alt gruplar içeren gerçek verilerde ÇGDFA ile ölçülen yapının ölçme değişmezliği belirlenerek maddelerin ölçme değişmezliğinin Bayes faktör ile incelenmesidir.

### **Araştırma Problemi**

Bayes faktör ile belirlenen ölçme değişmezliği sonuçları örneklem büyüklüğü ve madde güçlük parametresi farkına dayalı simülasyon koşullarına göre nasıl değişmektedir? TIMSS 2019 matematik alt testi ÇGDFA ve Bayes faktör sonuçlarına göre değişmezlik göstermekte midir?

### **Alt Problemler**

1. Bayes faktör ile belirlenen ölçme değişmezliği, örneklem büyüklüğü ( $n = 500, n = 1000, n = 1500, n = 2000$ ) ve madde güçlük parametreleri arasındaki farka göre ( $d_k = 0.0, d_k = 0.1, d_k = 0.3, d_k = 0.5, d_k = 0.7$ ) nasıl değişmektedir?
2. TIMSS 2019 uygulaması matematik alt testinde ÇGDFA ile Türkiye, Şili ve Singapur örneklemeleri arasında ölçme değişmezliği sağlanmakta mıdır?
  - 2.1. Şili-Türkiye örneklemelerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?
  - 2.2. Şili-Singapur örneklemelerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?
  - 2.3. Türkiye-Singapur örneklemelerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?

3. TIMSS 2019 uygulaması matematik alt testinde ikili puanlanan maddelerde Türkiye, Şili ve Singapur örneklemi arasında Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?

3.1. Şili-Türkiye örneklerinde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?

3.2. Türkiye-Singapur örneklerinde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?

3.3. Şili-Singapur örneklerinde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?

### **Sınırlılıklar**

1. Araştırma ikili puanlanan verilerle sınırlıdır.
2. Araştırmada kullanılan gerçek veriler TIMSS 2019 matematik alt testi 1 ve 14'nolu kitapçıkta yer alan ME01 bloğundaki 11 çoktan seçmeli madde ile sınırlıdır.
3. Araştırmada TIMSS 2019 uygulamasına katılan Şili, Türkiye ve Singapur örneklemi ile sınırlıdır.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Araştırmanın bu bölümünde ölçme değişmezliğine ilişkin kuramsal yapı aktarılarak geleneksel ölçme değişmezliği belirleme yöntemleri ve Bayesci madde tepki kuramı modellerinde ölçme değişmezliğinden bahsedilerek alanyazında yer alan ilgili araştırmalara yer verilmiştir.

#### Ölçme Değişmezliği

Ortalama farklılıklarının test edilmesindeki en önemli varsayım, ölçülmek istenen temel yapının gruplar arasında eşdeğer olmasıdır (Bejar,1980; Labouvie, 1980; Rock vd., 1978). Ölçme değişmezliği ve ölçme eşdeğerliği yerine kullanılan “ölçme eşdeğerliği”, “ölçme eşitliği” gibi kavramlar için çok sayıda tanım yapılmıştır (Vandenberg ve Lance, 2000). Johnson (1998) tarafından yapılan bir derleme çalışmasında ölçme eşdeğerliği için elliden fazla kavram kullanıldığı belirlenmiştir. Bu kavramlar yorumlama bağlamındaki eşdeğerlik kavramları için “anlam eşdeğerliği” ve kültürler arası karşılaştırmalarda kullanılan yöntemler için “yöntemsel eşdeğerlik” olmak üzere iki gruba indirgenmiştir. Bu tanımların ortak paydası, ölçülen özelliklerin tüm alt gruplarda karşılaştırılabilirliği ile ilgili olmasıdır. Ölçme değişmezliği bir ölçme aracının, ölçülmek istenen psikolojik yapının tüm alt gruplarda aynı şekilde ölçüldüğünü gösteren istatistiksel bir özelliğidir (Meredith,1993; Vandenberg ve Lance, 2000). Yani, ölçülmesi hedeflenen bir ya da daha fazla örtük özelliği ( $\theta$ ) ölçmek için tasarlanan  $X'e$  ait özelliklerin, tüm popülasyonlarda aynı olmasıdır.  $K$  tane popülasyonun ( $\Pi_k, k = 1,2,3, \dots, K$ ) olduğu varsayıldığında,  $X'in$   $k$ . popülasyona ait ölçülen tepki fonksiyonu  $P_k(X|\theta) = P(X|\theta)$  olmalıdır. Ölçme değişmezliğinin belirlenebilmesi için, hedef gizil değişken ve ölçme yapılacak popülasyonların net bir şekilde belirlenmesi gerekir (Millsap, 2011). Hedef gizil değişkeni ölçmek için tasarlanan  $X'e$  ait fonksiyonlar, gizil değişken dışında popülasyon özelliklerinden de etkilenebilir. Bu etkiler diğer etki kaynakları olarak adlandırılır (Ackerman, 1992). Söz konusu etki kaynakları ölçme değişmezliğinin popülasyon temelli ihlallerini tanımlar ve bu

fonksiyon  $P_k(X|\theta) \neq P(X|\theta)$  şeklinde ifade edilir. Bu fonksiyon herhangi bir  $k$  popülasyonundaki bireyler için,  $X$ 'e ait gözlenen puanların bazı  $\theta$  değerlerindeki koşullu dağılımlarının, diğer popülasyonların koşullu dağılımlarından farklı olacağı anlamına gelir (Millsap, 2011).

Bir ölçme aracının ölçme değişmezliğinin olmadığı durumlarda iki temel sorun vardır. Bunlardan ilki geçerlik sorunudur (Borsboom ve diğerleri, 2004). Eğer madde fonksiyonları gruba göre farklılaşıyorsa, ölçme aracının tüm alt gruplarda aynı yapıyı aynı şekilde ölçmediği söylenebilir. İkincisi sorun ise, ölçülen yapı ve yanıtlanan maddeler arasındaki ilişkinin her bir grup için farklılık göstermesi görmezden gelinerek değerlendirme ve karşılaştırmalar yapılmasıdır (Horn ve Mcardle, 1992). Ölçme değişmezliği testleri doğrulayıcı faktör analizi temelli yöntemler ve madde tepki kuramı temelli yöntemler olmak üzere iki farklı grupta incelenmektedir.

### **Doğrulayıcı Faktör Analizi Temelli Yöntemler**

Ölçme değişmezliği, kategorik değişkenlerle tanımlanan gruplarda çoklu-grup doğrulayıcı faktör analizi ile test edilir. ÇGDFA'da bir dizi yuvalanmış model, gizil faktör modelinin her kısıtlamasında test edilir (Jöreskog, 1971). Klasik bir DFA modelinde  $i$ . maddeye ( $i = 1, \dots, p$ ) verilen  $x_i$  yanıtına ait gizil değişkenin  $\xi_j$  ( $j = 1, \dots, J$ ) doğrusal fonksiyonu Denklem 1'de gösterilmiştir

( $\tau_i$ : kesişim parametresi,  $\delta_i$ : seçkisiz hata terimi,  $\lambda_{ij}$ :  $x_i$ 'nin  $\xi_j$  üzerindeki regresyon eğim katsayısı)

$$x_i = \tau_i + \lambda_{ij}\xi_j + \delta_i \quad (1)$$

Eğim katsayısı ya da faktör yükü  $\lambda_{ij}$ ,  $\xi_j$ 'de bir birimlik değişim olduğunda  $x_i$ 'nin beklenen değerini gösterir (Sörbom, 1974).

$p$  madde,  $m$  gizil değişken, ve her grup  $k$  ( $k = 1, \dots, K$ ) için faktör yapısının aynı olduğu varsayıldığında ölçme modeli Denklem 2'deki gibidir.

$$x^k = \tau^k + \Lambda^k \xi^k + \delta^k \quad (2)$$

$x^k$ :  $k$  grubunda gözlenen değişkenlerin  $px1$  vektörü  
 $\xi^k$ :  $k$  grubunda gizil değişkenlerin  $mx1$  vektörü  
 $\delta^k$ :  $k$  grubunda ölçme hatasının  $px1$  vektörü  
 $\tau^k$ :  $k$  grubunda madde kesişimlerinin  $px1$  vektörü  
 $\Lambda^k$ : faktör yüklerinin  $pxm$  matrisi

Denklem 2,  $p$ . maddelerde gözlenen puanların temel faktör puanlarının bir fonksiyonu olduğunu, ancak farklı kesişim katsayıları ( $\tau_i^k$ ) ve ölçek metrikleri ( $\lambda_{ij}^k$ ) nedeniyle gözlenen puanların gruplar arasında karşılaştırılabilir olmayabileceğini göstermektedir. Modeli tanımlamak için gizil yapılara bir ölçek atanması gerekir. Gruplar arası karşılaştırmalarda bu, faktör başına bir maddenin faktör yükünün 1'e sabitlenmesi ile yapılır. Faktör yükü sabitlenen maddeler, tüm gruplarda aynı madde olacak şekilde seçilerek referans madde olarak kabul edilir (Steenkamp ve Baumgartner, 1998). Denklem 3  $k$ . grupta gözlenen madde ortalaması ve gizil yapı ortalaması arasındaki ilişkiyi göstermektedir.

$$\mu^k = \tau^k + \Lambda^k G^k \quad (3)$$

$\mu^k$ : Madde ortalamalarına ait  $px1$  vektörü  
 $G^k$ : Gizil yapı ortalamasının  $mx1$  vektörü

Gizil değişkenlerin belirli bir başlangıç noktası yoktur. Bu belirsizliği ortadan kaldırmak için parametreler üzerinde bazı kısıtlamalar gereklidir. Olası durumlardan biri, her bir gizil değişkenin referans maddesine ait kesişim katsayısının sıfıra sabitlenmesidir. Bu, gizil değişkenlerin ortalamasını referans değişkenin ortalamasına sabitler. İkinci olası durum ise referans gruptaki gizil ortalamaların vektörünü sıfıra sabitleyerek her bir faktörün kesişim katsayısını gruplar arasında değişmez olacak şekilde 1'e sabitlemektir. Eğer modele daha fazla kısıtlama getirilirse parametreler aşırı tanımlanmış olur ve modelde ortalamaların uyumu incelenebilir.

Gizil değişkenin  $k$ . Gruptaki kovaryans matrisi ise Denklem 4'te gösterilmiştir.

$$\Sigma^k = \Lambda^k \Phi^k \Lambda^{k'} + \Theta^k \quad (4)$$

Ölçülen gizil değişken için koşullu ortalama ve kovaryans matrislerinin farklı gruplarda da aynı olması ancak  $\tau, \Lambda, \Phi$  parametrelerinin değişmez olması durumunda gerçekleşir (Steenkamp ve Baumgartner, 1998). ÇGDFA'nın en önemli avantajı ölçme değişmezliğinin tüm yönlerinin (faktör yükleri, kesişimler, faktör varyansı, faktör kovaryansı, artık varyans ve gizil ortalamalar) incelenebilmesidir. Yuvalanmış olarak oluşturulan modellerin değişmezlik testi sonuçları kendi aralarında karşılaştırılarak, her seviyeye ait çıkarımlar ayrı ayrı yapılabilir (Meredith, 1993). Daha yüksek değişmezlik seviyeleri, gruplar arasında daha katı parametre kısıtları gerektirdiği için elde edilmesi daha zordur.

### Ölçme Değişmezliğinin Türleri

**Yapısal Değişmezlik.** Yapısal değişmezlikte aynı sayıda faktörün ve faktör yüklerine ait örüntünün çalışılan tüm gruplarda aynı olması gerekir. Bu bağlamda yapısal değişmezlik sıfır olmayan elementleri tüm gruplarda aynı konumda olmaya zorlarken,  $\Lambda^k$  örüntü matrislerinin sıfır olmayan elementlerinin gruplar arasında değişmesine izin verir (Steenkamp ve Baumgartner, 1998). Bu değişmezlik modelinde tüm gruplarda aynı gizil yapının karşılaştırılması mümkündür ve en zayıf model olarak kabul edilir (Van de Vijver, 1998). Diğer değişmezlik modelleri için ön koşul olduğundan, temel değişmezlik modeli olarak kullanılır (Vandenberg ve Lance, 2000).

**Metrik Değişmezlik.** Metrik değişmezlik gruplar arasında eşit metrik ve ölçek aralıkları sunarak daha güçlü bir değişmezlik testi sağlar (Rock vd., 1978). Bir madde metrik değişmezliği sağlıyorsa farklı madde puanları gruplar arasında karşılaştırılabilir ve gözlenen bu madde farklılıkları, söz konusu yapının gruplar arasındaki farklılıklarının göstergesidir. Faktör yükleri, gizil puanlardaki değişikliklerin gözlenen puanlardaki değişikliklerle ilişkisini tanımlayan bilgi taşır. Bu doğrultuda faktör yükleri tüm gruplarda aynı olacak şekilde kısıtlanarak metrik değişmezlik testi edilir (Steenkamp ve Baumgartner, 1998).

$$\Lambda^1 = \Lambda^2 = \dots = \Lambda^K \quad (5)$$

**Skaler Değişmezlik.** Yapısal ve metrik değişmezlik yalnızca farklı gruptaki maddelerin kovaryansları hakkında bilgi içerir. Ancak grup ortalamalarını karşılaştırmak çoğu araştırma için önem taşır. Bu tür karşılaştırmaların anlamlı olabilmesi için skaler değişmezliğin test edilmesi gerekir (Meredith, 1993). Bu değişmezlik türünde gözlenen madde ortalamalarının gruplar arasındaki farklılıkları, ölçülen yapıdaki farklılıklardan kaynaklanır. Bu durum gruplar arasında gizil ortalamalar ve gözlenen ortalamalar arasındaki farklılığı işaret eder. Bir madde farklı gruplarda aynı metrikte olsa bile (metrik değişmezlik sağlandığı durumlar), madde puanları sistematik olarak yanlılık gösterebilir (Steenkamp ve Baumgartner, 1998). Skaler değişmezlik Denklem 6'daki gibi metrik değişmezlik modeline ek kısıtlama getirilerek test edilir, kesişim katsayıları ve faktör yükleri ( $\tau, \Lambda$ ) tüm gruplarda eşittir (Millsap, 2011). Skaler değişmezlik de gruplar arasında gizil ortalamaların anlamlı bir şekilde karşılaştırılmasına olanak sağlar (Widaman ve Reise, 1997).

$$\tau^1 = \tau^2 = \dots = \tau^K \quad (6)$$

**Katı Değişmezlik.** Katı değişmezlikte faktör varyans ve kovaryanslarında da değişmezlik testi yapılır ve her iki testte de gizil yapılar arasındaki ilişki değişmez ise katı değişmezlik sağlanır. Bu değişmezlik türünde ise kesişim katsayıları ( $\tau$ ) ve faktör yüklerine ek olarak ( $\Lambda$ ), ölçme hatalarının da tüm gruplarda eşit olması (Denklem 7) beklenir (Steenkamp ve Baumgartner, 1998).

$$\theta^1 = \theta^2 = \dots = \theta^K \quad (7)$$

### Madde Tepki Kuramı Temelli Modeller

Madde düzeyinde doğrusal bir ilişkinin olduğunu varsayan DFA'dan farklı olarak madde tepki kuramı madde düzeyinde gözlenen puan ve gizil yapı arasında doğrusal olmayan bir ilişki olduğunu varsayar (Millsap, 2011). Ölçme değişmezliğinin MTK'de karşılığı ise değişen madde fonksiyonudur. MTK modellerinden 1 parametrelili lojistik model ele alındığında  $i$ . bireyin ( $i = 1, 2, \dots, I$ )  $j$ . maddeye ( $j = 1, 2, \dots, J$ ) ait doğru yanıt verme olasılığı Denklem 8'deki gibidir ( $\theta_i$ : birey yetenek parametresi,  $b_j$ : madde güçlük parametresi).



$$P(Y_{ij} = 1 | \theta_i, b_j) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)} \quad (8)$$

Bir maddenin değişmez olması için farklı gruplarda madde parametrelerinin sabit olması gerekir. Yani, aynı yetenek düzeyine sahip bireylerin bir testte yer alan maddelere doğru yanıt verme olasılığı, grup üyeliğinden bağımsız şekilde yalnızca maddelere ait madde tepki fonksiyonlarının aynı olmasına bağlıdır. Ancak grup özelliklerine bağlı olarak bir maddeye ait madde tepki fonksiyonları aynı değilse bu maddenin değişen madde fonksiyonuna (DMF) sahip olduğunu gösterir (Lord, 1980). DMF, madde karakteristik eğrisindeki farklılıklar ve madde parametrelerindeki farklılıklar olarak iki grupta incelenebilir (Osterlind ve Everson, 2009).

### ***Madde Parametreleri Arasındaki Farka Dayalı Yöntemler***

MTK'de madde yanlılığının, madde parametrelerindeki grup farklılıkları tarafından tanımlandığı göz önüne alındığında, yanlılığı araştırmaya yönelik en doğrudan yaklaşım, gruplar arasında madde parametrelerine ait tahmin değerlerini karşılaştırmaktır. Örneklem büyüklükleri ve incelenen MTK modeli göz önüne alındığında beklenenden daha büyük olan grup farklılıkları, yanlılığın kanıtı olarak alınır (Millsap, 2011). Madde parametre kestirimlerinin karşılaştırılmasına dayalı olan bu yöntemde, odak ve referans gruptaki madde güçlük parametrelerinin eşit olduğu yokluk hipotezi kurulur. Madde güçlük parametreleri arasındaki fark ( $d$ ) Denklem 9'da gösterilmiştir (Lord, 1980).

$$d = \frac{\hat{b}_R - \hat{b}_O}{SE(\hat{b}_R - \hat{b}_O)} \quad (9)$$

$SE(\hat{b}_R - \hat{b}_O)$  : Referans grup ve odak grup arasındaki madde güçlük parametre farkına ait standart hata  
 $d \sim N(0,1)$

Lord (1980) ayrıca madde güçlük ve madde ayırıcılık parametrelerindeki farklılıkları aynı anda inceleyen ki-kare ( $\chi^2$ ) testini önermiştir. Bu test de yine her iki gruptaki tüm madde parametrelerinin eşit olduğu yokluk hipotezine dayanır. Lord "Wald" test (1980) gruplar arasında madde parametrelerinin vektörlerini karşılaştırır. Eğer karşılaştırılan maddeye ait

parametre vektörleri gruplar arasında farklılaşıyorsa, madde fonksiyonları da gruplar arasında farklılık gösterir.

**Olabilirlik Oran Testi.** Olabilirlik oran testi, genel model uyumu için bir testtir. Bu test madde düzeyinde farklılık gösteren iki model arasındaki olasılığı karşılaştırır. Modellerden biri, odak ve referans gruptaki madde parametrelerini eşit olacak şekilde kısıtlarken diğer modelde de bu parametreleri her bir grup için farklı kestirir, daha sonra da bu modeller kıyaslanır (Thissen vd., 1988, Camilli ve Shepard 1994). Her iki gruptaki olabilirlik oranları ki-kare testi için kullanılır ve DMF gösteren maddeler için ki-kare testinin anlamlı çıkması beklenir.

### ***Madde Karakteristik Eğrileri Arasındaki Alan Ölçülerine Dayalı Yöntemler***

Madde güçlük ve madde ayırıcılık parametrelerinin her iki grup için madde karakteristik eğrilerinin karşılaştırılmasına dayalı olan bu model ilk kez Raju (1988) tarafından önerilmiştir. DMF madde karakteristik eğrileri arasındaki alanın büyüklüğü ile doğru orantılıdır. Madde karakteristik eğrileri arasındaki farka dayalı olarak işaretli alan indeksi ve işaretli alan indeksi ile istatistikler çıkarımlar yapılabilir (Raju, 1988, Raju vd., 1995).

### **Bayesci Madde Tepki Kuramı Modellerinde Ölçme Değişmezliği**

Madde tepki kuramı modelleri bireylerin ölçülen yetenek düzeylerini ve bir maddeye verilen yanıt arasındaki ilişkiyi gösteren modellerdir (Demars, 2010). Bu çalışmada, Bayesci MTK modeli 1 parametrelili lojistik model çerçevesinde ele alınmıştır.

1 parametrelili lojistik modelde,  $i$  birey ( $i=1, \dots, N$ ) ve  $k$  madde ( $k=1, \dots, K$ ) bulunan bir testte ikili puanlanan bir maddeye ait doğru yanıt olasılığı ( $\theta_i$ : Gizil değişkene ilişkin birey parametresi,  $b_k$ : madde güçlük parametresi) Denklem 10'deki gibi tanımlanmaktadır.

$$P(Y_{ik} = 1 | \theta_i, b_k) = \frac{e^{(\theta_i - b_k)}}{1 + e^{(\theta_i - b_k)}} \quad (10)$$

Rastgele etki modeli olarak da bilinen Bayesci MTK modellerinde, model parametreleri rastgele değişkenlerdir ve verileri gözlemeden önce parametrelerin gerçek değerleri hakkındaki belirsizliği yansıtan önsel dağılımlara sahiptir. Verileri gözlemledikten sonra, önsel

bilgiler verilerden elde edilen bilgilerle birleştirilir ve bir sonsal dağılım oluşturulur. Bayes çıkarımları gözlemlenen verilere bağlı olarak yapılır ve parametrelerle ilgili çıkarımlar da doğrudan sonsal yoğunluk fonksiyonlarından yapılabilir (Fox, 2010). Bu modellerde madde parametreleri için önseller madde karakteristiklerinin varyasyonlarını belirler ve bir teste bulunan maddeler madde evreninin rastgele bir örnekleme olarak kabul edilir. Madde parametreleri için önseller tüm maddeler için ortak ortalama ve varyansla standart normal dağılım gösterirler (Janssen vd., 2000; De Boeck, 2008).

$$b_k \sim N(b_0, \sigma_{b_k}^2)$$

Her bir parametre için sonsal dağılımlar ise, söz konusu maddenin tüm gruptaki ortalama doğruluk yüzdesi ile  $b_0$  ve  $\sigma_{b_k}^2$  önsel dağılımının birleşiminden oluşan fonksiyondur. Birey parametrelerinin önsel dağılımı için ise  $N(0,1)$  standart normal önsel seçilir (Fox, 2010).

Bayesci MTK modelde ölçme değişmezliği testinde ise, çoklu grup karşılaştırmaları yapılacağı için çoklu-grup MTK kapsamında ele alınır. Çoklu-grup MTK model, test puanlarında ve madde karakteristiklerinde gruplar arasında farklılıklara izin verir (Bock ve Zimowski, 1997), böylece Bayesci MTK modelde maddeler arasındaki varyasyonun yanı sıra gruba özgü madde parametrelerinin gruplar arasındaki değişimi de hesaba katılarak bir ölçme modeli oluşturulur. Oluşturulan bu çoklu-grup 1 parametrelili lojistik modelde, bir gruptaki bir bireyin doğru yanıt olasılığı denklem 11'de gösterilmiştir (j. grup, i. birey,  $\theta_{ij}$ = gruba özgü birey parametresi,  $\tilde{b}_{kj}$ = gruba özgü madde parametresi).

$$P(Y_{ijk} = 1 | \theta_{ij}, \tilde{b}_{kj}) = \frac{e^{(\theta_{ij} - \tilde{b}_{kj})}}{1 + e^{(\theta_{ij} - \tilde{b}_{kj})}} \quad (11)$$

Gruba özgü birey parametresi olan  $\theta_{ij}$  hiyerarşik olarak modellenir ve grup ortalaması  $\mu_j$  etrafında normal dağılım gösterir.

$$\theta_{ij} \sim N(\mu_j, \sigma_j)$$

Çoklu-grup MTK modellerinde ölçme varyansını modellemek için gruba özgü madde parametrelerinin çok düzeyli bir yapıya sahip olduğu varsayılır (Fox, 2010). Tüm maddeler için

gruba özgü sapmalar ortalaması sıfır ve varyansı  $\sigma_{b_k}^2$  olacak şekilde normal dağılım gösterir. Bu varyans bileşeni madde fonksiyonlarının gruplar arasında değişkenliğini tanımlar ve söz konusu varyans sıfır olduğunda değişkenlik olmadığı için maddenin değişmez olduğu kabul edilir. Eğer az sayıda ya da belirlenmiş gruplar arasında bir ölçme değişmezliği çalışması yapılması isteniyorsa rastgele madde etkilerinin varyansını kestirmek güç olacaktır. Verhagen vd. (2016) iki grubu karşılaştırarak ölçme değişmezliğini Bayes faktörü ile test edebilen bir model ortaya koymuşlardır. Modelde farklı gruplardaki gruba özgü madde parametreleri ayrı ayrı kestirilirler, birbirlerinden bağımsızlardır ve birbirileri hakkında bilgi sağlamazlar. Böyle bir durumda grup ortalaması için olası bir önsel dağılım geniş bir varyansa sahip normal önsel dağılımdır. Gruba özgü madde karakteristiklerine çok değişkenli normal model uygulanır ve farklı grupların madde parametreleri arasındaki korelasyona dayalı oluşturulan kovaryans matrisleri kullanılır. Modelde tanımlanan gruba özgü madde parametreleri denklem 12'de gösterilmiştir ( $\mu_j$ : j. grubun ortalama madde güçlüğü,  $e_{kj}$ : hata terimi).

$$\tilde{b}_{kj} = \mu_j + e_{kj} \quad (12)$$

Modelde  $\mu_j$  sifıra eşittir ve  $e_{kj}$  gruptaki ortalama madde güçlüğünden sapma miktarına eşittir. Bu sapmaların her bir gruba ait madde parametrelerinin varyansından oluşan madde güçlükleri kovaryansı ( $\Sigma_b$ ) ile çok değişkenli normal dağılım gösterdiği varsayılır. Madde parametrelerinin varyansı gruplara göre değişebilir, bu da bir grubun madde güçlük parametrelerindeki varyansın diğerine göre fazla olduğu anlamına gelir. Gruplardaki, gruba özgü madde güçlükleri bağımsız olarak kestirildikleri için ölçme değişmezliği madde güçlükleri arasındaki farka dayalı olarak direkt kestirilebilir. İki grup arasında güçlük parametreleri farkı eşitlik 13'teki gibi tanımlanır (k. madde, grup j ve j', j < j').

$$d_{kjj'} = b_{kj} - b_{kj'} \quad (13)$$

İki grup için ölçme değişmezliği testinde ise hipotezler madde güçlük parametreleri arasındaki farka ( $d_k$ ) dayalı olarak kurulur.

$$H_0 : d_k = 0$$

$$H_1 : d_k \neq 0$$

Bayes hipotez testini kullanan Bayes faktör ile ölçme değişmezliğinin değerlendirilmesi tüm ölçme aracındaki maddelerin değişmezliğine ilişkin doğrudan bilgi sunabilmesinin yanı sıra, bu testi yaparken daha önceden değişmezliği kanıtlanmış maddelere ihtiyaç duymaması bakımından çok avantajlıdır (Verhagen vd., 2016). Bununla birlikte, frekansçı istatistikler gibi yalnızca sıfır hipotezi için kanıt sunmak yerine hem  $H_0$  hem de  $H_1$  için kanıtlar üretir.  $H_1$  için sunulan kanıtlar kullanılmadan yalnızca  $H_0$  için kanıt sunulduğunda, özellikle düşük düzeyde güce sahip olan çalışmalarda sadece sıfır hipotezine sahip kanıtların ele alınarak değerlendirilmesi  $H_0$  hipotezine karşı abartılı sonuçlara yol açar (Rouder vd., 2009; Wagenmakers vd.,2017). Ayrıca her iki hipotez testi için de kanıt sunulması hangi hipotezin tercih edilmesi gerektiğine dair de bilgi vermesi bakımından avantajlıdır.

Bayes çerçevesinde değişmezlik hipotezleri, değişmezlik kısıtlamaları olan ve olmayan modellerin marjinal olasılıklarını karşılaştıran Bayes faktörler ile test edilir. Böylece en genel modeli tahmin etmeye yönelik MCMC algoritması, modelin çeşitli kısıtlı versiyonlarını değerlendirmek için kullanılabilir. Bu şekilde belirli varsayımlar önceden değişmezliği belirlenmiş maddelere bağlı olmadan ve aynı anda test edilebilir.

Ölçme değişmezliği testinde  $H_1$  hipotezinin marjinal olabilirliği, alternatif hipotezin tüm olası değerleri üzerindeki ortalama olabilirlik önsel olasılık ile ağırlıklandırılır. Söz konusu ortalama olabilirlik değeri hipotezde bulunan parametrelerin önsel yoğunluk fonksiyonu tarafından ağırlıklandırılan olabilirlik fonksiyonunun integraline eşittir. Bayes faktör hem sıfır hipotez hem de alternatif hipotez sonuçları için marjinal olabilirlik oranını içerir. Test edilen hipotezler için görel kanıt sunan Bayes faktör denklem 14'teki gibidir.

$$BF_{01} = \frac{P(Y|H_0)}{P(Y|H_1)} = \frac{P(Y|d_k = 0)}{\int P(Y|d_k)p_1(d_k)dd_k} \quad (14)$$

$p_1(d_k)$  :  $H_1$  altında parametreler arası farka ( $d_k$ ) ait önsel dağılım

Bayes faktör ölçme değişmezliği testlerinde model karmaşıklığını düzenler ve aşırı model uyumunun (overfitting) önüne geçer. Model karmaşıklığını önsel olarak analize dâhil ederek veriyi en iyi betimleyen modelin seçilmesini sağlar (Hojitnik, 2012). Ayrıca Bayes faktör tam ölçme değişmezliği ve kısmi ölçme değişmezliği modelleri gibi karmaşıklığı değişen modellerin hipotezlerini karşılaştırmak için de kullanılabilir (Fox vd., 2017).

### **İlgili Araştırmalar**

Verhagen ve Fox (2013) çalışmalarında bir evrenden rastgele örneklenen gruplar için çoklu ölçme değişmezliği hipotezlerini test edebilen farklı Bayes yaklaşımlarının kullanılabilceğini göstermişlerdir. Çalışmalarında ölçme değişmezliği hipotezlerini test etmek için önerilen Bayes faktörün, değişmezliği saptamak için farklı örneklem büyüklüklerinde iyi derecede güç ve düşük Tip 1 hata oranına sahip olduğunu raporlamışlardır. Bayes faktör testi madde güçlük ve madde ayırt edicilik parametreleri için hesaplanabilmektedir. Ancak değişmezlik için önerilen bu model çoklu grup karşılaştırmalarına uygunken az sayıda grup olduğunda kullanılmaya uygun değildir.

Verhagen ve diğerleri (2016) yalnızca ikili gruplarda ölçme değişmezliğinin testi için Bayes faktör önermişlerdir. Doğrudan madde güçlük parametreleri arasındaki farka dayalı olarak hesaplanan Bayes faktörün geçerliği yaygın olarak DMF analizlerinde kullanılan Wald test ile karşılaştırılarak değerlendirilmiştir. Farklı simülasyon koşullarında ve gerçek verilerde hesaplanan BF sonuçları Wald test ile tutarlı sonuçlar göstermiştir. Çalışmalarında ikili gruplar söz konusu olduğunda ölçme değişmezliği testi için Bayes faktörün geçerli bir yöntem olduğunu ortaya koymuşlardır.

Wang (2017) Bayesci YEM ve rastgele madde etkisi modelini (MLIRT), PISA 2012 matematik okuryazarlığı maddelerini kullanarak karşılaştırdığı çalışmasında, maddelerin güçlük ve ayırt edicilik parametrelerine ait BF sonuçlarının, ölçme değişmezliğinin sağlanmadığı ya da değişmezlik konusunda kararsız sonuçlar ürettiğini belirlemiştir.

Naumann ve arkadaşları (2017), DMF analizi için maddelerin mutlak ve görelî duyarlılığını Bayes faktör ile test etmişlerdir. Mutlak duyarlılık, madde duyarlılığının gruptan bağımsız; görelî duyarlılık ise madde parametrelerinin grubun yetenek düzeyine bağılı tahmin edilmesi olarak tanımlanmıştır. Çalışmalarında mutlak duyarlılık için 35 madden 27'sine ait BF sonuçları, görelî duyarlılık için ise 35 maddeden 31'i ölçme değişmezliği lehine sonuç üretmiştir.

Fox ve arkadaşları (2017), kesirli Bayes faktör kullanarak kovaryans yapılarına dayalı ölçme değişmezliğini ikili (1-0) puanlanan iki farklı gerçek veri seti üzerinde test ederek yeni bir model önerdikleri çalışmalarında kesirli Bayes faktör sonuçlarının iyi nitelikte olduğunu ortaya koymuşlardır.

Thompson (2018), ölçme değişmezliği için Bayes faktör testi yaptıkları simülasyon sonuçlarında, BF sonuçlarının değişmez olan ve olmayan maddeleri ayırmada iyi performans gösterdiğini raporlamışlardır. Ayrıca BF testi için farklı önsel dağılımları karşılaştırmışlar ve Bayes faktörün en iyi performans gösterdiği dağılımı Cauchy dağılımı olarak belirlemişlerdir.

Fay ve diğerleri (2018) yürüttükleri araştırmalarında maddelerin psikometrik olarak izomorf olup olmadıklarını Bayes faktör ile değerlendirmişlerdir. Eğer maddeler madde güçlüğü, madde ayırt ediciliği gibi psikometrik özellikler açısından farklılaşmıyorsa bu maddelere "psikometrik izomorf" tanımını yapmışlardır ve değişmezlik terimi yerine kullanılabilir olduğunu ifade etmişlerdir. Elde edilen sonuçlarda test edilen maddelere ait BF sonuçlarının,  $H_0$  ve  $H_1$  lehine kanıt üretmekle beraber kararsız kaldığı durumların da olduğunu ifade etmişlerdir. Yapılan 66 karşılaştırmadan 45'inin sonuçlarının izomorfluk için değişmezlik lehine kanıt ürettiğini raporlamışlardır.

Martin ve diğerleri (2019) rastgele etkiler modelinin daha genişletilmiş hali olan ve değişmezlik testini Bayes faktör ile gerçekleştiren Bayesci hiyerarşik içirme modelini önermişlerdir. Bu modelde, parametreler arası varyansın sıfır olduğu ya da olmadığı durumlar göz önünde bulundurularak ölçme değişmezliği testi yapılır. Gruba özgü parametre karakteristiklerinin hiyerarşik olarak modele eklenmesiyle değişmezlik bilgileri parametreler

arasında toplanır ve değişmezliğin sağlandığı ya da sağlanmadığı durumlar ayırt edilebilir. Yürüttükleri simülasyon çalışmasında Bayesci hiyerarşik içirme modelinin değişmezliğe ilişkin rastgele madde etkileri modeline göre daha detaylı bilgi verdiğini belirlemişlerdir.

Ayvallı ve Kelecioğlu (2021) Bayesci hiyerarşik içirme modelini kullanarak TIMSS 2015 uygulaması fen ve matematik alt testlerindeki ölçme değişmezliğini BF ile test etmişlerdir. Elde edilen sonuçlarda 8 çoktan seçmeli fen maddesinin 3 tanesinde tüm parametrelerde ölçme değişmezliği lehine, 8 çoktan seçmeli matematik maddesinin 4 tanesinde tüm parametrelerde ölçme değişmezliği lehine sonuç elde etmişlerdir.

Montenegro-Montenegro (2020), Bayesci YEM ile çoklu grup ölçme değişmezliği testinde farklı koşullarda model uyum indekslerinin performansını incelemiştir. Simülasyon çalışmasının sonucunda değişmezlik testi için BCFI ile, bir model uyum indeksi olmamasına rağmen Bayes faktör sonuçlarının değişmezliğin belirlenmesinde iyi performans gösterdiğini raporlamıştır.

Fox ve diğerleri (2020), farklı tip yanıtlar bulunan bir veride ölçme değişmezliğinin Bayesci kovaryans yapısı matrisini kullanarak belirlenebilmesine ilişkin bir model önermişler ve bu modelde değişmezlik testi için kesirli Bayes faktörü kullanmışlardır. Gruba özgü parametreler ya da ortak maddeye ihtiyaç duyulmayan yöntemde kümelenmiş madde yanıtları arasındaki korelasyona dayalı bir ölçme değişmezliği testi uygulanır. Ölçme değişmezliğinin ihlali her gruptaki artıkların arasındaki korelasyona bakılarak değerlendirilir. Ampirik verilerle yürütülen bu çalışmada kesirli Bayes faktör sonuçları yanıt kategorileri ve gruplar arasında eş zamanlı değerlendirme yapmaya olanak sağlar.

Joo ve arkadaşları (2022), GGUM (generalized graded unfolding model) ile DMF belirlemede Bayes faktör ve DIC kriterlerini farklı koşullar altında karşılaştırdıkları çalışmalarında BF ve DIC'nin her ikisinin de DMF'nin belirlenmesinde düşük Tip 1 hataya ve yüksek düzeyde güce sahip olduğunu belirlemişlerdir.



Yapılan alanyazın taraması sonucunda Bayesci yaklaşımlar kullanılarak ölçme değişmezliğinin incelenmesi ile ilgili çalışmaların tarihinin çok eski olmadığı görülmüştür. Bayesci yöntemlerle yapılan ölçme değişmezliği çalışmalarında genellikle çoklu grup karşılaştırmalarının yapıldığı belirlenmiştir. Bununla birlikte, ikili grup karşılaştırmasına dayalı çalışmaların nadir olduğu görülmüştür. Özellikle ölçme değişmezliği testlerine ve DMF belirlemeye yönelik Bayes faktör kullanılan çalışmaların simülasyon veriler üzerinden yürütüldüğü saptanmış, gerçek veriler üzerinden Bayes faktör ile ölçme değişmezliğinin belirlenmesini amaçlayan çalışmalara rastlanamamıştır. Bu araştırmada Bayes faktör ile ölçme değişmezliğinin değerlendirilebilmesi için hem simülasyon veri hem de gerçek veri uygulaması yapılmıştır.

## Bölüm 3

### Yöntem

Bu bölümde araştırmanın türü belirtilerek araştırma verilerinin elde edilmesi, çalışma grubu ve verilerin analizi sunulmuştur.

#### Araştırmanın Türü

Bu araştırmada Bayesci madde tepki kuramı modelleri kapsamında, madde güçlük parametreleri arasındaki farka dayalı olarak ölçme değişmezliği belirlenmiş ve geleneksel ölçme değişmezliği belirleme yöntemi olarak kabul edilen çoklu-grup doğrulayıcı faktör analizi yöntemi ile de yapının değişmezliği belirlenmiştir. Ayrıca çalışmada, Bayesci MTK modelleri kapsamında madde güçlük parametresi farkına dayalı olarak ölçme değişmezliğine dair bir kesme noktası belirlemeye çalışılmış ve bu kesme noktası ile birlikte ortak madde gerektiren çalışmalarda, ortak madde seçimine dair bilgi sunmayı hedeflemiştir. Araştırmanın ikinci bölümünde TIMSS 2019 uygulaması matematik testinde çoklu-grup doğrulayıcı faktör analizi yöntemiyle yapının değişmezliği test edilmiş, maddelerin ölçme değişmezliği ise Bayesci MTK modeli kullanılarak Bayes faktör incelenmiştir. Bu bağlamda araştırma bir simülasyon çalışması ve betimsel bir çalışmadır.

#### Çalışma Grubu

TIMSS 2019 uygulamasına 64 ülkeden toplamda 510.000 öğrenci katılmıştır. Bu ülkelerin yanı sıra uygulamaya 8 karşılaştırma kuruluşu (benchmarking entities) dâhil olmuştur. 4. Sınıf düzeyinde 58 ülke ve 6 karşılaştırma katılımcısı, 8. Sınıf düzeyinde ise 39 ülke ve 7 karşılaştırma katılımcısı yer almıştır. TIMSS 2019 ile ilk kez dijital değerlendirmeye geçiş yapılmıştır. Tüm ülkeler dijital değerlendirme yapmaya hazır olmadığı için katılımcı ülkelerin bir kısmında eTIMSS uygulaması gerçekleştirilirken diğer ülkeler kâğıt-kalem formatında TIMSS uygulamasını gerçekleştirmişlerdir.

Araştırmaya eTIMSS uygulamasına katılan ülkelerden Singapur, Türkiye ve Şili örneklemi seçilmiştir. Örneklem seçilirken başarı kıstas alınarak en başarılı ülke olarak Singapur, orta düzey başarılı olarak Türkiye ve en başarısız ülke olarak Şili seçilmiştir. Ülkelerin başarı ortalamaları ve araştırmada kullanılan 1 ve 14 no'lu kitapçıkları alan bireylerin yer aldığı grup büyüklükleri Tablo 1'de gösterilmiştir.

**Tablo 1**

*Çalışma grubunda yer alan katılımcı sayıları ve ortalama başarı puanları*

Ülke	n	Ortalama Başarı Puanı
Singapur	582	616*
Türkiye	684	496
Şili	585	441*

\*TIMSS uygulaması ortalama puanından anlamlı şekilde farklılaşmıştır (TIMSS scale centerpoint=500)

## **Araştırma Verileri**

### ***Simülasyon Koşulları ve Verilerin Üretilmesi***

Bayes faktör kullanılarak ölçme değişmezliği değerlendirilen veriler üretilirken R yazılımı (R Core Team, 2018) kullanılmıştır. Veri üretiminde her bir grup için, madde eşik parametrelerinin ( $b_{kj}$ ) toplamı ve referans gruba ait yetenek parametresinin ortalaması ( $\mu_{\theta j}$ ) 0 olarak kabul edilmiştir. Gruplardaki örneklem büyüklükleri eşit olarak 500, 1000, 1500 ve 2000 olacak şekilde belirlenmiştir. Önceki çalışmalar incelendiğinde yanlı olmayan parametre kestirimi için minimum örneklem büyüklüğünün 500 olması gerektiği önerilmiştir (Thompson, 2018, Asparouhov & Muthén, 2014; De Boeck, 2008; Stark vd., 2006). Finch (2016) grup büyüklüğü 500'den 1000'e çıkarıldığında Tip I hata oranının düştüğünü, ancak grup büyüklüğü 1000'den 2000'e çıkarıldığında Tip I hata oranında önemli bir fark olmadığını belirtmiştir. Ayrıca 500 ve üzeri örneklem büyüklüklerinde Bayes faktörün iyi performans gösterdiği belirlenmiştir (Verhagen vd., 2016). Thompson (2018), ölçme değişmezliği için Bayes faktörün daha büyük örneklem büyüklüklerinde daha yüksek güç oranına sahip olduğunu belirterek örneklem büyüklüğünün en az 500 olarak belirlenmesini önermiştir. Ek olarak, PISA, TIMSS ve PIRLS

gibi gerçek test uygulamaları göz önüne alındığında, gruplara ait örneklem büyüklüklerinin genellikle 500 ve üzerinde olduğu bilinmektedir. Literatürdeki bulgular ve gerçek test uygulamaları dikkate alınarak çalışmada örneklem büyüklükleri her bir grupta eşit olacak şekilde 500, 1000, 1500 ve 2000 olarak belirlenmiştir.

Her bir simülasyon koşulu için veriler ikili olarak (1-0) puanlanan 10 madde için 1 PLM altında üretilmiştir. Gruplara ait güçlük parametreleri arasındaki fark  $d_k=0.0$ ,  $d_k=0.1$ ,  $d_k=0.3$ ,  $d_k=0.5$  ve  $d_k=0.7$  olarak belirlenmiştir.  $d_k=0.0$  (madde güçlük parametreleri arasında farkın olmadığı durum) ölçme değişmezliğinin sağlandığı madde olarak kabul edilmiştir ve madde güçlük parametreleri arasındaki fark kademeli olarak artırılmıştır (Verhagen vd., 2016).

Harwell vd. (1996) madde tepki kuramına dayalı simülasyon çalışmalarında 100 veya daha az replikasyonun yeterli güce sahip olacağını belirtmişlerdir ve en az 25 replikasyon yapılmasını önermişlerdir. Çalışmada her bir koşul için 100 replikasyon uygulanmıştır. Analizler her bir veri seti için ayrı ayrı gerçekleştirilmiştir. Gruplar için her bir maddeye ait parametre değerleri Ek-A'da verilmiştir.

### ***Gerçek Verilerin Elde Edilmesi***

Araştırmada, "Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study-TIMSS) 2019" uygulama verileri kullanılmıştır. Araştırma verilerine TIMSS & PIRLS Uluslararası Eğitim Merkezi'nin resmî Web sayfası üzerinden ulaşılmıştır. Veriler TIMSS 2019 uygulamasına ait <https://timss2019.org/international-database/> adresinden indirilmiştir.

TIMSS, 4. Sınıf ve 8. Sınıf düzeyindeki öğrencilerin matematik ve fen başarılarını tüm dünya ülkelerinde değerlendiren uluslararası karşılaştırmalı bir çalışmadır. 1995 yılından bu yana 4 yıllık periyotlarla uygulanan TIMSS, matematik ve fen başarısının yanı sıra bu başarıyı etkileyen eğitim faktörleri hakkında da bilgiler toplayarak uluslararası karşılaştırmalı perspektifler sunar. TIMSS 2019 uygulaması diğer uygulamalardan farklı olarak ilk kez dijital formatta (eTIMSS) uygulamaya geçiş yapmıştır. Katılımcı ülkelerin yarısından fazlası eTIMSS

uygulamasına geçiş yaparken geri kalan ülkeler kâğıt-kalem tabanlı uygulama yapmışlardır. TIMSS 2019 uygulamasında hem kâğıt-kalem hem de bilgisayar tabanlı uygulamada toplamda 14 kitapçık bulunmaktadır. Her bir kitapçık madde bloklarından oluşur. TIMSS uygulamasında her bir düzey için (4. Sınıf/8. Sınıf) toplamda 28 (14 blok Matematik, 14 Blok Fen) madde bloğu yer alır. Her kitapçık 2 matematik 2 fen olacak şekilde 4 tane madde bloğu içerir. Bloklarda çoktan seçmeli ve yapılandırılmış yanıt olmak üzere iki madde formatı bulunmaktadır. Tüm bloklardaki maddelerin kapsam ve bilişsel alanlardaki dağılımı genel madde havuzundaki dağılımla eşleşir. Kâğıt-kalem uygulamalı TIMSS’de yer alan her bloğun eTIMSS uygulamasında bir karşılığı vardır. TIMSS uygulamasına katılan ülkelerde, her bir maddeyi yeterli sayıda öğrencinin cevaplaması için, ülkelere en az 4000 öğrenciden oluşan bir örneklem seçilir ve 14 kitapçığa eşit oranda öğrenci cevap verecek şekilde örnekleme dağılımı sağlanır (Mullis ve Martin, 2017).

Araştırma 8. Sınıf düzeyindeki öğrencilere uygulanan eTIMSS matematik başarı maddeleri ile sınırlandırılmıştır. TIMSS 2019 8. sınıf matematik içerik alanları (Mullis ve Martin, 2017) Tablo 2’de gösterilmiştir.

**Tablo 2**

*TIMSS 2019 8. Sınıf matematik çerçevesi*

Matematik İçerik Alanları	Oranlar
Sayılar (Number)	%30
Cebir (Algebra)	%30
Geometri (Geometry)	%20
Veri ve Olasılık (Data and Probability)	%20

Araştırmada analizler 8. Sınıf düzeyi 1 ve 14 no’lu kitapçıklarda yer alan “ME01” bloğundaki, TIMSS 2015 uygulamasında da yer alan ve uygulama formatına göre (dijital/kâğıt-kalem) ölçme değişmezliği kanıtlanmış 11 çoktan seçmeli matematik maddesi üzerinden gerçekleştirilmiştir.

TIMSS 2019 teknik raporunda kâğıt-kalem uygulamasından bilgisayar tabanlı uygulamaya geçiş sürecinde özellikle zamana karşı karşılaştırmalar yapılırken ölçme değişmezliğinin ihlalinin söz konusu olabileceği ve kâğıt-kalem ve bilgisayar tabanlı olarak uygulanan testlerde yer alan maddelerin her iki uygulamada da aynı şekilde çalışmayabileceği belirtilmiştir. Bu durum ölçme değişmezliğini etkileyebileceğinden bu duruma mod etkisi adı verilerek eTIMSS ve kâğıt-kalem TIMSS uygulamalarında yer alan maddelerde ölçme değişmezliği testi yapılmıştır. İki uygulama arasında ölçme değişmezliğinin belirlenebilmesi için MTK modelleri genişletilerek bu modellere “mod etkisi” (genel mod etkisi, ortalama mod etkisi ve rastgele mod etkisi) parametreleri eklenmiştir. TIMSS 2019 uygulamasında daha önceden TIMSS 2015 uygulamasında da yer alan trend maddeler üzerinden yapılan analizler sonucunda maddelerin %80’inden fazlasının değişmez olduğu belirlenmiştir (von Davier, 2020).

Bu araştırmada ülkeler arası ölçme değişmezliği incelenmesi hedeflendiğinden, mod etkisi incelenmiş ve uygulama formatına göre değişmezliği kanıtlanmış maddelerin analizlere dâhil edilmesine karar verilmiştir.

Analize dâhil edilen maddeler Tablo 3’te gösterilmiştir.

**Tablo 3***Analize dâhil edilen çoktan seçmeli matematik maddeleri*

Maddeler	İçerik Alanı	Konu Alanı	Bilişsel Alan
ME52024	Sayılar	Kesirler ve ondalık sayılar	Bilme
ME52125	Sayılar	Tam sayılar	Akıl Yürütme
ME52063	Cebir	İfadeler, işlemler ve denklemler	Uygulama
ME52072	Cebir	İfadeler, işlemler ve denklemler	Bilme
ME52092	Cebir	İlişkiler ve fonksiyonlar	Uygulama
ME52046	Geometri	Geometrik şekiller ve ölçümler	Akıl Yürütme
ME52083	Geometri	Geometrik şekiller ve ölçümler	Uygulama
ME52082	Geometri	Geometrik şekiller ve ölçümler	Uygulama
ME52161	Veri ve Olasılık	Olasılık	Uygulama
ME52418A	Veri ve Olasılık	Veri	Uygulama
ME52418B	Veri ve Olasılık	Veri	Uygulama

**Verilerin Analizi*****Çoklu-grup doğrulayıcı faktör analizi***

Analizlere başlamadan önce Singapur, Türkiye ve Şili örneklemelerine ait veri setleri analize uygun hale getirilmiştir. Öncelikle veri setlerindeki kayıp değerler incelenmiş, maddelerdeki kayıp veri oranının %1 ile %3.6 arasında değiştiği belirlenmiştir. Kayıp değer olarak yalnızca TIMSS 2019 Uluslararası Veritabanı Kullanım Klavuzu'nda (Fishbein vd., 2021) "atlanmış (omitted)" olarak belirtilen değerler tespit edilmiştir. Bu kayıp veriler öğrencilerin cevap vermesi gerektiği halde boş bırakılmış ya da birden fazla cevap seçeneği işaretlenmiş yanıtlardır. Klavuzda bu yanıtların genellikle yanlış olarak kabul edildiği belirtildiğinden (Fishbein vd., 2021) veri setindeki kayıp değerler "0" olarak yeniden kodlanmış ve yanlış yanıt olarak değerlendirilmiştir.

Veri analizinde uç değerler istatistiksel sonuçları etkileyerek hatalı sonuçlara yol açabileceğinden tek değişkenli ve çok değişkenli uç değerler incelenmiştir (Tabachnick ve Fidell, 2013). Tek değişkenli uç değer taramasında, puanlar standart Z puanlarına

dönüştürülerek  $\pm 3$  (Raykov ve Marcoulides, 2008) kesme puanı olarak belirlenmiştir. Bu aralık dışında kalan herhangi bir veri tespit edilememiştir. Çok değişkenli uç değerleri belirlemek için ise  $\chi^2$  dağılımını kullanan Mahalanobis uzaklığı incelenmiş,  $p < 0.001$  ölçüt olarak kabul edilmiştir (Tabachnick ve Fidell, 2013). Bu ölçüte göre veri setlerinde çok değişkenli uç değer tespit edilmemiştir.

Yapısal eşitlik modellemelerinde diğer bir önemli unsur örneklem büyüklükleridir. Korelasyon ya da kovaryans matrisleri kullanılarak kestirimler yapıldığından küçük örneklerle model kestirimlerinde daha az tutarlı sonuçlar üretilir (Tabachnick ve Fidell, 2013). Bu modellerde örneklem büyüklüğü arttıkça güç miktarı da artmaktadır (MacCallum vd., 1996). Meade ve Bauer (2007) grup büyüklükleri için kesin bir sayı önermese de grup büyüklüğü 100 olduğunda gücün düşük olduğunu, grup büyüklüğü 200 olduğunda sonuçların değişken olduğunu, grup büyüklüğü 400 olduğunda ise gücün yüksek olduğunu raporlamışlardır. Araştırmada kullanılan örneklem büyüklükleri 400'ün üzerinde olduğundan ölçme değişmezliği modellerini test etmek için yeterli sayıda olduğu kabul edilmiştir.

Araştırmada model kestirim yöntemlerinden ağırlıklı en küçük kareler ortalaması ve varyansı (WLSMV) yöntemi kullanılmıştır. Değişkenler kategorik olduğunda (ikili puanlanan, çoklu puanlanan ya da sıralı) doğrulayıcı faktör analizi modellerinde kestirim yapmak için sıklıkla tercih edilen en çok olabilirlik yöntemi (ML) kullanılması önerilmemektedir (Li, 2016). DFA'da kategorik değişkenleri sürekli değişken olarak kabul etmek, özellikle taban ya da tavan etkisi olan durumlarda göstergeler arasındaki korelasyonlara ilişkin zayıf kestirimler üretmesi, hatalı test istatistikleri ve standart hata üretmesi gibi sonuçlara yol açabilir. Bu nedenle modelde kategorik veriler olduğunda ve normal dağılım varsayımı karşılanmadığında ağırlıklı en küçük kareler (WLS) ya da ağırlıklı en küçük kareler ortalaması ve varyansı (WLSMV) yöntemleri tercih edilebilir (Brown, 2006; Finney ve DiStefano, 2006). Söz konusu dayanıklı kestirim yöntemleri farklı yazılım programlarına göre değişkenlik gösterse de (WLSMV- Mplus, DWLS-Lisrel) tüm teknikler WLS kestirim yönteminin formülü üzerine inşa edilmişlerdir. WLSMV ve DWLS yöntemlerinde WLS kestirim yönteminde olduğu gibi ağırlık matrisinin



tümünü tersine çevirmek yerine sadece ağırlık matrisinin köşegenleri tersine çevrilir (Muthen ve Kaplan, 1985; Bollen, 1989). Bu nedenle bu yöntemler kullanılan yazılım programından bağımsız olarak çapraz ağırlıklı en küçük kareler yöntemi (DWLS) olarak isimlendirilir (DiStefano ve Morgan, 2014). Bu yöntemde ikili puanlanan göstergeler için kovaryans matrisleri yerine tetrakorik korelasyon matrisi, 2'den fazla kategorisi olan ya da sıralı göstergeler için ise polikorik korelasyon matrisi kullanılır (Brown, 2006; Kline 2016).

Yapısal eşitlik modellerinde model uyumunun değerlendirilmesinde uyum indekslerinden faydalanılır. Uyum indeksleri mutlak uyum, model tutumluluğu ve artan uyum indeksleri olmak üzere üç kategoride incelenir (Brown, 2006). Kline (2016)  $\chi^2/_{sd}$ , RMSEA, CFI ve SRMR'nin de içerisinde bulunduğu minimum dört uyum indeksinin raporlanmasını önermiştir. Hu ve Bentler (1999) CFI, RMSEA, SRMR ve TLI indekslerinin model değerlendirilmeleri yapılırken iyi performans gösterdiğini belirtmişlerdir. Araştırmada  $\chi^2/_{sd}$ , RMSEA, CFI ve TLI indeksleri temel alınarak model uyumu değerlendirilmiş, SRMR ise kategorik değişkenlere dayalı DFA modellerinde iyi performans göstermediği için değerlendirmede kullanılmamıştır (Yu, 2002).

Ölçme değişmezliğinin belirlenebilmesi için model karşılaştırmaları yapılır. Değişmezlik en az kısıtlayıcı modelden en kısıtlayıcı modele doğru hiyerarşik (yuvalanmış) bir yapı oluşturur. Örneğin metrik değişmezliğin sağlanabilmesi için yapısal değişmezlik, skaler değişmezlik için ise metrik değişmezlik bir ön koşuldur (Meredith, 1993). Gruplar arası anlamlı ve doğru karşılaştırmaların yapılabilmesi için katı değişmezliğin sağlanması önerilirken (Meredith, 1993) metrik değişmezliğin ya da en iyi ihtimalle skaler değişmezliğin sağlanmasının ölçme değişmezliği için yeterli kanıt oluşturduğu düşünülmektedir (Horn ve McArdle, 1992; Little, 1997; Steenkamp ve Baumgartner, 1998). Çalışmada model karşılaştırmalarında, yuvalanmış yapısal eşitlik modellerini karşılaştırmak için kullanılan ve  $\chi^2$  farkının anlamlı olup olmadığını gösteren Mplus DIFFTEST (Asparouhov ve Muthén, 2006) prosedüründen faydalanılmıştır. Ayrıca yuvalanmış modellerin değişmez olarak

değerlendirilebilmesi için CFI, TLI ve RMSEA indekslerinden faydalanılır ve bu indekslerde mutlak farkın 0.01'den küçük olması ( $\Delta CFI \leq 0.010$ ,  $\Delta TLI \leq 0.010$ ,  $\Delta RMSEA \leq 0.010$ ) önerilir (Cheung ve Rensvold, 2002). Araştırmada model karşılaştırılmaları için bu kesme puanları temel alınmıştır.

Çoklu-grup doğrulayıcı faktör analizi kapsamında ölçme değişmezliği modelleri Mplus (Muthén ve Muthén, 1998-2017) programında teta parametreleri kullanılarak gerçekleştirilmiştir. Öncelikle her bir modele ilişkin doğrulayıcı faktör analizi yapılmıştır. Ölçme modellerine ilişkin DFA sonuçları bulgular bölümünde verilmiştir.

### **Bayesci Analizler**

Araştırmada her bir maddenin güçlük parametreleri arasındaki fark için Bayes faktör oluşturularak maddelerin ölçme değişmezliğine ilişkin kanıt sunulmaya çalışılmıştır.

Bayes kuralında " $H_0$ " test edilen hipotez, " $H_1$ " karşıt hipotez olduğu durumda araştırmacılar hipotezler için  $P(H_0)$  ve  $P(H_1)$  olarak tanımlanan önsel olasılıklara sahiptir. X bir olayın olasılığı olarak tanımlandığında  $P(H_1|X)$  ve  $P(H_0|X)$  olabilirlik fonksiyonları olarak tanımlanır. Her bir hipoteze ait Bayes kuralı sonsal olasılıklardır. Bu olasılıklara ait formüller Denklem 16 ( $P(X|H)$ ) ve Denklem 17'de ( $P(X|H_0)$ ) gösterilmiştir (Etz ve Vandekerckhove, 2018).

$$P(H_1|X) = \frac{P(H_1)P(X|H_1)}{P(X)} \quad (16)$$

$$P(H_0|X) = \frac{P(H_0)P(X|H_0)}{P(X)} \quad (17)$$

Bayes faktör ise birbirine karşıt bu hipotezlerin performansını belirler. Söz konusu karşıt hipotezlere dayalı Bayes faktör Denklem 18'de gösterilmiştir (van Doorn vd., 2021).

$$\frac{P(H_1)}{P(H_0)} \times \frac{P(X|H_1)}{P(X|H_0)} = \frac{P(H_1|X)}{P(H_0|X)} \quad (18)^*$$

\*Önsel odds x Bayes Faktör<sub>10</sub> = Sonsal odds

Yani Bayes faktör, “ $H_0$  ve  $H_1$ ’in sonsal olasılıkların oranıyla,  $H_0$  ve  $H_1$ ’in önsel olasılıklarının oranının birbirlerine oranına eşittir.” şeklinde tanımlanır (Kass ve Raftery, 1995).

Bayes faktör  $H_0$  ve  $H_1$  hipotezleri için göreceli bir tahmin performansı ortaya koyar. Bir başka ifadeyle hipotezin tahmin kalitesine ilişkin göreceli bir ölçü sunar. Örneğin  $BF_{01} = 5$  olduğu bir durumda,  $H_0$ ’in  $H_1$ ’den 5 kat daha olası olduğu anlamına gelir. Ancak Bayes faktör’ün  $H_0$  lehine olması  $H_0$ ’in verileri daha iyi tahmin ettiği anlamına gelmez (van Doorn vd., 2021). Bayes faktör, veriler tarafından sağlanan kanıtları özetler. Örneğin  $BF_{01} = 2$  değeri, sıfır hipotezinin veriler tarafından alternatif hipotezin iki katı kadar fazla desteklendiği anlamına gelir (Fox, 2010).

Ölçme değişmezliği testi için oluşturulan Bayes faktörde herhangi iki grup için madde güçlük parametreleri arasındaki fark,

$$d_k = b_{k1} - b_{k2} = 0 \quad (19)$$

olarak tanımlandığında  $H_0 = d_k = 0$  hipotezinin  $H_1 = d_k \neq 0$  hipotezine göre göreceli desteğini değerlendirirken, Bayes faktör  $H_0$  hipotezini  $H_1$  hipotezindeki parametreler arası farkın önsel ve sonsal dağılımlarına indirger. Daha basit bir anlatımla  $H_0 = 0$  olacak şekilde ayarlanarak alternatif hipotez üzerinden elde edilir.

$$BF_{01} = \frac{P(Y|H_0)}{P(Y|H_1)} = \frac{P(d_k = 0 | H_1, Y)}{P(d_k = 0 | H_1)} \quad (20)$$

Böylece Bayes faktör yokluk hipotezi için alternatif hipotezden daha fazla kanıt üretir (Verhagen vd., 2016).

Bayes faktör sonuçlarının değerlendirilebilmesi için kesme puanları sıfır hipotezinin reddedilmesi, kabul edilmesi ve belirsizlik olmak üzere üç farklı kategoriye ayrılabilir. Jeffreys (1961)’e göre 1 ile 3 arasındaki Bayes faktör yokluk hipotezi ve alternatif hipotez için eşit kanıt üretir. Bu değerler zayıf kanıt olarak kabul edilir ve hipoteze ilişkin belirsizlik olduğunu ortaya

koyar. 3 ile 10 arasındaki bir Bayes faktörü  $H_0$  hipotezi için yeterli derecede kanıt olarak kabul edilir. Eğer Bayes faktör 10'dan büyük bir değere sahipse  $H_0$  hipotezi için güçlü kanıt olarak kabul edilir. Bayes faktörün 0.33 ile 0.10 arasında değer aldığı durumlarda alternatif hipotez için yeterli derecede kanıt, 0.10'dan küçük olduğu durumlarda ise alternatif hipotez için güçlü düzeyde kanıt olarak kabul edilir (Jeffreys, 1961; Kass ve Raftery, 1995).

Çalışmada Bayes faktör için kesme puanı değişmezliğin sağlandığı durumlar için 3, sağlanmadığı durumlar için 0.33 olarak belirlenmiştir.

Hipotez testlerinde yokluk hipotezinin önsel ve sonsal dağılımlar altındaki yoğunluğunun oranı Bayes faktör test sonuçlarını etkiler. Bayes faktörlere ait test sonuçları değerlendirmeye alınacak parametreler için seçilen önsellere bağlıdır. Madde parametreleri farkı ( $d_k$ ) için önsel dağılımlar, her bir grubun madde parametrelerinin ayrı önsel dağılımlarından oluşur. Önseller, parametre değerleri için kabul edilen varsayımlara göre seçilebilirler. Gruba özgü madde parametreleri arasındaki fark çok değişkenli Cauchy dağılımda daha eşit dağılır. Bu nedenle Bayes faktörün ölçme değişmezliğini desteklemesi daha olası olacağından ve değişmezlik için daha tutarlı sonuçların elde edilebilmesi için analizler çok değişkenli Cauchy dağılımı kullanılarak gerçekleştirilmiştir. (Verhagen vd., 2016; Thompson, 2018). Bayes faktör hesaplamasında MCMC örnekleme yöntemlerinden biri olan Savage–Dickey yoğunluk oranı kullanılmıştır. Bu metot yuvalanmış modellerde uygulanır ve test edilen parametre için Bayes faktörün hesaplanması sonsallığın yüksekliğini ve önsel dağılımın yüksekliğini gerektirir ve özellikle yuvalanmış yapılar gibi karmaşık modellerde ölçme değişmezliği testi için kullanılabilir. En basit haliyle düşünüldüğünde yokluk hipotezi ilgilenilen parametre değerinin belirlenen bir değere sabitlendiği, alternatif hipotez de bu parametrenin serbest bırakıldığı hipotezdir. Dolayısıyla yokluk hipotezi alternatif hipotez altında yuvalanır. (Wagenmaker vd., 2010). Savage-Dickey yoğunluk oranı özel bir hesaplama (Wetzels vd., 2010). Madde parametrelerinin rastgele gruplar üzerindeki varyansının  $\sigma_{b_k}^2$  sifira eşit olduğu durumda  $H_0$  hipotezini test etmek için hesaplanan Bayes faktör denklem 21'deki gibidir.

$$BF_{01} = \frac{P(Y|H_0)}{P(Y|H_1)} = \frac{P(\sigma_{b_k}^2 < \delta | H_1, Y)}{P(\sigma_{b_k}^2 < \delta | H_1)} \quad (21)$$

Bayesci analizler R'da yazılan kodlar kullanılarak WINBUGS üzerinden gerçekleştirilmiştir. Analizler için "R2WinBUGS", "R2OpenBUGS" (Sturtz vd., 2005), "polspline" (Kooperberg, 2015), "MASS" (Venables ve Ripley, 2002) paketleri kullanılmıştır. MCMC süreçlerinin verimli bir biçimde tamamlanması için analizler ısınma periyodu (burn-in) 300 olacak şekilde 3000 iterasyonla gerçekleştirilmiştir. Analizlere ilişkin örnek yakınsama grafikleri Ek-B'de gösterilmiştir. Yakınsama grafikleri "coda" paketi ile elde edilmiştir (Plummer vd., 2006). Ülkelere ait madde güçlük parametreleri ise R yazılımında "ltn" paketi kullanılarak kestirilmiştir (Rizopoulos, 2006).

## Bölüm 4

### Bulgular, Yorumlar ve Tartışma

Araştırmanın bu bölümünde, araştırma problemleri sonucunda elde edilen bulgulara ve bu bulguları dayalı olarak yorumlara ve tartışmaya yer verilmiştir.

#### Araştırmanın Birinci Alt Problemine İlişkin Bulgular

Araştırmanın birinci alt probleminde “Simülasyon verilerde grup büyüklükleri ( $n = 500, n = 1000, n = 1500, n = 2000$ ) ve madde güçlük parametreleri arasındaki farkın ( $d_k = 0.0, d_k = 0.1, d_k = 0.3, d_k = 0.5, d_k = 0.7$ ) değişkenlik gösterdiği durumlarda Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?” sorusuna yanıt aranmıştır. Elde edilen bulgular madde güçlük parametreleri arasındaki farklılara göre her bir örneklem büyüklüğü için ayrı ayrı raporlanmıştır.

Madde güçlük parametreleri arasındaki farkın olmadığı durumda  $d_k = 0.0$  tüm grup büyüklüklerinde ölçme değişmezliğinin sağlanıp sağlanmadığına ilişkin Bayes faktör sonuçları Tablo 4’de verilmiştir.

**Tablo 4**

*Madde güçlük parametreleri arasında farkın olmadığı durumlarda ( $d_k = 0.0$ ) ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları*

	$BF_{01}$			
	$N=1000$	$N=2000$	$N=3000$	$N=4000$
	(Her grup için $n=500$ )	(Her grup için $n=1000$ )	(Her grup için $n=1500$ )	(Her grup için $n=2000$ )
Madde 1	4.59773	6.30162	10.37240	18.34086
Madde 2	10.85618	10.0564	7.82427	16.83027
Madde 3	10.67626	11.04705	7.63375	11.0673
Madde 4	11.57481	9.56177	16.39697	15.62178
Madde 5	6.24642	11.08648	21.14449	14.84418
Madde 6	5.50773	6.36254	10.24370	19.08634
Madde 7	10.97618	11.0004	8.84272	17.80372
Madde 8	12.67626	10.99905	7.75363	12.00662
Madde 9	15.57481	9.59548	17.12697	16.16278
Madde 10	6.43642	13.08868	22.00443	15.73325

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Elde edilen sonuçlar incelendiğinde güçlük parametreleri arasında fark olmadığı, yani  $d_k = 0.0$  olduğu durumlarda tüm örneklem büyüklüklerinde madde parametrelerine ait  $BF_{01}$  değerlerinin ölçme değişmezliği lehine yeterli düzeyde kanıt ürettiği ( $BF_{01} > 3$ ) görülmektedir. Gruplardaki kişi sayıları 500 olduğu durumda Bayes faktör sonuçlarının dört maddenin ölçme değişmezliğinin lehine yeterli düzeyde ( $3 < BF_{01} < 10$ ), altı maddenin ise ölçme değişmezliği lehine ilişkin güçlü düzeyde kanıt ürettiği ( $BF_{01} > 10$ ) belirlenmiştir. Grup büyüklükleri  $n=1000$  ve  $n=1500$  olduğu durumlarda Bayes faktör sonuçlarının yine dört maddenin ölçme değişmezliğinin sağlandığına ilişkin yeterli düzeyde ( $3 < BF_{01} < 10$ ), altı maddenin ise ölçme değişmezliği lehine güçlü düzeyde kanıt ürettiği ( $BF_{01} > 10$ ) belirlenmiştir. Gruplardaki birey

sayısı 2000 olduğunda ise tüm maddelere ait  $BF_{01}$  değerlerinin 10'dan büyük olduğu ve ölçme değişmezliği lehine güçlü düzeyde kanıt oluşturduğu görülmektedir.

Madde güçlük parametreleri arasındaki farkın  $d_k = 0.1$  olduğu durumlarda ölçme değişmezliğinin sağlanıp sağlanmadığına ilişkin Bayes faktör sonuçları Tablo 5'te verilmiştir.

**Tablo 5**

*Madde güçlük parametreleri arasındaki farkın  $d_k = 0.1$  olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları*

	$BF_{01}$			
	<i>N=1000 (Her grup için n=500)</i>	<i>N=2000 (Her grup için n=1000)</i>	<i>N=3000 (Her grup için n=1500)</i>	<i>N=4000 (Her grup için n=2000)</i>
Madde 1	9.53619	13.53914	3.94931	9.53863
Madde 2	3.32114	10.49564	10.93055	4.85995
Madde 3	8.24058	8.81567	20.05434	6.65487
Madde 4	11.84632	9.00784	12.1836	16.94075
Madde 5	7.99764	12.38298	9.74235	6.89683
Madde 6	9.78869	15.91453	3.00491	9.63375
Madde 7	5.32114	10.56449	11.04056	4.99502
Madde 8	8.42058	7.99815	21.45434	7.48765
Madde 9	13.87632	9.78400	13.18360	17.93081
Madde 10	8.19765	12.29809	9.35945	7.96828

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Tablo 5'te yer alan sonuçlar değerlendirildiğinde  $d_k = 0.1$  olduğu durumlarda da tüm örneklem büyüklüklerinde  $BF_{01}$  değerlerinin 3'ten büyük olduğu ve ölçme değişmezliği lehine kanıt ürettiği görülmektedir.  $n=500$  olduğu durumda sekiz madde için Bayes faktörün değişmezlik lehine yeterli düzeyde kanıt ürettiği ( $3 < BF_{01} < 10$ ), iki maddenin ise ölçme değişmezliği lehine güçlü ( $BF_{01} > 10$ ) kanıt ürettiği belirlenmiştir.  $n=1000$  ve  $n=1500$  olduğu durumlarda Bayes faktör sonuçlarının yine dört maddenin ölçme değişmezliğinin sağlandığı



yönünde yeterli düzeyde ( $3 < BF_{01} < 10$ ), altı maddenin ise değişmezliğin sağlandığı yönünde güçlü düzeyde kanıt ürettiği ( $BF_{01} > 10$ ) belirlenmiştir. Grup büyüklükleri 2000 olduğunda ise 8 madde için Bayes faktörün  $H_0$  lehine yeterli düzeyde kanıt ürettiği ( $3 < BF_{01} < 10$ ), iki maddenin ise ölçme değişmezliği lehine yani  $H_0$  lehine güçlü düzeyde ( $BF_{01} > 10$ ) kanıt ürettiği belirlenmiştir.

Madde güçlük parametreleri arasındaki farkın  $d_k = 0.3$  olduğu durumlarda ölçme değişmezliğinin sağlanıp sağlanamadığına ilişkin Bayes faktör sonuçları her bir grup büyüklüğüne göre Tablo 6'da verilmiştir.

**Tablo 6**

*Madde güçlük parametreleri arasındaki farkın  $d_k = 0.3$  olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları*

	$BF_{01}$			
	$N=1000$	$N=2000$	$N=3000$	$N=4000$
	(Her grup için $n=500$ )	(Her grup için $n=1000$ )	(Her grup için $n=1500$ )	(Her grup için $n=2000$ )
Madde 1	7.19235	10.3738	2.61409	0.02789
Madde 2	12.04128	1.12155	0.29342	0.05877
Madde 3	8.74905	0.20997	0.32324	0.09858
Madde 4	1.92266	4.18721	0.01816	0.27239
Madde 5	1.32949	0.0207	0.01307	0.13994
Madde 6	8.19235	13.3738	2.96104	0.02789
Madde 7	11.94128	1.15585	0.19388	0.05877
Madde 8	7.14901	0.19997	0.23564	0.09858
Madde 9	0.99266	4.72118	0.02817	0.27239
Madde 10	2.31742	0.20700	0.01509	0.13994

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Madde güçlük parametreleri arası fark  $d_k = 0.3$  olduğunda ise, grup büyüklüğü 500 iken 4 maddenin ölçme değişmezliğinin sağlandığı yönünde yeterli kanıt ( $3 < BF_{01} < 10$ ), iki

maddenin deęişmezlięine iliřkin g¼çlü kanıt ( $BF_{01} > 10$ ) sunduęu belirlenirken dięer 4 maddenin deęişmezlięine iliřkin Bayes fakt¼r¼n kararsız kaldıęı ( $1 < BF_{01} < 3$ ) belirlenmiřtir. Her bir grup b¼y¼kl¼ę¼ 1000 olduęunda ise yalnızca iki maddenin ölçme deęişmezlięinin saęlandıęı yön¼nde yeterli düzeyde kanıt ( $3 < BF_{01} < 10$ ) sunduęu, iki maddenin deęişmezlięine iliřkin g¼çlü düzeyde ( $BF_{01} > 10$ ) kanıt sunduęu belirlenmiřtir. 2 maddenin deęişmezlięine iliřkin Bayes fakt¼r¼n kararsız kaldıęı ( $1 < BF_{01} < 3$ ) ve 4 maddenin ise ölçme deęişmezlięinin saęlanmadıęına ( $BF_{01} < 0.33$ ) iliřkin kanıt ürettięi belirlenmiřtir. Grup b¼y¼kl¼klerinin 1500 olduęu durum incelendięinde ise iki madde (madde 6 ve madde 1) için hesaplanan Bayes fakt¼r¼n her iki hipotez için de eřit düzeyde kanıt sunduęu ve maddenin deęişmezlięi hakkında kararsız kaldıęı ( $1 < BF_{01} < 3$ ) belirlenmiřtir. Üç madde için hesaplanan Bayes fakt¼r¼n ölçme deęişmezlięinin saęlanmadıęına iliřkin yeterli düzeyde ( $BF_{01} < 0.33$ ), geriye kalan dört madde için hesaplanan Bayes fakt¼r¼n ise deęişmezlięin saęlanmadıęına iliřkin g¼çlü düzeyde ( $BF_{01} < 0.10$ ) kanıt ürettięi belirlenmiřtir. Grup b¼y¼kl¼kleri 2000 olduęunda yalnızca üç maddenin Bayes fakt¼r¼ sonuçlarının deęişmezlięin saęlanmadıęına iliřkin yeterli düzeyde kanıt ürettięi ( $BF_{01} < 0.33$ ), kalan yedi maddenin Bayes fakt¼r¼ sonuçlarının ise maddelerin ölçme deęişmezlięinin saęlanmadıęına iliřkin g¼çlü düzeyde ( $BF_{01} < 0.10$ ) kanıt ürettięi belirlenmiřtir.

Madde g¼çlükl¼k parametreleri arasındaki farkın  $d_k = 0.5$  olduęu durumlarda ölçme deęişmezlięi testi için elde edilen Bayes fakt¼r¼ sonuçları Tablo 7'de gör¼lmektedir.

**Tablo 7**

*Madde güçlük parametreleri arasındaki farkın  $d_k = 0.5$  olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları*

	$BF_{01}$			
	$N=1000$ (Her grup için $n=500$ )	$N=2000$ (Her grup için $n=1000$ )	$N=3000$ (Her grup için $n=1500$ )	$N=4000$ (Her grup için $n=2000$ )
Madde 1	0.15524	0.06085	0.00168	0.00016
Madde 2	0.18284	0.00891	0.01364	0.03911
Madde 3	4.09778	0.00085	0.00116	0.00415
Madde 4	0.35232	0.02586	0.06630	0.00147
Madde 5	4.31605	0.03253	0.00393	0.00135
Madde 6	0.23245	0.06857	0.01368	0.00012
Madde 7	0.17294	0.00981	0.02264	0.01368
Madde 8	4.00038	0.00508	0.01015	0.00307
Madde 9	0.29852	0.01258	0.06730	0.00112
Madde 10	3.93167	0.03534	0.00298	0.00129

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Madde güçlük parametreleri arasındaki fark  $d_k = 0.5$  olduğunda yalnızca  $n=500$  olduğu durumlarda ölçme değişmezliği lehine Bayes faktör sonuçları elde edilmiş, diğer grup büyüklüklerinde  $H_0$  lehine Bayes faktör sonucu tespit edilmemiştir. Gruplardaki birey sayısı  $n=500$  iken dört maddenin ölçme değişmezliğinin sağlandığına ilişkin  $H_0$  lehine sonuç ürettiği, bir maddenin ölçme değişmezliği konusunda Bayes faktörün kararsız kaldığı, kalan beş maddenin ise ölçme değişmezliğinin sağlanmadığına ilişkin  $H_1$  lehine yeterli düzeyde ( $BF_{01} < 0.33$ ) kanıt ürettiği belirlenmiştir.

Grup büyüklüğü  $n=1000$ ,  $n=1500$  ve  $n=2000$  olduğu durumlarda tüm maddelere ilişkin elde edilen  $BF_{01}$  değerlerinin tümünün  $0.10$ 'dan küçük olduğu ve  $H_1$  hipotezi lehine ölçme değişmezliğinin sağlanmadığına dair güçlü kanıt oluşturduğu belirlenmiştir.

Madde güçlük parametreleri arasındaki farkın  $d_k=0.7$  olduğu durumlarda ölçme değişmezliği testi için elde edilen Bayes faktör sonuçları Tablo 8’de görülmektedir.

**Tablo 8**

*Madde güçlük parametreleri arasındaki farkın  $d_k = 0.7$  olduğu durumlarda ölçme değişmezliğine ilişkin elde edilen Bayes faktör sonuçları*

	$BF_{01}$			
	$N=1000$ (Her grup için $n=500$ )	$N=2000$ (Her grup için $n=1000$ )	$N=3000$ (Her grup için $n=1500$ )	$N=4000$ (Her grup için $n=2000$ )
Madde 1	0.32990	0.01297	0.00030	0.00003
Madde 2	0.00047	0.00172	0.00156	0.00020
Madde 3	0.00328	0.00136	0.00077	0.00113
Madde 4	0.00232	0.02296	0.00175	0.00008
Madde 5	0.06634	0.00223	0.00026	0.00026
Madde 6	0.26890	0.02129	0.00140	0.00002
Madde 7	0.00007	0.00147	0.00185	0.00009
Madde 8	0.00285	0.00163	0.00113	0.00126
Madde 9	0.00292	0.03295	0.00156	0.00017
Madde 10	0.07654	0.02019	0.00102	0.00032

\* $BF_{01}>3$  :  $H_0$  lehine,  $1<BF_{01} <3$ : Kararsız,  $BF_{01}<0.33$  :  $H_1$  lehine

Tablo 8’deki sonuçlar incelendiğinde tüm grup büyüklüklerinde  $BF_{01}$  değerleri  $H_1$  lehine kanıt üretmiştir. Grup büyüklüğü  $n=500$  olduğu durumda yalnızca 2 maddeye ilişkin  $BF_{01}$  değeri 0.33’ten küçük olduğu için değişmezliğin sağlanmadığına dair yeterli düzeyde kanıt üretirken diğer tüm durumlardaki Bayes faktör sonuçlarının ölçme değişmezliğin sağlanmadığı yönünde  $H_1$  lehine güçlü düzeyde kanıt ( $BF_{01} < 0.10$ ) ürettiği görülmektedir.

## Araştırmanın İkinci Alt Problemine İlişkin Bulgular

Ölçme değişmezliği analizleri yapılmadan önce, ölçme modelinin ülke gruplarında doğrulanıp doğrulanmadığını belirlemek için Şili, Türkiye ve Singapur örneklemelerinde DFA ile modeller test edilmiş, model uyumuna ilişkin elde edilen sonuçlar Tablo 9'da gösterilmiştir.

**Tablo 9**

*Ülkelere ait doğrulayıcı faktör analizi sonuçları*

	$\chi^2$	sd	$\chi^2/sd$	CFI	TLI	RMSEA
Şili	41.868	44	0.95	1.00	1.00	0.027
Türkiye	110.443	44	2.51	0.96	0.947	0.051
Singapur	85.837	44	1.95	0.983	0.979	0.037

Sonuçlar incelendiğinde tüm ülkelere ait ölçme modellerinde  $\chi^2/sd$  değerlerinin 2'nin altında olduğu ve modellerin mükemmel uyum gösterdiği belirlenmiştir (Tabachnick ve Fidell, 2013). CFI ve TLI değerlerinin 0.95'in üzerinde olduğu ve modelin mükemmel uyum gösterdiği belirlenmiştir (Hu ve Bentler, 1999). RMSEA ve SRMR değerleri de 0.05 kesme noktasına göre mükemmel uyum göstermiştir (Brown, 2006). Genel olarak değerlendirildiğinde, ölçme modellerinin doğrulandığı görülmüştür.

### 2.1. Şili-Türkiye örneklemelerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?

Şili ve Türkiye örneklemeleri için ölçme değişmezliğinin sağlanıp sağlanmadığını test etmek amacıyla yapılan ÇGDFA sonuçları Tablo 10'da gösterilmiştir.

**Tablo 10***Şili ve Türkiye örneklemi için çoklu grup doğrulayıcı faktör analizi sonuçları*

	$\chi^2$	sd	$\chi^2/sd$	DIFFTTEST*	CFI	$\Delta CFI$	TLI	$\Delta TLI$	RMSEA	$\Delta RMSEA$
Yapısal Değişmezlik	147.200	88	1.67	-	0.971	-	0.964	-	0.034	-
Metrik Değişmezlik	171.273	98	1.75	0.009	0.964	-0.007	0.960	-0.004	0.036	0.002
Skaler Değişmezlik	252.220	108	2.34	0.000	0.930	-0.034	0.929	-0.031	0.048	0.012
Katı Değişmezlik	316.118	107	2.95	0.000	0.899	-0.031	0.896	-0.033	0.058	0.010

\*Ki-kare fark anlamlılık testi sonucu

Sonuçlar değerlendirildiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA değerlerinin tümünün mükemmel uyum gösterdiği ve yapısal değişmezlik modelinin doğrulandığı görülmektedir.

Metrik değişmezlik modeli incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA indekslerinin yine mükemmel uyum gösterdiği belirlenmiştir. Metrik değişmezlik ve yapısal değişmezlik modeli karşılaştırıldığında  $\chi^2$  karşılaştırma testinin sonucunun anlamlı çıktığı görülmüş ( $p < 0.05$ ) ve daha basit olan modelin kabul edilmesi yönünde sonuç üretmiştir. Diğer uyum indekslerindeki değişimler incelendiğinde ise her üç indekste de mutlak farkın kabul edilen kesme noktasından küçük olduğu görülmektedir ( $\Delta CFI < 0.01$ ,  $\Delta TLI < 0.01$ ,  $\Delta RMSEA < 0.01$ ).  $\chi^2$  testinin sonucu yapısal değişmezlik modelinin kabul edilmesi gerektiği anlamına gelse de  $\chi^2$  örneklem büyüklüğüne duyarlı bir istatistik olduğundan (Hirschfeld ve von Brachel, 2014), diğer üç uyum indeksinin sonuçlarına dayalı olarak metrik değişmezliğin sağlandığına karar verilmiştir.

Skaler değişmezlik modeli incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA indekslerinin iyi uyum gösterse de modelleri karşılaştıran  $\chi^2$  testinin sonucu anlamlı çıkmış, CFI ( $\Delta CFI > 0.01$ ), TLI ( $\Delta TLI > 0.01$ ) ve RMSEA ( $\Delta RMSEA > 0.01$ ) değerlerindeki değişimlerin de belirlenen kesme noktasından büyük olduğu belirlenmiştir. Dolayısıyla skaler değişmezliğin sağlanmadığı kanısına varılmıştır. Yuvalanmış modellerde bir önceki model kendisinden sonraki modele ön koşul oluşturduğundan katı değişmezlik modeli değerlendirilmemiştir.

Sonuç olarak Şili-Türkiye örneklerinde ölçme değişmezliği metrik değişmezlik düzeyinde sağlanmıştır.

## 2.2. Şili-Singapur örneklerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?

Şili ve Singapur örnekleri için ölçme değişmezliğinin sağlanıp sağlanmadığını test etmek amacıyla yapılan ÇGDFA sonuçları Tablo 11'de gösterilmiştir.

**Tablo 11**

*Şili ve Singapur örnekleri için çoklu grup doğrulayıcı faktör analizi sonuçları*

	$\chi^2$	sd	$\chi^2/sd$	DIFFTEST*	CFI	$\Delta CFI$	TLI	$\Delta TLI$	RMSEA	$\Delta RMSEA$
Yapısal	124.753	88	1.41	-	0.988	-	0.985	-	0.026	-
Değişmezlik										
Metrik	170.211	98	1.73	0.0004	0.976	-0.012	0.973	-0.012	0.034	0.008
Değişmezlik										
Skaler	279.086	108	2.58	0.0000	0.943	-0.033	0.942	-0.031	0.050	0.016
Değişmezlik										
Katı	336.155	107	3.14	0.0000	0.923	-0.020	0.921	-0.021	0.058	0.008
Değişmezlik										

\*Ki-kare fark anlamlılık testi sonucu

Tablo 15'teki sonuçlar değerlendirildiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA değerlerinin tümünün mükemmel uyum gösterdiği ve yapısal değişmezlik modelinin doğrulandığı belirlenmiştir.

Metrik değişmezlik modeli incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA indekslerinin yine mükemmel uyum gösterdiği belirlenmiştir. Metrik değişmezlik ve yapısal değişmezlik modeli karşılaştırıldığında  $\chi^2$  testinin sonucunun anlamlı çıktığı görülmüştür ( $p < 0.05$ ). Diğer uyum indekslerindeki değişimler incelendiğinde ise  $\Delta CFI > 0.01$ ,  $\Delta TLI > 0.01$  ve  $\Delta RMSEA > 0.01$  olarak belirlenmiştir. Yalnızca  $\Delta RMSEA$  belirlenen kesme noktasından küçük bir değer alırken,  $\chi^2$  testinin sonucu, CFI ve TLI'daki değişimler göz önünde bulundurulduğunda metrik

değişmezliğin sağlanamadığına karar verilmiştir. Dolayısıyla Şili ve Singapur için ölçme değişmezliği yapısal değişmezlik düzeyinde sağlanmıştır.

2.3. Türkiye-Singapur örneklerinde ÇGDFA ile ölçme değişmezliği sağlanmakta mıdır?

Türkiye ve Singapur örnekleri için ölçme değişmezliğinin sağlanıp sağlanmadığını test etmek amacıyla yapılan ÇGDFA sonuçları Tablo 12'de gösterilmiştir.

**Tablo 12**

Türkiye ve Singapur örnekleri için çoklu grup doğrulayıcı faktör analizi sonuçları

	$\chi^2$	sd	$\chi^2/sd$	DIFFTEST*	CFI	$\Delta CFI$	TLI	$\Delta TLI$	RMSEA	$\Delta RMSEA$
Yapısal	196.425	88	2.23	-	0.973	-	0.966	-	0.044	-
Değişmezlik										
Metrik	241.649	98	2.46	0.000	0.964	-0.009	0.960	-0.006	0.048	0.004
Değişmezlik										
Skaler	415.739	108	3.84	0.000	0.923	-0.041	0.922	-0.038	0.067	0.019
Değişmezlik										
Katı	608.814	107	5.68	0.0000	0.875	-0.048	0.872	-0.050	0.086	0.019
Değişmezlik										

Analiz sonuçları incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA değerlerinin tümünün mükemmel uyum gösterdiği ve yapısal değişmezlik modelinin doğrulandığı belirlenmiştir.

Metrik değişmezlik incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA indekslerinin yine mükemmel uyum gösterdiği görülmüştür. Metrik değişmezlik ve yapısal değişmezlik modeli karşılaştırılırken incelenen uyum indekslerindeki değişimler  $\Delta CFI < 0.01$ ,  $\Delta TLI < 0.01$  ve  $\Delta RMSEA < 0.01$  olarak belirlenmiştir. Model karşılaştırmalarına ilişkin  $\chi^2$  testinin sonucunun ise anlamlı çıktığı görülmüş ( $p < 0.05$ ) ancak diğer uyum indekslerinin sonuçlarına dayalı olarak metrik değişmezliğin sağlandığına karar verilmiştir.

Skaler değişmezlik modeli incelendiğinde  $\chi^2/sd$ , CFI, TLI ve RMSEA indekslerine göre modelin iyi uyum gösterdiği belirlenmiştir. Ancak skaler değişmezlik ve metrik değişmezlik



modelleri arasındaki farka ilişkin  $\chi^2$  testinin sonucunun anlamlı çıktığı belirlenmiştir. CFI ( $\Delta CFI > 0.01$ ), TLI ( $\Delta TLI > 0.01$ ) ve RMSEA ( $\Delta RMSEA > 0.01$ ) değerlerindeki değişimlerin de belirlenen kesme noktasından büyük olduğu sonucuna ulaşılmıştır. Tüm sonuçlar genel olarak değerlendirildiğinde skaler değişmezliğin sağlanmadığı belirlenmiştir ve skaler değişmezliğin ön koşul oluşturduğu diğer modeller değerlendirilmemiştir.

Sonuç olarak Türkiye-Singapur örneklemelerinde ölçme değişmezliği metrik değişmezlik düzeyinde sağlanmıştır.

### **Araştırmanın Üçüncü Alt Problemine İlişkin Bulgular**

Araştırmanın üçüncü alt probleminde “TIMSS 2019 uygulaması matematik alt testinde ikili puanlanan maddelerde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?” sorusuna yanıt aranmıştır. Araştırmada madde güçlük parametreleri arasındaki farka ( $d_k$ ) dayalı olarak kurulan hipotezlerin testi için Bayes faktör hesaplanarak ölçme değişmezliği belirlenmeye çalışılmıştır. Bayes faktör hesaplamaları yapılmadan önce seçilen örneklemelerde ayrı ayrı madde güçlük parametreleri kestirilmiş ve sonuçları Tablo 13’te verilmiştir. Daha sonra ikili grup kombinasyonları için ölçme değişmezliği testi Bayes faktör sonuçları ile değerlendirilmiş, grupların madde güçlük parametreleri arasındaki fark ve Bayes faktör sonuçlarının hangi hipotez lehine kanıt ürettiği beraber ele alınarak yorumlanmıştır. Madde parametreleri arasındaki fark değerlendirilirken gerçek verilere ait örneklem büyüklükleri dikkate alınmıştır. Ülkelerin örneklem büyüklükleri simülasyon verilerde belirlenen 500 örneklem büyüklüğüne yakın olduğu için, gerçek veriden elde edilen sonuçlar simülasyon veride 500 örneklem büyüklüğü olduğu durumda elde edilen sonuçlarla karşılaştırılmıştır. Madde parametreleri arasındaki fark için simülasyon veri sonuçlarında ölçme değişmezliğinin sağlanmadığı yönünde tutarlı kanıt üreten  $d_k=0.7$  kesme puanı olarak belirlenerek değerlendirmeler yapılmıştır.

**Tablo 13***Ülkelere ait madde güçlük parametre kestirimleri*

	Maddeler	Şili	Türkiye	Singapur
	Madde 1	1.025	0.112	-0.948
	Madde 2	0.819	-0.139	-0.887
	Madde 3	1.572	-0.443	-1.608
	Madde 4	0.178	-0.068	-1.328
	Madde 5	1.827	0.939	0.848
<b><math>b_i</math></b>	Madde 6	2.110	0.948	0.632
	Madde 7	1.717	1.186	-0.854
	Madde 8	0.652	-0.312	-1.448
	Madde 9	-0.187	-0.810	-1.494
	Madde 10	1.690	0.487	-0.900
	Madde 11	0.756	0.271	-1.024

3.1. “Şili-Türkiye örneklemelerinde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?”

Bayesci MTK kapsamında ölçme değişmezliği testi için hesaplanan Şili-Türkiye gruplarına ilişkin Bayes faktör sonuçları ( $BF_{01}$ ) ve grupların madde güçlük parametreleri arasındaki fark ( $d_S - d_T$ ) Tablo 14’te gösterilmiştir.

**Tablo 14**

*Şili-Türkiye Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark*

Maddeler	$d_S - d_T$	$BF_{01}$	Karar*
Madde 1	0.913	<b>8.967</b>	$H_0$ lehine
Madde 2	0.958	0.043	$H_1$ lehine
Madde 3	2.015	<b>3.951</b>	$H_0$ lehine
Madde 4	0.246	0.576	Kararsız
Madde 5	0.888	<b>6.667</b>	$H_0$ lehine
Madde 6	1.162	<b>6.694</b>	$H_0$ lehine
Madde 7	0.531	2.212	Kararsız
Madde 8	0.964	0.266	$H_1$ lehine
Madde 9	0.623	<b>3.728</b>	$H_0$ lehine
Madde 10	1.203	1.493	Kararsız
Madde 11	0.485	<b>5.196</b>	$H_0$ lehine

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Bayes Faktör ( $BF_{01}$ ) sonuçları incelendiğinde, madde 1, madde 3, madde 5, madde 6, madde 9 ve madde 11'in ölçme değişmezliğinin sağlandığı hipotezine ilişkin yeterli derecede kanıt ürettiği ( $BF_{01} > 3$ ) ancak ölçme değişmezliğine ilişkin güçlü kanıt üretmediği belirlenmiştir. Madde güçlük parametreleri arası fark ve Bayes faktör sonuçları birlikte değerlendirildiğinde, parametreler arası farkın 0.7'nin üzerinde olduğu durumlarda bile Bayes faktör sonuçları ölçme değişmezliğine yeterli düzeyde kanıt oluşturmuştur.

Madde 2 ve madde 8'in her ikisi için de hesaplanan güçlük parametreleri arasındaki fark 0.7'den büyük olarak belirlenmiş, madde 2'ye ait Bayes faktör sonuçları ölçme değişmezliğinin sağlanmadığına ilişkin güçlü düzeyde kanıt üretirken, madde 8'e ait Bayes faktör sonuçlarının  $H_1$  lehine yeterli düzeyde kanıt ürettiği belirlenmiştir.

Madde 4, madde 7 ve madde 10 için Bayes faktör  $H_0$  hipotezine ilişkin yeterli düzeyde kanıt üretmediğinden ( $1 < BF_{01} < 3$ ) maddelerin ölçme değişmezliğine ilişkin yorum yapılamamaktadır. Bu maddelerin güçlük parametreleri arasındaki fark değerlendirildiğinde, yalnızca madde 10'a ait farkın 0.7'den büyük olduğu diğer maddelerin ise bu değerden küçük olduğu görülmektedir.

3.2. "Türkiye-Singapur örneklemelerinde Bayes faktör ile ölçme değişmezliği sağlanmakta mıdır?"

Maddelerin ölçme değişmezliği hipotez testi için hesaplanan Türkiye-Singapur gruplarına ilişkin Bayes faktör sonuçları ( $BF_{01}$ ) ve grupların madde güçlük parametreleri arasındaki fark ( $d_T - d_S$ ) Tablo 15'de gösterilmiştir.

**Tablo 15**

*Türkiye-Singapur Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark*

Maddeler	$d_T - d_S$	$BF_{01}$	Karar*
Madde 1	<b>1.060</b>	<b>5.837*</b>	$H_0$ lehine
Madde 2	0.748	0.298	$H_1$ lehine
Madde 3	1.165	0.000	$H_1$ lehine
Madde 4	1.260	0.749	Kararsız
Madde 5	0.091	1.536	Kararsız
Madde 6	<b>0.316</b>	<b>5.666*</b>	$H_0$ lehine
Madde 7	2.040	0.743	Kararsız
Madde 8	1.136	0.100	$H_1$ lehine
Madde 9	0.684	0.000	$H_1$ lehine
Madde 10	1.387	2.611	Kararsız
Madde 11	1.295	0.141	$H_1$ lehine

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Bayes Faktör ( $BF_{01}$ ) sonuçları incelendiğinde, madde 1 ve madde 6'nın ölçme değişmezliğinin sağlandığına ilişkin yeterli derecede kanıt ürettiği ( $BF_{01} > 3$ ) ancak ölçme değişmezliğine ilişkin güçlü kanıt üretmediği belirlenmiştir. Madde güçlük parametreleri arasındaki fark ve Bayes faktör sonuçları birlikte değerlendirildiğinde, madde 1'e ilişkin parametreler arası farkın 0.7'nin üzerinde olduğu, madde 6 için hesaplanan parametre farkının ise yaklaşık olarak 0.3 görülmektedir.

Madde 2, madde 3, madde 8, madde 9 ve madde 11'in  $BF_{01}$  sonuçları ölçme değişmezliğinin sağlanmadığına ilişkin güçlü düzeyde ( $BF_{01} < 0.10$ ) kanıt üretmiştir. Parametreler arası fark değerlendirildiğinde, madde 9'a ait farkın yaklaşık olarak 0.7 olduğu diğer tüm maddelerin ise 0.7'den büyük olduğu görülmektedir.

Madde 4, madde 5, madde 7 ve madde 10'nun ölçme değişmezliğine ilişkin Bayes faktör kararsız kaldığından ( $1 < BF_{01} < 3$ ) değişmezliğin sağlanıp sağlanmadığına ilişkin yorum yapılamamaktadır. Bayes faktörün kararsız kaldığı bu maddelerin güçlük parametreleri arasındaki fark ise madde 5 için yaklaşık olarak 0.1 iken diğer maddelerde 0.7'den büyük olarak belirlenmiştir.

3.3. "Şili-Singapur örneklerinde Bayes faktör ile ölçme değişmezliği sağlanmaktadır mıdır?"

Maddelerin ölçme değişmezliği hipotez testi için hesaplanan Şili-Singapur gruplarına ilişkin Bayes faktör sonuçları ( $BF_{01}$ ) ve grupların madde güçlük parametreleri arasındaki fark ( $d_T - d_S$ ) Tablo 16'da gösterilmiştir.

**Tablo 16**

*Şili-Singapur Bayes Faktör sonuçları ve madde güçlük parametreleri arasındaki fark*

Maddeler	$d_S - d_T$	$BF_{01}$	Karar*
Madde 1	1.973	<b>4.311</b>	$H_0$ lehine
Madde 2	1.706	<b>5.904</b>	$H_0$ lehine
Madde 3	3.180	0.000	$H_1$ lehine
Madde 4	1.506	0.817	Kararsız
Madde 5	0.979	<b>5.503</b>	$H_0$ lehine
Madde 6	1.478	2.480	Kararsız
Madde 7	2.571	2.653	Kararsız
Madde 8	2.100	<b>6.021</b>	$H_0$ lehine
Madde 9	1.307	0.870	Kararsız
Madde 10	2.590	<b>4.512</b>	$H_0$ lehine
Madde 11	1.780	2.105	Kararsız

\* $BF_{01} > 3$  :  $H_0$  lehine,  $1 < BF_{01} < 3$  : Kararsız,  $BF_{01} < 0.33$  :  $H_1$  lehine

Bayes Faktör ( $BF_{01}$ ) sonuçlarına göre, madde 1, madde 2, madde 5, madde 8 ve madde 10'nun ölçme değişmezliğinin sağlandığına ilişkin yeterli derecede kanıt ürettiği ( $BF_{01} > 3$ ) ancak ölçme değişmezliğine ilişkin güçlü kanıt üretmediği söylenebilir. Madde güçlük parametreleri arası fark ve Bayes faktör sonuçları birlikte değerlendirildiğinde, tüm maddeler için hesaplanan parametre farkının 0.7'den büyük olduğu görülmektedir.

Madde 3 için hesaplanan Bayes faktör ölçme değişmezliğinin sağlanmadığına ilişkin güçlü düzeyde kanıt üretmiştir ve parametreler arasındaki fark 0.7'den büyüktür.

Madde 4, madde 6, madde 7, madde 9 ve madde 11'in Bayes faktör sonuçları ölçme değişmezliğine ilişkin kararsız kaldığından değişmezliğin sağlanıp sağlanmadığına ilişkin kanıt üretilmemiştir. Bayes faktörün kararsız kaldığı bu maddelerin güçlük parametreleri arasındaki fark ise tüm maddelerde 0.7'den büyük olarak belirlenmiştir.

## Tartışma

Simülasyon çalışmasından elde edilen sonuçlarda, gruplarda madde güçlük parametreleri arasındaki farkın olmadığı ve 0.1 olduğu durumlarda tüm grup büyüklüklerinde BF'nin ölçme değişmezliği lehine kanıt ürettiği görülmüştür. Madde güçlük parametreleri arası fark  $d_k=0.3$ , örneklem büyüklüğü 1500 ve daha küçük olduğu durumlarda ölçme değişmezliği lehine kanıt üreten Bayes faktör sonuçlarının giderek azaldığı,  $n=2000$  olduğunda ise ölçme değişmezliği lehine kanıt üreten BF sonucu olmadığı gözlemlenmemiştir.  $d_k=0.5$  olduğu durumda, yalnızca  $n=500$  iken ölçme değişmezliği lehine kanıt üreten maddelere rastlanırken diğer örneklem büyüklüklerinde BF sonuçlarının ölçme değişmezliğinin sağlanamadığı yönünde kanıt ürettiği belirlenmiştir. Madde güçlük parametreleri arasındaki fark 0.7 olduğunda ise örneklem büyüklüğünden bağımsız tüm gruplarda ölçme değişmezliği lehine kanıt üreten BF değerine rastlanmamıştır.

Alanyazında Bayesci YEM, Alignment yöntemi gibi diğer Bayesci yaklaşımlarla ilgili birçok çalışma bulunmasına rağmen ölçme değişmezliğinin Bayes faktör ile incelendiği sınırlı sayıda çalışma yer almaktadır. Araştırmadan elde edilen sonuçlara benzer olarak Verhagen (2013) madde güçlük parametreleri arasındaki fark büyük olduğunda ( $d_k \geq 0.5$ ) Bayes faktörün ölçme değişmezliğinin belirlenmesinde iyi performans gösterdiğini raporlamıştır. Madde güçlük parametreleri arasındaki fark daha küçük olduğunda ise ( $d_k < 0.1$  ve  $d_k < 0.3$ ) Bayes faktörün ölçme değişmezliği hakkında yeterli kanıt üretmediği ve kararsız kaldığını belirtmiştir. Verhagen ve diğerleri (2016) madde güçlük parametreleri arasındaki fark 0.5'ten büyük ise Bayes faktörün ölçme değişmezliğinin belirlenmesinde geçerli bir yöntem olduğunu ve örneklem büyüklüğü arttıkça ölçme değişmezliğine ait sonuçların doğru tespit edilmesinin daha olası olduğunu raporlamışlardır. Benzer şekilde çalışmada madde güçlük parametreleri arasındaki fark 0.5 olduğu durumlarda BF ölçme değişmezliğinin belirlenmesi konusunda büyük örneklemelerde daha iyi performans göstermiştir.

BF testinde, geçerli çıkarımlar yapmak için her iki ölçme değişmezliği hipotezinin de ( $H_0$  ve  $H_1$ ) doğru olması gerekmez. BF testi, sıfır hipotezinin doğru olduğu varsayılarak

çıkarımların yapıldığı frekansçı hipotez testlerinin aksine, sadece hangi hipotezin daha olası olduğu hakkında bilgi verir ve kararda her iki hipotez de dikkate alınır (Fox, 2020). BF kesme noktalarından yapılan Bayesci yorumlar, teorik bir kanıttan oluşturulmamıştır ve yapılabilecek istatistiksel yorumlar, araştırma sorularına ve çıkarımlarına bağlı olarak farklılık gösterebilir. (Joo vd., 2022). Konjin ve diğerleri (2015), BF sonuçlarının frekansçı istatistiklerdeki “p değeri” gibi yorumlanmaması gerektiği konusunda araştırmacıların dikkatli olmasını önermişlerdir. Bu açıdan elde edilen BF sonuçlarından yola çıkarak kesin yorumlamalar yapmak yerine yalnızca test edilen hipotezlerin hangisine daha fazla destek oluşturduğunun yorumlanması gerektiğini ifade etmişlerdir. Elde edilen sonuçlar genel olarak değerlendirildiğinde alanyazındaki çalışmaların araştırma sonuçları ile tutarlı olduğu belirlenmiştir.

Araştırmada TIMSS 2019 Şili, Türkiye ve Singapur örneklemi matematik alt testinde yer alan maddelerin ölçme değişmezliği Bayesci MTK modelleri kapsamında Bayes faktör ile incelenmiş, elde edilen sonuçlarda Şili-Türkiye (düşük düzey başarı-orta düzey başarı) ve Türkiye-Singapur (orta düzey başarı-yüksek düzey başarı) örneklemelerinde elde edilen BF değerlerinin ölçme değişmezliğinin belirlenmesinde simülasyon verilerden elde edilen sonuçlarla tutarlı olduğu görülmüştür. Ancak Şili-Singapur (düşük düzey başarı-yüksek düzey başarı) örneğinde sonuçların tutarlı olmadığı belirlenmiştir. Şili-Türkiye ve Türkiye-Singapur örneklemelerinde testin yaklaşık olarak %45’inde madde güçlük parametreleri arası fark arttıkça ölçme değişmezliğinin sağlanmadığı yönünde kanıt üretilmediği ya da Bayes faktörün kararsız kaldığı görülmüştür. Bununla birlikte her üç örneklem de Madde 1’de madde güçlük parametreleri arasındaki fark 0.7’den büyük olmasına rağmen ölçme değişmezliğinin sağlandığı yönünde kanıt üretilmiştir. Ayrıca Madde 4 ve Madde 7’de yine her üç örnekte de Bayes faktörün kararsız kaldığı görülmüştür. Bu durumda madde güçlük parametreleri arasındaki farkın yanı sıra, ölçme değişmezliğini etkileyebilecek madde içeriği, birey özellikleri arasındaki farklılıklar gibi çeşitli faktörlerin olabileceği yorumu yapılabilir. MTK modellerde ölçme değişmezliği testi olan DMF aynı yetenek düzeyindeki bireylerin bir testteki maddeye ait yanıtlama olasılıkları farklılaştığında görülür (Holland ve Wainer, 1993; Penfield ve Camilli,



2007; Osterlind ve Everson, 2009). Çalışmalar benzer yetenek düzeyine sahip olmayan gruplarda DMF tespitinin doğruluğunu ve yakınsama oranını önemli derecede etkileyeceğini göstermiştir. (Stark vd., 2006; De Boeck 2008; Finch 2016). Bununla birlikte Lin (2020) ölçme değişmezliği testlerinden rastgele madde etkileri modeli ile Alignment yöntemini karşılaştırdığı çalışmada, her iki yöntemde de ortalama grup yeteneğinin parametre kestirimleri üzerinde oldukça büyük bir etkiye sahip olduğunu ve grup yetenek ortalamaları eşit olmadığında hata oranı yüksek sonuçlar elde edildiğini raporlamıştır. Bu sebeple grup başarı düzeyleri arasındaki farkın ikili grup karşılaştırmalarında da Bayes faktörün performansını etkileyebileceği düşünülmektedir. Araştırmada kullanılan Bayesci MTK kapsamındaki Bayes faktör sonuçlarının Şili (eTIMSS uygulamasında en düşük puan alan ülke) ve Singapur (eTIMSS uygulamasında en yüksek puan alan ülke) örneklerinde tutarsız sonuçlar üretmesi grupların ortalama yetenek düzeyinin birbirinden çok farklı olması ile açıklanabilir. TIMSS 2019 Şili, Türkiye ve Singapur örnekleri matematik alt testine ait yapının ölçme değişmezliği ÇGDFA ile ortaya koyulmuş, Şili-Türkiye ve Türkiye-Singapur örneklerinde ölçme değişmezliği metrik değişmezlik düzeyinde, Şili-Singapur örneklerinde yapısal değişmezlik düzeyinde sağlamıştır. ÇGDFA ile ölçme değişmezliği testinde, model uyumları grup düzeyinde karşılaştırılırken, Bayesci MTK modellerde madde düzeyinde ölçme değişmezliği testi sonuçları sunulur. Araştırmada bu iki yöntem için herhangi bir karşılaştırma analizi yapılmamıştır. Ancak genel olarak değerlendirildiğinde, Şili-Türkiye ve Türkiye-Singapur örneklerinde metrik değişmezlik düzeyi, Şili-Singapur örneğinde yapısal değişmezlik düzeyi ölçme değişmezliği sağlanabildiği için testte yer alan maddelerin çoğunluğunda madde düzeyinde değişmezlik gösterme ihtimalinin de düşük olduğu, buna dayalı olarak da ÇGDFA ve Bayes faktör sonuçlarının tutarlı olduğu yorumu yapılabilir.

## Bölüm 5

### Sonuç ve Öneriler

Bu bölümde araştırmadan elde edilen bulgulara dayalı ulaşılan sonuçlara ve elde edilen bu sonuçlar çerçevesindeki önerilere yer verilmiştir. Öncelikle simülasyon verisine ait sonuçlar, daha sonra gerçek verilerden elde edilen sonuçlar paylaşılmıştır.

#### Sonuçlar

##### *Simülasyon Verilerden Elde Edilen Sonuçlar*

Araştırmada Bayesci MTK model kapsamında bir test uygulamasındaki maddelerin ölçme değişmezliğine ilişkin madde güçlük parametreleri arasındaki farka dayalı bir kesme puanı önerilmesi amaçlanmıştır. Bu amaç doğrultusunda yürütülen simülasyon çalışmasında parametreler arası farka dayalı kurulan hipoteze ilişkin Bayes faktör sonuçları farklı örneklem büyüklüklerinde incelenmiştir. Simülasyon çalışmasından elde edilen sonuçlara göre;

- 1) Madde güçlük parametreleri arasında farkın olmadığı ( $d_k=0.0$ ) durumda maddelerin tümü için hesaplanan Bayes faktör sonuçlarının tüm grup büyüklüklerinde kesme puanı olarak belirlenen 3'ten büyük olduğu belirlenmiştir ( $BF_{01}>3$ ). Gruplardaki kişi sayıları 500, 1000 ve 1500 olduğu durumlarda Bayes faktör için hem yeterli düzeyde kanıt hem de güçlü düzeyde kanıt bulunan maddeler varken, gruplardaki kişi sayısı 2000 olduğu durumda ise Bayes faktör sonuçlarının maddelerde ölçme değişmezliğine ilişkin güçlü düzeyde kanıt oluşturduğu belirlenmiştir.
- 2) Gruplarda madde güçlük parametreleri arasındaki farkın  $d_k=0.1$  olduğu durumda tüm grup büyüklüklerinde bütün maddelerde ölçme değişmezliğinin sağlandığı belirlenmiştir. Elde edilen Bayes faktör sonuçlarına göre tüm grup büyüklüklerinde ölçme değişmezliğine ilişkin hem yeterli düzeyde ( $BF_{01}>3$ ), hem de güçlü düzeyde ( $BF_{01}>10$ ) kanıt üreten maddeler olduğu ortaya koyulmuştur.

- 3) Madde güçlük parametreleri arasındaki farkın  $d_k = 0.3$  olduğu durumlarda yalnızca gruplarda 500 ve 1000 kişi olduğunda ölçme değişmezliğinin sağlandığı maddelere rastlanmaktadır. Grup büyüklüğü 2000 olduğunda ise tüm maddelerdeki Bayes faktör sonuçlarının ölçme değişmezliğinin sağlanmadığına ilişkin kanıt ürettiği belirlenmiştir. Genel olarak  $d_k = 0.3$  olduğu durumlar değerlendirilirse, grup büyüklükleri 500 ve 1000 olduğunda, değişmezliğin sağlandığı, sağlanmadığı ve kararsız sonuçlar üretilirken, 1500 ve 2000 olduğu durumlarda hiçbir maddede ölçme değişmezliğinin sağlanmadığı belirlenmiştir. Genel olarak değerlendirildiğinde grup büyüklüğü arttıkça 0.3 düzeyindeki bir fark için Bayes faktör sonuçlarının grup büyüklüğünden etkilendiği ve grup büyüklüğü arttıkça değişmez madde sayısının azaldığı belirlenmiştir.
- 4) Madde güçlük parametreleri arasındaki fark  $d_k = 0.5$  olduğunda yalnızca grup büyüklüğü 500 olduğunda ölçme değişmezliği lehine yeterli düzeyde ( $BF_{01} > 3$ ) kanıt bulunan dört madde belirlenirken, bir madde için Bayes faktör kararsız kalmış, diğer beş maddenin değişmezliğinin sağlanmadığına ilişkin kanıt üretmiştir. Grup büyüklükleri 1000, 1500 ve 2000 olduğu durumlarda ise Bayes faktörün tüm maddelerde  $H_1$  lehine güçlü kanıt ( $BF_{01} < 0.10$ ) ürettiği belirlenmiştir. Bu durumda örneklem büyüklüğü arttıkça BF'nin doğru karar verme olasılığının yükseldiği belirlenmiştir.
- 5) Madde güçlük parametreleri arasındaki farkın  $d_k = 0.7$  olduğu durumda ise tüm grup büyüklüklerinde ölçme değişmezliği için  $H_1$  lehine ( $BF_{01} < 0.10$ ) değişmezliğin sağlanmadığına ilişkin kanıt üretildiği belirlenmiştir.

### **Gerçek Verilerden Elde Edilen Sonuçlar**

Araştırmada TIMSS 2019 uygulaması matematik alt testinde Şili, Türkiye ve Singapur örneklemelerinde Bayes faktör ve çoklu-grup doğrulayıcı faktör analizi ile ölçme değişmezliği testleri yapılmıştır. Çoklu-grup faktör analizi testleri ölçülen yapının ölçme değişmezliğinin hangi düzeyde sağlandığını belirleme amacıyla yapılmıştır. Madde düzeyinde ölçme değişmezliği ise Bayes faktör ile incelenmiştir. Bayes faktör ile ölçme değişmezliği testi

sonuçları simülasyon çalışması ile karşılaştırılmıştır. Elde edilen sonuçlar maddeler halinde sıralanmıştır.

- 1) Şili-Türkiye örnekleme için BF ile ölçme değişmezliği sonuçları değerlendirildiğinde, matematik alt testinde yer alan 11 maddenin 6 tanesi için ölçme değişmezliği lehine yeterli düzeyde kanıt üretildiği tespit edilmiştir ( $BF_{01} > 3$ ). Bu maddelerin 2 tanesinde madde güçlük parametreleri arası fark değişmezlik için kesme puanı olarak belirlenen  $d_k = 0.7$ 'den daha küçük olduğu, diğer 4 maddenin güçlük parametreleri arası farkın bu kesme puanından büyük olduğu belirlenmiştir. 2 maddede ölçme değişmezliğinin sağlanmadığı ve bu maddelerin parametreler arası farkının  $0.7$ 'den büyük olduğu belirlenmiş, 3 maddenin değişmezliği için ise BF kararsız kalmıştır. Genel olarak değerlendirildiğinde 6 maddenin değişmezlik sonuçları simülasyon verilerden elde edilen sonuçlar ile yaklaşık olarak %55'lik bir oranla tutarlılık göstermiştir.
- 2) Şili-Türkiye örnekleme için ÇGDFA sonuçlarına göre ölçme değişmezliği metrik değişmezlik sağlanmıştır. Bu durumda Şili ve Türkiye örneklemlerinin eşit faktör yüklerine sahip olduğu söylenebilir. MTK'de madde güçlük parametreleri DFA'da eşik parametrelerine karşılık geldiğinden tüm maddelerde eşik parametresi eşit olmadığından skaler değişmezlik sağlanamamıştır. ÇGDFA sonuçları ile BF sonuçlarının tutarlılık gösterdiği söylenebilir.
- 3) Türkiye-Singapur örnekleme için BF sonuçları 2 maddede ölçme değişmezliği lehine kanıt üretmiştir. Bu maddelerden bir tanesinin güçlük parametreleri arasındaki fark  $0.7$ 'den küçük, diğerinde ise  $0.7$ 'den büyüktür. Kalan 5 maddede ise BF sonuçları değişmezliğin sağlanmadığına yönelik kanıt üretmiş ve bu maddelerin tümünde madde güçlük parametreleri arası fark  $d_k \geq 0.7$  olarak belirlenmiştir. 4 madde için ise BF kararsız sonuçlar üretmiştir. Genel olarak değerlendirildiğinde 6 maddeye ait sonucun yaklaşık olarak %55 oranda simülasyon veriden elde edilen sonuçlar ile tutarlı olduğu belirlenmiştir.

- 4) Türkiye-Singapur örnekleme ÇGDFA sonuçlarına göre ölçme değişmezliğinin metrik değişmezlik düzeyinde sağlandığı belirlenmiştir. Bu sonuç BF ile ölçme değişmezliği testi sonuçları ile tutarlılık göstermektedir.
- 5) Şili-Singapur örnekleme için BF sonuçları 5 madde için tüm maddelerin güçlük parametreleri arasındaki fark 0.7'den büyük olmasına rağmen ölçme değişmezliği lehine kanıt üretmiştir. Yalnızca 1 maddede değişmezliğin sağlanmadığına dair güçlü kanıt üretmiş ( $d_k = 3.180$ ), diğer bütün maddeler için BF sonuçları ölçme değişmezliği hipotezi için kararsız kalmıştır ( $d_k > 0.7$ ). İki ülke arasında tüm maddelerde güçlük parametreleri arasındaki farkın oldukça yüksek olduğu göz önüne alındığında BF sonuçlarının simülasyon çalışmasında elde edilen sonuçlarla yalnızca %9'luk bir benzerlik yakaladığı ve simülasyon sonuçları ile tutarlılık göstermediği belirlenmiştir.
- 6) Şili-Singapur örnekleme ÇGDFA sonuçlarına göre ölçme değişmezliğinin en zayıf model olan yapısal değişmezlik düzeyinde sağlandığı belirlenmiştir. Elde edilen bu sonucun BF sonuçları ile tutarlı olmadığı belirlenmiştir.

## Öneriler

### *Uygulayıcılara Yönelik Öneriler*

- 1) Bayes faktör ile ölçme değişmezliği testi, frekansçı yaklaşımların aksine ölçme değişmezliği hipotezlerinin herhangi birinin doğru olmasını gerektirmez. BF, yalnızca hangi hipotezin daha olası olduğuna ilişkin bilgi sağladığından kararda her iki hipotez de dikkate alınmalıdır.
- 2) BF, tüm maddeler için ölçme değişmezliği testini aynı anda herhangi bir kısıtlama olmaksızın yapabildiğinden ortak madde bulunmayan durumlarda ölçme değişmezliğinin belirlenmesinde kullanılabilir.
- 3) BF ile ölçme değişmezliği sonuçları ortak madde ihtiyacı olan durumlarda, ortak maddeyi belirleme amacıyla kullanılabilir.

- 4) Simülasyon veriden elde edilen sonuçlara göre örneklem büyüklüğü arttıkça BF sonuçlarının daha iyi performans gösterdiği ve ölçme değişmezliğine ilişkin karar verirken daha iyi performans gösterdiği belirlenmiştir. BF ile ölçme değişmezliği testi sonuçları değerlendirilirken örneklem büyüklüğü dikkate alınarak yorumlamalar yapılmalıdır.
- 5) Grup yetenek düzeyleri arasında farkın çok olduğu durumlarda BF ile ölçme değişmezliği sonuçları yanlış sonuçlar üretebileceğinden, BF ile ölçme değişmezliği testinin tercih edilmemesi, tercih edildiği durumlarda ise sonuçlarının dikkatli yorumlanması gerekliliği göz önünde bulundurulmalıdır.
- 6) TIMSS 2019 uygulamasında, kültürlere ya da yapılan uygulamanın türüne göre ölçme değişmezliği sonuçları farklılık gösterebileceğinden, gruplar arası karşılaştırmalar ve yorumlamalar yapılmadan önce, ölçme değişmezliğine yönelik çalışmalarının gerçekleştirilmesi gerekliliği göz ardı edilmemelidir.

### ***Araştırmacılara Yönelik Öneriler***

- 1) Araştırma 1 parametrelili lojistik model ile sınırlandırılmıştır. Gelecek araştırmalarda diğer MTK modellerinde BF ile ölçme değişmezliği testi ele alınabilir.
- 2) Çalışmada yalnızca ikili puanlanan (1-0) maddeler ele alınarak BF testi yapılmıştır. Gelecek çalışmalarda çoklu puanlanan maddelerle BF ile ölçme değişmezliği ele alınabilir.
- 3) Araştırmada simülasyon veride BF ile ölçme değişmezliği grup büyüklüklerinin eşit olduğu durumlar için incelenmiştir, gelecek çalışmalarda karşılaştırılan grup büyüklüklerinin eşit olmadığı durumlarda da BF ile ölçme değişmezliği sonuçları incelenebilir.
- 4) Çalışmadan elde edilen sonuçlarda grupların yetenek düzeyi birbirinden çok farklı olduğu durumlarda BF sonuçlarının tutarlı olmadığı belirlenmiştir. Gelecek çalışmalar

benzer yetenek düzeyine sahip gruplar ile BF ile ölçme değişmezliği sonuçlarını inceleyebilirler.

- 5) Araştırmada ÇGDFA sonucunda hangi düzeyde ölçme değişmezliğinin sağlandığı belirlenerek BF ile madde düzeyi ölçme değişmezliği sonuçları incelenmiştir. Gelecek araştırmalar BF sonuçları ile ÇGDFA temelli, MTK temelli ya da diğer ölçme değişmezliği yöntemlerini (Alignent yöntemi, Bayesci YEM vs.) kıyaslayabilirler.
- 6) Araştırma Bayes yaklaşımlarından yalnızca BF ile sınırlandırılmıştır. Gelecek çalışmalarda DIC gibi diğer Bayesci bilgi kriterleri kullanılarak ölçme değişmezliğinin sonucu incelenebilir ve BF ile bu sonuçlar kıyaslanabilir.
- 7) Araştırmada yalnızca farklı kültürlere sahip gruplar karşılaştırılarak BF ile ölçme değişmezliği incelenmiştir. Gelecek araştırmalar aynı kültüre sahip alt gruplarda ölçülen gizil değişkene ait BF ile ölçme değişmezliği testleri incelenebilir.
- 8) Çalışmada farklı grupların ölçümlerine ilişkin ölçme değişmezliği incelenmiştir. Gelecek araştırmalarda aynı grubun farklı zamanlardaki ölçüm sonuçlarına ilişkin ölçme değişmezliği BF ile incelenebilir.

## Kaynaklar

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67-91. <https://doi.org/10.1111/j.1745-3984.1992.tb00368.x>
- Asparouhov, T., & Muthén, B. O. (2006). Robust chi square difference testing with mean and variance adjusted test statistics (*Mplus web notes no. 10*). <http://www.statmodel.com/download/webnotes/webnote10.pdf>
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural equation modeling: A Multidisciplinary Journal*, 21(4), 495-508. <https://doi.org/10.1080/10705511.2014.919210>
- Ayvallı, M., & Kelecioğlu, H. (2021, 1-4 Eylül). *Bayesian hiyerarşik modeller ölçme değişmezliğinin incelenmesi*. [Kongre Bildiri Özeti]. Yedinci Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, <https://epodder.org/kitap/epod-2020-bildiri-kitabi.pdf>
- Bejar, I. I. (1980). Biased assessment of program impact due to psychometric artifacts. *Psychological Bulletin*, 87(3), 513. <https://doi.org/10.1037/0033-2909.87.3.513>
- Bock, R. D., & Zimowski, M. F. (1997). The multiple groups IRT. In Wim J. van der Linden, & Ronald K. Hambleton (Eds.), *Handbook of modern item response theory*. Springer-Verlag. [https://doi.org/10.1007/978-1-4757-2691-6\\_25](https://doi.org/10.1007/978-1-4757-2691-6_25)
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.



- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*: Sage.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25(1), 1-27. <https://doi.org/10.1177/014920639902500101>
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212. <https://doi.org/10.1177/0022022100031002003>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40, 55-75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73(4), 533-559. <https://doi.org/10.1007/s11336-008-9092-x>
- Demars, C. (2010). *Item response theory: Understanding statistics measurement*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>
- Dimitrov, D.M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.

- DiStefano, C. & Morgan, G. B., (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data, *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 425-438. <https://doi.org/10.1080/10705511.2014.915373>
- Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, 25(1), 5-34. <https://doi.org/10.3758/s13423-017-1262-3>
- Fay, D. M., Levy, R., & Mehta, V. (2018). Investigating psychometric isomorphism for traditional and performance-based assessment. *Journal of Educational Measurement*, 55(1), 52-77. <https://doi.org/10.1111/jedm.12163>
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT Likelihood Ratio. *Applied Psychological Measurement*, 29(4), 278–295. <https://doi.org/10.1177/0146621605275728>
- Finch, W. H. (2016). Detection of differential item functioning for more than two groups: A monte carlo comparison of methods. *Applied Measurement in Education*, 29(1), 30-45. <https://doi.org/10.1080/08957347.2015.1102916>
- Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: Information Age.
- Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 User Guide for the International Database* (2<sup>nd</sup> ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-database/>
- Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4419-0742-4>
- Fox, J. P., Mulder, J., & Sinharay, S. (2017). Bayes factor covariance testing in item response models. *Psychometrika*, 82, 979-1006. <https://doi.org/10.1007/s11336-017-9577-6>

- Fox, J. P., Koops, J., Feskens, R., & Beinhauer, L. (2020). Bayesian covariance structure modelling for measurement invariance testing. *Behaviormetrika*, *47*, 385-410. <https://doi.org/10.1007/s41237-020-00119-3>
- Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, *20*(2), 101-125. <https://doi.org/10.1177/014662169602000>
- Hirschfeld, G., & Von Brachel, R. (2014). Improving Multiple-Group confirmatory factor analysis in R—A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, and Evaluation*, *19*(1), 7. <https://doi.org/10.7275/qazy-2946>
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117-144. <https://doi.org/10.1080/03610739208253916>
- Hojtink, H. (2012). *Informative hypotheses. Theory and practice for behavioral and social scientists*. Chapman & Hall/CRC.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical irt model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, *25*, 285–306. <https://doi.org/10.3102/10769986025003285>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

- Johnson, T. P. (1998). *Approaches to equivalence in cross-cultural and cross-national survey research* (Vol. 3, pp. 1-40). DEU. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49730-6>
- Joo, S. H., Lee, P., & Stark, S. (2022). Bayesian approaches for detecting differential item functioning using the generalized graded unfolding model. *Applied Psychological Measurement*, 46(2), 98-115. <https://doi.org/10.1177/01466216211066606>
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426. <https://doi.org/10.1007/BF02291366>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of The American Statistical Association*, 90(430), 773-795. <http://dx.doi.org/10.1080/01621459.1995.10476572>
- Kline, R. B. (2016). *Principles and practice of structural equation modelling*. New York: Guilford Press.
- Konijn, E. A., van de Schoot, R., Winter, S. D., & Ferguson, C. J. (2015). Possible solution to publication bias through Bayesian statistics, including proper null hypothesis testing. *Communication Methods and Measures*, 9(4), 280-302. <https://doi.org/10.1080/19312458.2015.1096332>
- Kooperberg, C. (2015). *Polspline: polynomial spline routines*. <https://CRAN.R-project.org/package=polspline>. R package version 1.1.12.
- Labouvie, E. W. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological Issues* (pp. 493–502). American Psychological Association. <https://doi.org/10.1037/10050-036>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619-7>

- Lin, L. (2020). *Evaluate Measurement invariance across multiple groups: A comparison between the alignment optimization and the random item effects model* (Publication No. 28367620) [Doctoral dissertation, University of Pittsburgh]. ProQuest Dissertations & Theses Global.
- Little, T.D. (1997). Mean and covariance structures (macs) analyses of crosscultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53-76. [https://doi.org/10.1207/s15327906mbr3201\\_3](https://doi.org/10.1207/s15327906mbr3201_3)
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490–504. <https://doi.org/10.1037/0033-2909.111.3.490>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Martin, S. R., Williams, D. R., & Rast, P. (2019). Measurement invariance assessment with Bayesian hierarchical inclusion modeling. *PsyArXiv*. <https://doi.org/10.31234/osf.io/qbdjt>
- Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23(3), 524–545. <https://doi.org/10.1037/met0000113>
- McDonald, R. P. (2013). *Test theory: A unified treatment*: Psychology Press.
- Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling*, 14, 611–635. <https://doi.org/10.1080/10705510701575461>

- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://doi.org/10.1007/BF02294825>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY:Routledge. <https://doi.org/10.4324/9780203821961>
- Montenegro-Montenegro, E. (2020). *Improving the guidelines to conduct multigroup invariance test in Bayesian SEM*. [Unpublished doctoral dissertation]. Texas Tech University.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, L.K. & Muthén, B.O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén. [https://www.statmodel.com/download/usersguide/MplusUserGuideVer\\_8.pdf](https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf)
- Muthén, B., & Asparouhov, T. (2013). New methods for the study of measurement invariance with many groups. (*Mplus web notes* no.18). <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=aa3b622e3d67c400edd36e92ae186284cd77f0ab>
- Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research*, 47(4), 637-664. <https://doi.org/10.1177/0049124117701488>
- Naumann, A., Hartig, J., & Hochweber, J. (2017). Absolute and relative measures of instructional sensitivity. *Journal of Educational and Behavioral Statistics*, 42(6), 678-705. <https://doi.org/10.3102/1076998617703649>

- Organisation for Economic Co-operation and Development. (2019). *Invariance analyses in large-scale studies*. OECD Education Working Paper No. 201. [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP\(2019\)9&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP(2019)9&docLanguage=En)
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Sage Publications.
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay & C. R. Rao (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics* (pp. 125– 167). New York : Elsevier.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. In *R News* (Vol. 6, Issue 1, pp. 7–11). <https://journal.r-project.org/archive/>
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika* 53, 495– 502. <https://doi.org/10.1007/BF02294403>
- Raju, N. S., Van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19, 353–368. <https://doi.org/10.1177/014662169501900405>
- Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge. <https://doi.org/10.4324/9780203809532>
- Rizopoulos D (2006). “lrm: An R package for latent variable modelling and item response theory analyses.” *Journal of Statistical Software*, 17(5), 1– 25. <https://doi.org/10.18637/jss.v017.i05>.

- Rock, D. A., Werts, C. E., & Flaughner, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, 13(4), 403-418. [https://doi.org/10.1207/s15327906mbr1304\\_3](https://doi.org/10.1207/s15327906mbr1304_3)
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- 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. <https://doi.org/10.1111/j.2044-8317.1974.tb00543.x>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the Effects of Differential Item (Functioning and Differential) Test Functioning on Selection Decisions: When Are Statistically Significant Effects Practically Important? *Journal of Applied Psychology*, 89(3), 497–508. <https://doi.org/10.1037/0021-9010.89.3.497>
- 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. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90. <https://doi.org/10.1086/209528>
- Sturtz S, Ligges U, Gelman A (2005). "R2WinBUGS: A Package for Running WinBUGS from R." *Journal of Statistical Software*, 12(3), 1–16. <http://www.jstatsoft.org>.



- Tabachnick, B.G. and Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). New Jersey: Pearson.
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393-408. <https://doi.org/10.1007/BF02294363>
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99(1), 118–128. <https://doi.org/10.1037/0033-2909.99.1.118>
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Lawrence Erlbaum Associates, Inc.
- Thompson, Y. T. (2018). *Bayesian and frequentist approaches for factorial invariance test* [Unpublished doctoral dissertation], University of Oklahoma.
- Van de Vijver, F. J. (1998). Towards a theory of bias and equivalence. *50967*, 3, 41-65. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49731-1>
- 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(1), 4-70. <https://doi.org/10.1177/109442810031002>

- Van Doorn, J., van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š. L., Marsman, A., Matzke, M., Gupa, D., R. A., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813–826. <https://doi.org/10.3758/s13423-020-01798-5>
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Verhagen, A. J. (2012). *Bayesian item response theory models for measurement variance*. [Unpublished doctoral dissertation]. University of Twente.
- Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, 66(3), 383-401. <https://doi.org/10.1111/j.2044-8317.2012.02059.x>
- Verhagen, J., & Wagenmakers, E.-J. (2014). "Bayesian tests to quantify the result of a replication attempt": Correction to Verhagen and Wagenmakers (2014). *Journal of Experimental Psychology: General*, 143(6), 2073. <https://doi.org/10.1037/a0038326>
- Verhagen, J., Levy, R., Millsap, R. E., & Fox, J. P. (2016). Evaluating evidence for invariant items: A Bayes factor applied to testing measurement invariance in IRT models. *Journal of Mathematical Psychology*, 72, 171-182. <https://doi.org/10.1016/j.jmp.2015.06.005>
- von Davier, M. (2020). TIMSS 2019 scaling methodology: Item Response Theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 11.1-11.25). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html>

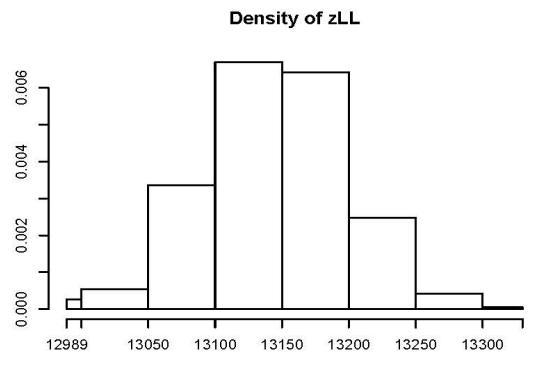
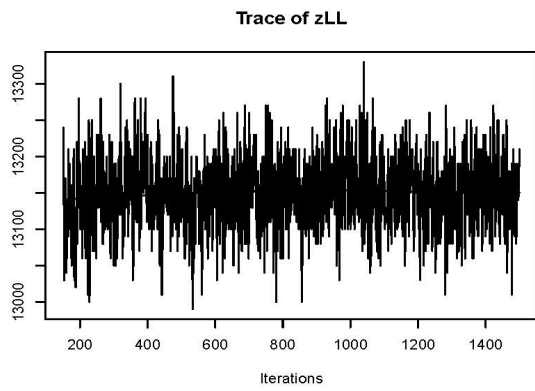
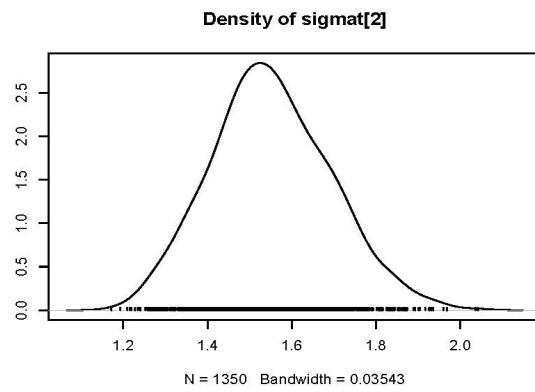
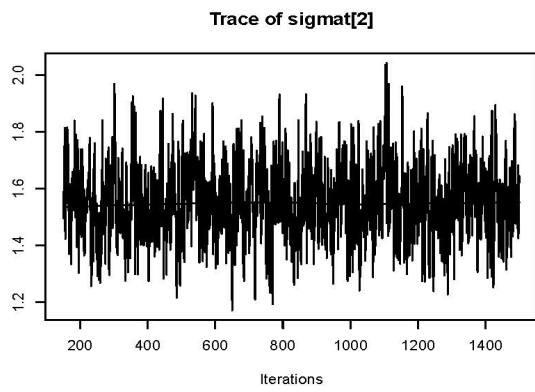
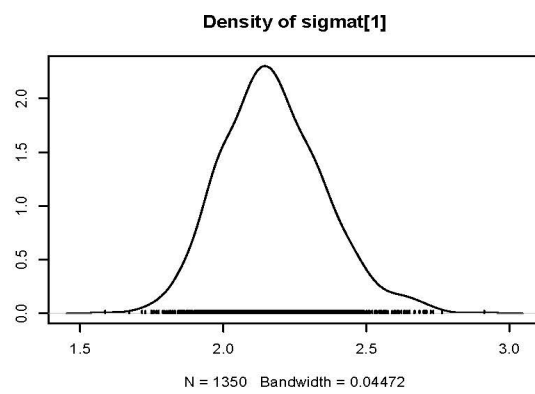
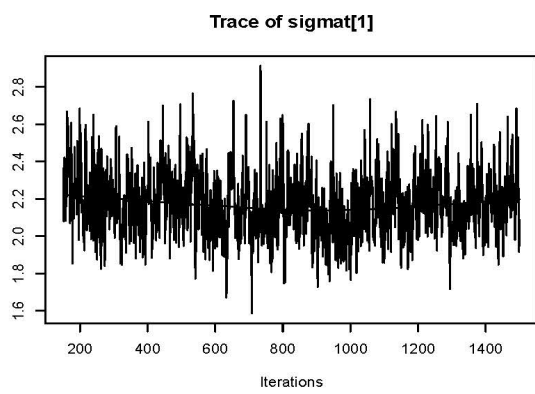
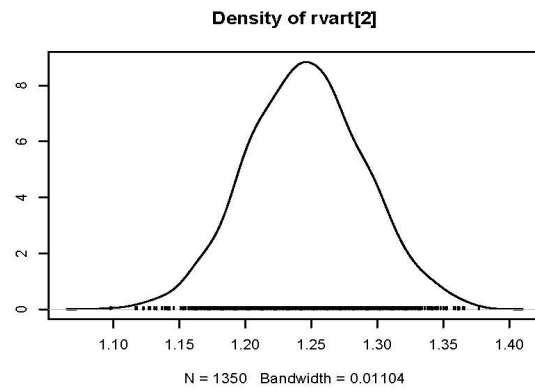
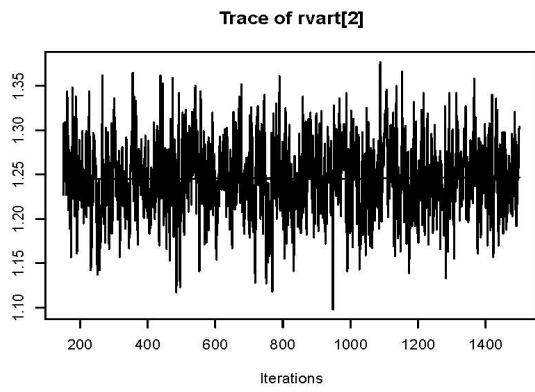
- Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3), 158-189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Wagenmakers, E. J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 123-138. <https://doi.org/10.1002/9781119095910.ch8>
- Wang, S. (2017). *Bayesian approximate measurement invariance approach* (Publication No. 27539277) [Doctoral dissertation, University of Cincinnati]. ProQuest Dissertations & Theses Global.
- Wetzels, R., Grasman, R. P., & Wagenmakers, E. J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio. *Computational Statistics & Data Analysis*, 54(9), 2094-2102. <https://doi.org/10.1016/j.csda.2010.03.016>
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097-1126. <https://doi.org/10.1111/1468-0262.00152>
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association. <https://doi.org/10.1037/10222-009>
- Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57. <https://doi.org/10.1177/0146621607314044>
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339–361. <https://doi.org/10.1177/0146621611405984>

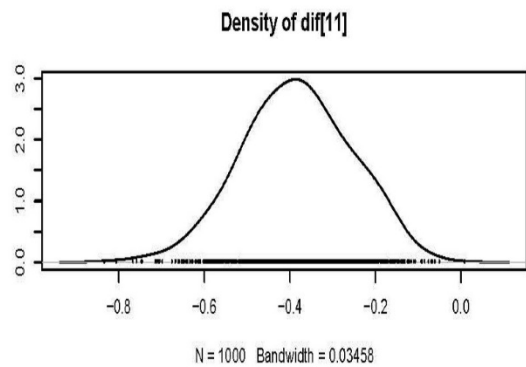
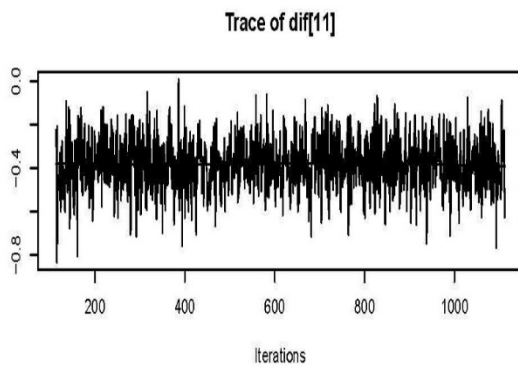
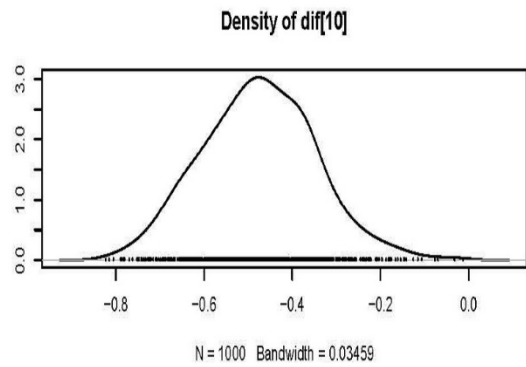
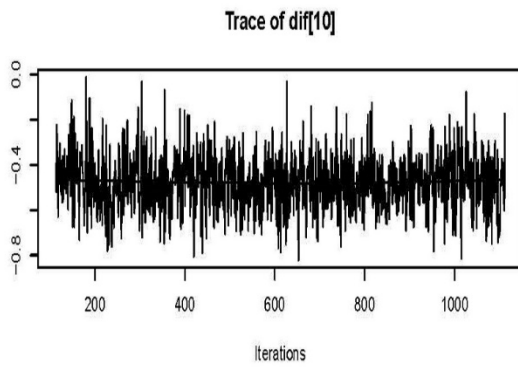
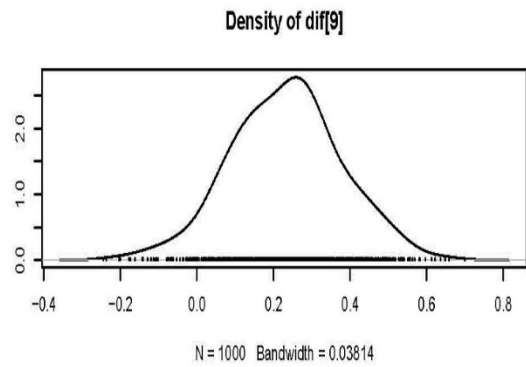
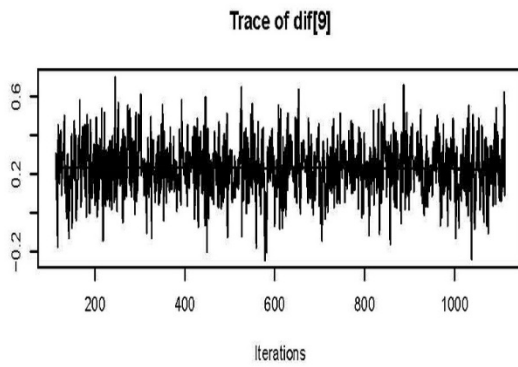
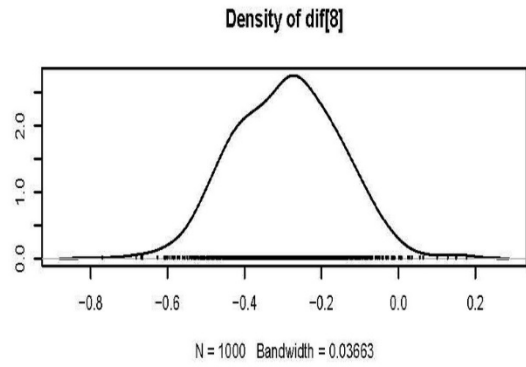
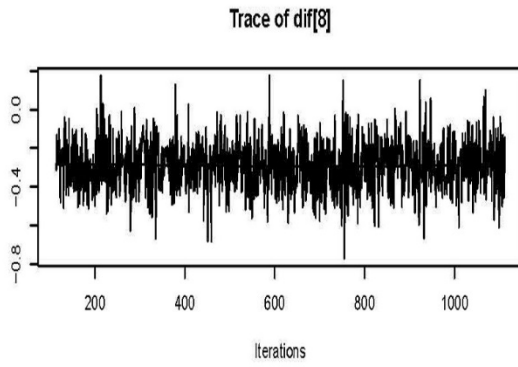
- Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, 73(3), 532-547. <https://doi.org/10.1177/0013164412464875>
- Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling*, 14(3), 435–463. <https://doi.org/10.1080/10705510701301677>
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Publication No. 3066425) [Doctoral dissertation, University of California, Los Angeles]. ProQuest Dissertations & Theses Global.

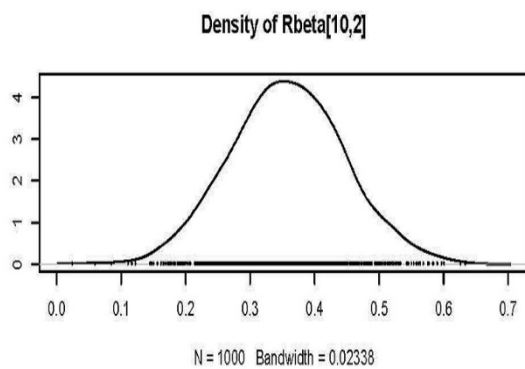
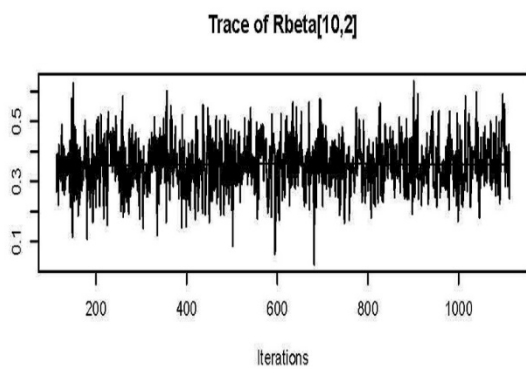
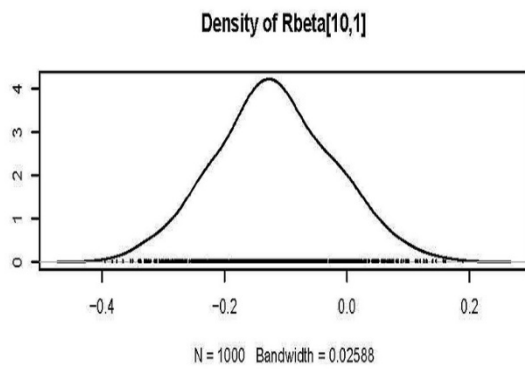
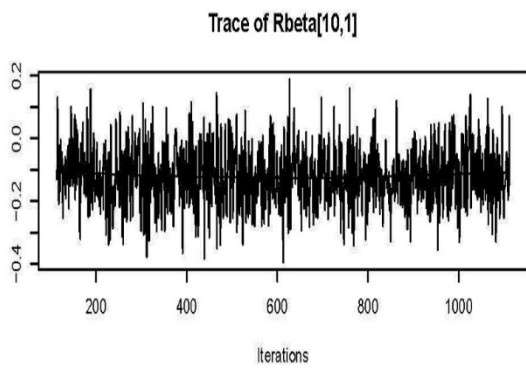
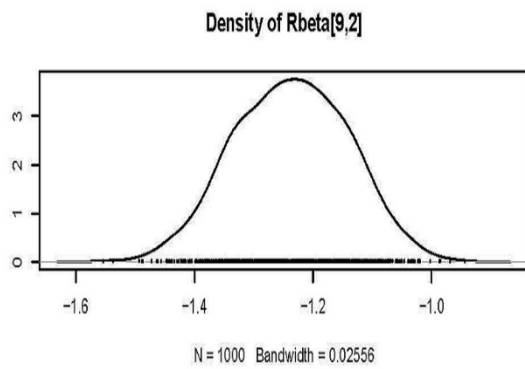
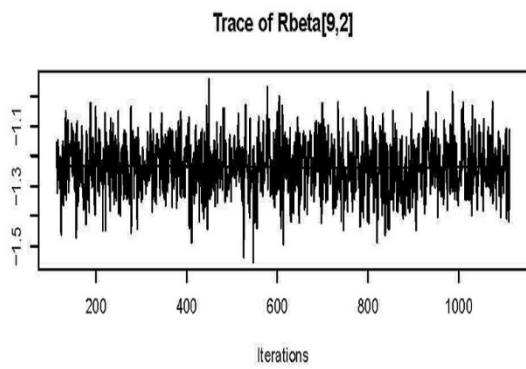
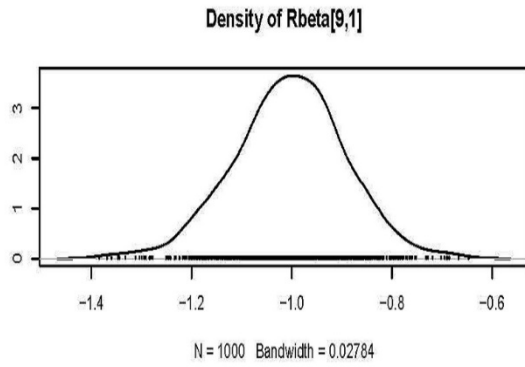
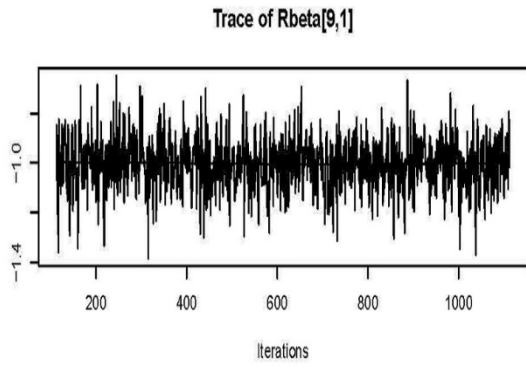
## EK-A: Simülasyon veriye ait madde güçlük parametre kestirimleri

Madde Parametreleri Arası Fark	Maddeler	Grup 1 N= 500	Grup 2 N= 500	Grup 1 N= 1000	Grup 2 N= 1000	Grup 1 N= 1500	Grup 2 N= 1500	Grup 1 N= 2000	Grup 2 N= 2000
$d_k=0.0$	Madde 1	-1.51524	-1.51524	1.35099	1.35099	-0.14897	-0.14897	0.61867	0.61867
	Madde 2	1.15225	1.15225	0.15923	0.15923	0.25329	0.25329	-1.20775	-1.20775
	Madde 3	-0.58729	-0.58729	-1.04742	-1.04742	-0.13534	-0.13534	-0.08178	-0.08178
	Madde 4	-1.17103	-1.17103	-2.31015	-2.31015	1.17636	1.17636	-0.95174	-0.95174
	Madde 5	-0.19137	-0.19137	1.41825	1.41825	-0.44912	-0.44912	-1.66571	-1.66571
	Madde 6	1.00173	1.00173	-0.8556	-0.8556	-0.85863	-0.85863	1.11071	1.11071
	Madde 7	0.92033	0.92033	-0.22294	-0.22294	0.10577	0.10577	-0.18306	-0.18306
	Madde 8	-0.19002	-0.19002	-0.1573	-0.1573	-0.77499	-0.77499	-0.09381	-0.09381
	Madde 9	0.13939	0.13939	1.84483	1.84483	0.7638	0.7638	1.10559	1.10559
	Madde 10	0.44124	0.44124	-0.17991	-0.17991	0.06784	0.06784	1.34887	1.34887
$d_k=0.1$	Madde 1	0.11130	0.21130	1.47472	1.57472	-1.14341	-1.04341	-1.25033	-1.15033
	Madde 2	-1.42535	-1.52535	-0.72646	-0.82646	0.64185	0.54185	-0.98874	-1.08874
	Madde 3	1.0506	1.1506	1.28423	1.38423	-0.23042	-0.13042	0.46737	0.56737
	Madde 4	0.61857	0.51857	0.72544	0.62544	-0.14181	-0.24181	0.28588	0.18588
	Madde 5	-1.17563	-1.07563	-1.6525	-1.5525	0.33662	0.43662	-0.4916	-0.3916
	Madde 6	0.85522	0.75522	-0.6442	-0.7442	1.94682	1.84682	1.52708	1.42708
	Madde 7	-0.72103	-0.62103	-0.2051	-0.1051	0.04659	0.14659	-0.2995	-0.1995
	Madde 8	0.67932	0.57932	1.21579	1.11579	-0.46611	-0.56611	-1.12467	-1.22467
	Madde 9	-1.13605	-1.03605	-0.84939	-0.74939	0.98127	1.08127	0.72206	0.82206
	Madde 10	1.14304	1.04304	-0.62251	-0.72251	-1.97141	-2.07141	1.15243	1.05243
$d_k=0.3$	Madde 1	-0.83200	-0.53200	1.39128	1.69128	-0.24920	0.05080	0.31251	0.61251
	Madde 2	2.23221	1.93221	-1.75713	-2.05713	-0.48799	-0.78799	-0.34488	-0.64488
	Madde 3	1.19172	1.49172	-0.58153	-0.28153	0.23554	0.53554	0.28424	0.58424
	Madde 4	-0.15327	-0.45327	1.3534	1.0534	0.28922	-0.01078	0.39478	0.09478
	Madde 5	-0.85613	-0.55613	-0.49663	-0.19663	-1.10461	-0.80461	-1.6253	-1.3253
	Madde 6	-1.24592	-1.54592	1.27965	0.97965	-0.53536	-0.83536	0.34952	0.04952
	Madde 7	0.30058	0.60058	-0.97319	-0.67319	0.65091	0.95091	1.91602	2.21602
	Madde 8	-0.37292	-0.67292	-0.18041	-0.48041	0.55902	0.25902	-0.05588	-0.35588
	Madde 9	-0.49208	-0.19208	-0.22857	0.07143	1.61207	1.91207	-1.14867	-0.84867
	Madde 10	0.22780	-0.07220	0.19312	-0.10688	-0.96959	-1.26959	-0.08233	-0.38233
$d_k=0.5$	Madde 1	1.01674	0.51674	0.69355	0.19355	0.38955	-0.11045	0.91923	0.41923
	Madde 2	-0.78838	-0.28838	1.6497	2.1497	-0.76968	-0.26968	-0.24888	0.25112
	Madde 3	-0.08844	-0.58844	-0.18656	-0.68656	1.59203	1.09203	-1.80859	-2.30859
	Madde 4	-0.82591	-0.32591	0.92077	1.42077	0.83879	1.33879	-0.5192	-0.0192
	Madde 5	0.88156	0.38156	0.36453	-0.13547	1.4836	0.9836	3.13911	2.63911
	Madde 6	0.4986	0.9986	-0.801	-0.301	0.43809	0.93809	-1.70057	-1.20057
	Madde 7	-0.47966	-0.97966	-0.60329	-1.10329	-0.74047	-1.24047	0.49179	-0.00821
	Madde 8	-0.3249	0.1751	0.09697	0.59697	-1.30876	-0.80876	-0.14801	0.35199
	Madde 9	-1.01617	-1.51617	-0.97367	-1.47367	-0.76574	-1.26574	0.23068	-0.26932
	Madde 10	1.12655	1.62655	-1.16101	-0.66101	-1.15741	-0.65741	-0.35558	0.14442
$d_k=0.7$	Madde 1	-0.02227	-0.72227	-0.06218	-0.76218	0.99252	0.29252	1.78028	1.08028
	Madde 2	0.29443	0.99443	2.12972	2.82972	0.43886	1.13886	0.3036	1.0036
	Madde 3	-0.88435	-1.58435	-0.77638	-1.47638	0.32672	-0.37328	0.74875	0.04875
	Madde 4	0.98951	1.68951	0.12361	0.82361	-0.37344	0.32656	-0.46008	0.23992
	Madde 5	0.00746	-0.69254	-0.58109	-1.28109	0.21024	-0.48976	-0.46803	-1.16803
	Madde 6	-1.44076	-0.74076	-0.1527	0.5473	0.35615	1.05615	-1.92719	-1.22719
	Madde 7	-0.21819	-0.91819	-1.40278	-2.10278	-0.45817	-1.15817	0.59344	-0.10656
	Madde 8	2.1256	2.8256	0.84659	1.54659	-1.23148	-0.53148	0.20459	0.90459
	Madde 9	-0.54605	-1.24605	0.42517	-0.27483	-0.31452	-1.01452	0.34729	-0.35271
	Madde 10	-0.30538	0.39462	-0.54997	0.15003	0.05311	0.75311	-1.12264	-0.42264

### EK-B: Örnek İz Grafikleri










## EK-C: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu Onay

## Bildirimi

	<b>Hacettepe Üniversitesi</b> <b>Eğitim Bilimleri Enstitüsü</b> <b>Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu</b>	<b>F46</b>
26 / 12 / 2022		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı Başkanlığına		
<b>Tez/Araştırma Başlığı</b>	Çok Düzeyli Madde Tepki Modellerinde Bayes Yaklaşımı ile Ölçme Değişmezliğinin İncelenmesi	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır. 2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir. 3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir. 4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir. 5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.		
Çalışmada kullanacağım veriler: (X) Kamusal erişime açık (buraya yazınız): Çalışmada gerçek veri olarak TIMSS 2019 verileri kullanılmıştır. Araştırmada kullanılan veriler kamusal erişime açık olup TIMSS & PIRLS Uluslararası Eğitim Merkezi'nin resmî Web sayfası üzerinden ulaşılmıştır. Veriler TIMSS 2019 uygulamasına ait <a href="https://timss2019.org/international-database/">https://timss2019.org/international-database/</a> adresinden indirilmiştir. ( ) Özel izin ve onaya tabi (buraya yazınız): ..... (X) Üretilmiş veri (buraya yazınız): Çalışmada araştırmacı tarafından R yazılımı kullanılarak üretilen simülasyon verisi kullanılmıştır. ( ) Diğer (buraya yazınız): .....		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
<b>Merve AYVALLI KARAGÖZ</b> <i>(Araştırmacı Adı Soyadı, İmzası)</i>		
<b>Araştırmacı Bilgileri</b>		
<b>Adı Soyadı</b>	<b>Merve AYVALLI KARAGÖZ</b>	
<b>Öğrenci ise No</b>	N17147258	
<b>Ana Bilim Dalı</b>	Eğitim Bilimleri Ana Bilim Dalı	
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme	
<b>Statüsü</b>	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
<b>Danışman Görüşü ve Onayı*</b>		
Çalışmada simülasyon veri ve TIMSS 2019 verileri kullanılmış olup, kişiler üzerinde bir uygulama yapılmamıştır		
<b>Prof. Dr. Hülya KELECİOĞLU</b> <i>(İmza)</i> <i>(Danışmanın Ünvanı, Adı ve Soyadı)</i>		
*Tez ve tezden üretilen yayınlarda gerekli		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: <a href="http://ebe.hacettepe.edu.tr/">http://ebe.hacettepe.edu.tr/</a> e-Posta: <a href="mailto:ebe@hacettepe.edu.tr">ebe@hacettepe.edu.tr</a>		

**EK-D: Etik Beyanı**

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- \* tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- \* görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- \* başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- \* atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- \* kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- \* bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

17/01/2023

(İmza)

Merve AYVALLI KARAGÖZ

**EK-E: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu**

13/02/2023

HACETTEPE ÜNİVERSİTESİ  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Bayesci Madde Tepki Kuramı Modellerinde Ölçme Değişmezliğinin İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
13/02/2023	62	92557	17/01 /2023	%3	2013045163

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

**Ad Soyadı:** Merve AYVALLI KARAGÖZ

**Öğrenci No.:** N17147258

**Ana Bilim Dalı:** Eğitim Bilimleri Ana Bilim Dalı

İmza

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

**DANIŞMAN ONAYI**

UYGUNDUR.

Prof. Dr. Hülya KELECİOĞLU

## EK-F: Thesis/Dissertation Originality Report

13/02/2023

HACETTEPE UNIVERSITY  
Graduate School of Educational Sciences  
To The Department of Educational Sciences

Thesis Title: Examining Measurement Invariance In Bayesian Item Response Theory Models

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
13/02/2023	62	92557	17/01/2023	%3	2013045163

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

**Name Lastname:** Merve AYVALLI KARAGÖZ

**Student No.:** N17147258

**Department:** Educational Sciences

**Program:** Educational Measurement and Evaluation

**Status:**  Masters  Ph.D.  Integrated Ph.D.

Signature

### ADVISOR APPROVAL

APPROVED  
Prof. Dr. Hülya KELECİOĞLU

## EK-G: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. <sup>(1)</sup>
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir. <sup>(3)</sup>

..... / ..... / .....

(imza)

Merve AYVALLI KARAGÖZ

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6.1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezimin erişime açılmasının ertelenmesine karar verebilir.
  - (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3 şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezimin erişime açılması engellenebilir.
  - (3) Madde 7.1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir\*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
- Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

\*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir