

BÖLÜM 2

TIMSS 2019 MATEMATİK ÖĞRENMEYİ SEVME ÖLÇEĞİ ÖLÇME DEĞİŞMEZLİĞİNİN BAYESİAN YÖNTEMİ İLE İNCELENMESİ¹

Abdullah YARADILMIŞ², Ragıp TERZİ³

1. GİRİŞ

Günümüzde, okul öncesi dönemden başlayarak eğitimin tüm kademelerinde hedeflere ulaşıp ulaşılmadığını belirlemek amacıyla ölçme

¹Bu çalışma, 29 Ekim – 1 Kasım 2025 tarihinde Antalya’da XII. Uluslararası Stratejik Araştırmalar Kongresi’nde online sözlü bildiri olarak sunulmuş ve birinci yazarın ikinci yazar danışmanlığında hazırladığı tezin bir kısmından üretilmiştir.

² Öğretmen, Bitlis Güroymak Ortaokulu, Bitlis, Türkiye. E-posta: ayaradilms@gmail.com | ORCID: <https://orcid.org/0009-0007-1433-6911>.

³ Doç. Dr., Eğitim Bilimleri Bölümü, Eğitim Fakültesi, Harran Üniversitesi, Şanlıurfa, Türkiye. E-posta: terziragip@harran.edu.tr | ORCID: <https://orcid.org/0000-0003-3976-5054>.



Lisans: Creative Commons Atıf–Ticari Olmayan–Türetilemez 4.0 Uluslararası / (**License:** *Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International*) (CC BY-NC-ND 4.0).

Açık Erişim / Open Access.

DOI (kitap): <https://doi.org/10.63127/tuaz.nd.35>

Tavsiye Edilen Atıf / Recommended Citation:

Yaradılmış, A., ve Terzi, R. (2025). TIMSS 2019 matematik öğrenme keyfi ölçeğinin ölçüm değişkenliğinin Bayes yöntemi kullanılarak incelenmesi. In E. H. Atılmış ve N. Yıldırım (Eds.), *Veriden insana eğitim yolculuğu: Ölçme, çevre ve değerler* (s. 10-25). Turkuaz Yayınları.

<https://doi.org/10.63127/tuaz.nd.35.c140>

ve değerlendirme çalışmaları yürütülmektedir. Ulusal düzeydeki bu çalışmalar, bireylerin sınıflandırılması, karşılaştırılması ve istihdam süreçlerinde kullanılmasının yanı sıra, onların eğitim yolculuğundaki gelişimlerini izlemeye ve desteklemeye de katkı sağlamaktadır. Uluslararası ölçme ve değerlendirme uygulamaları ise ülkelerin birbirlerinin eğitim sistemlerinden örnek almasına, kendi yapısına uyarlamasına ve başarı düzeylerini karşılaştırmasına olanak tanımaktadır. Bu bağlamda, uluslararası düzeyde en yaygın kullanılan ölçme ve değerlendirme uygulamalarından biri Uluslararası Matematik ve Fen Eğilimleri Araştırması'dır (TIMSS).

TIMSS, öğrenci başarı testlerinin yanı sıra uyguladığı ölçekler aracılığıyla, öğrencilerin derslere karşı tutumlarını, kendilerine güven düzeylerini ve öğrenmeye yönelik eğilimlerini içeren duyuşsal becerilere ilişkin veriler toplamaktadır. Bu ölçekler aracılığıyla toplanan verilerin karşılaştırılmasında sıkça karşılaşılan metodolojik sorunların giderilmesine yönelik artan bir gereksinim bulunmaktadır. Uluslararası karşılaştırmalı çalışmaların metodolojik sorunlarından biri, ölçme değişmezliğinin sağlanmasıdır. Ölçme değişmezliği, bir yapıyı grup üyeliğinden bağımsız olarak aynı şekilde ölçtüğünde geçerlidir ve anlamlı çapraz grup karşılaştırmaları için bir ön koşul niteliği taşımaktadır. Ön koşulun sağlandığı ampirik olarak test etmek yerine varsaymak, metodolojik açıdan riskli olup hatalı sonuçlara yol açabilir. Bu nedenle, yapılan çalışmaların ölçümlerinde ölçüm değişmezliğinin ampirik bir testi gereklidir.

1.1. Araştırmanın Amacı ve Önemi

Bu çalışmanın amacı, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan Matematik Öğrenmeyi Sevme Ölçeği'nin (MÖSÖ) örtük yapısını, belirlenen ülkelere (Kuveyt, Portekiz, Suudi Arabistan, Singapur, Türkiye ve Tayvan) göre Bayesian Ölçme Değişmezliği (BÖD) yöntemiyle incelemektir.

Eğitimde ve psikoloji alanında ölçme ve değerlendirmeye konu olan değişkenler (başarı, ilgi, tutum gibi) doğrudan gözlemlenemeyen örtük yapılar şeklinde karşımıza çıkar. Bu nedenle, bireylerin sahip olduğu diğer özelliklerden bağımsız olarak bu değişkenleri hassas ve doğru bir biçimde ölçmek oldukça zordur. Özellikle geniş ölçekli uluslararası testlerin (TIMSS

ve PISA gibi) yaygınlaşmasıyla birlikte, kullanılan ölçme araçlarının farklı gruplarda aynı yapıyı temsil edip etmediğini belirlemek veya belirli bir grubun lehine sistematik farklılıklar gösterip göstermediğini incelemek için birçok yeni yöntem ve teknik geliştirilmiştir. Alan yazına bakıldığında, araştırmaların genellikle ölçme değişmezliğinin sağlanıp sağlanmadığına odaklandığı görülmektedir. Bu çalışmada, belirsizliklerin yüksek olduğu durumlarda güçlü tahminler yapabilme ve önsel bilgileri analize katabilme avantajı nedeniyle Bayesian yöntemi tercih edilmiştir. Bu tercih ile alan yazına metodolojik açıdan katkı sağlanması beklenmektedir.

2. KURAMSAL ÇERÇEVE

2.1. Ölçme Değişmezliği

Ölçme ve değerlendirme çalışmalarının kökleri antik dönemlere kadar uzansa da (Turgut ve Baykul, 2012), ölçme değişmezliği kavramı akademik yazına ancak 20. yüzyılın ikinci yarısında girmiştir (Meredith, 1964; Struening ve Cohen, 1963). Son elli yılda ölçme değişmezliği değerlendirme yöntemleri önemli gelişmeler katetmiştir (Van de Schoot vd., 2012) ve bu kavramın gelişimine Karl Jöreskog, Roger Millsap, Michael Meredith, Gregor Muthén ve Peter Bentler gibi bilim insanları büyük katkıda bulunmuştur. Faktör yapılarının değişmezliğini detaylı bir şekilde elen alan ilk bilim insanı Jöreskog (1971) olmuştur. Ölçme değişmezliği kavramı ise Byrne, Shavelson ve Muthén (1989) tarafından tanıtılmış ve daha sonra test edilmiştir. Son yıllarda ölçme değişmezliği testlerinde kullanılan istatistiksel tekniklerin daha erişilebilir hale gelmesi, araştırmacılar için bu testleri uygulamayı kolaylaştırmış ve yaygınlaştırmıştır (Putnick ve Bornstein, 2016).

Ölçme değişmezliği, psikolojik bir ölçme aracının güvenilirlik ve yapı geçerliliği gibi psikometrik özelliklerinin farklı gruplarda biçimsel olarak eşitliğini değerlendirme süreci olarak tanımlanmaktadır (Herdman vd., 1998). Widaman ve Reice (1997), bu kavramı örtük değişkenlerle gözlenen değişkenler arasındaki ilişkinin gruplar arasında tutarlılığını ifade eden bir yapı olarak açıklamıştır. Ayrıca ölçme değişmezliği ile aynı zamanda ölçme araçlarının önemli bir özelliği olan yapı geçerliliği test eder ve her bir gruptan elde edilen faktör yapısı ile uyum göstermelidir (Başusta ve Gelbal, 2015).

Bir yapının ya da bir ölçme modelinin tüm gruplar için aynı yapıya sahip olması demek söz konusu ölçek maddelerinin faktör yüklerinin, madde sabitlerinin ve hata varyanslarının aynı olması demektir (Byrne, 2013).

2.2. Bayesian Teoremi ve İstatistiği

Bayes teoremi, adını İngiliz matematikçi ve istatistikçi olan Thomas Bayes'ten (1702-1761) alır ve ilk kez, ölümünden sonra arkadaşı Richard Price'ın 1763 yılında yayınladığı Şans Öğretisinde Bir Problemin Çözümüne Yönelik Bir Makale (An Essay Towards Solving a Problem in the Doctrine of Chances) adlı çalışmayla ortaya konmuştur (Stigler, 1983). Bu teorem, koşullu ve genel olasılıklar arasındaki ilişkiyi ele alarak, evrensel doğrular ve gözlemler aracılığıyla çeşitli durumları modellemeyi amaçlar (Çelebi, 2019). Bayesian yaklaşımın temelinde ise elde edilen yeni bilgiler doğrultusunda olasılıkların yeniden hesaplanarak kararların güncellenmesi; yani ilk tahminin ek bilgiler ışığında yeniden değerlendirilmesi gerektiği düşüncesi yer alır (Mlodinov, 2009). Bu tarihsel ve kuramsal çerçevede, Bayesian istatistiğinin temel bileşenlerini anlamak için gerekli zemini hazırlamaktadır.

Bayesian istatistiğinin üç temel bileşeni vardır: Ön dağılım (prior), olabilirlik fonksiyonu (likelihood) ve artık dağılım (posterior). Ön dağılım, modeldeki parametrelerle ilgili mevcut bilgi ya da varsayımları ifade eder. Olabilirlik fonksiyonu, gözlemlenen verilerden gelen bilgileri yansıtarak model parametrelerine dair olasılık dağılımı oluşturur. Artık dağılım ise veriler dikkate alındıktan sonra elde edilen güncellenmiş bilgi düzeyini yansıtır. Bu dağılım, ön dağılım ile olabilirlik fonksiyonu arasında bir denge kurar. Ön dağılım ne kadar bilgilendirici olursa, artık dağılım da o kadar ona benzeyecektir.

2.3. Bayesian Ölçme Değişmezliği ve Model Uyum İndeksleri

Ölçme değişmezliği, Bayesain yöntemi ile aşamalı ve hiyerarşik bir yapıda ele alınmaktadır. Her aşamada belirlenen hipotezler, model uyum katsayıları aracılığıyla test edilmektedir (Başusta ve Gelbal, 2015). Test sürecinin başlayabilmesi için öncelikle, kullanılan ölçme aracının faktör

yapısına dayalı temel modelinin karşılaştırılan grupların her birinde araştırma verisiyle yeterli düzeyde uyum gösterdiğinin kanıtlanması gerekir. Daha sonra her bir değişmezlik düzeyindeki model uyumunun, daha az kısıtlama içeren önceki model uyumu karşılaştırılması yoluyla belirlenir (Chen, 2007). Bu çalışmamızda, ölçme değişmezliği, hiyerarşik olarak üç aşamada test edilmiştir. Bu aşamalar; yapısal (configural) değişmezlik, metrik (metric) değişmezlik ve skalar (scalar) değişmezliktir. Belirlenen bu değişmezlik düzeylerinin test edilmesinin ardından, model uyumu değerlendirilmektedir.

BÖD modelinin uyumu, Bayesian Yapısal Eşitlik Modellemesi (BSEM) analitik çerçevesi kapsamında mevcut olan çeşitli model değerlendirme stratejileriyle incelenebilir. Model uyumu, sonsal tahmin p-değeri (Posterior Predictive p-value; PPP) ve %95 güven aralığının (Confidence Interval; CI) değerlendirilmesi gibi yöntemlerle analiz edilmektedir (Gelman vd., 2004). Araştırmacılar, bu uyum göstergelerine ek olarak model karşılaştırmalarında Sapma Bilgisi Kriteri'nden (Deviation Information Criterion; DIC) (Spiegelhalter vd., 2002) faydalanmaktadır. Genel olarak, ölçüm modelinin parametrelerine (faktör yükleri, madde sabitleri ve hata varyansları) eşitlik kısıtlaması getirerek model uyumunu değerlendirmeyi amaçlanmaktadır (Şen, 2023).

3.YÖNTEM

3.1. Araştırmanın Modeli

Bu araştırma, TIMSS 2019 uygulamasına katılan sekizinci sınıf öğrencilerinden MÖSÖ'yle elde edilen gerçek veriler üzerinden, ölçme değişmezliğini incelemeyi amaçlamaktadır. Bu veriler ile herhangi bir müdahale olmaksızın farklı tekniklerle değerlendirilmesi ve örneklemelerden elde edilen bilgilerle evren hakkında çıkarımlar yapılması nedeniyle, çalışma betimsel araştırma niteliği taşımaktadır (Fraenkel ve Wallen, 2006).

3.2. Evren ve Örneklem

Araştırmanın örneklemini, TIMSS 2019 uygulamasında sekizinci sınıf matematik başarı ortalamaları esas alınarak belirlenmiştir. Bu kapsamda; yüksek başarı düzeyine sahip ülkeler Tayvan (612 puan) ve Singapur (616

puan), orta düzeyde başarı gösteren ülkeler Türkiye (496 puan) ve Portekiz (500) ile düşük başarı düzeyine sahip ülkeler Suudi Arabistan (394) ve Kuveyt (403) olmak üzere toplam altı ülke örnekleme dahil edilmiştir. Araştırmanın evrenini ise sekizinci sınıf düzeyinde MÖSÖ'nün uygulandığı 39 katılımcı ülkeden oluşturmaktadır. Araştırmanın örneklemini oluşturan öğrencilerin ülkelere göre dağılımına ilişkin bilgiler Tablo 1'de sunulmuştur.

Tablo 1. Araştırma Örneklemini Oluşturan Öğrencilerin Ülkelere Göre Dağılımı

Ülkeler	N	%
Kuveyt	4574	16,6
Portekiz	3377	12,3
Suudi Arabistan	5680	20,7
Singapur	4853	17,7
Türkiye	4077	14,8
Tayvan (Çin)	4915	17,9
Toplam	27476	100

Tablo 1 incelendiğinde, altı ülkeden toplam 27.476 öğrencinin çalışmaya dahil edildiği görülmektedir. Öğrenci dağılımı incelendiğinde, en fazla katılımcıya sahip ülkenin 5.680 öğrenci (%20,7) ile Suudi Arabistan olduğu, en az katılımcıya sahip ülkenin ise 3.377 öğrenci (%12,3) ile Portekiz olduğu belirlenmiştir. Diğer ülkelerin katılımcı sayıları ise şu şekildedir: Tayvan 4.915 öğrenci (%17,9), Singapur 4853 öğrenci (%17,7), Kuveyt 4.574 öğrenci (%16,6) ve Türkiye 4.077 öğrenci (%14,8). Bu dağılım, ülkelerin örneklem büyüklüklerinin birbirine yakın olduğunu ve analizlerde karşılaştırılabilirlik açısından dengeli bir yapı sunduğunu göstermektedir.

3.3. Verilerin Elde Edilmesi

Bu araştırmada kullanılan veriler, TIMSS 2019 uygulamasında sekizinci sınıf öğrencilerinin MÖSÖ'ye verdikleri yanıtları içermektedir. Veriler, Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'nun (IEA) resmi internet sitesinden indirilmiştir. Aşağıda, araştırmada kullanılan MÖSÖ'ye ait maddelere Tablo 2'de yer verilmiştir.

Tablo 2. Matematik Öğrenmeyi Sevme Ölçeği Maddeleri

Madde Kodu	Madde İfadesi
BSBM16A	Matematik öğrenmekten hoşlanırım.
BSBM16B*	Keşke matematik çalışmak zorunda olmasam.
BSBM16C*	Matematik sıkıcıdır.
BSBM16D	Matematikte birçok ilginç şey öğrenirim.
BSBM16E	Matematiği severim.
BSBM16F	Sayıları içeren okul çalışmalarını severim.
BSBM16G	Matematik problemlerini çözmeyi severim.
BSBM16H	Matematik derslerini sabırsızlıkla beklerim.
BSBM16I	Matematik en sevdiğim derslerden biridir.

Not: *Ters kodlanmış maddeler

Öğrencilerin matematik öğrenmeye yönelik tutumlarını ölçmeyi amaçlayan matematik öğrenmeyi sevme ölçeği, 7'si olumlu ve 2'si olumsuz olmak üzere toplam 9 maddeden oluşmaktadır. Ölçek, "1=Kesinlikle Katılıyorum", "2=Biraz Katılıyorum", "3=Biraz Katılmıyorum" ve "4=Kesinlikle Katılmıyorum" şeklinde dört seçenekli Likert tipi yanıt formatıyla uygulanmaktadır. MÖSÖ'ye ait her madde için ortalama, standart sapma, çarpıklık ve basıklık gibi betimsel istatistik verileri bulgular bölümünde yer verilmiştir.

3.4. Verilerin Analizi

Veri analizi süreci aşamalı olarak yürütülmüştür. İlk aşamada, analiz öncesi sayıltılar kapsamında kayıp veriler, uç değerler, normallik ve çoklu bağlantılılık incelenmiştir. Ardından, TIMSS 2019 döngüsünde 8. sınıf öğrencilerinin katıldığı MÖSÖ ile oluşturulan model DFA ile test edilerek doğrulanmıştır. Son aşamada, BÖD yöntemiyle ölçme değişmezliği incelenmiştir. Araştırmada veri analizleri SPSS 26.0 (IBM Corp, 2013) ve Mplus 8.5 (Muthén ve Muthén, 2012) yazılımlarıyla gerçekleştirilmiştir.

MÖSÖ'ye ait tüm maddelere hiçbir veri girişi yapılmadığı tespit edilen altı ülkeye ait toplam 96 veri, tüm maddelere yalnızca kayıp veri kodu (9) girildiği belirlenen altı ülkeye ait toplam 168 veri ve 2. ile 3. maddelerin ters kodlanmış olmasına rağmen tüm maddelere aynı şekilde işaretleme yapılan, rastgele doldurulduğu düşünülen altı ülkeye ait toplam 256 veri silinmiştir; bu işlemler sonrasında geriye kalan kayıp verilere Beklenti Maksimizasyon (EM) yöntemi ile değer ataması gerçekleştirilmiştir. Veri setlerindeki tek değişkenli uç değerlerin belirlenmesi için puanlar z standart puanlarına dönüştürülmüş ve -3 ile +3 aralığı dışındaki puanlar uç değer olarak

değerlendirilirken (Kline, 2016), çok değişkenli uç değerlerin belirlenmesinde Mahalanobis uzaklıkları hesaplanmıştır. Normal dağılımın değerlendirilmesinde istatistiksel yöntem tercih edilmiş ve değişkenlerin çarpıklık ile basıklık değerleri dikkate alınmış, George ve Mallery (2010) ile Tabachnick ve Fidell'in (2007) belirttiği üzere katsayıların -1,5 ile +1,5 arasında olması durumunda veri setinin normal dağılım gösterdiği kabul edilerek ilgili sınır değerler göz önünde bulundurulmuştur. Son olarak, çoklu bağlantı probleminin incelenmesinde her bir ölçekteki maddeler için Varyans Artış Faktörü (VIF), Tolerans Değeri (TD) ve Koşul İndeksleri (CI) değerlendirilmiş, VIF değerinin 10'dan küçük, CI değerinin 30'dan küçük ve tolerans değerinin 0,10'dan büyük olması gerektiği dikkate alınmıştır (Tabachnick ve Fidell, 2007). Ölçme değişmezliği çalışmaları her grup için ayrı ayrı DFA yapılarak başlatılmış, çok değişkenli normallik sağlanamadığı için DFA'da parametre tahmini yöntemi olarak güçlü en çok olabilirlik (Robust Maximum Likelihood – MLR) tercih edilmiştir (Brown, 2015). DFA değerlendirilmelerinde önerilen uyum indeksleri ve uyum ölçütleri Tablo 3'te verilmiştir.

Tablo 3. Önerilen Uyum İndeksleri ve Uyum Ölçütleri

Uyum İndeksi	İyi Uyum Ölçütü	Kabul Edilebilir Uyum Ölçütü
Ki kare	$p > 0,05$	Manidar olmayan değerler
RMSEA	$0 \leq RMSEA \leq 0,05$	$0,05 \leq RMSEA \leq 0,08$
CFI	$0,95 \leq CFI \leq 1$	$0,90 \leq CFI \leq 0,90$
TLI	$0,95 \leq TLI \leq 1$	$0,90 \leq TLI \leq 0,95$
SRMR	$0 \leq SRMR \leq 0,08$	$0,08 \leq TLI \leq 1$

Kaynak: Hu ve Bentler, 1999; Kline, 2016; Şen, 2023

Muthén ve Asparouhov'a (2012) göre, Bayesian bir modelin veriye iyi uyduğunu söyleyebilmek için üç temel koşul dikkate alınmalıdır. Bunlardan ilki, sonsal tahmin p -değeri (PPP) anlamlı olmamalıdır; yani aşırı uç değerlerde yer almamalı ve orta değerlere yakın olmalıdır. İkincisi, ki kare farkı gibi uyum ölçütlerinin güven aralığı (CI) sıfır değerini içermelidir; bu durum, gözlenen ve simüle edilen değerler arasında anlamlı bir fark olmadığını ve dolayısıyla modelin veriye uyumlu olduğunu gösterir. Üçüncü olarak, model değerlendirmesinde, daha düşük Sapma Bilgisi Kriteri (DIC) değeri, görece olarak daha iyi uyum sağlayan ve daha az karmaşık bir modeli işaret eder.

Bayesian yöntemi ile yapılan tüm analizler, Mplus yazılımında mixture modelleme çerçevesinde gerçekleştirilmiş ve ülke değişkeni bilinen sınıf

olarak tanımlanarak her ülke önceden belirlenmiş altı sınıftan birine atanmıştır. Modelde faktör varyansı 1'e, ortalaması ise 0'a sabitlenmiştir. Bayesci tahmin yöntemiyle dört Markov zinciri üzerinden 150.000 iterasyon yürütülmüş ve hesaplama sürecinde dört işlemci çekirdeği paralel olarak kullanılmıştır. Bayesian analizde modelin esnekliği ile önsel bilginin gücünü dengelemek amacıyla Inverse Wishart dağılımı kullanılmış; bu çerçevede, gözlenen değişkenlerin hata (artık) varyansları için serbestlik derecesi (1) ve ölçek matrisi (15) kullanılmıştır ve gözlenen değişkenler arasındaki kovaryanslar için ise ortalama (0) ve varyans (15) olarak tanımlanmıştır.

4. BULGULAR

Bu bölümde, ilk aşamada, analiz sürecine ilişkin varsayımlar ile ölçek maddelerine ait betimsel istatistiksel bilgiler sunulmuştur. Devamında, çalışmaya dahil edilen altı ülkeye yönelik DFA sonuçları paylaşılmıştır. Son olarak, BÖD yöntemi kullanılarak gerçekleştirilen ölçme değişmezliği analizlerine ilişkin bulgulara yer verilmiştir.

4.1. Analiz Varsayımları

Veri seti üzerinde gerçekleştirilen inceleme sonucunda, tek değişkenli uç değerlerin belirlenmesinde z standart puanları hesaplanmış ve tüm ülkelerdeki değerlerin -2,04 ile +2,04 aralığında yer aldığı görülerek uç değer bulunmadığı sonucuna ulaşılmıştır (Kline, 2016). Çok değişkenli uç değerler için Mahalanobis uzaklıkları hesaplanmış, ancak yüksek sayıda uç değer tespiti model uyumunu olumsuz etkileyebileceği gerekçesiyle bu veriler silinmemiştir. Tek değişkenli normallik analizinde çarpıklık ve basıklık değerleri sırasıyla -0,48 ile +1,24 ve -1,53 ile -0,24 arasında değişmiş, bu durum tüm ülkelerde normalliğin sağlandığını göstermiştir (George ve Mallery, 2010; Tabachnick ve Fidell, 2007); buna karşın Mardia Çok Değişkenli Normallik testi sonuçları, hiçbir ülkenin çok değişkenli normalliği sağlamadığını ortaya koymuştur. Son olarak, çoklu bağlantı probleminin incelenmesinde VIF, TD, CI ve maddeler arası korelasyon değerleri analiz edilmiş; VIF(1,304–5,175), TD(0,178–0,767), CI(15,2–18,8) ve korelasyon (0,270–0,856) aralıklarında elde edilen sonuçlar, çoklu

bağlantı probleminin bulunmadığını göstermiştir (Tabachnick ve Fidell, 2007).

Ölçek maddelerine ilişkin betimsel istatistikte; maddelerin ortalama değerleri 1,69 ile 2,85 arasında değiştiği, maddelerin standart sapma değerlerinin ise 0,82 ile 1,19 arasında değiştiği görülmektedir. Çarpıklık ve basıklık değerleri yukarıda belirtilmiştir.

4.3. Doğrulayıcı Faktör Analizi Sonuçları

Analiz öncesinde varsayımlar değerlendirilmiş ve uygunlukları doğrulandıktan sonra, çalışmaya dahil edilen her ülkeye ait veriler kullanılarak DFA uygulanmış; bu kapsamda oluşturulan ölçüm modelinin geçerliliğini değerlendirmek amacıyla model uyum iyiliği indeksleri hesaplanarak Tablo 4'te verilmiştir.

Tablo 4. Ölçüm Modeline İlişkin Ülkelerin DFA Model Uyum İyiliği İndeksleri

Ülkeler	RMSEA (CI)	SRMR	CFI	TLI	Ki Kare	p
Kuveyt	,064(.059-.069)	,027	,973	,964	509,35	,00
Portekiz	,085(.080-.091)	,024	,964	,952	678,19	,00
Suudi A.	,079(.075-.083)	,033	,958	,945	965,35	,00
Singapur	,075(.071-.080)	,025	,967	,956	764,05	,00
Türkiye	,086(.081-.091)	,033	,948	,931	819,09	,00
Tayvan	,089(.084-.093)	,032	,954	,939	1062,73	,00

Tablo 4 incelendiğinde, tüm ülkelerde elde edilen ki-kare değerlerinin istatistiksel olarak anlamlı olduğu görülmektedir ($p < .05$). Bu durum, modellerin verilerle iyi uyum göstermediğini düşündürse de ki-kare testinin örneklem büyüklüğüne duyarlılığı göz önünde bulundurulduğunda, modelin genel uyumunu değerlendirmek için diğer uyum indekslerine odaklanmak daha anlamlı olacaktır (Şen, 2023). Nitekim RMSEA, CFI, TLI ve SRMR gibi alternatif uyum indekslerinin tüm ülkelerde kabul edilen eşik değerler içerisinde ya da bu değerlere oldukça yakın olduğu görülmektedir. Bu bulgular, ki-kare testi dışında kalan uyum ölçütlerinin modelin veriye yeterli düzeyde uyum sağladığını göstermektedir. Dolayısıyla, söz konusu modele ait parametrelerin güvenle yorumlanabileceği söylenebilir. DFA sonucunda ölçüm modeli tüm ülkelerde doğrulandıktan sonra, ölçme değişmezliğini belirlemek amacıyla BÖD'ye geçilmiştir.

Araştırmanın amacı, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BÖD yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini incelemektir. Bu kapsamda, BÖD yöntemi ile

ölçme değişmezliği aşamaları (yapısal, metrik ve skalar) test edilmiştir. Ölçüm modelinin her bir aşamanın model uyum indeksleri Tablo 5'te sunulmuştur.

Tablo 5. Ölçüm Modeline İlişkin Ülkelerin BÖD Model Uyum İyiliği İndeksleri

Model	PPP	CI	DIC
Yapısal	,503	(-70,80-70,18)	518820
Metrik	,310	(-53,99-90,162)	518842
Skalar	,000	(4722,60-4863,699)	523569

Tablo 5 incelendiğinde, yapısal değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin PPP değerinin ,503 olması ve güven aralığını sıfırı içermesi, model uyumu sağladığını göstermektedir. Yapısal değişmezliğin sağlandığına dair kanıt elde edilmesiyle birlikte, analiz süreci metrik değişmezlik aşamasına geçilerek devam edilmiştir.

Metrik değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin PPP değerinin ,310 olması ve güven aralığının sıfırı içermesi, model uyumun sağladığı anlamına gelmektedir. Metrik değişmezliğin sağlandığına dair elde edilen kanıt doğrultusunda analiz süreci skalar değişmezlik aşamasına geçilerek devam edilmiştir.

Skalar değişmezlik modeli aşamasında elde edilen model uyum iyiliği indekslerinin PPP değerinin ,000 olması ve güven aralığının sıfırı içermemesi, modelin bu düzeyde uyum sağlamadığını göstermektedir. Sonuç olarak, araştırmaya dahil edilen ülkelerle oluşturulan ölçme modeli yapısal ve metrik düzeyde ölçme değişmezliği sağlarken, skalar düzeyde ölçme değişmezliği sağlamamaktadır.

5. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde, öncelikle araştırma sonuçları sunulmuş, ardından bu sonuçlar ilgili literatürle karşılaştırılarak tartışılmıştır. Son olarak, çalışmanın bulguları doğrultusunda çeşitli önerilere yer verilmiştir.

Sonuç ve Tartışma

Bu çalışmada, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BÖD yöntemi ile belirlenen ülkeler arasında ölçme değişmezliği incelenmiştir. İlk aşamada, ilgili ülkelerin veri setleri üzerinde analiz öncesi varsayımlar değerlendirilmiş ve betimsel istatistikler

hesaplanmıştır. Analiz sonuçları, çalışmaya dahil edilen tüm ülkelerde tek değişkenli normallik koşulunun sağlandığı, çok değişkenli normalliğin ise hiçbir ülkede sağlanamamıştır. Çalışmada, ölçeğin tek boyutlu ve dokuz maddeden oluşan ölçme modelinin geçerliliğinin test etmek amacıyla DFA uygulanmıştır. DFA sonuçları, elde edilen modelin veri uyum iyiliği indekslerinin tüm ülkelerde kabul edilen eşik değerler içerisinde yer aldığını göstermektedir. Bu bulgu, ölçme modelinin incelenen tüm ülkelerde doğrulandığını ortaya koymaktadır. Ölçeğin faktör yapısının belirlenmesinin ardından, ölçme değişmezliği analizleri gerçekleştirilmiştir.

Bu çalışmanın amacı, TIMSS 2019 sekizinci sınıf öğrenci anketlerinde yer alan MÖSÖ'nün örtük yapısı, BÖD yöntemine göre belirlenen ülkeler arasında ölçme değişmezliğini incelenmesidir. BÖD yöntemi ile yapılan analizler sonucunda, yapısal ve metrik değişmezlik sağlandığı; ancak skalar değişmezliğin sağlanamadığı sonucuna ulaşılmıştır. Bu elde edilen sonuç ile ülkeler arasında gösterge sabitlerinin eşit olmadığı ve örtük değişkenin tüm ülkelerde aynı şekilde yorumlanmadığını göstermektedir. Dolayısıyla, ülkeler arası karşılaştırmalarda gözlenen değişken ortalamalarındaki farklılıkların, ölçeğin özelliklerinden kaynaklanabileceği göz önünde bulundurulmalıdır.

Literatür incelendiğinde, TIMSS verileri kullanılarak cinsiyet, kültürlerarası, bölgelerarası ve dil gibi değişkenler üzerinden ölçme değişmezliği araştırmalarının gerçekleştirildiği görülmektedir. Bazı araştırmalarda skalar düzeyde ölçme değişmezliği sağlanmıştır (Atılğan, 2022; Polat, 2019; Yiğiter, 2023a), bazı araştırmalarda ise bu düzeyde ölçme değişmezliği sağlanamamıştır (Dokur, 2023; Gucciardi vd., 2016; Koçak, 2024; Ölçüoğlu ve Çetin, 2016; Sedding ve Leitgöb, 2018; Stevanovic vd., 2021; Yiğiter, 2023b). Bayesian yöntemi ile yapılan bu çalışmada, yapısal ve metrik düzeyde ölçme değişmezliği sağlanmış, ancak skalar düzeyde sağlanamamıştır. Elde edilen bulgular literatürdeki sonuçlarla tutarlılık göstermektedir.

Öneriler

Araştırma süreci ve elde edilen sonuçlar değerlendirildiğinde, uygulayıcılara ve araştırmacılara yönelik öneriler yer almaktadır.

1. Bu çalışmada, TIMSS 2019 uygulamasına ait sekizinci sınıf öğrenci anket verileri kullanılmıştır. Bu çalışmada ele alınan problem, farklı eğitim kademeleri, yıllar ve sınav türlerine ait veriler üzerinde de yürütülebilir.

3. Bu çalışmada, ölçme değişmezliği ülkeler arası (kültürlerarası) değişkenine göre incelenmiştir. Gelecek çalışmalarda cinsiyet, bölgesel farklılıklar, sosyoekonomik düzey ve dil gibi farklı değişkenler dikkate alınarak ölçme değişmezliği analizleri genişletilebilir.

4. Araştırma kapsamında, öğrencilerin matematiği öğrenmeyi sevmeye boyutuna ilişkin ölçme değişmezliği değerlendirilmiştir. Benzer çalışmalar diğer boyutlar için de gerçekleştirilebilir.

5. Ölçme değişmezliği analizleri, BÖD yöntemiyle yürütülmüştür. Benzer çalışmalar farklı analiz yaklaşımları (örtük sınıf analizi, hizalama metodu) kullanarak tekrarlanabilir; ayrıca birden fazla yöntemle yapılan analizler arasında yöntem karşılaştırmaları yapılabilir.

KAYNAKÇA

- Atılğan, M. (2022). *TIMSS 2019 matematik başarısına ilişkin duyuşsal özelliklerin cinsiyete göre ölçme değişmezliğinin incelenmesi* (Yüksek lisans tezi). Ankara Üniversitesi. <https://avesis.ankara.edu.tr/yonetilen-tez/dce954bd-58cf-4bc3-96ad-994820aff421>
- Başusta, N. B. ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90. <https://doi.org/10.21031/hunefd.94663>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications. <https://doi.org/10.1521/978.1462544646>
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press. <https://doi.org/10.4324/9780203774762>
- 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>

- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504. <https://doi.org/10.1080/10705510701301834>
- Çelebi, V. (2019). Bayes teoremi bağlamında olasılıkçı bayes epistemolojisinin kapsamı üzerine bir inceleme. *Felsefe ve Sosyal Bilimler Dergisi*, 28, 319-343. <https://doi.org/10.5281/zenodo.622357>
- Dokur, Y. (2023). *TIMSS 2019 sekizinci sınıf matematiğe yönelik güven ölçeğinin hizalama metoduna göre ölçme değişmezliğinin incelenmesi* (Yüksek lisans tezi). Gazi Üniversitesi.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). McGraw-Hill.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Chapman & Hall/CRC.
- George, D., & Mallery, M. (2010). *SPSS for windows step by step: a simple guide and reference*, 17.0 update (10a ed.). Pearson
- Gucciardi, D. F., Peeling, P., Ducker, K. J., & Dawson, B. (2016). When the going gets tough: Mental toughness and its relationship with behavioural perseverance. *Journal of Science and Medicine in Sport*, 19(1), 81-86. <https://doi.org/10.1016/j.jsams.2014.12.005>
- Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research*, 7(4), 323-335. <https://doi.org/10.1023/A:1024985930536>
- Hu, L., & 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>
- IBM Corp. (2013). *IBM SPSS Statistics for Windows* (Version 26.0). IBM Corp.
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika* 36(4), 409–426. <https://doi.org/10.1007/BF02291366>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. (4th ed.). Guilford Press.
- Koçak, F. (2024). *TIMSS 2019 matematik başarı testi ve matematiğe ilişkin duyuşsal özelliklerin ölçme değişmezliğinin incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29(2), 177–185. <https://doi.org/10.1007/BF02289699>
- Mlodinow, L. (2009). *Ayyaş yürüyüşü*, (M. Kayı, Çev.). April Yayıncılık.

- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335. <https://doi.org/10.1037/a0026802>
- Muthén, L. K., ve Muthén, B. O. (2012). Mplus: Statistical analysis with latent variables; user's guide [version 7]. Muthén et Muthén.
- Ölçüoğlu, R. ve Çetin, S. (2016). TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 202-220.
- Polat, M. (2019). TIMSS-2015 matematik ve fen duyuşsal özellik modellerinin kültürle, cinsiyete ve bölgelere göre ölçme değişmezliğinin incelenmesi (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara. <http://hdl.handle.net/11655/8068>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Seddig, D., & Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Survey Research Methods*, 12(1), 29–41. <https://doi.org/10.18148/srm/2018.v12i1.7210>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., ... Knez, R. (2021). Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries. *Autism Research*, 14(12), 2544-2554. <https://doi.org/10.1002/aur.2586>
- Stigler, S. M. (1983). Who discovered Bayes's theorem? *The American Statistician*, 37(4a), 290-296. <https://doi.org/10.1080/00031305.1983.10483110>
- Struening, E. L., & Cohen, J. (1963). Factorial invariance and other psychometric characteristics of five opinions about mental illness factors. *Educational and Psychological Measurement*, 23(2), 289-298. <https://doi.org/10.1177/001316446302300206>
- Şen, S. (2023). *Mplus ile yapısal eşitlik modellemesi uygulamaları* (2. Basım). Nobel Akademik Yayıncılık.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA* (Vol. 724). Thomson/Brooks/Cole.

- Turgut, M. F., ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme* (4. baskı). Pegem Akademi.
- Van de Schoot R., Lugtig P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- 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>
- Yiğiter, M. S. (2023). Matematik duyuşsal özellik faktörlerinin cinsiyete göre ölçme değişmezliğinin incelenmesi: TIMSS 2019 Türkiye örneği. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi*, 7(4), 859–882. <https://doi.org/10.34056/aujef.1198134>
- Yiğiter, M. S. (2024). Matematik içsel motivasyonunun ülkeler arası ölçümü: On dört ülke arasında MG-CFA ve hizalama yöntemi ile ölçme değişmezliğinin incelenmesi. *Kuramsal Eğitimbilim Dergisi*, 17(1), 1–27. <https://doi.org/10.30831/akukeg.1207350>