



Matematik Duyuşsal Özellik Faktörlerinin Cinsiyete Göre Ölçme Değişmezliğinin İncelenmesi: TIMSS 2019 Türkiye Örneği

Investigation of Measurement Invariance of Mathematics Affective Characteristic Factors According to Gender: TIMSS 2019 Turkey Sample

Mahmut Sami YİĞİTER¹

Makale Türü: Araştırma Makalesi

Başvuru Tarihi: 02.11.2022

Kabul Tarihi: 07.08.2023

Atıf İçin: Yiğiter, M.S. (2023). Matematik duyuşsal özellik faktörlerinin cinsiyete göre ölçme değişmezliğinin incelenmesi: TIMSS 2019 Türkiye örneği. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi (AUJEF)*, 7(4), 859-882.

ÖZ: Uluslararası geniş ölçekli değerlendirmelerin temel hedeflerinden biri göre farklı ülkeler veya altgruplar arasında karşılaştırmalar yaparak eğitim politikaları veya eğitim sistemleri hakkında çıkarımlarda bulunmaktır. Farklı gruplar arasında karşılaştırma yapmanın temel kriterlerinden biri de ölçme değişmezliğinin sağlanmasıdır. Ölçme değişmezliği, ölçülen yapının gruplar arasında psikometrik olarak eşdeğer olduğunu göstermektedir. Ölçme değişmezliği kanıtı sunulmadan yapılan karşılaştırmalardaki farklılıklara dair iddialar güvenilir olamaz. Bu çalışmanın amacı matematik duyuşsal özellikleri ile oluşturulan modelin cinsiyete göre ölçme değişmezliğinin sınanmasıdır. Bu amaçla TIMSS 2019 döngüsünde yer alan matematik öğrenmeyi sevme (MOS), matematik öğretiminin netliği (MON), matematik dersinde disiplinsiz davranış (MDDD), matematikte kendine güven (MKG) ve matematiğe değer verme (MDV) ölçekleri ile Matematik Duyuşsal Özellikleri Modeli oluşturulmuştur. Çalışmanın örneklemini TIMSS 2019 döngüsüne 8. Sınıf düzeyinde Türkiyeden katılan 3658 öğrenci oluşturmaktadır. Araştırmanın ilk bölümünde matematik duyuşsal özellikler modelinin faktör yapısını incelemek için Doğrulayıcı Faktör Analizi (DFA) yapılmıştır. DFA modeli sonuçları model veri uyumunun sağlandığını göstermektedir (RMSEA=0.046, SRMR=0.051, CFI=0.973 ve TLI=0.975). Ölçme değişmezliği analizinde Çok Gruplu DFA (ÇG-DFA) analizi ile aşamalar arasında hiyerarşik olarak test edilmiştir. Bulgular, matematik duyuşsal özellikler modelinin sırasıyla yapısal, metrik, ölçek ve katı değişmezlik aşamalarını karşıladığını göstermektedir. Dolayısıyla matematik duyuşsal özellikler modelinin cinsiyete göre faktör yükleri, varyansları, hata varyansları ve kovaryansları eşdeğer olup gruplar arasında anlamlı karşılaştırmalar yapılabileceği sonucuna ulaşılmıştır. Ölçme değişmezliğinin incelenmesinin ardından modelde yer alan değişkenlerin cinsiyete göre anlamlı farklılıklarını incelemek için t testi analizleri gerçekleştirilmiştir. Sonuçlar, MON ölçeğinde erkekler lehine, MKG ve MDDD ölçeklerinde kızlar lehine anlamlı farklılık olduğuna işaret ederken, MDV ve MOS değişkenlerinde cinsiyete göre anlamlı farklılık bulunmamaktadır.

Anahtar sözcükler: Duyuşsal özellikler, matematik, motivasyon, ölçme değişmezliği, değişmezlik

¹ Öğr. Gör. Dr., Ankara Sosyal Bilimler Üniversitesi, mahmutsamiyigiter@gmail.com, ORCID: 0000-0002-2896-0201

ABSTRACT: One of the main objectives of large-scale assessments is to draw conclusions about education policies or education systems by making comparisons between different countries or subgroups. One of the main criteria for making comparisons between different groups is to satisfy measurement invariance. Measurement invariance indicates that the measured construct is psychometrically equivalent between groups. Claims of differences in comparisons without evidence of measurement invariance can be unreliable. The aim of this study was to test the measurement invariance of the model created with mathematics affective characteristics according to gender. For this purpose, the Mathematics Affective Characteristics Model was created with the scales of Like Learning Mathematics (MOS), Instructional Clarity in Mathematics Lessons (MON), Disorderly Behavior During Mathematics Lessons (MDDD), Students Confident in Mathematics (MKG) and Students Value Mathematics (MDV) in the TIMSS 2019 cycle. The sample of the study consists of 3658 students from Turkey who participated in the TIMSS 2019 cycle at the 8th grade level. In the first part of the study, Confirmatory Factor Analysis (CFA) was conducted to examine the factor structure of the mathematics affective characteristics model. DFA model results show that model data fit is reached (RMSEA=0.046, SRMR=0.051, CFI=0.973 and TLI=0.975). In the measurement invariance analysis, it was tested hierarchically between the stages with Multi-Group CFA (MG-CFA) analysis. The findings show that the mathematics affective characteristics model meets the configural, metric, scalar, and strict invariance stages, respectively. Therefore, the factor loadings, variances, error variances and covariances of the mathematics affective characteristics model were equivalent according to gender, and it was concluded that significant comparisons could be made between the groups. After examining measurement invariance, t-test analyses were conducted to examine the significant differences of the variables in the model according to gender. The results indicate that there is a significant difference in favor of boys in the MON scale, in favor of girls in the MKG and MDDD scales, while there is no significant difference in the MDV and MOS variables according to gender.

Keywords: Affective characteristics, math, motivation, measurement invariance, invariance

1. GİRİŞ

Uluslararası geniş ölçekli değerlendirme (International Large-Scale Assessment-ILSA) programları, 20. yy'ın sonlarından itibaren ülkelerin eğitim sistemleri üzerinde karşılaştırma yapmak ve öğrenci başarılarını belirlemek için pek çok ülkenin katılımıyla gerçekleştirilmeye başlanmıştır (Cardoso, 2020). Uluslararası Öğrenci Değerlendirme Programı (PISA), Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) ve Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) bu amaçla gerçekleştirilen sınavlardandır. PISA, 15 yaşındaki öğrencilerin okuma becerileri, matematik okuryazarlığı ve fen okur yazarlığı becerilerine odaklanan bir değerlendirme programıdır (Schleicher, 2019). PISA, OECD tarafından gerçekleştirilmekte olup üç yılda bir uygulanmaktadır. TIMSS ve PIRLS ise Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA) tarafından yürütülmektedir. PIRLS, 4. Sınıf öğrencilerinin okuma becerilerini belirlemek üzere her beş yılda bir gerçekleştirilmektedir (Mullis ve Martin, 2019). TIMSS ise her dört yılda bir olmak üzere 4. ve 8. sınıf öğrencilerinin matematik ve fen becerileri alanları üzerinden öğrencilerin akademik başarılarını değerlendiren, zaman içerisinde ülkeler arası başarı farklılıklarını ve ülkelerin başarı düzeyini arttırmaya yönelik girişimlerinin sonuçlarını izleyen bir değerlendirme programıdır (Mullis ve Martin, 2017). TIMSS, matematik ve fen becerilerini ölçmenin yanında öğrenciler, öğretmenler ve okullara ait bilişsel ve duyuşsal özellikleri ölçen pek çok ölçek ve anket içermektedir (Yin ve Fishbein, 2019). Öğrencilerin matematik başarısına yönelik duyuşsal özelliklerini ise 1995'ten beri ölçmektedir. TIMSS 2019 öğrenci anketinde matematik duyuşsal özelliklerini ölçen beş farklı ölçek yer almaktadır. Bu ölçekler, içsel motivasyon ile ilgili olan matematik öğrenmeyi sevmeye, öğretimin açıklığını ölçen matematik öğretiminin netliği, ders içerisindeki disiplinsiz davranışları belirleyen matematik dersinde disiplinsiz davranış, matematik özgüvenini ölçen matematikte kendine güven ve öğrencinin matematiğe olan bakışımı değerlendiren matematiğe değer verme ölçeğidir (Mullis, Martin, Foy, Kelly ve Fishbein, 2020).

Duyuşsal özelliklerin matematik başarısı üzerinde anlamlı etkileri vardır (Ma, 1997; Van der Bergh, 2013; Yagan, 2021). Örneğin içsel motivasyonun matematik başarısı üzerinde pozitif yönde anlamlı etkileri olduğunu bildiren pek çok çalışma yer almaktadır (Akben-Selcuk, 2017; Guo vd., 2015; Hooper vd., 2020). Benzer şekilde bazı çalışmalar da özgüven ile matematik başarısının ilişkili olduğunu belirtmektedir (Van der Bergh, 2013; Çiftçi ve Yıldız, 2019). Dolayısıyla bu duyuşsal özellikleri ölçen ölçeklerin geçerli ve güvenilir ölçümler gerçekleştirmeleri, duyuşsal özellikler ile matematik başarısı arasındaki ilişkilerin doğru ve tarafsız biçimde ortaya konulması açısından önemlidir.

ILSA'ların ana hedeflerinden biri farklı ülkeler, eğitim sistemleri, altgruplar ve bireyler arasında karşılaştırmalar yapmaktır (Engel ve Rutkowski, 2021). ILSA'lardan elde edilen verilerin karşılaştırma yapmak için kullanılmasına ilişkin metodolojik tartışmalar giderek artmaktadır (Gustafsson, 2018; Kim ve diğerleri, 2017; Rutkowski ve Svetina, 2014). Farklı gruplar arasında karşılaştırma yapmanın temel ve kritik ölçütlerinden biri ölçme değişmezliğinin kurulmasıdır (Putnick ve Bornstein, 2016). Ölçme değişmezliği, ölçülen yapının gruplar arasında eşdeğer olduğunu gösteren istatistiksel bir özelliktir (Raykov, 2004). Diğer bir deyişle, yapıdan elde edilen varyans, ölçme aracı ile ilgisiz olmalı, sadece ölçeğin uygulanma koşulları arasındaki farklılıkları yansıtmalıdır. Ölçme değişmezliği kanıtı sunulmadan, katılımcıların ölçülen yapıyı anlamadaki kültürel ve çeviri farklılıkları göz önüne alınır, yapılan karşılaştırmalardaki farklılıklara dair iddialar güvenilir hâle gelebilir (Rutkowski ve Svetina, 2014).

1.1. Ölçme Değişmezliği

Ölçme değişmezliği, kullanılan ölçme aracının farklı gruplarda aynı anlama karşılık gelip gelmemesidir. Diğer bir deyişle, aynı ölçek yapılan ölçmede farklı gruplarda ölçekten elde edilen psikometrik değerler farklılaşıyorsa ölçme değişmezliğinin sağlanamadığı söylenebilir. Farklı gruplarda yer alan ve ölçülen özellik (yapı) açısından birbirine eşdeğer olan bireylerin bir testten aynı gözlenen puanı almaları beklenir. Bireyler ölçülen yapı açısından aynı, fakat gözlenen puanları farklıysa ölçeğin (testin) ölçme değişmezliğinin sağlanamadığı söylenebilir (Schmitt ve Kuljanin, 2008).

Her ölçme aracının geliştirilmesinde temel bir ilke olarak uygulandığı her grupta aynı özelliği ölçtüğü varsayılarak geliştirilmektedir. Fakat pratikte uygulandıkları gruplara göre sonuçlar farklılık gösterebilmektedir. Gruplardan elde edilen sonuçlar eşit/eşdeğer psikometrik niteliklere sahip değilse sonuçları gruplara genellemek doğru olmayacaktır (Başusta ve Gelbal, 2015). Dolayısıyla ölçme aracı, her bir alt grupta aynı yapıyı ölçmelidir. Ölçme değişmezliği ile bir ölçme modelinin faktör yükleri, boyutlar arası korelasyonları ve hata varyanslarının her grupta aynı olduğunun gösterilmesi, ölçme aracının farklı gruplarda aynı yapıya sahip olduğunu gösterir (Jöreskog ve Sörbom, 1993). Araştırmacılar, ölçeğin alt gruplarda aynı yapıyı ölçüp ölçmediğine dair kanıt elde ederler (Millsap ve Olivera-Ogilar, 2012; Uyar ve Doğan, 2014). Ölçme değişmezliğinin sağlanamaması ölçme aracı için bir geçerlilik problemidir. Dolayısıyla böyle bir ölçme aracına dayalı grup karşılaştırmalarının sonuçlarına ilişkin yorumlar da hatalı olabilir (Vandenberg ve Lance, 2000). Ölçme değişmezliği olduğunun gösterilmesi aynı zamanda ölçme aracına yönelik bir geçerlik kanıtı sağlayacaktır. He ve diğerleri (2019), TIMSS ve PISA verisi ile kültürler arası karşılaştırılabilirlik üzerine yaptığı araştırmada ölçme değişmezliği incelenmeden yapılan karşılaştırmaların hatalı sonuçlara yol açabileceğini, dolayısıyla ölçme değişmezliğini test etmenin önemini belirtmektedir. Ölçme değişmezliği kanıtlanamazsa gruplar arası karşılaştırmaların sonuçlarını yorumlamak doğru olmaz. Bunun nedeni gruplar arası ortaya çıkan farklılığın gerçek bir yapı farklılığından mı yoksa ölçek maddelerine verilen yanıtlar arasındaki farklılıktan mı kaynaklandığı bilinemez (Horn ve McArdle, 1992). Dolayısıyla iki veya daha fazla gruptan elde edilen ölçmeleri karşılaştırmadan önce ölçme değişmezliğini incelemek önemlidir.

Ölçme değişmezliğini test etmek için literatürde en çok kullanılan yöntemlerden biri Doğrulamalı Faktör Analizi'dir (DFA) (Schmitt ve Kuljanin, 2008; Van De Schoot vd., 2015). Ölçme değişmezliği, Çok Gruplu DFA (ÇG-DFA) yöntemi ile hiyerarşik olan dört aşamada test edilir. Bu aşamalar; yapısal (configural) değişmezlik, metrik (metric) değişmezlik, ölçek (scalar) değişmezliği ve katı (strict) değişmezliktir (Meredith, 1993).

1.1.1. Yapısal (Configural) Değişmezlik

Ölçme değişmezliğinin ilk hiyerarşik aşamasıdır. Bu aşamada grupların aynı faktör yapısına sahip olup olmadıkları test edilir. Bu amaçla, bu aşamada faktörlerin denkliği ve faktör yükleri örüntüsü analiz edilir (Taris, Bok ve Meijer, 1998). Bu aşamada herhangi bir parametre kısıtlaması yapılmaz. Yapısal değişmezlik sağlanıyorsa grupların aynı yapıyı ölçtüğünü ifade edilebilir (Wu, Li ve Zumbo, 2007). Yapısal değişmezlik sağlanamazsa grupların farklı yapıyı ölçtüğü belirtilir ve ölçme değişmezliğinin ileri aşamalarına geçilmez. Yapısal değişmezlik, literatürde şekilsel veya biçimsel değişmezlik olarak da adlandırılmaktadır.

1.1.2. Metrik (Metric) Değişmezlik

Yapısal değişmezliğin sağlandığı gösterildiğinde metrik değişmezlik testi yapılabilir (Milfont ve Fischer, 2010). Metrik değişmezlikte farklı gruplardaki faktör yüklerinin eşitliği test edilir. Diğer bir deyişle, bir gruptan kestirilen faktör yükleri diğer gruba sabitlenir ve modelin uyum indeksleri incelenir. Metrik değişmezlik sağlanırsa faktör yüklerine dayalı gruplar arası yapılacak karşılaştırmalar savunulabilir (Gregorich, 2006). Metrik değişmezlik, zayıf (weak) değişmezlik olarak da bilinmektedir (Meredith, 1993).

1.1.3. Ölçek (Scaler) Değişmezliği

Metrik değişmezliğin sağlanması durumunda ölçek değişmezliği aşamasına geçilir. Bu aşamada hem faktör yüklerinin hem de regresyon sabitlerinin gruplar arasında eşitliği test edilir. Diğer bir deyişle, faktör varyansı ve kovaryanslarının gruplar arası eşdeğerliğinin test edildiği bu aşamada, faktör yüklerinin eşdeğerliği de incelenir. Ölçek değişmezliği sağlanırsa gözlenen değişkenlerin ortalamalarının ve faktör yüklerinin karşılaştırılabileceği anlamına gelir (Gregorich, 2006). Ölçek değişmezliği, güçlü (strong) değişmezlik olarak da bilinmektedir.

1.1.4. Katı (Strict) Değişmezlik

Ölçme değişmezliğinin son adımı olan bu aşamada, önceki aşamalardaki kısıtlamalara ek olarak, hata varyanslarının da gruplar arasında eşitliği test edilir (Vandenberg ve Lance, 2000). Gruplar arasında aynı yapıyı ölçtüğünü iddia eden ölçeklerin katı değişmezliği sağlanması gerekir. Katı değişmezliğin sağlanması ile ölçme değişmezliği tam olarak sağlanmış olacaktır.

1.2. Matematik Duyuşsal Özellik Ölçekleri

Bu bölümde TIMSS 2019 döngüsünde yer alan matematik duyuşsal özellik ölçekleri ile ölçülen yapılar ve ölçekler hakkında bilgiler verilmiştir.

1.2.1. Matematik Öğrenmeyi Sevme (MOS)

İçsel motivasyon, bir aktivite ile uğraşırken alınan keyif ve bu aktiviteyi gerçekleştirme dürtüsü olarak tanımlanmaktadır (OECD, 2013). Öğrencilerin matematiği öğrenmeye istekli olmaları, matematiği ilginç ve eğlenceli bulmalarından kaynaklanmaktadır (Ryan ve Deci, 2009). İçsel motivasyon, enerji kaynağıdır ve davranışa geçmenin ön koşuludur (Yığıter, 2019; Malone ve Lepper, 2021). İçsel motivasyon ile matematik başarısı arasında güçlü bir ilişki olduğunu gösteren çalışmalar vardır (Tavani ve Losh, 2003; Hooper vd., 2020; Mullis et al., 2017). Matematik Öğrenmeyi Sevme (MOS) ölçeği ile öğrencilerin matematik öğrenmede içsel motivasyonları ölçülmektedir. TIMSS 2019 uygulamasında yer alan MOS ölçeği 9 maddeden oluşmaktadır.

1.2.2. Matematik Öğretiminin Netliği (MON)

Eğitimde öğretmenlerin müfredatı öğretmesindeki açıklığı (netliği), öğrencilerin öğrenmesinde önemli etkilere sahiptir. Matematik Öğretiminin Netliği (MON) ölçeğinde öğrencilere öğretmenin anlaşılır olup olmadığı, sorularına net yanıt verip vermediği, matematiği öğrencilerine açıklamada iyi olup olmadığı, yeni bilgileri eski bilgilerle ilişkilendirerek öğrenmeyi kolaylaştırıp kolaylaştırmadığı gibi sorular yer almaktadır. MON ölçeği 7 maddeden oluşmaktadır. TIMSS sonuçları, öğretmenin netliği arttıkça başarının da arttığını göstermektedir (Mullis vd., 2020).

1.2.3. Matematik Dersinde Disiplinsiz Davranış (MDDD)

Öğrenme, öğrencilerin dikkatini derse verebildiği ve öğrenmeye odaklanan elverişli bir sınıf ortamında gerçekleşebilir (Radovan ve Makovec, 2015). Matematik Dersinde Disiplinsiz Davranış (MDDD) ölçeği matematik dersi esnasında öğrencilerin disiplinsiz davranışlarını belirlemek için geliştirilmiştir. Bu ölçekte öğrencilerin matematik dersi sırasında öğretmenin sözünü kesmesi, rahatsızlık veren gürültü çıkarılması, öğrencilerin öğretmenin söylediklerini dinlememesi ve dersin başında öğretmenin öğrencileri uzun süre beklemesi gibi sınıf ortamındaki disiplinsiz davranışlara ilişkin maddeler yer almaktadır (Yin ve Fishbein, 2019). Bu ölçekte 6 madde bulunmaktadır. MDDD ölçeği ile matematik başarısı arasında negatif yönlü ilişkilerin olduğu bildirilmektedir (Mullis vd., 2020).

1.2.4. Matematikte Kendine Güven (MKG)

Akademik benlik, bir öğrencinin belirli bir akademik uğraşı karşısında, diğer öğrencilere göre kendinin ne kadar yetenekli olduğuna ilişkin geliştirdiği kanısı olarak tanımlanmaktadır (Arseven, 1986,18). Öğrencilerin, akademik benlik kavramı üzerine değerlendirmeleri genellikle kendi geçmiş deneyimlerine ve akranlarıyla kıyaslamaları ile kendilerini nasıl gördüklerine dayanır (Marsh ve Craven, 2006). Literatürde akademik benlik kavramı ile matematik başarısı arasında güçlü bir ilişki olduğuna dair çalışmalar yer almaktadır (Mullis et al., 2017). 9 maddeden oluşan Matematikte Kendine Güven (MKG) ölçeği, akademik benlik kavramını ölçmektedir (Mullis vd., 2020).

1.2.5. Matematiğe Değer Verme (MDV)

Dışsal motivasyon, övgü, kariyer başarısı, para ve diğer teşvikler gibi dış ödüllerden gelen dürtüyü ifade eder (Zeng vd., 2022). 9 maddeden oluşan Matematiğe Değer Verme (MDV) ölçeği ise öğrencilerin matematiğe yönelik dışsal motivasyonunu ölçmektedir. Literatürde matematiğe değer verme ile matematik başarısı arasında güçlü bir ilişki olduğunu gösteren çalışmalar yer almaktadır (Güvendir, 2016; Mullis et al., 2017).

1.3. Çalışmanın Amacı ve Önemi

Matematik duyuşsal özelliklerinin akademik başarı gibi değişkenlerle ilişkisini inceleyen farklı araştırmalar yapılmaktadır. Bu araştırmalar yapılmadan önce farklı grup, bölge, ülke veya kıtalardan katılan katılımcılara göre gizil değişkenlerin aynı şekilde anlaşıldığının belirlenmesi için faktör yapılarının ölçme değişmezliği ile irdelenmesi gerekir. TIMSS uygulamasından elde edilecek bulgular ile doğru ve tarafsız karşılaştırmalar yapmak için uygulama yapılan gruplar arasında ölçme

değişmezliğinin sağlandığını gösterilmelidir. Bu çalışmanın amacı, TIMSS 2019 döngüsünde yer alan matematik duyuşsal özellik ölçeklerinin cinsiyete göre ölçme değişmezliğini incelemektir. Literatürdeki değişmezlik çalışmaları incelendiğinde kültür, bölge, dil ve cinsiyete göre ölçme değişmezliği çalışmalarının gerçekleştirildiği görülmüştür (Uyar ve Doğan, 2014; Alatlı, 2020; Bağdu Söyler vd., 2021). Duyuşsal değişkenlere göre gerçekleştirilen ölçme değişmezliği çalışmaları ise oldukça sınırlıdır (Bofah ve Hannula, 2015; Ertürk ve Erdinç-Akan, 2018; Polat, 2019; Reynolds vd., 2022). TIMSS 2019 döngüsü verisi ile matematik duyuşsal özelliklerinin ölçme değişmezliğini sınavan herhangi bir çalışma ise bulunmamaktadır. Bu çalışma hem öğrencilerin matematiğe yönelik duyuşsal özelliklerinin değişmezliğine odaklanması hem gelecekte yapılacak çalışmalar için ölçme değişmezliği kanıtı sunması açısından önemlidir.

1.3. İlgili Araştırmalar

Literatürde matematik duyuşsal özellikler ölçekleri üzerine ölçme değişmezliğini inceleyen çalışmalar mevcuttur. Ertürk ve Erdinç-Akan (2018), TIMSS 2015 döngüsünden elde ettikleri veri ile matematiği sevme, matematiğe olan ilgi ve matematiğe ilişkin özgüven değişkenleri üzerinden ölçme değişmezliğini incelemiştir. Analiz sonuçlarına göre matematiği sevme değişkeninin ölçme değişmezliğini sağladığını, fakat diğer iki değişkenin yapı değişmezliği düzeyinde kaldığını bildirilmektedir. Polat (2019), TIMSS 2015 8. sınıf öğrenci anketini kullanarak matematik ve fen duyuşsal özellik modelleri oluşturmuştur. Oluşturulan model ile cinsiyete ve kültürlere (Türkiye, Suudi Arabistan ve Singapur) göre ölçme değişmezliğini incelemiştir. Çalışmanın sonuçları hem matematik hem de fen duyuşsal özellikler modelinde cinsiyete göre katı değişmezliğin sağlandığını, fakat bölgelere göre ölçek değişmezliğinin sağlandığını bildirmektedir. Çakıcı Eser (2021) ise TIMSS 2015 4. Sınıf verileri ile ev kaynakları (internet bağlantısı, ısıtma sistemi, soğutma sistemi, bulaşık makinesi) değişkenleri ile ölçme değişmezliği çalışması gerçekleştirmiştir. Bu çalışmanın sonuçları tüm ev kaynakları değişkenlerine göre katı ölçme değişmezliğinin sağlandığını belirtilmektedir. Ding, Yang Hansen ve Klapp (2022), PISA 2003 ve 2012 döngülerinden matematik benlik kavramı ve matematik öz-yeterliliğinin 40 ülke arasında ölçme değişmezliğini incelemiştir. Bu çalışmanın sonuçları tüm ülkeler arasında katı değişmezliğin sağlanmadığını ve ülkeler arası karşılaştırmaların yapılamayacağını göstermektedir.

2. YÖNTEM

Bu çalışma, TIMSS 2019 öğrenci anketinde yer alan matematik başarıları ile ilgili duyuşsal alt ölçekler (matematik öğrenmeyi sevme, matematikte kendine güven, matematik dersinde disiplinsiz davranış, matematiğe değer verme, matematik öğretiminin netliği) ile oluşturulan Matematik Duyuşsal Özellik Modeli'nin cinsiyete göre değişmediğini belirlemeyi amaçladığından betimsel bir çalışmadır (Büyüköztürk vd., 2017).

2.1. Evren ve Örneklem

Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement- IEA) tarafından uygulanan TIMSS 2019'a Türkiye'den 4. Sınıf düzeyinde 180 okuldan 4028 öğrenci, 8. sınıf düzeyinde 181 okuldan 4077 öğrenci katılmıştır (MEB, 2020). Bu araştırmanın örneklemi ise TIMSS 2019'a Türkiye'den katılan ve 8. Sınıfa devam eden 3658

öğrenci oluşturmaktadır. Çalışmanın gerçekleştirildiği örneklemin cinsiyete göre dağılımı Tablo 4'te verilmiştir.

Tablo 1: Örneklemin Cinsiyete Göre Dağılımı

Cinsiyet	N	%
Kız	1848	50,51
Erkek	1810	49,49
Toplam	3658	100

2.2. Veri Kaynağı

Veriler <https://timss2019.org/international-database/> adresindeki veri tabanından elde edilmiştir. TIMSS uygulamasında matematik ve fen başarı testlerinin yanında, öğrenci, öğretmen, okul ve ev anketleri de yer almaktadır. Bunun yanında matematik ve fen bilimleri alanlarına ilişkin öğrencinin duyuşsal özelliklerini inceleyen maddeler de yer almaktadır. Bu çalışma matematik alanına ilişkin duyuşsal özellikleri ölçen ölçekler ile sınırlandırılmıştır. TIMSS 2019'da 8. sınıf düzeyinde Matematik Öğrenmeyi Sevme (MOS), Matematikte Kendine Güven (MKG), Matematik Dersinde Disiplinsiz Davranış (MDDD), Matematiğe Değer Verme (MDV), Matematik Öğretiminin Netliği (MON) olmak üzere matematik alanında beş farklı duyuşsal ölçek sunulmaktadır. 4. sınıflara bu ölçeklerden MDV ölçeği uygulanmadığı için 4. Sınıflar kapsam dışı bırakılmış olup sadece 8. Sınıf düzeyi ile araştırma gerçekleştirilmiştir. Çalışmaya dahil edilen değişkenlerin adları ve tanımları Tablo 2'de verilmiştir.

Tablo 2: Örneklemin Cinsiyete Göre Dağılımı

Değişken Kodu	Değişken Adı	Madde Sayısı
BSBG01	Cinsiyet	1
BSBM16	Matematik Öğrenmeyi Sevme (MOS)	9
BSBM17	Matematik Öğretiminin Netliği (MON)	7
BSBM18	Matematik Dersinde Disiplinsiz Davranış (MDDD)	6
BSBM19	Matematikte Kendine Güven (MKG)	9
BSBM20	Matematiğe Değer Verme (MDV)	9

Tablo 2'de görüldüğü üzere TIMSS 2019 8. sınıf uygulamasında matematiğe yönelik duyuşsal özelliklere ilişkin 4'lü Likert tipine göre puanlanan beş farklı ölçekte toplamda 40 madde bulunmaktadır.

2.3. Veri Analizi

Bu çalışmada yer alan tüm analizler açık kaynak kodlu R programı ile yapılmıştır. Veri analizine başlamadan önce, tüm çok değişkenli istatistiksel tekniklerde sağlanması gereken varsayımlar incelenmiştir. TIMSS 2019 veritabanından elde edilen veride 8. Sınıf düzeyinde Türkiye'den katılan 4077 öğrenci yer almaktadır. Öncelikle öğrenciler tarafından cevaplanmayan kayıp veriler incelenmiştir. Değişken bazında kayıp veri oranlarının %0.07 (29/4077) ile %3,6 (148/4077) arasında değiştiği görülmüştür. Tüm veri bazında ise kayıp veri oranı %15,2'ye (619/4077) ulaşmaktadır. Kayıp verinin rassallığını incelemek için Little MCAR testi yapılmıştır. Little MCAR testi sonuçları kayıp verinin rassal olduğunu, herhangi bir örüntü içermediğini göstermektedir (LittleMCAR=804; sd=5942; p=0.99). Kayıp verinin %10'dan fazla olması sebebiyle analiz gücünün kaybolmaması için liste bazında silme yöntemi yerine kayıp veri ataması yapılması tercih edilmiştir. Kayıp veri atamasından önce cinsiyet değişkenindeki kayıp veriler liste bazında silinerek 29 katılımcı örneklemden çıkarılmıştır. Kalan 4048 katılımcının verisine EM algoritması ile kayıp veri ataması yapılmıştır. Uçdeğer analizi ise üç aşamada yapılmıştır. Uçdeğer analizinden önce dikkatsiz yanıtlayan katılımcılar tespit edilmiştir. Ters kodlamalı maddelerin yer aldığı MOS ve MKG ölçeklerinde normal ve ters kodlanan maddelere aynı uç tepkileri veren katılımcılar dikkatsiz yanıtlayan davranışı sergilemektedir (Woods, 2006). Bu katılımcıların faktör yapısına bozucu etkileri olduğu belirtilmektedir (Kam, 2019). Dikkatsiz yanıtlayan 231 katılımcı belirlenmiş ve örneklemden çıkarılmıştır ve 3817 katılımcının verisi ile analize devam edilmiştir (Kam ve Meyer, 2015). İkinci aşamada alt ölçeklere göre toplam ve standartlaştırılmış z puanları hesaplanmıştır. Z puanına göre [-3,+3] aralığı dışında kalan katılımcılar uçdeğer olarak belirlenmiştir. Z puanlarına göre bu aralık dışında yer alan MON ölçeğindeki 78 birey ve MDV ölçeğindeki 49 birey olduğu görülmüştür. Toplamda uçdeğer olarak belirlenen 127 birey örneklemden çıkarılmıştır ve 3690 katılımcının verisi ile analizler sürdürülmüştür. Üçüncü aşamada Mahalanobis uzaklıkları ile çok değişkenli uç değerlerin olup olmadığı incelenmiştir. Yapılan mahalanobis uzaklık hesaplamasında 4 serbestlik derecesi ve $p < .001$ anlamlılık düzeyinde χ^2 'nin kritik değerinin 18.466 olduğu görülmüştür. Bu değer üzerinde χ^2 değerine sahip olan 32 birey çok değişkenli uçdeğer olarak belirlenerek örneklemden çıkarılmış ve 3658 katılımcı ile analize devam edilmiştir. Verilerin normal dağılıma sahip olup olmadığına karar vermek için değişkenlerin çarpıklık ve basıklık değerleri incelenmiştir. Tüm ölçeklerin basıklık ve çarpıklık katsayıları [-1.5,+1.5] aralığında olduğundan normal dağıldığına karar verilmiştir (Tabachnick ve Fidell, 2013). Çoklu bağlantı durumunun incelenmesi için tüm maddeler için VIF (variance inflation factor) değeri hesaplanmıştır. En yüksek VIF değerinin 3.56 olarak MOS ölçeğindeki BSBM16E kodlu maddede yer aldığı görülmüştür. VIF değerinin 5'ten büyük olması durumunda çoklu bağlantı sorunu oluşturulacağı için tüm ölçeklerde çoklu bağlantı sorununun olmadığına karar verilmiştir (Kline, 2011). Varyansların homojenliği varsayımını incelemek için Box-M testi yapılmıştır. Test sonuçlarına göre varyansların homojen dağılmadığı sonucuna ulaşılmıştır ($\chi^2 = 52.434$, $sd = 15$, $p = 0.000$).

Bu çalışmada kategorik ve sıralı değişkenler kullanılmıştır. Tek değişkenli normallik varsayımı sağlanırken, çok değişkenli normallik ve varyansların homojenliği varsayımı sağlanmamaktadır. Çok değişkenli normallik sayılısının ihlal edildiği durumlarda WLS (weighted least squares), WLSMV (robust weighted least squares) veya ULS (unweighted least squares) yöntemlerinin kullanılması gerekmektedir (Brown, 2006; Koğar ve Yılmaz Koğar, 2015). Dolayısıyla bu çalışmada kestirim yöntemi olarak Çoklu Grup DFA modelinde iyi sonuçlar verdiği belirtilen ULS kestirim yöntemi kullanılmıştır (Forero, Maydeu-Olivares ve Gallardo-Pujol, 2009). Çoklu Grup DFA analizleri R'da lavaan (Rosseel, 2012) paketi kullanılarak gerçekleştirilmiştir. Bu paketin tercih edilmesinin bir diğer

sebebi Çoklu Grup DFA modeli ile kestirim yaparken örneklem ağırlıklarının kullanılmasına olanak sağlamasıdır. Geniş ölçekli testlerde zaman ve maddi açılarından mümkün olmamasından dolayı evrende yer alan tüm katılımcılar (ilgili sınıf düzeyindeki tüm öğrenciler) örnekleme dahil edilememektedir. Bu sınırlılığı aşmak ve örneklemin evrene genellenebilirliğini sağlamak amacıyla örneklem ağırlıkları kullanılmaktadır (Arıkan vd., 2020). Bu çalışmada TIMSS 2019 verisinde yer alan öğrenci ağırlıkları modele örneklem ağırlığı olarak eklenmiştir.

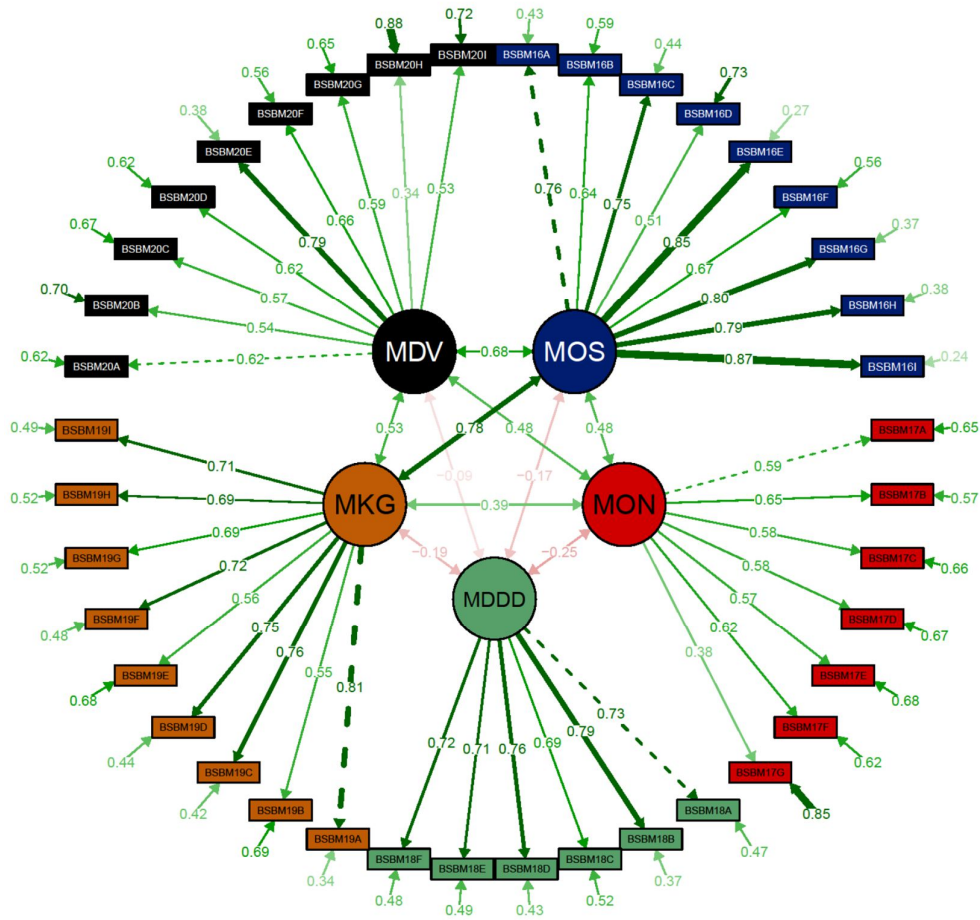
ÇG-DFA'daki ölçme değişmezliği, dört hiyerarşik model test edilerek incelenmiştir. Bu dört model sırasıyla yapısal değişmezlik, metrik değişmezlik, ölçek değişmezliği, katı değişmezliktir (Vandenberg ve Lance, 2000). Aşamalar arasındaki model veri uyumunu incelemek için χ^2 , χ^2/df , *RMSEA* (*RMSEA* < .08), *SRMR* (*SRMR* < .08), *TLI* (*TLI* > .95), *CFI* (*CFI* > .95) ve ΔCFI değerleri dikkate alınmıştır (Hu ve Bentler, 1999; Schermelleh Engel, Moosbrugger ve Müller, 2003; Schumacker ve Lomax, 2004). Bir aşamadaki uyumun yeterli düzeyde olduğuna karar verildiğinde bir sonraki aşamaya geçilmiştir. Aşamalar arası değişmezliğin sağlanıp sağlanmadığının belirlenmesinde ki-kare değerlerinin farkına ($\Delta\chi^2$) göre anlamlılığının test edilebileceğini öneren çalışmalar bulunmaktadır (Schmitt ve Kuljanin, 2008). Ancak ki-kare fark testi, örneklem büyüklüğü arttıkça sıfır hipotezini çok fazla güçle reddetmektedir. Dolayısıyla Cheung ve Rensvold (2002) $\Delta\chi^2$ 'ye alternatif olarak *CFI* değerindeki (ΔCFI) değişikliğin incelenmesini önermiştir. Bu çalışmada aşamalar arası değişmezliğin sağlanıp sağlanmadığının belirlenmesinde ΔCFI farkının 0.01'den küçük veya buna eşit olması değişmezliğin sağlandığını belirlemede bir kriter olarak kullanılmıştır (Cheung ve Rensvold, 2002).

3. BULGULAR

Bu bölümde öncelikle kurulan model ile DFA analizi gerçekleştirilmiş ve model-veri uyumu incelenmiştir. Ardından cinsiyete göre ÇD-DFA modeli ile ölçme değişmezliği bulgularına yer verilmiştir.

3.1. Matematik Duyuşsal Özellikler Modelinin DFA Bulguları

Bu çalışmada beş alt ölçek ile kurulan modelin model veri uyumunu incelemek için Doğrulayıcı Faktör Analizi kullanılmıştır. DFA sonucunda elde edilen standartlaştırılmış faktör yükleri, hata varyansları ve standartlaştırılmış kovaryanslar (korelasyonlar) Şekil 1'deki diyagramda görülmektedir.



MOS: Matematik Öğrenmeyi Sevme, MKG: Matematikte Kendine Güven, MDDD: Matematik Dersinde Disiplinsiz Davranış, MDV: Matematiğe Değer Verme, MON: Matematik Öğretiminin Netliği.

Şekil 1: Matematik Duyuşsal Özellikler Modelinin DFA Sonuçları

Şekil 1’de dikdörtgen şekiller maddeleri, daire şekiller ise faktörleri temsil eder. Modelde faktörlerden maddelere giden oklar ile standartlaştırılmış faktör yükleri belirtir. Faktörler arasındaki değerler ise faktörler arası standartlaştırılmış kovaryansları (korelasyon) belirtmektedir. Şeklin dış kısmından maddelere giden oklar ise standartlaştırılmış hata varyanslarıdır. Kurulan bu modelin uyum iyiliği istatistikleri Tablo 3’te sunulmuştur.

Tablo 3: DFA Modelinin Uyum İstatistikleri

χ^2	sd	χ^2/sd	RMSEA	SRMR	TLI	CFI
6339,6	730	8,68	0,046	0,051	0,973	0,975

Model veri uyumunun sağlanması için χ^2/df değerinin 5'ten küçük olması beklenmektedir. Fakat χ^2/df oranının özellikle örneklem büyüklüğünden etkilenmesinden dolayı model veri uyumunda diğer indekslerden faydalanılması önerilmektedir. RMSEA'nın 0.046, SRMR'nin 0.051, CFI indeksinin 0.973 ve TLI indeksinin 0.975 olduğu görülmektedir. Bu değerlere göre model veri uyumunun kabul edilebilir aralıklarda olduğu söylenebilir ($RMSEA < .08$, $SRMR < .08$, $TLI > .95$, $CFI > .95$) (Schreiber vd., 2006). Dolayısıyla model veri uyumunun sağlandığı kararına varılmıştır.

DFA modelinden elde edilen standartlaştırılmış faktör yükleri, AVE (average variance extracted) değerleri ve güvenilirlik katsayıları Tablo 4'te sunulmuştur.

Tablo 4: Standartlaştırılmış Faktör Yükleri, AVE Değerleri ve Güvenirlik Katsayıları

Ölçekler	Maddeler	Faktör Yükleri	AVE	Cronbach Alfa	McDonald's Omega
Matematik Öğrenmeyi Sevme (MOS)	BSBM16A	0,755	55,71	0,913	0,916
	BSBM16B	0,638			
	BSBM16C	0,745			
	BSBM16D	0,515			
	BSBM16E	0,853			
	BSBM16F	0,667			
	BSBM16G	0,797			
	BSBM16H	0,787			
	BSBM16I	0,871			
Matematik Öğretiminin Netliği (MON)	BSBM17A	0,594	34,01	0,778	0,775
	BSBM17B	0,655			
	BSBM17C	0,585			
	BSBM17D	0,576			
	BSBM17E	0,567			
	BSBM17F	0,618			
	BSBM17G	0,384			
Matematik Dersinde Disiplinsiz Davranış (MDDD)	BSBM18A	0,726	53,90	0,874	0,875
	BSBM18B	0,793			
	BSBM18C	0,690			
	BSBM18D	0,756			
	BSBM18E	0,712			
	BSBM18F	0,722			
Matematikte Kendine Güven (MKG)	BSBM19A	0,811	48,33	0,891	0,893
	BSBM19B	0,554			
	BSBM19C	0,761			
	BSBM19D	0,750			
	BSBM19E	0,561			
	BSBM19F	0,723			
	BSBM19G	0,693			
	BSBM19H	0,694			
	BSBM19I	0,714			
Matematiğe Değer Verme (MDV)	BSBM20A	0,619	38,92	0,839	0,838
	BSBM20B	0,544			
	BSBM20C	0,573			
	BSBM20D	0,619			
	BSBM20E	0,789			
	BSBM20F	0,665			
	BSBM20G	0,591			
	BSBM20H	0,343			
	BSBM20I	0,532			

DFA, ölçme aracının veya oluşturulan modelin geçerliğini kanıtlamak için kullanılan bir yöntemdir. Aynı yapıyı ölçmek için bir araya getirilen maddelerin kendi içerisinde ve ölçülen örtük değişkenle iyi düzeyde ilişkili olması beklenir. Bu beklenen ilişki yakınsak geçerlik olarak bilinir (Kline, 2011). Tablo 4 incelendiğinde MOS, MDDD ve MKG ölçeklerinin faktör yüklerinin yüksek olduğu görülmektedir. MON ve MDV ölçeklerinde ise faktör yüklerinin daha düşük olduğu göze çarpmaktadır. Harrington (2009), faktör yüklerinin 0.30 ve üzerinde olması gerektiğini belirtmektedir. Genel olarak faktör yüklerinin 0.34 ile 0.87 aralığında değiştiği göz önüne alındığında yeterli düzeyde olduğu söylenebilir. Tabloda yer alan AVE (ortalama varyans-average variance extracted) değerleri faktör yüklerinden hesaplanmaktadır ve faktörün maddelerin varyansının yüzde kaçını açıkladığı bilgisini verir. Genellikle bu değer %50 ve üzerinde olması istenir (Hair ve diğerleri, 2014). Fakat Fornell ve Larcker (1981) %50'den daha azının açıklanmasının da kabul edilebilir olduğunu bildirmektedir. MOS, MDDD ölçeklerinin %50 ve üzerinde AVE değerleri vardır; MON, MKG ve MDV ölçeklerinin kabul edilebilir düzeyde AVE değerine sahip olduğu söylenebilir. Diğer taraftan içtutarlılık katsayıları, güvenilirliğe ait bir kanıt oluşturmanın yanında homojen bir yapının ölçülüp ölçülmediğine dair kanıt oluşturmaktadır (Aybek, 2022). İchtutarlılığa dair tabloda ölçeklerin Cronbach Alfa ve McDonald's Omega katsayıları değerleri yer almaktadır. Büyüköztürk'e (2011) göre güvenilirlik katsayısının 0.70 ve üzerinde olması ölçek güvenilirliği için yeterli kabul edilmektedir. Kline (2011) ise 0.90 ve üzeri mükemmel, 0.80 ve üzeri iyi, 0.70 ve üzeri ise kabul edilebilir olarak sınıflandırmaktadır. Tablo 4'te görüldüğü üzere hem Cronbach Alfa hem de McDonald's Omega katsayıları 0.70'ten fazladır ve yeterli düzeyde olduğu kabul edilmiştir.

3.2. Matematik Duyuşsal Özellikler Modelinin Cinsiyete Göre Ölçme Değişmezliği Bulguları

Ölçme değişmezliği; yapısal değişmezlik, metrik değişmezlik, ölçek değişmezliği ve katı değişmezlik hiyerarşik sıralamasına göre test edilmiştir. Hiyerarşik iki aşama arasındaki değişmezliğin sağlanıp sağlanmadığının saptanmasında ΔCFI değeri dikkate alınmıştır. Bu değer aşamalar arasında $\Delta CFI < 0.01$ olması ilgili aşamada ölçme değişmezliğinin sağlandığı şeklinde yorumlanmaktadır (Cheung ve Rensvold, 2002).

Matematik duyuşsal alt ölçekleri ile oluşturulan modelin cinsiyete göre ölçme değişmezliği bulguları Tablo 5'te sunulmuştur.

Tablo 5: Cinsiyete Göre Ölçme Değişmezliği

	χ^2	sd	χ^2/sd	RMSEA	SRMR	TLI	CFI	ΔCFI
Yapısal	6838,19	1460	4,68	0,045	0,052	0,975	0,977	-
Metrik	7304,98	1495	4,89	0,046	0,054	0,974	0,975	0,002
Ölçek	7433,11	1530	4,86	0,046	0,055	0,974	0,974	0,001
Katı	7651,72	1570	4,87	0,046	0,056	0,974	0,973	0,001

Yapısal değişmezlik bulguları incelendiğinde cinsiyete göre her iki grupta da uyum değerleri iyi düzeyde olduğundan yapısal değişmezliğin sağlandığı görülmektedir ($RMSEA < .08$, $SRMR < .08$, $TLI > .95$, $CFI > .95$). Yapısal değişmezliğin ardından metrik değişmezlik incelenmiştir. Yapısal değişmezlik

ile metrik değişmezlik aşamalar arasındaki CFI değeri farkı (ΔCFI) 0.002 olduğundan metrik değişmezliğin sağlandığına karar verilmiştir. Ayrıca metrik değişmezlik aşamasında diğer uyum indekslerinin de iyi düzeyde olduğu görülmektedir ($\chi^2/sd=4,89$, $RMSEA=0,046$, $SRMR=0,054$, $TLI= 0,974$). Metrik değişmezlikten sonra ölçek değişmezliği incelenmiştir. Metrik değişmezlik ile ölçek değişmezliği arasındaki ΔCFI değeri 0,001 olduğundan ölçek değişmezliğinin sağlandığına karar verilmiştir ($\chi^2/sd=4,86$, $RMSEA=0,046$, $SRMR=0,055$, $TLI= 0,973$). Son aşama olan katı değişmezlik sonuçları incelendiğinde, ölçek değişmezliği ile katı değişmezlik arasındaki ΔCFI değerinin 0,001 olduğu, dolayısıyla katı değişmezliğin sağlandığı görülmektedir.

ÇG-DFA yöntemi ile yapılan analiz sonuçlarına göre ΔCFI ($\Delta CFI < 0.01$) değerleri ve uyum iyiliği istatistikleri ($RMSEA < .08$, $SRMR < .08$, $TLI > .95$, $CFI > .95$) dikkate alındığında cinsiyete göre tam ölçme değişmezliği sağlanmaktadır. Bu sonuca göre, bu araştırmada madde-faktör yapısının ve faktör yükleri, varyanslar, kovaryanslar ve hata varyanslarının kız ve erkek gruplarına göre eşdeğer olduğu görülmüştür.

3.3. Matematik Duyuşsal Özellikler Değişkenlerinin Cinsiyete Göre Karşılaştırılması

Bu bölümde matematik duyuşsal özellikler modelinde yer alan değişkenlerin cinsiyete göre farklılıkları karşılaştırılmıştır. Cinsiyete göre karşılaştırmalar örneklem ağırlıkları dâhil edilerek t testi ile gerçekleştirilmiştir. Bulgular Tablo 6'da sunulmuştur.

Tablo 6: Değişkenlerin Cinsiyete Göre Farklılığı

	Ortalama		Standart Sapma		t	Standart Hata	p
	Kız	Erkek	Kız	Erkek			
Matematik Öğrenmeyi Sevme (MOS)	18,533	18,496	7,38	6,775	0,157	0,234	0,875
Matematik Öğretiminin Netliği (MON)	9,551	10,035	3,156	3,288	-4,617	0,105	0,000
Matematik Dersinde Disiplinsiz Davranış (MDDD)	17,089	16,355	4,848	4,927	4,547	0,162	0,000
Matematikte Kendine Güven (MKG)	21,782	21,332	7,237	6,479	1,985	0,227	0,047
Matematiğe Değer Verme (MDV)	14,746	15,072	5,15	5,128	-1,917	0,17	0,055

Tablo 6 incelendiğinde matematik duyuşsal özellikler modelinde yer alan beş değişkenin cinsiyete göre anlamlı farklılıkları incelendiğinde üç değişkende cinsiyete göre anlamlı farklılık varken ($p < .05$), iki değişkende ise anlamlı farklılık yer almamaktadır ($p > .05$). Matematik öğrenmeyi sevme değişkeninin cinsiyete göre kız ve erkek öğrenciler arasında anlamlı olarak farklılaşmadığı görülmektedir ($t=0.157$; $p=0.875$). Matematik öğretiminin netliği değişkeninde ise cinsiyete göre erkek öğrenciler lehine anlamlı farklılık olduğu görülmektedir ($t=-4.617$; $p=0.000$). Diğer bir deyişle erkek öğrenciler kız öğrencilere göre matematik öğretiminin netliğinin daha yüksek olduğunu belirtmektedirler. Matematik dersinde disiplinsiz davranış değişkeninde cinsiyete göre kız öğrenciler

lehine anlamlı farklılık olduğu görülmektedir ($t=4.547$; $p=0.000$). Yani kız öğrenciler erkek öğrencilere göre matematik dersinde disiplinsiz davranışların daha yüksek olduğunu ifade etmektedirler. Matematikte kendine güven değişkeninde cinsiyete göre erkek öğrenciler lehine anlamlı farklılık olduğu görülmektedir ($t=1.985$; $p=0.047$). Diğer bir deyişle kız öğrencilerin matematikte kendine güvenleri erkek öğrencilere göre daha yüksektir. Matematiğe değer verme değişkeninde ise cinsiyete göre gruplar arasında anlamlı farklılık olmadığı görülmektedir ($t=-1.917$; $p=0.055$).

4. TARTIŞMA ve SONUÇ

Bu çalışmada TIMSS 2019 öğrenci anketinde yer alan ve matematik duyuşsal özellikleri ile ilgili olan beş ölçek ile matematiksel duyuşsal özellikler modeli kurulmuştur. Bu modelin cinsiyete göre ölçme değişmezliği gösterip göstermediği araştırılmıştır. Analizler ÇG-DFA modeli ile gerçekleştirilmiştir. Ölçme değişmezliğinde aşamalar yuvalanmış (nested) modeller olduğundan modellerin karşılaştırılmasında literatürde genellikle $\Delta\chi^2$ ve ΔCFI değerleri kullanılmaktadır. Bu çalışmada, aşamalar arasında ölçme değişmezliğine karar vermek için ΔCFI değeri kullanılmıştır. Ayrıca uyum iyiliği istatistikleri de raporlanmıştır. Çalışma sonuçlarına göre matematik duyuşsal özellikler modelinin cinsiyete göre sırasıyla yapısal, metrik, ölçek ve katı ölçme değişmezliğini sağladığı görülmektedir. Kız ve erkek gruplarına göre modelin ortalamaları, varyansları, kovaryansları ve hata varyansları her iki grupta da eşdeğerdir. Sonuçlar, matematiksel duyuşsal özellikler modelinden elde edilen gözlenen puanların cinsiyete göre karşılaştırılabileceğini göstermektedir. Literatürde matematik duyuşsal özellikler ölçekleri üzerine ölçme değişmezliğini inceleyen çalışmalar mevcuttur. Ertürk ve Erdiñ-Akan (2018), Polat (2019) ve Çakıcı Eser (2021) çalışmalarında TIMSS 2015 uygulamasının matematiksel duyuşsal özellikler ölçeklerine odaklanmışlardır. Ertürk ve Erdiñ-Akan (2018), TIMSS 2015 4. Sınıf verisi ile üç farklı matematik duyuşsal özellik ölçeğinin cinsiyete göre ölçme değişmezliğini incelediği çalışmada “Matematiği Sevme” ölçeğinin katı ölçme değişmezliğini sağladığını, fakat diğer iki ölçeğin ölçek değişmezliği ve yapı değişmezliği aşamalarında kaldığını belirtmektedir. Polat (2019), TIMSS 2015 8. Sınıf verileri ile matematik ve fen duyuşsal özellikler modelleri ile cinsiyet ve bölgelere göre ölçme değişmezliğini araştırmıştır. Bu çalışmanın sonuçlarına göre hem matematik hem de fen duyuşsal özellikler modelinde cinsiyete göre katı değişmezlik sağlanırken, bölgelere göre ölçek değişmezliğinin sağlandığı görülmüştür. Çakıcı Eser (2021) ise TIMSS 2015 4. Sınıf verileri ile ev kaynakları değişkenlerine göre ölçme değişmezliği çalışması gerçekleştirmiştir. Çalışmanın sonuçları tüm ev kaynakları değişkenlerine göre katı ölçme değişmezliğinin sağlandığını belirtilmektedir (Çakıcı Eser, 2021). Uyar (2021) ise matematik duyuşsal özellikleri ile Açıklayıcı SEM analizi gerçekleştirdiği çalışmada matematik duyuşsal özelliklerinin cinsiyete göre katı değişmezliği sağladığını belirtmektedir. Bu çalışmanın sonuçları Polat (2019), Çakıcı Eser (2021) ve Uyar (2021)’in çalışmalarının sonuçlarına benzer şekilde duyuşsal özellikler ile cinsiyet arasında katı değişmezlik olduğunu göstermektedir. Ertürk ve Erdiñ-Akan (2018)’in çalışmasının sonuçları ile de kısmi benzerlik gösterdiği söylenebilir.

Literatürde cinsiyete göre farklı duyuşsal özellikler üzerine ölçme değişmezliğini inceleyen çalışmalar da yer almaktadır. Başusta ve Gelbal (2015), Kıbrıslıođlu (2015), Uyar ve Uyanık (2019) ve Güngör ve Kabasakal (2020) tarafından yapılan çalışmalar PISA örneğinde ele alınmıştır. Başusta ve Gelbal (2015), PISA 2006 verileri ile oluşturduğu fen duyuşsal özellikler modelinin cinsiyete göre ölçme değişmezliğini incelemiştir. Çalışmanın sonuçları, fen duyuşsal özellikler modelinin cinsiyete göre katı değişmezliği sağladığını göstermektedir. Kıbrıslıođlu (2015) ise PISA 2012 verisi ile matematiksel öğrenme modelinin ülkelere ve cinsiyete göre ölçme değişmezliğini araştırmıştır.

Çalışmanın sonuçları, modelin kültürler arasında yapısal değişmezliği sağladığını, cinsiyete göre ise katı değişmezliğin sağlandığını belirtmektedir. Uyar ve Uyanık (2019) PISA 2015 uygulaması ile kurduğu fen öğrenme modelinin cinsiyete ve kültürlere göre ölçme değişmezliğini sınamıştır. Çalışmanın sonuçları kurulan modelin cinsiyete göre metrik değişmezliği, kültürlere göre yapısal değişmezliği sağladığı sonucuna ulaşılmıştır. Güngör ve Kabasakal (2020), PISA 2015 uygulaması verileri ile fen motivasyonu ve fen özyeterlik ölçekleri ile oluşturduğu modelin cinsiyete ve bölgelere göre ölçme değişmezliğini incelemiştir. Çalışmanın sonuçları cinsiyete göre yapısal değişmezliğin sağlandığını, bölgelere göre ise metrik değişmezliğin sağlandığını bildirmektedir. Tekin ve Cobanoğlu-Aktan (2021) ise PISA 2015 verisi ile işbirlikli problem çözme ölçeği üzerine Türkiye, Norveç ve Singapur ülkelerini karşılaştıran ölçme değişmezliği çalışması gerçekleştirmiştir. Çalışmanın sonuçları, bu üç ülkenin sadece yapısal değişmezlik aşamasını sağladığını bildirmektedir. Ersozlu, Usak ve Blake (2022), matematik kaygısının Avustralya ve Rusya grupları arasında ölçme değişmezliğini incelediği çalışmasında tam metrik ve kısmi ölçek değişmezliği sağladığını raporlamaktadır. Bu çalışmaların sonuçlarına göre uluslararası büyük ölçekli sınavlarda cinsiyete göre katı değişmezliğin genellikle sağlandığı, fakat kültürlere göre katı değişmezliği sağlayan çalışmanın olmadığı görülmektedir. Bu çalışmanın sonuçları, belirtilen çalışmaların çoğunun sonuçlarına benzer şekilde cinsiyete göre katı değişmezliği sağlandığını ortaya koymaktadır.

Matematik duyuşsal özellikler modelinin cinsiyete göre ölçme değişmezliği sınıandıktan sonra modelde yer alan değişkenlerin cinsiyete göre anlamlı farklılık gösterip göstermediği incelenmiştir. Modelde yer alan beş değişkenin üçü cinsiyete göre anlamlı farklılık gösterirken, iki değişken ise cinsiyete göre anlamlı farklılık göstermemektedir. Matematik öğretiminin netliği (MON) değişkenine göre erkek öğrenciler kız öğrencilere göre matematik öğretiminin netliğinin daha iyi olduğunu belirtmektedirler. Matematik dersinde disiplinsiz davranış (MDDD) değişkenine göre kız öğrenciler erkek öğrencilere göre matematik dersinde disiplinsiz davranışların daha yüksek olduğunu ifade etmektedirler. Matematikte kendine güven (MKG) değişkenine göre ise kız öğrencilerin matematikte kendine güvenlerinin erkeklerden daha yüksek olduğu görülmektedir. Matematik öğrenmeyi sevme (MOS) ve Matematiğe değer verme (MDV) değişkenleri ise cinsiyete göre gruplar arasında anlamlı farklılık göstermemektedir.

Literatürde yer alan pek çok araştırma, matematik duyuşsal özelliklerinin matematik başarılarında etkili olduğunu desteklemektedir (Pajares ve Miller, 1994; Thien ve Ong, 2015). Matematik duyuşsal özellikleri modelinin cinsiyete göre ölçme değişmezliğini sağlaması, oluşturulan modelin Türk öğrencilerinde yanlı davranmadığını ve geçerli-güvenilir ölçümler sağladığını kanıtlar niteliktedir. Dolayısıyla ölçme değişmezliğinin sınıandıği ölçme modelinin kız ve erkek öğrenci grupları arasında aynı şekilde ölçüm yaptığını göstermektedir. Matematik duyuşsal özellikler modeli ile kurulacak Yapısal Eşitlik Modeli gibi ileri düzey istatistiksel modellerde cinsiyete göre yapılacak karşılaştırmalarda bu yapının güvenle kullanılabileceği sonucuna ulaşılmıştır.

Bu araştırmanın yalnızca Türkiye örneğinde gerçekleştirilmesi bu çalışmanın sınırlılığıdır. Gelecek çalışmalarda uluslararası geniş ölçekli değerlendirmelerden elde edilecek öğrenci, öğretmen veya okul anketleri ile cinsiyet, bölge, ülke veya kıtalara göre karşılaştırmalar yapılabilir. Yapılacak çalışmalar ile bölge, ülke veya kıtalara göre ölçülen yapının daha iyi anlaşılacağı, ortaya çıkan farklılıkların çeviri, kültürel veya sosyolojik farklılıklar ile açıklanabilmesi açısından yararlı olacaktır.

EKLER

EK 1. Araştırmada Kullanılan R Kodları

```

# Paketlerin Kutuphaneden Çağırılması
library("lavaan")
library("dplyr")
library("semPlot")
library("haven")
# Masaüstündeki "bsgturm7.sav" adlı TIMSS Verisinin R'a aktarılması
bsgturm7 <- read_sav("C:/Users/user/Desktop/bsgturm7.sav")
# Veriden gerekli değişkenlerin çekilmesi
data <- bsgturm7 %>%
select(starts_with(c("BSBG01", "BSBM16", "BSBM17", "BSBM18", "BSBM19",
"BSBM20", "TOTWGT")))

##### DFA ANALIZI #####
# DFA Modeli
model <- 'MOS =~ BSBM16A + BSBM16B + BSBM16C + BSBM16D + BSBM16E+
BSBM16F + BSBM16G + BSBM16H + BSBM16I
MON =~ BSBM17A + BSBM17B + BSBM17C + BSBM17D + BSBM17E+
BSBM17F + BSBM17G
MDDD =~ BSBM18A + BSBM18B + BSBM18C + BSBM18D + BSBM18E+
BSBM18F
MKG =~ BSBM19A + BSBM19B + BSBM19C + BSBM19D + BSBM19E+
BSBM19F + BSBM19G + BSBM19H + BSBM19I
MDV =~ BSBM20A + BSBM20B + BSBM20C + BSBM20D + BSBM20E+
BSBM20F + BSBM20G+BSBM20H + BSBM20I
MOS~~MON
MOS~~MDDD
MOS~~MKG
MOS~~MDV
MON~~MDDD
MON~~MKG
MON~~MDV
MDDD~~MKG
MDDD~~MDV
MKG~~MDV'
# DFA Analizi
fit <- cfa(model, data = data, estimator = "ULS", sampling.weights =
"TOTWGT")
# DFA Analizinden model uyum indekslerinin ve katsayıların elde
edilmesi
summary(fit, standardized = TRUE, fit.measures = TRUE)
##### OLCME DEĞİŞMEZLİĞİ ANALİZLERİ #####
class(data$BSBG01) <- "numeric"
# Yapısal değişmezlik modeli
Yapısal <- cfa(model, data = data, sampling.weights = "TOTWGT",
estimator = "ULS", group = "BSBG01")
# Yapısal değişmezlik modelinden uyum indekslerinin ve katsayıların
elde edilmesi
summary(Yapısal, fit.measures = TRUE, standardized = TRUE)

# Metrik değişmezlik modeli
Metrik <- cfa(model, data = data, sampling.weights = "TOTWGT",
estimator = "ULS", group = "BSBG01",
group.equal = "loadings")
# Metrik değişmezlik modelinden uyum indekslerinin ve katsayıların
elde edilmesi
summary(Metrik, fit.measures = TRUE, standardized = TRUE)

```

```
# Olcek degismezligi modeli
Olcek <- cfa(model, data = data, sampling.weights = "TOTWGT",
estimator = "ULS", group = "BSBG01",
group.equal = c("loadings","intercepts"))
# Olcek degismezligi modelinden uyum indekslerinin ve katsayilarin
elde edilmesi
summary(Olcek, fit.measures = TRUE, standardized = TRUE)

# Kati degismezlik modeli
Kati <- cfa(model, data = data, sampling.weights = "TOTWGT", estimator
= "ULS", group = "BSBG01",
group.equal = c("loadings","intercepts",
"residuals"))
# Kati degismezlik modelinden uyum indekslerinin ve katsayilarin elde
edilmesi
summary(Kati, fit.measures = TRUE, standardized = TRUE)

# DFA modeline göre grafigin cizilmesi
semPaths(fit, what = "std", style = "lisrel", whatLabels = "std",
rotation = 3, layout = "circle", mar = c(2, 2, 2, 2),
nCharNodes = 15,
sizeMan = 6, sizeMan2 = 2, groups = "latents",
color = c("#011D70","#D10000","#58A069","#BF5B04","black"),
residuals = TRUE)
```

KAYNAKLAR

- Akben-Selcuk, E. (2017). Personality, motivation, and math achievement among Turkish students: Evidence from PISA data. *Perceptual and Motor Skills*, 124(2), 514–530. <https://doi.org/10.1177/0031512516686505>
- Alatlı, B. (2020). Cross-cultural measurement invariance of the items in the science literacy test in the Programme for International Student Assessment (PISA-2015). *International Journal of Education and Literacy Studies*, 8(2), 16. <https://doi.org/10.7575/aiac.ijels.v.8n.2p.16>
- Arseven, D. A. (1986). Çocukta Benlik Gelişimine Ailenin Etkisi ve Çocuğun Okuldaki Başarısı. *Eđitim ve Bilim*, 10 (60), 11-17.
- Aybek E.C. (2022). Doğrulatory Faktör Analizi. Göçer Şahin S. ve Buluş, M. (Ed.), Adım Adım Uygulamalı İstatistik içinde (343-372). Pegem Yayınevi.
- Bađdu Söyler, P., Aydın, B. ve Atılgan, H. (2021). PISA 2015 reading test item parameters across language groups: A measurement invariance study with binary variables. *Eđitimde ve Psikolojide Ölçme ve Deđerlendirme Dergisi*, 112–128. <https://doi.org/10.21031/epod.800697>
- Başusta, N. B. ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme deđermezliğinin test edilmesi: PISA öğrenci anketi örneđi. *Hacettepe Üniversitesi Eđitim Fakültesi Dergisi*, 30(4), 80-90.
- Bofah, E. A.-T. ve Hannula, M. S. (2015). TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties. *Large-Scale Assessments in Education*, 3(1). <https://doi.org/10.1186/s40536-015-0014-y>
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Büyüköztürk, Ş. (2011). *Sosyal Bilimler İçin Veri Analizi El Kitabı*, 14. Baskı, Ankara: Pegem yayıncılık.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2017). Bilimsel araştırma yöntemleri. *Pegem*, 1-360.
- Cardoso, M. E. (2020). Policy evidence by design: International large-scale assessments and grade repetition. *Comparative Education Review*, 64(4), 598-618.
- Cheung, G. W. ve Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- Çakici Eser, D. (2021). Investigation of measurement invariance according to home resources: TIMSS 2015 mathematical affective characteristics questionnaire. *International Journal of Assessment Tools in Education*, 633–648. <https://doi.org/10.21449/ijate.817168>
- Çiftçi, Ş. K. ve Yıldız, P. (2019). The Effect of Self-Confidence on Mathematics Achievement: The Meta-Analysis of Trends in International Mathematics and Science Study (TIMSS). *International Journal of Instruction*, 12(2), 683-694. <https://doi.org/10.29333/iji.2019.12243a>
- Ding, Y., Yang Hansen, K. ve Klapp, A. (2022). Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFA and the alignment method. *European Journal of Psychology of Education*. <https://doi.org/10.1007/s10212-022-00623-y>
- Engel, L. C. ve Rutkowski, D. (2021). Costs of big data. In *Digital Disruption In Teaching And Testing* (pp. 124–135). Routledge.
- Ersozlu, Z., Usak, M. ve Blake, D. (2022). Using Multi-Group Invariance analysis in exploring cross-cultural differences in mathematics anxiety: A comparison of Australia and Russia. *Journal of Ethnic and Cultural Studies*, 9(1), 1–18. <https://doi.org/10.29333/ejecs/987>
- Ertürk, Z. ve Erdiñç-Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı deđerşkenlerin cinsiyete göre ölçme deđermezliğinin incelenmesi. *Journal of Theoretical Educational Science*, 204-226.
- F. Hair, J., Jr, Sarstedt, M., Hopkins, L. ve G. Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106–121. <https://doi.org/10.1108/eb-10-2013-0128>
- Forero, C. G., Maydeu-Olivares, A. ve Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 625–641. <https://doi.org/10.1080/10705510903203573>

- Fornell, C. ve Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *JMR, Journal of Marketing Research*, 18(3), 382–388. <https://doi.org/10.1177/002224378101800313>
- Guo, J., Parker, P. D., Marsh, H. W. ve Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, 51(8), 1163–1176. <https://doi.org/10.1037/a0039440>
- Gustafsson, J.-E. (2018). International large scale assessments: Current status and ways forward. *Scandinavian Journal of Educational Research*, 62(3), 328–332. <https://doi.org/10.1080/00313831.2018.1443573>
- Güngör, M. ve Atalay Kabasakal, K. (2020). Investigation of measurement invariance of science motivation and self-efficacy model: PISA 2015 turkey sample. *International Journal of Assessment Tools in Education*, 207–222. <https://doi.org/10.21449/ijate.730481>
- Güvendir, M. A. (2016). Students' extrinsic and intrinsic motivation level and its relationship with their mathematics achievement. *International Journal for Mathematics Teaching and Learning*, 17(1).
- He, J., Barrera-Pedemonte, F. ve Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education Principles Policy and Practice*, 26(4), 369–385. <https://doi.org/10.1080/0969594x.2018.1469467>
- Hooper, M., Mullis, I. V., Martin, M. O. ve Fishbein, B. (2020). TIMSS 2019 context questionnaire framework. *Timss*, 59-78.
- Horn, J. L. ve McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3–4), 117–144. <https://doi.org/10.1080/03610739208253916>
- Hu, L.-T. ve 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>
- Jöreskog, K. G. ve Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific software international.
- Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, 10, 1258. <https://doi.org/10.3389/fpsyg.2019.01258>
- Kam, C. C. S. ve Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 18(3), 512–541. <https://doi.org/10.1177/1094428115571894>
- Kıbrıslıoğlu, N. (2015). The investigation of measurement invariance PISA 2012 mathematics learning model according to culture and gender: Turkey - China (Shangai) – Indonesia [Master Thesis] Hacettepe University.
- Kline, R. B. 2011. “Convergence of Structural Equation Modeling and Multilevel Modeling.” In *The SAGE Handbook of Innovation in Social Research Methods*, edited by M. Williams and W. P. Vogt, 562–589. SAGE Publications. doi:10.4135/9781446268261.
- Koçar, H. ve Yılmaz Koçar, E. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2). <https://doi.org/10.21031/epod.94857>
- Ma, X. (1997). Reciprocal relationships between attitude toward mathematics and achievement in mathematics. *The Journal of Educational Research*, 90(4), 221–229. <https://doi.org/10.1080/00220671.1997.10544576>
- Malone, T. W. ve Lepper, M. R. (2021). Making learning fun: A taxonomy of intrinsic motivations for learning. In *Aptitude, learning, and instruction* (pp. 223-254). Routledge.
- Marsh, H. W. ve Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 1(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.

- Milfont, T. L. ve Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of psychological research*, 3(1), 111-130.
- Milli Eğitim Bakanlığı (2020). TIMSS 2019 Türkiye Ön Raporu. URL : https://odsgm.meb.gov.tr/meb_iys_dosyalar/2020_12/10175514_TIMSS_2019_Turkiye_On_Raporu_.pdf
- Millsap, R. E. ve Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle, (Ed.) *Handbook of structural equation modeling*, (pp. 380-392), Guilford.
- Mullis, I. V. S. ve 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/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L. ve Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results/>
- Mullis, I. V. ve Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.
- OECD (2013) PISA 2012 results: excellence through equity: giving every student the chance to succeed, vol II. PISA, OECD Publishing, Paris
- Polat, M. (2019). TIMSS-2015 Matematik Ve Fen Duyuşsal Özellik Modellerinin Kültürlere, Cinsiyete ve Bölgelere Göre Ölçme Değişmezliğinin İncelenmesi. Ankara: Hacettepe Üniversitesi.
- Putnick, D. L. ve Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review: DR*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Radovan, M. ve Makovec, D. (2015). Relations between students' motivation, and perceptions of the learning environment. *Center for Educational Policy Studies Journal*, 5(2), 115–138. <https://doi.org/10.26529/cepsj.145>
- Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, 35(2), 299–331. [https://doi.org/10.1016/s0005-7894\(04\)80041-8](https://doi.org/10.1016/s0005-7894(04)80041-8)
- Reynolds, K., Khorramdel, L. ve von Davier, M. (2022). Can students' attitudes towards mathematics and science be compared across countries? Evidence from measurement invariance modeling in TIMSS 2019. *Studies in Educational Evaluation*, 74(101169), 101169. <https://doi.org/10.1016/j.stueduc.2022.101169>
- Rosseel, Y. (2012). lavaan: AnRPackage for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Rutkowski, L. ve Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>
- Ryan RM, Deci EL (2009) Promoting self-determined school engagement: motivation, learning, and well-being. In: Wentzel KR, Wigfield A (eds) *Handbook on motivation at school*. Routledge, New York, pp 171–196.
- Schermelleh-Engel, K., Moosbrugger, H. ve Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23-74.
- Schleicher, A. (2019). *PISA 2018: Insights and Interpretations*. OECD Publishing.
- Schmitt, N. ve 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>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. ve King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/joer.99.6.323-338>
- Schumacker, R. E. ve Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Psychology Press.
- Şeyma, Uyar. ve Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 3, 30-43.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6th edition). Northridge: Pearson.

- Taris, T. W., Bok, I. A. ve Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *The Journal of Psychology*, 132(3), 301–316. <https://doi.org/10.1080/00223989809599169>
- Tavani C.M., Losh S.C. (2003) Motivation, self-confidence, and expectations as predictors of the academic performances among our high school students. *Child Study J* 33(3):141–151.
- Tekin, Y. T. ve Çobanoğlu Aktan, D. (2021). Investigation of measurement invariance of PISA 2015 collaborative problem solving skills: Turkey, Norway and Singapore. *International Journal of Assessment Tools in Education*, 90–105. <https://doi.org/10.21449/ijate.690576>
- Thien, L. M. ve Ong, M. Y. (2015). Malaysian and Singaporean students' affective characteristics and mathematics performance: evidence from PISA 2012. *Springer Plus*, 4(1), 563. <https://doi.org/10.1186/s40064-015-1358-z>
- Uyar, Ş. ve Kaya Uyanık, G. (2019). Fen Bilimlerine Yönelik Öğrenme Modelinin Ölçme Değişmezliğinin İncelenmesi: PISA 2015 Örneği. *Kastamonu Eğitim Dergisi*, 497–507. <https://doi.org/10.24106/kefdergi.2570>
- Uyar, Y. D. D. Ş. (2021). Factor structure and measurement invariance of the TIMSS 2015 mathematics attitude questionnaire: Exploratory structural equation modelling approach. *International Journal of Assessment Tools in Education*, 8(4), 855–871. <https://doi.org/10.21449/ijate.796862>
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K. ve Zondervan-Zwijenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Van der Bergh, E. (2013). *The influence of academic self-confidence on mathematics achievement* (Doctoral dissertation, North-West University).
- Vandenberg, R. J. ve 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.
- Vandenberg, R. J. ve 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>
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186-191.
- Wu, A. D., Li, Z. ve Zumbo, B. D. (2007). *Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data*. University of Massachusetts Amherst. <https://doi.org/10.7275/MHQA-CD89>
- Yagan, S. A. (2021). The relationships between instructional clarity, classroom management and mathematics achievement: Mediator role of attitudes towards mathematics. *University of South Florida M3 Center Publishing*, 3, 7.
- Yandı, A., Köse, İ. A. ve Uysal, Ö. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 243-253.
- Yiğiter, M. S. (2019). *Öğretmenlerin teknoloji kullanımının mesleki motivasyonlarına etkisi: Çankaya ilçesi örneği*. Yayımlanmamış Yüksek Lisans Tezi, Ankara Hacı Bayram Veli Üniversitesi, Lisansüstü Eğitim Enstitüsü.
- Yin, L. ve Fishbein, B. (2019). Creating and interpreting the TIMSS 2019 context questionnaire scales. *Methods and procedures: TIMSS*, 16-1.
- Zeng, D., Takada, N., Hara, Y., Sugiyama, S., Ito, Y., Nihei, Y. ve Asakura, K. (2022). Impact of intrinsic and extrinsic motivation on work engagement: A cross-sectional study of nurses working in long-term care facilities. *International Journal of Environmental Research and Public Health*, 19(3), 1284. <https://doi.org/10.3390/ijerph19031284>

EXTENDED ABSTRACT

Since the end of the 20th century, International Large-Scale Assessment (ILSA) programs have started to be carried out with the participation of many countries in order to make comparisons on the education systems of countries and to determine student achievements (Cardoso, 2020). One of the main aims of international large-scale assessments is to make comparisons between different countries, education policies, education systems or subgroups. TIMSS is an evaluation program that evaluates the academic achievement of 4th and 8th grade students over the fields of mathematics and science skills, every four years, and monitors the differences in achievement between countries over time and the results of countries' attempts to increase the level of success (Mullis & Martin, 2017). In addition to measuring math and science skills, TIMSS includes many scales and questionnaires that measure cognitive and affective characteristics of students, teachers, and schools (Yin & Fishbein, 2019). There are five different scales measuring mathematics affective characteristics in the TIMSS 2019 student questionnaire.

One of the main criteria for making comparisons between different groups is to satisfy measurement invariance. Measurement invariance indicates that the measured construct is psychometrically equivalent between groups. Claims of differences in comparisons without evidence of measurement invariance can be unreliable. He et al. (2019) stated that in their research on TIMSS and PISA data and cross-cultural comparability, comparisons made without examining measurement invariance may lead to incorrect results, therefore it is important to testing measurement invariance.

One of the most used methods in the literature to test measurement invariance is Confirmatory Factor Analysis (CFA) (Schmitt and Kuljanin, 2008; Van De Schoot et al., 2015). Measurement invariance is tested in four hierarchical stages with the Multi-Group CFA (MG-CFA) method. These stages are; configural invariance, metric invariance, scalar invariance and strict invariance (Meredith, 1993).

The aim of this study was to test the measurement invariance of the model created with mathematics affective characteristics according to gender. For this purpose, the Mathematics Affective Characteristics Model was created with the scales of Like Learning Mathematics (MOS), Instructional Clarity in Mathematics Lessons (MON), Disorderly Behavior During Mathematics Lessons (MDDD), Students Confident in Mathematics (MKG) and Students Value Mathematics (MDV) in the TIMSS 2019 cycle.

Before starting the data analysis, the assumptions that must be tested in all multivariate statistical techniques are examined. In the data obtained from the TIMSS 2019 database, there are 4077 students from Turkey at the 8th grade level. Before proceeding to the analysis, missing data, univariate and multivariate outlier analysis, multicollinearity analysis, normal distribution analysis, multiple normal distribution analysis and homogeneity of variances were tested. The data provide the normal distribution, but not the assumptions of multivariate normal distribution and homogeneity of variances. Therefore, the "ULS" estimation method, which is resistant to the violation of these assumptions, was used. The sample of the study consists of a total of 3658 students, 1848 girls and 1810 boys.

Multi-Group CFA analyzes were performed in R using the lavaan (Rosseel, 2012) package. Reason for choosing this package is that it allows the use of sample weights when estimating with the Multi-Group CFA model. Since it is not possible in terms of time and money in large-scale tests, all participants in the universe (all students at the relevant grade level) cannot be included in the sample. In

order to overcome this limitation and to provide the generalizability of the sample to the population, sample weights are used (Arıkan et al., 2020).

Measurement invariance in MG-CFA was examined by testing four hierarchical models. These four models are configural invariance, metric invariance, scalar invariance, and strict invariance (Vandenberg & Lance, 2000). The values of χ^2 , χ^2/df , *RMSEA* (*RMSEA* < .08), *SRMR* (*SRMR* < .08), *TLI* (*TLI* > .95), *CFI* (*CFI* > .95), and ΔCFI to examine the model data fit between stages taken into account (Hu & Bentler, 1999; Schermelleh Engel, Moosbrugger & Müller, 2003; Schumacker & Lomax, 2004). Cheung and Rensvold (2002) suggested examining the different in CFI value (ΔCFI) as an alternative to $\Delta\chi^2$. In this study. The ΔCFI difference being less than or equal to 0.01 was used as a criterion to determine whether invariance was achieved between the stages (Cheung & Rensvold, 2002).

The sample of the study consists of 3658 students from Turkey who participated in the TIMSS 2019 cycle at the 8th grade level. In the first part of the study, Confirmatory Factor Analysis (CFA) was conducted to examine the factor structure of the mathematics affective characteristics model. DFA model results show that model data fit is reached (*RMSEA*=0.046, *SRMR*=0.051, *CFI*=0.973 and *TLI*=0.975). In the measurement invariance analysis, it was tested hierarchically between the stages with Multi-Group CFA (MG-CFA) analysis. The findings show that the mathematics affective characteristics model meets the configural, metric, scalar, and strict invariance stages, respectively. Therefore, the factor loadings, variances, error variances and covariances of the mathematics affective characteristics model were equivalent according to gender, and it was concluded that significant comparisons could be made between the groups. In the literature, there are studies examining measurement invariance on mathematics affective characteristics scales. Ertürk and Erdiñç-Akan (2018), Polat (2019) and Çakıcı Eser (2021) focused on the mathematical affective characteristics scales of the TIMSS 2015 application in their studies. The results of this study show that there is a strict invariance between affective characteristics and gender, similar to the results of the studies of Polat (2019), Çakıcı Eser (2021), and Uyar (2021). It can be said that it shows partial similarity with the results of the study of Ertürk and Erdiñç-Akan (2018).

After testing the measurement invariance of the mathematics affective characteristics model according to gender, it was examined whether the variables in the model showed significant differences according to gender. While three of the five variables in the model showed significant differences according to gender, two variables did not show significant differences according to gender. According to Instructional Clarity in Mathematics Lessons (MON) variable, male students stated that the clarity of mathematics teaching was better than female students. According to the variable of Disorderly Behavior During Mathematics Lessons (MDDD), female students state that undisciplined behavior in mathematics lesson is higher than male students. According to the variable of Students Confident in Mathematics (MKG), it is seen that female students have higher self-confidence in mathematics than male students. The variables of Like Learning Mathematics (MOS) and Students Value Mathematics (MDV) do not show significant differences between the groups according to gender.