TIMSS 2015 Sekizinci Sınıf Matematik Başarı Testinin OECD Ülkelerine Göre Ölçme Değişmezliğinin İncelenmesi

Öğrt. Özge Öncü

ORCID ID: <https://orcid.org/0000-0002-8642-0000>

T. C. Milli Eğitim Bakanlığı, Antalya – TÜRKİYE

Doç. Dr. Hakan Koğar

ORCID ID: <https://orcid.org/0000-0001-5749-9824>

Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Antalya – TÜRKİYE

Makale Geçmişi

Geliş: 16.11.2020

Kabul: 12.06.2021

On-line Yayın: 30.06.2021

Anahtar Kelimeler

TIMSS 2015

Matematik Başarısı

Ölçme Değişmezliği

Çoklu-Grup Doğrulamalı Faktör Analizi

Araştırma Makalesi

* Bu çalışma, birinci yazar Özge Öncü tarafından ikinci yazar Doç. Dr. Hakan Koğar danışmanlığında 2019 yılında hazırlanan "TIMSS 2015 Sekizinci Sınıf Matematik Başarı Testinin OECD Ülkelerine Göre Ölçme Değişmezliğinin İncelenmesi" başlıklı Yüksek Lisans Tezi'nden üretilmiştir.

DOI:

<http://dx.doi.org/10.29228/JASSS.47644>

Öz

Bu araştırmanın amacı Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS-Trends in International Mathematics and Science Study) 2015 verileri kullanılarak, matematik başarı testine katılan 8. Sınıf öğrencilerinin matematik başarılarının İktisadi İşbirliği ve Kalkınma Teşkilatı (The Organisation For Economic Co-Operation and Development OECD) ülkelerine göre ölçme değişmezliğinin incelenmesidir. Araştırma 11 no'lu kitapçığı almış olan 7.223 öğrenciye ait veri seti üzerinden yürütülmüş olup, ülkelerin 8. Sınıf matematik başarı testi puanları için, test istatistikleri, normallik testleri ve güvenilirlik katsayılarına bakılmış, merkezi eğilim ve dağılım ölçülerinin birbirine yakın olduğu sonucuna varılmıştır. Her ülke için ayrı ayrı doğrulamalı faktör analizi yapılarak χ^2/sd oranının oldukça iyi olduğu sonucuna ulaşılmıştır. Uyum indeksleri genel olarak değerlendirildiğinde tüm ülkelere ait ölçme modelleri tüm gruplar için ayrı ayrı doğrulanmıştır. OECD üyesi ülkelerin matematik başarı testi puanları için kovaryans matrislerinin eşitliği testi sonuçlarını ele aldığımızda, uyum indeksleri oldukça iyi olduğu görülmektedir. Ölçme değişmezlik çoklu-grup doğrulamalı faktör analizi ile test edilmiştir, yapısal değişmezlik için uyum indekslerinin değerleri kabul edilebilir düzeylerde ve model doğrulanmıştır. Araştırma sonuçları OECD üyesi ülkeler için ölçme değişmezliğinin sağlanmadığını, ölçme değişmezliğine ilişkin modeller arasında en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğunu göstermiştir. Tüm sonuçlar değerlendirildiğinde, TIMSS 2015 uygulamasında matematik başarısına ait elde edilen sonuçların farklı ülkeler, diller, cinsiyetler ve kültürlerle sahip katılımcılardan elde edilen sonuçların karşılaştırılabilirliği ve buna bağlı olarak yapılan yorumların ölçme değişmezliğinin sağlanmadığı görülmektedir. Uygulanan ölçme aracının tüm katılımcılar için aynı anlamı ifade etmemesi nedeniyle, TIMSS 2015 uygulaması tartışmaya açık hale gelmektedir. Bu araştırma geniş ölçekli sınav uygulamalarının sonuçları değerlendirilirken ölçme değişmezliğinin sağlanamıyor oluşuna, özellikle ülkeler arası karşılaştırmalar ve yorumlamalar yapılırken bu durumun titizlikle göz önünde bulundurulmasına dikkat çekmektedir.

Atıf Bilgisi / Reference Information

Öncü, Ö. ve Koğar, H. (2021). TIMSS 2015 Sekizinci Sınıf Matematik Başarı Testinin OECD Ülkelerine Göre Ölçme Değişmezliğinin İncelenmesi. *Jass Studies-The Journal of Academic Social Science Studies*, 14(85): 1-30.

An Investigation Into The Measurement Invariance According to OECD Countries of TIMSS 2015 Eighth Grade Math Achievement Test

Teach. Özge Öncü

T. C. Ministry of National Education, Antalya – TURKEY

Assoc. Dr. Hakan Koğar

Akdeniz University, Faculty of Education, Department of Educational Sciences, Antalya – TURKEY

Article History

Submitted: 16.11.2020

Accepted: 12.06.2021

Published Online: 30.06.2021

Keywords

TIMSS 2015

Mathematics Achievement

Measurement Invariance

Multi-Group Confirmatory Factor Analysis

Abstract

The research, by using Trends in International Mathematics and Science Study (TIMSS) 2015 data, was conducted on the data set of 7,223 students who received the booklet number 11. 8th grade math achievement test scores of the countries, test statistics, normality tests and reliability coefficients were examined. It is concluded that the central tendency and distribution measures are close to each other. Confirmatory factor analysis was performed for each country separately and it was concluded that χ^2 / sd ratio was quite good. When the fit indexes were evaluated in general, all countries' measurement models were verified separately for all groups. When we take the results of the equality test of covariance matrices for mathematics achievement test scores of The Organisation For Economic Co-Operation and Development (OECD) member countries, the fit indices are quite good. Measurement invariance was tested by multi-group confirmatory factor analysis, the values of the fit indices for structural invariance were acceptable and the model was confirmed. The results of the study showed that the measurement invariance could not be achieved for OECD member countries and that the best working model among the measurement invariance models was the strong factorial invariance model. When all the results are evaluated, it is seen that the results of mathematics achievement in TIMSS 2015 are comparable to the results obtained from participants from different countries, languages, genders and cultures and the measurement invariance related to these results cannot be achieved. TIMSS 2015 is open to discussion as the measurement tool does not have the same meaning for all participants. This research draws attention to the fact that measurement invariance cannot be achieved when evaluating the results of large-scale examination applications, and that this situation is taken into consideration carefully when making comparisons and interpretations between countries.

Research Article

DOI:

<http://dx.doi.org/10.29228/JASSS.47644>

GİRİŞ

Çağımızda her geçen gün teknoloji başta olmak üzere bütün alanlarda kendisini iyice hissettiren rekabet ortamı, gelişmekte olan ve gelişmiş ülkelerin en önemli hedefini olan sürdürülebilir gelişmeyi gerçekleştirmeye odaklanmaktadır. Dünya üzerinde gerçekleşen bu gelişmelerin aynı zamanda eğitim sistemini de derinden etkilediği söylenebilir.

Sovyetlerin uzaya ilk kozmonotlarını yollamaya başlaması ile dünyanın birçok ülkesinde teknoloji hızlı bir şekilde ilerlemiştir. Bulunduğumuz çağa damgasını vuran bilgi teknolojileri, özellikle 1900'lü yılların ikinci yarısından itibaren başta Amerika Birleşik Devletleri olmak üzere dünyanın birçok ülkesinde fen ve matematik bilimlerine ve bunların eğitime verilen önemin giderek artmasına sebep olmuştur ve eğitim hedefleri zaman içerisinde büyük değişimlere uğramıştır (Korkmaz, 2004). Ülkelerin statüsünü belirleyen en önemli unsurlardan biri olan bilim ve teknoloji toplumsal ilerlemenin başlangıç noktası olarak kabul edilmektedir (Atalay, Bayır ve Çakıcı, 2016). Özellikle teknolojik yarışta geri kalmak istemeyen bu ülkeler çareyi fen bilimleri ve ilgili alanlara önem vermekte ve bu önem doğrultusunda öğretimi geliştirmekte bulmuşlardır (Ayas, Çepni, Johnson ve Turgut, 1997). Teknolojik gelişmeler, insan kaynakları ve bu alanda verilen eğitim, ülkelerin gelişmesini, yaşam standartlarının yükselmesini, daha bilgili nesiller yetişmesini sağlamasından dolayı, kalite ve verimin merkezinde bulunması sebebiyle de bilgi toplumlarında oldukça önemli bir konumdadır (Ergül, 1999). Bu duruma günümüzde de oldukça önem verilmektedir ve artan bir şekilde devam etmektedir.

İnsan gücü kaynağının çağın gereklerine uygun bilgi ve becerilere sahip olarak yetiştirilmesi ülkelerin sosyal, ekonomik ve kültürel açıdan ilerlemesinin vazgeçilmez gereklerindendir. Bilgi ve teknolojilerin sürekli gelişmesi nedeniyle insan kaynağının gerekli donanımlarla yetiştirilmesi için eğitim ve öğretimin sürekli güncellenmesi gerekmektedir. Bu nedenle önemli bazı uluslararası kuruluşlar eğitim alanında araştırmalar yapmaktadır. Sonuçlarının ülkelerin eğitim politikalarına yön verdiği uluslararası düzeyde yapılan bu araştırmalar, ülkelerin eğitim sistemleri üzerine getirebilecekleri önemli birer eleştiri ve geri bildirim niteliğindedir.

Küreselleşen dünyada, eğitim politikalarının belirlenmesinde ve yeniden düzenlenmesinde etkili olan ulusal ölçme ve değerlendirme çalışmalarının yanı sıra uluslararası çalışmalar da önem kazanmaktadır. Uluslararası düzeyde ülkelerin eğitim sistemi açısından hangi konumda olduğunun, diğer ülkelerle olan farklılıklarının ve eğitim eksiklerinin belirlenebilmesi için birçok ülke tarafından referans olarak kabul edilen ve uluslararası düzeyde uygulanan sınavlar bulunmaktadır. Öğrenci başarıları göz önünde bulundurularak pek çok ülkede eğitim sistemi ve öğelerine yönelik eksikliklerin belirlenmesi ve gereksinim duyulan alanların geliştirilmesi çalışmaları yürütülmektedir (MEB, 2014). Bunların yüksek önem taşıyan araştırmalar olduğu düşünülmektedir çünkü değerlendiriciler, öğrenci performansını karşılaştırmak ve katılımcı ülkelerdeki eğitim politikalarının ve uygulamalarının etkililiğini değerlendirmek için bu araştırmaların sonuçlarını kullanmaktadır (Gierl, 2000).

Gelişmiş ülkelerin zamanın şartlarına uygun fen bilimleri ve matematik eğitim programlarını geliştirme çabası içinde olmaları, ülkemizin de bu doğrultuda eğitim sistemini sürekli gözden geçirmesi gerektiğini düşünmesine yol açmıştır. Avrupa ve Orta Asya bölgelerinde 29 ülkeyi içeren güncel bir araştırmaya göre eğitim, Türk vatandaşlarının endişelendiği en önemli konudur. (Çelebi, Güner, Kaya ve Korumaz, 2014). Bu doğrultuda Türkiye'de, fen bilimleri ve matematik eğitimindeki gelişmelerini daha iyi görebilmek için TIMSS (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment) ve PIRLS (Progress in International Reading Literacy Study) gibi uluslararası düzeyde karşılaştırma sınavlarına katılmıştır.

Türkiye'nin de katıldığı uluslararası düzeyde öğrenci başarılarını belirlemek amacıyla eğitim araştırmaları yürütülmektedir. Bu araştırmalardan biri de Uluslararası Matematik ve Fen Eğilimleri Araştırmasıdır (TIMSS-Trends in International Mathematics and Science Study). Öğrencilerin

matematik ve fen alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik bir tarama araştırması olan TIMSS, Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement) IEA'nın bir projesidir ve dört yıllık aralıklarla dördüncü ve sekizinci sınıf düzeyindeki öğrencilere uygulanmaktadır.

TIMSS'de yer alan matematik ve fen soruları daha önceden belirlenen kazanımlar çerçevesinde ülke temsilcileri tarafından ortak olarak hazırlanır. Hazırlanan bu sorular IEA'nın Fen ve Matematik maddeleri inceleme komitesi tarafından incelenir ve açık uçlu soruların puanlama anahtarları hazırlanır. Daha sonra oluşturulan yedek ve esas sorular, taslak bloklar halinde incelenir ve sorulara son şekli verilir. Hazırlanan sorular katılımcı ülkelerde çeviri, uyarlama işlemlerinden sonra pilot uygulama ile test edilir. İşleyen sorular önceki uygulama soruları ile birleştirilerek pilot uygulamadan bir yıl sonra nihai uygulamada yer alır.

Uluslararası Eğitim Başarılarını Değerlendirme Derneğine (IEA) göre TIMSS, genel olarak fenden çok uluslararası bir matematik değerlendirmesi olarak kabul görmektedir ve dördüncü ve sekizinci sınıflarda 1995'ten bu yana her dört yılda bir yapılmaktadır (Arora, Foy, Martin ve Mullis, 2012).

Ülkeler 4. ve 8. sınıf düzeyinde uygulanan sınava sadece 4. sınıf veya sadece 8. sınıf düzeyinde katılma hakkına sahip olmasının yanında, her iki düzeyde de katılabilme şansına sahiptir. Öğrencilerin bulunduğu okul ve sınıflar büyük bir titizlikle her bir ülkeyi temsil etmesi sağlanacak şekilde seçilmektedir. Toplam 28'e ayrılan parçalardan 14'ü matematik ve geriye kalan 14'ü fen bilimleridir. PISA'nın matematik ve fen okuryazarlığını ölçtüğü bir sınav olmasının aksine TIMSS'in genel olarak matematik ve fen başarısını ölçtüğü bir sınav olması sebebi ile her bir bloktaki sorular 4. ve 8. sınıf seviyesinde ve bilgiyi anlamadan çok bilme ve uygulama basamakları yoğun bir şekilde işlenen bir sınav olması ile tanınır.

TIMSS şablonu bir bütün olarak değerlendirilirse, her ülkenin farklı eğitim sistemi, farklı kültürü, farklı demografik yapısı, farklı düşünce tarzı ve farklı eğitim sistemi olduğu aşikârdır. Bir ülke eğitimde ahlâğa ve çalışkanlığa önem verirken, başka bir ülke bilgiyi bilmeye, daha başka bir ülke ise bilgiyi yorumlama ve anlamlandırmaya ve daha birçok ülke birçok kritere göre eğitimine o alanda önem vermektedir. Mesela bu kriterleri ülkeler ile ilişkilendirildiğinde, eğitimde ahlak ve çalışkanlığı genel olarak Japonya'nın yansıttığını, eğitimde bilgiyi bilmeye Türkiye'nin önem verdiğini ve verilen bilgiyi yorumlama ve anlamlandırmaya da Finlandiya'nın önem verdiği söylenebilir. TIMSS uygulamasına katılan 4. ve 8. sınıf öğrencilerinin bulunduğu okullarda bu sınavın nasıl bir süreçten geçtiği, her ülkenin eğitim sisteminde nasıl ele alındığı, okul uygulamalarındaki benzer ve farklı yönleri, ne gibi dönütler aldığı, ne kadar verimle yordama yapabildiği vb. bilgileri TIMSS havuzunda karşılaştırmaş bilgi düzeyinde toplanır ve bu bilgiler değerlendirilerek sınav sonu raporu hazırlanır. Tüm bu bilgiler TIMSS'in amacını ifade etmektedir ve matematik ve fen bilimleri başarısını yordayan bu proje belirtilen karşılaştırmalı bilgilerden yola çıkarak her bir ülkenin eğitim sisteminde bulunan eksiklikleri, farklılıkları ve olumlu yanları analiz ederek, öğrencilerin bu sınavlardaki başarılarının hangi kritere daha yaklaştığını göstermektedir. Bununla birlikte bu proje uygulamasında öğrencilere çeşitli anketler ve başarı testleri uygulanmış, bu anket ve başarı testlerine bağlı olarak, okulların yönetim, eğitim-öğretim düzeyi şeması, öğrencilerin karakteristik özellik ve bilgi düzeyi şeması, eğitim ve öğretim programları, bu programlara dayalı olarak işlenen ders kazanımları, sınava katılan öğrencilerin matematik ve fen başarıları ölçülmüştür. Bu bilgiler ışığında tüm veriler TIMSS havuzunda toplanarak her bir ülkenin eğitim durumuna değinilmiştir. (EARGED, 2014; TIMSS, 2015.).

Bu projenin bir diğer amacı, matematik ve fen alanında öğretimi ve öğrenmeyi geliştirmek için kullanılabilecek önemli bir arka plan bilgisi sağlamaktır. Örneğin TIMSS, müfredat ve müfredat uygulaması, öğretim uygulamaları ve okul kaynakları hakkında ayrıntılı bilgi toplar. Bu amaç

doğrultusunda başarı testleri ve çeşitli anketler kullanılarak öğrencilerin fen ve matematik alanındaki performansları, eğitim sistemleri, öğretim programları, öğrenci özellikleri, öğretmen ve okulların karakteristik özellikleri ile ilgili bilgiler toplanmaktadır. TIMSS, düzenli aralıklarla başarıdaki değişiklikleri izlediğinden, yeni veya gözden geçirilmiş eğitim politikalarının başarıyı etkileyip etkilemediğini araştırmak için değerli bir araçtır.

TIMSS 2015 uygulaması, eğitim sistemindeki performansların daha iyi anlaşılması ve bu konudaki uzmanlara yön vermesi amacıyla yapılmakta olup, matematik ve fen alanında iki ayrı eğitim seviyesindeki (4. ve 8. sınıf) öğrencilerin bilgilerini karşılaştırmaktadır. Bu karşılaştırmalar yapılırken bir takım veriler de toplanır. Bunlar: eğitim sistemi, idareciler, öğretmen, okul bilgileri ve öğrenci anketleri gibi verilerdir. Böylece, matematik ile fen başarısında eğitim sisteminin nasıl işlediği konusunda bilgiler sağlanırken, eğitimde yapılacak reform ve gelişmeler için de değerli birer kaynak oluşturulmaktadır (Arora, Foy, Martin and Mullis, 2012).

Böylesine detaylı olan TIMSS uygulamasının merkezini Boston College ve TIMSS, PIRLS Uluslararası Çalışma Merkezi oluşturur. Bu birimler tarafından ele alınan her ülkeye dair matematik ve fen başarısı, anketler vb. datalar bu birimler bünyesinde depolanır ve yorumlanır. Ayrıca bu ulusal kurum uluslararası kurumlarla bağlantı halinde olup gerekli olan verileri işbirliği içerisinde devam ettirmektedir. Türkiye’de TIMSS projesinin başında Milli Eğitim Bakanlığına bağlı Ölçme ve Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü (ÖDSGM) vardır.

Uluslararası sınavlar kadar bu sınavların ölçme kriterlerine uygunluğu da çok önemlidir. Kültürler arası karşılaştırmaların geçerliliği, psikoloji ve eğitimle ilgili araştırmalardaki birçok uygulama için öneme sahiptir. Kültürler arası araştırmanın geçerlilik kanıtı farklı ülkelerden elde edilen test puanlarının aynı metriğe ait aynı yapıyı ölçmesidir (Li, Wu and Zumbo, 2007). TIMSS ülkelerin fen ve matematik başarılarını tek bir ölçme aracıyla değerlendirip karşılaştıran bir uygulamadır. Yapılan karşılaştırmaların anlamlılığından bahsedebilmek için kullanılan ölçme aracının farklı gruplarda aynı özelliği ölçtüğünün doğrulanması gerekmektedir (Öğretmen ve Uzun, 2010).

50 yıldan uzun bir süre önce literatüre girmiş olan ölçme değişmezliği testinin önemi ancak değişmezlik testi için kullanılan istatistiksel teknikler daha çok son yıllarda ulaşılabilir hale geldiğinden araştırmacıların ölçme değişmezliğini test etmesi günümüzde daha kolay ulaşılabilir ve uygulanabilir hale gelmiştir (Bornstein and Putnick, 2016). Faktör yapılarının denkliliğini araştıran ve yazan ilk araştırmacı Jöreskog (1971) olmuştur. Ölçme değişmezliği kavramı ise ilk olarak Byrne, Muthe’n ve Shavelson (1989) tarafından tanıtılmış ve daha sonra test edilmiştir.

Byrne ve Watkins’e (2003) göre ölçme değişmezliği, tüm gruplarda ölçme aracındaki maddelerin tamamıyla aynı biçimde algılanıp yorumlanması şeklinde tanımlanmaktadır. Ölçme değişmezliğinin temelinde, karşılaştırma yapılan gruplar için yapılan ölçümlerin geçerliliğinin sağlanması yatar (Tyson, 2004). Herdman’a (1998) göre ölçme değişmezliği bir psikolojik ölçme aracının güvenilirlik ve yapı geçerliliği gibi psikometrik niteliklerinin farklı gruplardaki eşitliğinin biçimsel olarak değerlendirilmesi şeklinde ifade edilmektedir (Akt: Moraes ve Reichenheim, 2002). Ölçme değişmezliği, gizil bir yapıyı ortaya koymak için geliştirilmiş bir ölçeğin psikometrik bir özelliğidir. Aynı yapı ülkeler, kültürel birimler, zaman hususları veya ülkeler içindeki bölgeler gibi farklı gruplar arasında aynı şekilde ölçüldüğünde ölçümün değişmez olduğu kabul edilir (Davidov ve diğerleri, 2014; Horn and McArdle, 1992; Lance and Vandenberg, 2000; Meredith, 1993; Millsap, 2011). Ölçme değişmezliği sağlandığında gizil değişken üzerinde yapılan gruplar arası karşılaştırmaların geçerliliği sağlanabilir; aynı zamanda gelecekte yapılacak olan çalışmalarla gizil değişken puanlarının oluşumu, belirleyicileri ve sonuçları da kıyaslanarak değerlendirilebilir (Hox, Lugtig and Schoot, 2012).

Uzun ve Öğretmen (2010), TIMSS 1999 verileri üzerinde yürüttükleri çalışmalarında, öğrencilerin fen bilimleri başarısını belirleyen değişkenleri belirlemişler ve bu değişkenlerin cinsiyet

bazında değişmezliğini incelemişlerdir. Özyeterlik, önem, tutum ve sınıf içi öğrenci etkinlikleri için aşamalı şekilde ölçme değişmezliği testleri yapılmış ve elde edilen uyum indeksleri sonucunda modelde yer alan değişkenlerin yalnızca metrik değişmezlik modeline kadar uyum gösterdikleri görülmüştür. Cinsiyet alanında oluşturulan grupların ise katı değişmezlik koşullarını sağlayamadıkları belirlenmiştir.

Ölçülüoğlu (2015), TIMSS 2011 Türkiye sekizinci sınıf matematik başarısını etkileyen faktörlerin bölgelere göre incelenmesini ele alan yüksek lisans çalışmasında, TIMSS 2011 verileri ile Türkiye'deki sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenleri modellemiş ve modelin coğrafi bölgelere göre ölçme değişmezliğini inceleyerek değişkenlerin bölgeler bakımından farklılık gösterip göstermediğini analiz etmiştir. Matematik başarısının yordandığı yapısal model, yedi coğrafi bölgede ayrı ayrı test edilmiş ve modelin bütün bölgelerde geçerli olduğu bulunmuştur. YEM ile kurulmuş model sonucunda duyuşsal özellikler değişkeninin matematik başarısını belirlemede en yüksek paya sahip olduğunu göstermiştir. Daha sonra modele alınan değişkenlerin coğrafi bölgeler arasında karşılaştırılabilmesi için matematik başarısına ilişkin kurulan modelin çoklu grup doğrulayıcı faktör analizi ile ölçme değişmezliği incelenmiştir. Modele alınan değişkenlerinin bölgelere ait regresyon katsayılarındaki farklılıklar incelendiğinde okul gizil değişkeninin İç Anadolu ve Güneydoğu Anadolu bölgesi hariç ülke genelindeki gibi matematik başarısını yordamadığı saptanmıştır.

Arıkan, Sarı ve Yıldızlı (2017), TIMSS 2015 sekizinci sınıf matematik akademik başarısını yordayan faktörler çalışmalarında, matematik başarısındaki farklılıklarda öğrenci ile ilgili değişkenlerin (öğrenci, öğretmen ve okul faktörlerinin) %34'ünün açıkladığını dile getirmişlerdir. Duyuşsal özellik olan öz-yeterlik inancı 8.sınıf öğrencilerinin matematik başarılarını yordamada en önemli değişken olduğu ve diğer önemli değişkenin öğrencilerin evde sahip oldukları eğitimsel kaynakları olduğunu ifade etmiştir. Duyuşsal alan içerisinde yer alan tutum ile matematik başarısı arasında negatif yönde ilişki gözlemlendiğini söylemişlerdir. Araştırmada zorbalık, okula aidiyet ve öğretim etkinlikleri ile matematik başarısı arasında anlamlı bir ilişkinin olup bu ilişki matematik başarısını yordamada daha az öneme sahip olduğunu vurgulamışlardır. Tüm değişkenler bir aradayken okullardaki 8. sınıflar arası başarı farkının %29'unu açıklanmaktadır. Öğrenmede çevre faktöründe yer alan öğrencilerden kaynaklı sorunlar ve okulda başarıya verilen önem öğrencilerin matematik başarısını yordamadaki en önemli iki değişken olduğu sonucuna ulaşmışlardır.

Ayvallı (2016), PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğinin incelenmesini ele aldığı yüksek lisans çalışmasında, matematik okuryazarlığı testinin Türkiye'deki cinsiyet, bölge grupları ve OECD ülkeleri arasında ölçme değişmezliğinin sağlanıp sağlanmadığını çoklu grup doğrulayıcı faktör analizi aracılığıyla test etmiştir. Tüm gruplarda ayrı ayrı doğrulayıcı faktör analizi sağlaması yapıldıktan sonra kovaryans matris eşitliği de sağlanmış, ardından ölçme değişmezliği testleri yapılmıştır. Yapılan analizler sonucunda Türkiye'de cinsiyet ve bölge grupları için ölçme değişmezliğinin sağlandığı, OECD üyesi ülkeler için ise ölçme değişmezliğinin sağlanmadığı görülmüştür.

Karakoç Alatl (2016), PISA 2012 okuryazarlık testlerinin ölçme değişmezliğinin incelenmesi doktora çalışmasında, PISA 2012 matematik ve fen okuryazarlık testleri ile okuma becerileri alt testlerinin teste katılan Avustralya, Şangay-Çin, Türkiye ve Fransa grupları arasında belirli kriterlere göre ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeyi amaçlamıştır. Dil değişkenine göre incelenen çalışmada; Çoklu Grup Doğrulayıcı Faktör Analizi (ÇGDFA), Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM), Simultaneous Item Bias Test (SIBTEST) ve Madde Tepki Kuramı'na dayalı Olabilirlik Oran (MTK-OO) gibi farklı teknikleri kullanmıştır. Araştırmada ÇGDFA sonuçlarına göre, PISA 2012 matematik ve fen okuryazarlığı testleri ile okuma becerileri testlerinin belirlenen ülke örneklemeleri arasında dil değişkenine göre ölçme değişmezliğinin sağlanmadığını belirlemiştir. Ayrıca

temel model olan yapısal değişmezliğin sağlandığını ancak metrik değişmezlik modelinin sağlanmadığını gözler önüne sermişlerdir.

Köse, Uysal ve Yandı, (2017), PISA 2012’de bulunan veriler üzerinden farklı yöntemlerle ölçme değişmezliğinin araştırılmasını ele aldıkları çalışmalarında, PISA 2012 verileri ile istatistiksel varsayımlar açısından farklı olan yöntemler kullanılarak ölçme değişmezliği ile ilgili sonuçların incelenmesini amaçlanmışlardır. Ayrıca normallik varsayımı yapılması gereken ve gerekmeyen yöntemlerin normallik varsayımı sağlanamayan durumlarda farklı sonuçlar gösterip göstermediğine bakılmıştır. Bu amaca göre PISA 2012 alt ölçeklerinden beş maddeden oluşan Problem Çözmeye Açıklık ölçeği Türkiye ve Finlandiya örneklemi veri seti üzerinde yapısal eşitlik modellemesi içerisindeki ortalama kovaryans yapılarının değişmezliği analizi ve örtük sınıf analizi çatısı altındaki çoklu grup örtük sınıf analizi yöntemli analizler gerçekleştirilmiştir. Sonuçlara göre farklı metodlar kullanıldığında ölçme değişmezliği bulguları oldukça değişiklik göstermiştir. Bu durumda yöntemler için gerekli varsayımların mutlaka incelenmesi ve gizil ve gözlenen değişken yapıları göz önünde bulundurularak uygun yöntem seçmeye dikkat edilmesi hususuna dikkat çekmişlerdir.

Başusta ve Gelbal (2015), gruplararası kıyaslamalarda PISA öğrenci anketi örneğinde ölçme değişmezliğinin test edilmesini ele aldıkları çalışmalarında, PISA 2009 öğrenci anketinde yer alan fen bilgisi ve teknolojileri ile ilgili maddelere ait faktörleri belirleyerek bu faktörlerin cinsiyet grupları arasında ölçme değişmezliğine sahip olup olmadığını incelemişler ve Türkiye örneğini ele almışlardır (Büyüköztürk, Çakan, Tan ve Atar, 2014). Anket verilerinin analizi için yapısal eşitlik modellemesi (YEM) kullanılmış ve ölçme modeli aşama aşama sınırlandırılıp test edilmiştir. Değişmezlik testi olarak uyum indeksleri farkı alınmış ve analiz sonuçlarından elde edilen uyum indeksleri standart değerler baz alınarak karşılaştırılmış ve ölçme değişmezliğinin cinsiyet değişkeni bakımından sağlandığı sonucuna ulaşılmıştır. Buradan hareketle, ölçme modelinin kız ve erkek öğrenci grupları için aynı ve doğru şekilde ölçüm yaptığı söylenebilir.

Kıbrışoğlu (2015), kültürlere ve cinsiyete göre PISA 2012 matematik öğrenme modelinin ölçme değişmezliğinin incelenmesinin ele alındığı yüksek lisans çalışmasında, matematik öğrenme alt boyutu anketleri ile oluşturulan matematik öğrenme modelinin ülke ve cinsiyet grupları arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalışmıştır. Türkiye, Endonezya ve Çin-Şangay örneklemelerini ele almıştır. Açımlayıcı faktör analizi (AFA) ile oluşturduğu matematik öğrenme algısı modelini doğruladıktan sonra, ölçme değişmezliği analizlerinde çoklu grup doğrulayıcı faktör analizini kullanarak modeli incelemiştir. Elde edilen araştırma sonuçlarına göre model ülkeler arasında ölçme değişmezliğinin tüm koşullarını sağlayamamış olup cinsiyet grupları arasında ise model ölçme değişmezliğini her aşamada sağlamıştır. Buradan hareketle modele ait ortalama, varyans ve kovaryansların yalnızca cinsiyet grupları arasında karşılaştırılabilir olduğu sonucuna ulaşmışlardır. Ayrıca farklılıkların cinsiyet farkından dolayı olduğunu da belirtmiştir.

Asil ve Brown (2015), PISA 2009 okuma becerileri testinin dil, kültür ve ekonomik gelişim düzeylerinde ölçme değişmezliğini test etmişlerdir. Araştırmaya, örneklem büyüklükleri göz önünde bulundurulduğundan 55 ülke ele alınmıştır. Avustralya’nın referans olarak seçildiği çalışmada, İngilizce konuşulan yalnızca üç ülkenin Avustralya ile skaler değişmezliği olduğu sonucuna ulaşılmıştır. Ayrıca elde edilen verilerden hareketle, eğitimde sosyo-ekonomik kaynakların ölçme değişmezliğinde önemli bir rol oynadığı ancak dil faktörünün ve dil eğitimi uygulamalarının etkilerinin daha az olduğu görülmüştür.

Doğan ve Uyar (2014), PISA 2009 Türkiye örneğinde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesinin ele alındığı yüksek lisans çalışmasında bazı bölümlerin içinde yer aldığı aynı isimli makaleden birkaç ekleme yaparak, PISA 2009 Türkiye örneğinde öğrenci anketinde yer alan öğrenme stratejilerine ilişkin bir modeli; cinsiyet, okul türü ve

istatistiksel 12 bölge gruplarına göre incelemiş ve ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalışmışlardır. Sınırlı ve sınırlı olmayan modelden elde edilen değerler arasındaki farklara bakılıp, parametrelerin değişmez olup olmadığı incelenmiştir. Yaptıkları analizler sonucunda; modelin cinsiyet ve okul türleri gruplarında ölçme değişmezliğinin sağlanmadığını, istatistiksel bölge gruplarında ise tüm değişmezlik koşullarını yerine getirerek değişmezliğin sağlandığını belirlemişlerdir.

Bahadır (2012), PISA 2009'a verileri ile Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesinin ele alındığı yüksek lisans tezi çalışmasında, okuma becerisini etkilediği düşünülen aile ve öğrenci özelliklerini faktör analizi sonucunda üç boyutta ele almıştır. Anne-baba eğitim düzeyi ve öğrenciye sağlanan imkanların olduğu aile değişkeni, evde bulunan kültürel varlıkların olduğu kulvar değişkeni ve öğrencilerin okumaya karşı tutumlarının olduğu tutum değişkeni çalışmanın boyutlarını oluşturmaktadır. Öğrencilerin okuma becerileriyle modele alınan değişkenler arasındaki uyumu test etmek için Yapısal Eşitlik Modellemesi (YEM) kullanılmıştır. Ölçme değişmezliği test edilerek modeldeki değişkenler coğrafi bölgeler arasında kıyaslanmıştır. Modelin skaler değişmezlik şartlarını sağlamaması sebebi ile kısmi değişmezlikten faydalanılarak değişmezlik aşamaları test edilerek bölgeler arasında en büyük farklılığın anne-baba eğitim düzeyi ve öğrenciye sağlanan olanaklarda olduğu görülmüştür.

Gülleroğlu (2017), PISA 2012 matematik okuryazarlığı testine katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesini ele aldığı çalışmasında, matematiğe yönelik ilgi, matematik kaygısı, matematiğe yönelik benlik algısı ve matematik öz yeterliliği boyutlarının cinsiyete göre değişmezliğini tespit etmeyi amaçlamıştır ve verilere Çoklu Grup Doğrulamalı Faktör Analizi (ÇG-DFA) uygulamıştır. Matematiğe yönelik ilgi, matematik kaygısı ve matematik benlik algısı değişkenlerinde hem kız hem de erkekler için aynı yapının var olduğunu ve bu örtük değişkenlerin gruplar için bile değişmez olduğunu ancak matematik öz yeterliliği değişkeninin bunu sağlamadığını gözler önüne sermiştir. Matematik kaygısı ve matematiğe ilgi boyutları için kız ve erkek gruplarının maddelere benzer şekilde cevap verdiği ve iki gruptan elde edilen puan karşılaştırmalarının anlamlı olduğu sonucu vurgulanmıştır.

Erdinç ve Ertürk Akan (2018), TIMSS 2015 matematik başarıları ile ilgili bazı değişkenlerin cinsiyet açısından ölçme değişmezliğinin incelenmesini ele aldığı çalışmalarında, Çok Gruplu Doğrulamalı Faktör Analizi (ÇG-DFA) yöntemi kullanılarak öğrencilerin matematik performansı ve ilgili değişkenler yorumlanmıştır. Araştırmada yer alan matematiği sevmeye, matematiğe olan ilgi ve matematiğe ilişkin özgüven gizil değişkenleri için cinsiyet açısından sırasıyla biçimsel, metrik, skalar ve katı değişmezlik testleri yapılmıştır. Analizlerde biçimsel değişmezlik testine hiçbir sınırlamanın yapılmamış ve diğer değişmezlik testlerine kademeli olarak fazla sınırlama getirilerek arasındaki CFI karşılaştırmalı uyum iyiliği indeks farkları kıyaslanmıştır. Araştırmada cinsiyet gruplarında katı değişmezlik koşulunu sadece "matematiği sevmeye" gizil değişkeninin karşıladığı ancak test edilen üç gizil değişkenin biçimsel değişmezlik koşulunu karşıladığı bulgusu gözler önüne serilmiştir.

Koçar ve Yılmaz Koçar (2015), (2017), TIMSS 2015 matematik başarıları ile öğretmenlerin matematik konularına yönelik hazırlıklı gelme düzeylerinin ilişkisini ele aldıkları çalışmalarında, Türkiye ve Singapur örneğinde bulunan başarı testi ve öğretmen anketinden yola çıkarak matematik öğretmenlerinin matematik öğrenme alanlarını, öğretmedeki yeterlik düzeylerini ortaya koyarak 8. sınıf öğrencilerinin bu alandaki başarılarının öğretmenlerin konuya hazırlıklı gelmesinden mi yoksa öğrencilerin bu konu hakkındaki başarılarından kaynaklı mı farklı sonuçlar verdiğini ayrıca ülkeden ülkeye farklılaşıp farklılaşmadığını analiz etmişlerdir. Singapur'daki öğretmenlerin Türkiye'deki öğretmenlere kıyasla matematik konularına daha iyi hazırlandıkları sonucuna varılmıştır ve derse hazırlık ne oranda artarsa öğrenci başarıları da aynı oranda artmaktadır. TIMSS'te bulunan

matematik öğrenme alanları ile TEOG'ta bulunan matematik öğrenme alanları örtüşmektedir ve konulara hazırlık düzeyi ile matematik başarısı arasında pozitif uyum bulunan bu araştırmada öğretmenin konulara hazırlıklı gelmesindeki artışın öğrencilerin TEOG matematik başarılarında da yükseliş sağlayabileceği düşünülmektedir.

Alanyazında yapılan araştırmalar incelendiğinde ölçme değişmezliği ile ilgili birçok çalışmanın özellikle psikolojik yapıların kültürler arası genellenabilirliğini incelemek ve bu yapıların karşılaştırılabilirliğini belirlemek için yapıldığı açıkça anlaşılmaktadır.

Kültürler arası karşılaştırma çalışmalarının hem psikometrik hem de kavramsal bir tarafı vardır. Bu bağlamda ölçme değişmezliği incelenirken test yanlılığı ve test eşitliği gibi kavramsal ve yöntemsel durumlardan da bahsedilebilir (Ayvalli, 2016).

Bu doğrultuda araştırmanın problemini, çalışmalarda daha çok cinsiyetler, bölgeler, gelenekler gibi değişkenler çalışılmış olup, bilişsel alanların ölçülmesinde kullanılan çoktan seçmeli ve açık uçlu testler ve öğrenci ölçekleri üzerinden yürütülen çalışmaların özellikle de fen bilimlerine oranla matematik alanında yapılan araştırmaların azlığı saptanması dolayısı ile, TIMSS 2015 uygulamasındaki matematik başarısının OECD üyesi ülkeler için ölçme değişmezliğinin test edilmesi oluşturmaktadır.

1.1. Araştırmanın Amacı

Bu araştırmanın amacı, TIMSS 2015 matematik testinin sekizinci sınıf uygulamasının OECD üyesi ülkeler için ölçme değişmezliğinin test edilmesidir. Bu amaç doğrultusunda araştırmanın alt problemleri ise aşağıdaki gibi oluşturulmuştur:

TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında;

ölçme değişmezliğini sağlamakta mıdır?

şekil değişmezliği sağlamakta mıdır?

metrik değişmezliği sağlamakta mıdır?

ölçek değişmezliğini sağlamakta mıdır?

katı değişmezliği sağlamakta mıdır?

1.2. Araştırmanın Önemi

Belirli gruplar arası karşılaştırmaların yapıldığı birçok çalışmada araştırmacılar genellikle ölçme araçlarının, ölçme değişmezliğini/eşitliğini sağladığı varsayımına dayanarak söz konusu karşılaştırmaları yaparlar. Ancak bilişsel özelliklerin analiz edilip ölçülmesi sürecinde, en çok da farklı dil, kültür, cinsiyet ve geleneğe sahip gruplar için bilişsel özellikleri ölçen testlerin aynı sonucu verip vermediği, aynı anlama gelip gelmediği gibi sorulara yanıt bulmadan bu karşılaştırmaları yaparak sonuçlardan bazı çıkarımlar elde etmek yanlış bulgulara ve yorumlamalara sebep olabilir (Lance and Vandenberg, 2000).

TIMSS ile ilgili yapılan çalışmalar incelendiğinde, duyuşsal özellikleri ölçen araçlara ilişkin ölçme değişmezliğinin incelendiği literatür taraması sonucu ulaşılabilen çalışma sayısı oldukça fazladır.

Öğretmen ve Uzun (2010), TIMSS 1999 verileri üzerinde yürüttükleri çalışmalarında, öğrencilerin fen bilimleri başarısını belirleyen değişkenleri belirlemişler ve bu değişkenlerin cinsiyet bazında değişmezliğini incelemişlerdir.

Karakoç Alatl (2016), doktora çalışmasında PISA 2012 okuryazarlık testlerinin ölçme değişmezliğini incelemiştir.

Ayvalli (2016), yüksek lisans çalışmasında PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğini incelemiştir.

Arıkan, Sarı ve Yıldızlı (2017), çalışmalarında TIMSS 2015 sekizinci sınıf matematik akademik başarısını yordayan faktörleri incelemişlerdir.

Ölçülüoğlu (2015), yüksek lisans çalışmasında TIMSS 2011 Türkiye sekizinci sınıf matematik başarısını etkileyen faktörlerin bölgelere göre incelenmesini ele almıştır.

Köse, Uysal ve Yandı (2017), çalışmalarında PISA 2012’de bulunan veriler üzerinden farklı yöntemlerle ölçme değişmezliğinin araştırılmasını incelemişlerdir.

Başusta ve Gelbal (2015), çalışmalarında gruplar arası kıyaslamalarda PISA öğrenci anketi örneğinde ölçme değişmezliğinin test edilmesini ele almışlardır.

Kıbrıslıoğlu (2015), yüksek lisans çalışmasında kültürlere ve cinsiyete göre PISA 2012 matematik öğrenme modelinin ölçme değişmezliğinin incelenmesini ele almıştır.

Asil ve Brown (2015), PISA 2009 okuma becerileri testinin dil, kültür ve ekonomik gelişim düzeylerinde ölçme değişmezliğini test etmişlerdir.

Doğan ve Uyar (2014), yüksek lisans çalışmasında PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesini ele almışlardır.

Bahadır (2012), yüksek lisans tezi çalışmasında PISA 2009’a verileri ile Türkiye’deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesini ele almıştır.

Gülleroğlu (2017), çalışmasında PISA 2012 matematik okuryazarlığı testine katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesini ele almıştır.

Erdinç ve Ertürk Akan (2018), çalışmalarında TIMSS 2015 matematik başarısı ile ilgili bazı değişkenlerin cinsiyet açısından ölçme değişmezliğinin incelenmesini ele almışlardır.

Koğar ve Yılmaz Koğar (2017), çalışmalarında TIMSS 2015 matematik başarısı ile öğretmenlerin matematik konularına yönelik hazırlıklı gelme düzeylerinin ilişkisini ele almışlardır.

Tüm bu alanyazı araştırmalarına bağlı olarak çalışmalarda daha çok cinsiyetler, bölgeler, gelenekler gibi değişkenler çalışılmış olup, bilişsel alanların ölçülmesinde kullanılan çoktan seçmeli ve açık uçlu testler ve öğrenci ölçekleri üzerinden yürütülen çalışmaların özellikle de fen bilimlerine oranla matematik alanında yapılan araştırmaların azlığı saptanmıştır. Ayrıca değişmezlik çalışmalarında OECD ülkelerinin ele alındığı çalışmalara daha az ulaşılmış olup OECD ülkeleri arası matematik başarı performanslarında farklılıklar tespit edilmiştir (TIMSS, 2015). Buradan yola çıkarak araştırmada farklı gruplarda farklı bireysel özelliklere sahip kişiler arasında ölçme değişmezliği çalışmasının yapılması, geniş ölçekli testler ve bu testlerin sonuçlarının doğru yorumlanabilmesi gerekmektedir.

2. YÖNTEM

Bu bölümde araştırmanın modeli, evren ve örneklem, verilerin elde edilmesi ve verilerin analizine ilişkin bilgiler yer almaktadır.

2.1.Araştırma Modeli

Bu araştırmanın modelini, TIMSS 2015 uygulamasında yer alan sekizinci sınıf matematik başarı testi yapılarının OECD üyesi ülkeler arasında ölçme değişmezliğini sağlama durumunun belirlenmesi oluşturmaktadır. Bu bağlamda çalışma, TIMSS 2015 değerlendirmesine ait bilişsel testin geçerlilik düzeyini belirlemeye yönelik olduğu için ilişkisel tarama modelinde yürütülmüştür (Balcı, 2013).Tarama modelinde yürütülen çalışmalarda, evreni temsil eden belirli bir örneklem grubu üzerinde yapılan çalışmalarla evren hakkında genel bir yargıya varabilme amaçlanır (Bailey, 1982).

2.2.Evren ve Örneklem

Araştırmanın evrenini OECD ülkelerinden TIMSS 2015 uygulamasına katılan 8. sınıf öğrencileri oluşturmaktadır. TIMSS 2015 uygulamasında örneklem arası katılımcı sayılarının diğer kitapçıklara göre birbirine daha yakın olması sebebi ile 11 nolu kitapçık seçilmiştir. Bu kitapçığı alan bireylerin OECD ülkeleri için dağılım tablosu Tablo 1’deki şekildedir.

Tablo 1.
OECD Ülkelerine Göre Örneklem Dağılımı

Ülke	N	%
Amerika	741	10.26
Avusturalya	732	10.13
Hong Kong	288	3.99
İngiltere	342	4.73
İrlanda	339	4.69
İsrail	393	5.44
İsveç	303	4.19
İtalya	321	4.44
Japonya	345	4.78
Kanada	631	8.74
Kore	384	5.32
Macaristan	354	4.90
Norveç	342	4.73
Şili	348	4.82
Slovenya	301	4.17
Türkiye	441	6.11
Yeni Zelanda	618	8.56
TOPLAM	7.223	100

Tablo 1 incelendiğinde 8. sınıf öğrencisi bulunan 17 OECD üyesi ülkede 7.223 kişiden çalışmaya dâhil edilen bireylerin dağılımı incelendiğinde en fazla katılımcıya sahip ülke %10,26 (741 kişi) ile Amerika, en az katılımcıya sahip olan ülke 3,99 (288 kişi) ile Hong Kong'dur. Örneklem büyüklüklerinin tümü kullanılan istatistiksel teknik için yeterlidir. Bu nedenle örneklemlemlerle analizlerde bir sorun oluşmamıştır.

Diğer ülkelerin katılımcı oranları ise %10,13 ve %4,17 (732 ve 301 kişi) arasında değişmektedir.

2.3.Araştırma Verileri

Verilere TIMSS'e ait resmi web sitesinin International Database - SPSS Data ilgili bölümünden ulaşılmış ve <http://timssandpirls.bc.edu/timss2015/international-database/> adresinden indirilmiştir.

TIMSS döngüleri ve Türkiye'nin katılımına Tablo 2'ye bakacak olursak; TIMSS'e 41 ülkenin katılım gösterdiği 1995 ve 46 ülkenin katılım gösterdiği 2003 yıllarında ülkemiz katılmamıştır. 38 ülkenin katılım gösterdiği 1999 senesinde sadece 8. sınıf düzeyinde katılmış olup, 1999 katılımı ilk katılımımızı oluşturmuştur. Daha sonra 59 ülkenin katılım gösterdiği 2007 senesi 8. sınıf düzeyinde, 66 ülkenin katılım gösterdiği 2011 senesi ilk kez 4. ve 8. sınıf düzeyinde ve en yakın tarih olan 2015 senesinde 61 ülkenin katılımı ile 4. ve 8. sınıf düzeyinde sınava katılmıştır. 2019 senesinde yapılan TIMSS'in verileri henüz yayınlanmamıştır.

Tablo 2.

TIMSS'e Ait Yıllara Göre Temel Bilgiler ve Katılım Durumu

Yıl	Katılımcı Ülke Sayısı	Türkiye'nin Katılım Durumu
1995	41	Katılmamıştır.
1999	38	Sekizinci sınıf düzeyinde katılmıştır.
2003	46	Katılmamıştır.
2007	59	Sadece sekizinci sınıf düzeyinde katılmıştır.
2011	66	Dördüncü ve sekizinci sınıf düzeyinde katılmıştır.
2015	61	Dördüncü ve sekizinci sınıf düzeyinde katılmıştır.

TIMSS değerlendirmesinde çoktan seçmeli test ve uçlu soru türleri kullanılmaktadır. 14 kitapçıktan seçilen 11 nolu kitapçıkta açık uçlu ve çoktan seçmeli maddeler bulunmaktadır. Bu araştırmada maddelere ait puanlamalar madde türüne göre farklılık gösterdiği için sadece matematik başarısına ait, 16 çoktan seçmeli ve 13 açık uçlu olmak üzere toplam 29 madde analize dâhil edilmiştir.

TIMSS matematik maddelerinin bilişsel alanlara göre yüzdesel dağılımı Tablo 3'de verilmiştir.

Tablo 3.

TIMSS Matematik Maddelerinin 8. Sınıf Bilişsel Alanlara Göre Yüzdesel Dağılımı

Bilişsel Alan	Maddeler
Bilgi	%35
Uygulama	%40
Akıl Yürütme	%25

Sorulara bilişsel alanlar bağlamında bakıldığında Türkiye'de eskiden kullanılan ezbere dayalı sorular değil, artık yenilikçi eğitim anlayışıyla günümüzde uygulanan anlamaya, keşfetmeye, yorumlama gücüne dayalı sorular kullanılmakta olduğu gözler önündedir. Ayrıca sorular bilgi, kavrama, uygulama basamaklarını yoklamaktadır. TIMSS uygulamasında, sekizinci sınıf seviyesinde matematik ve fen bilimleri cevap anahtarında bilişsel alan ve öğrenme alanı olmak üzere iki kısım bulunmaktadır. TIMSS 2015 verilerine dayanarak 8. sınıf matematik sorularının uygulama ağırlıklı olduğu görülmektedir. Sonrasında bilgi ve en son sırada da akıl yürütme becerisi gelmektedir.

Tablo 4.

TIMSS Matematik Maddelerinin Sekizinci Sınıf Matematik Öğrenme Alanlarına Göre Yüzdesel Dağılımı

Öğrenme Alanı	Alt Öğrenme Alanı	Yüzde
Sayılar	Doğal Sayılar	%30
	Kesirler, Ondalık Sayılar ve Tam Sayılar	
	Oran, Orantı ve Yüzde	
Cebir	Örüntüler	%30
	Cebirsel İfadeler ve İşlemler	
	Eşitlik ve Eşitsizlikler	
	Formüller ve Fonksiyonlar	
Geometri	Geometrik Şekiller	%20
	Geometrik Ölçme	
	Dik Koordinat Sistemi (Bir Noktanın Yeri, Yansıma, Öteleme ve Dönme)	
Veri ve Olasılık	Veri Kümesinin Özellikleri	%20
	Verileri Yorumlama	
	Olasılık	

Tablo 4'e bakıldığında 8. sınıf matematik sorularının öğrenme alanı ve alt öğrenme alanları gözükmekte, sayılar ve cebir öğrenme alanlarının daha yoğun olduğu, geometri ve veri olasılık öğrenme alanlarının kısmen daha az olduğu belirtilmektedir. Alt öğrenme alanı bakımından çoğu kazanıma yer verilmeye çalışıldığı gözükmektedir.

2.4.Verilerin Analizi

Aritmetik Ortalama: Aritmetik ortalama, en çok kullanılan merkezi eğilim ölçüsüdür. Birimlerin belirli bir değişken bakımından aldıkları değerlerin toplamının birim sayısına bölümü olarak tanımlanır. Eşit aralıklı ve oran ölçme düzeyinde ölçülen değişkenler için kullanılır. Aritmetik ortalama hem kitle hem de örneklem için hesaplanır.

Tepe Değer (Mod): Bir veri grubunda en çok tekrarlanan değere tepe değer(mod) denir. Tepe değerini hesaplanmasında birimlerin büyüklük sırasına konulması şart değilse de, bu işlemin yapılması tepe değerini bulunmasında kolaylık sağlar.

Ortanca (Medyan): Bir veri grubundaki değerlerin küçükten büyüğe sıralandığında tam ortaya düşen değer ortanca değeridir. Kitledeki birimlerin sayısı çok fazla ise verilerin özetlenmesinde merkezi eğilim ölçüsü olarak ortanca kullanılabilir. Ortanca, sınıflama ölçme düzeyi ile ölçülen değişkenler için kullanılmaz. Eşit aralıklı, oran ve sıralama ölçme düzeyinde ölçülen değişkenler için kullanılır.

Varyans ve Standart Sapma: Varyans, birim değerlerinin ortalamadan sapmalarının kareler toplamının birim sayısına bölünmesi ile elde edilir. Varyans gözlem sonuçlarının aritmetik ortalamadan ne ölçüde farklı olabileceğini ortaya koyan bir dağılım ölçüsüdür.

Değişim Genişliği (Ranj): Bir veri grubunda en büyük değer ile en küçük değer arasındaki farka değişim genişliği denir, R ile gösterilir.

KR-20 ve KR-21: Güvenirliği testteki soruların birbirleriyle uyumunu, tutarlılığını gösteren güvenilirlik hesaplama yöntemleridir. Testin bir kere uygulanması yeterlidir. KR-20 hesaplanırken testteki sorular 1 ve 0 (doğru-yanlış) şeklinde puanlanır ve madde güçlük indeksini içeren formül kullanılarak hesaplanır. KR-21 hesaplanırken madde güçlük indeksi yerine test puanlarının ortalaması kullanılır. (Akgün, Büyüköztürk, Çakmak, Demirel ve Karadeniz 2008; Büyüköztürk, 2002).

Veriler analiz edilmeden önce sınava katılan ülkelerin öğrenci sayısı, ortalaması, mod, medyan, standart sapma, ranj, K_y , B_s , KR-20 değerleri kısacası merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 güvenilirlik katsayıları hesaplanmıştır. Ölçeğin literatür sonuçlarıyla değerlendirilerek normal dağılıma uygun olduğu belirlenmiştir. Verilere Açımlayıcı Faktör Analizi (AFA) yapılmış ve tek faktörlü yapıda olduğu bulunmuştur. Yapı geçerliliğine ilişkin kanıt elde edebilmek amacıyla DFA (Doğrulayıcı Faktör Analizi) yapılmıştır. Daha sonra kovaryans matrislerinin eşitliği test edilerek, maddelere ait ölçme değişmezliği OECD üyesi ülkelere göre ÇGDFA (Çok Gruplu Doğrulayıcı Faktör Analizi) yöntemi ile test edilmiştir. Analizler çeşitli

veri analizi paket programları kullanılarak yapılmış, asimtotik kovaryanslar matrisi üzerinden model parametrelerinin kestirilmesinde en çok ağırlıklandırılmamış küçük kareler yöntemi (unweighted least square-ULS) kullanılmıştır.

İki kategorili veriler için tetrakorik korelasyon kullanılırken, polikorik korelasyon ikiden fazla kategoriye sahip veriler için kullanılır. Ordinal veriler için geliştirilen tahmin yöntemlerinden bazıları en küçük kareler ağırlıklı (WLS), ağırlıklandırılmamış en küçük kareler (ULS) ve çapraz ağırlıklandırılmış en küçük kareler (DWLS) yöntemleridir (Babakus, 1985).

Alanyazında yapılan araştırmalar sonucunda normalde istatistiksel yöntemler içerisinde yer alan regresyon çözümlemesi için en çok kullanılan yöntem en küçük kareler yöntemidir. Olası birçok regresyon yöntemlerinin dışında, genellikle matematiksel hesaplamalardaki kolaylığından dolayı veri analizi uygulamalarında en küçük kareler yöntemi (WLS) kestiricileri yaygın olarak tercih edilmektedir. (Alma ve Vupa, 2008)

Fakat alanyazında yapılan bazı araştırmalar sonucunda ağırlıklandırılmamış küçük kareler (ULS) yöntemlerinin, ordinalde daha az standart hata olan parametreleri daha iyi tahmin ettiği bulunmuştur. Parametrelerin tahmininde ağırlıklandırılmamış en küçük kareler (ULS) yöntemi daha düşük örneklem boyutlarında daha iyi sonuçlar verdiği ve çapraz ağırlıklandırılmış en küçük kareler (DWLS) yönteminin ise yüksek örneklem boyutlarında daha iyi sonuçlar verdiği gözlenmiştir. Ayrıca ağırlıklandırılmamış küçük kareler (ULS) yöntemi, bir kerede tüm parametreleri tahmin yeteneği gibi bazı özelliklere sahiptir ve en az yinleme sayısına sahip parametreleri tahmin etmiştir bu durumda tüm veri kümeleri ve ilgili örnek verinin parametrelerini tahmin etmek için daha doğru bir yöntemdir (Koğar ve Koğar, 2015). Bu yüzden analizde ilgili çalışmada OECD ülkeleri ile analiz yapılması sebebi ile örneklem azalmış ve bu sebeple ağırlıklandırılmamış en küçük kareler (ULS) yöntemi kullanılmıştır.

Analizlerin yapılabilmesi için veri setinin temel varsayımları karşılayıp karşılamadığı test edilmiştir. Bu varsayımlar ait sonuçlara Tablo 5'te yer verilmiştir.

Tablo 5.
OECD Üyesi Ülkelerin 8. Sınıf Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları

Ülke	N	$\bar{x}$	Mod	Medyan	S	Ranj	K _y	B _s	KR-20
Amerika	741	14.21	19	14	6.14	29	.109	-.733	.87
Avusturalya	732	13.87	11	13	6.42	28	.073	-.890	.89
Hong Kong	288	19.99	25	22	5.93	28	-1.155	.744	.89
İngiltere	342	15.47	12	15	6.25	28	.132	.749	.88
İrlanda	339	15.64	21	16	6.34	27	-.220	-.795	.88
İsrail	393	14.10	10	13	7.54	28	.110	-1.197	.92
İsveç	303	13	6	13	6.45	28	.191	-.697	.89
İtalya	321	12.57	9	12	5.70	28	.136	-.360	.86
Japonya	345	12.24	24	19	6.58	26	-.427	-.830	.90
Kanada	631	15.70	17	17	5.65	28	-.238	-.611	.85
Kore	384	22.70	28	24	5.34	27	-1.008	.532	.83
Macaristan	354	15.40	16	15	6.85	28	-.031	-.929	.90
Norveç	342	12.62	12	12	5.43	27	.351	-.329	.84
Şili	348	10.25	11	10	5.28	26	.598	.055	.84
Slovenya	301	14.41	13	14	5.61	25	.204	.589	.85
Türkiye	441	10.95	5	10	6.42	29	.669	-.404	.89
Yeni Zelanda	618	13.84	9	14	6.28	28	.148	-.849	.88

K_y: Çarpıklık katsayısı

B_s: Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

Tablo 5'te yer alan sonuçlara göre, matematik okur-yazarlığı test puanlarının OECD ülkelerine ait merkezi eğilim ölçülerinin birbirine yakın olduğu söylenebilir. Normallik için basıklık ve çarpıklık katsayıları incelendiğinde 17 ülkeden 14 ülkeye ait değerlerin ± 1 arasında yer alıyor olması veri setinin normale yakın bir dağılım gösterdiği anlamına gelir. Ülkelere ait güvenilirlik katsayıları (KR-20) incelendiğinde ise güvenilirliğin .83 ile .92 arasında değiştiği görülmüştür. Bazı ülkeler için sonuçların kabul edilebilir düzeyin üstünde, bazı ülkeler içinse altında olması, yapılan araştırma bir ölçme aracı geliştirme çalışması olmadığı için sadece raporlaştırılmış ve herhangi bir işlem yapılmamıştır. ÇGDFA yapılmadan önce ülkelere ait kovaryans matrislerinin eşitliği test edilmiştir.

Araştırmada kullanılacak uyum indekslerine bakacak olursak Ki-Kare uyum iyiliği testi, GFI, AGFI, RMR, SRMR, NFI, CFI ve RMSEA genel olarak kullanılmaktadır. Aşağıda bu indekslerin daha detaylı açıklamalarına yer verilmiştir:

Ki-kare değeri (χ^2): Ki-kare değeri genel olarak model uyumunun genelini değerlendirmek için kullanılan geleneksel bir ölçümdür ve örneklem ile uyumlu kovaryans matrisleri arasındaki uyumsuzluk kısmını inceler (Hu ve Bentler, 1999: 2). İyi bir model uyumu .05 düzeyinde sonucu manidar değildir (Barrett, 2007). Böylece ki-kare istatistik testi literatürde çoğu zaman “kötülük uyumu” (Kline, 2005) veya “uyum eksikliği” (Mulaik ve diğerleri, 1989) şeklinde tanımlanır. Alwin, Muthen, Summers ve Wheaton, (1977) tarafından önerilen örneklem büyüklüğünün etkisini azaltan normlaştırılmış ki-kare istatistiği (χ^2/sd) bulunmaktadır ve bu değer kabul edilebilir düzeyinin en fazla 5, en az 2 (Fidell and Tabachnick, 2013) kabul edilebileceği belirtilmiştir.

RMSEA: Modelde en uygun değeri veren fakat modelin bilinmeyen parametre tahminleriyle birlikte evrenlerin kovaryans matrisine ne oranda uyduğunu gösterir (Byrne, 1998). RMSEA istatistiğinde uyum indeksini belirlemek için .06'ya yakın bir kesme noktası iyi uyum olarak değerlendirilirken (Hu ve Bentler, 1999) .05 ile 0 arasındaki değerler modelin mükemmel uyum gösterdiği anlamına gelir (Jöreskog and Sörbom, 1993; Brown, 2006; Marcoulides and Raykov, 2008).

GFI: Geliştirilmesi Jöreskog'la beraber bu alanda çalışan Sörbom tarafından yapılan GFI, tahmini evren kovaryansı ile açıklanan varyans oranını hesaplamakta yaygın olarak kullanılan bir uyum indeksidir (Fidell and Tabachnick 2013). GFI 0 ile 1 arasında değerler alır ve örneklem büyüdükçe GFI değeri artar (Bollen, 1990; Miles and Shevlin, 1998). Genel olarak GFI için kesme noktası olarak .90 önerilmektedir ancak elde edilen sonuçlar simülasyon çalışmaları, faktör yükleri ve örneklem büyüklüğü küçük olduğunda .95 değerindeki kesme noktasının daha uygun olduğunu göstermektedir (Miles ve Shevlin, 1998). Bu genel ifadeler ışığında, AGFI da örneklem büyüklüğünün artmasıyla birlikte artış gösterir ve 0 ile 1 arasında değerler alır. AGFI, .90 ve üzerinde değer alıyorsa bu modelin mükemmel uyum gösterdiğini ifade etmektedir.

SRMR ve RMR: Veri seti içerisindeki örneklemde bulunan kovaryans matrisi ve varsayımda bulunulan kovaryans modelinin artık değerlerinin arasındaki farkın karekökü alınarak SRMR değeri bulunur. Bir ölçek, değişik düzeyde maddeler içeriyorsa (bazı maddeler 1 ile 5 arasında değişirken, diğerleri 1 ile 7 arasında değişiklik gösteriyorsa) SRMR'nin yorumlanması gittikçe zorlaşır (Kline, 2005). Çünkü SRMR değer aralığı her bir göstergenin değer ölçeğine bağlıdır Standardize edilmiş RMR (SRMR) bu sorunu çözdüğü için yorumlanması daha kolaydır. SRMR değerlerine genel olarak bakacak olursak 0 ve 1 düzeylerinde değiştiğini ve .05'ten küçük değerler model uyumunun iyi olduğunu göstermektedir (Byrne, 2013). Buna ilaveten .08'e kadar olan SRMR değerleri de kabul edilebilir düzeyde bulunmaktadır (Bentler ve Hu, 1999). SRMR'nin 0 değerine yaklaşması uyumun iyileştiğini, 0 değerini alması mükemmel uyumu işaret etmektedir ancak SRMR modelde çok sayıda parametre olduğunda ve büyük örneklem büyüklüklerine sahip modellerde diğer uyum indekslerinde olduğu gibi düşük değer göstermeye eğilimlidir.

NFI: Yokluk hipoteziyle oluşturulmuş modelin χ^2 değeriyle, normal modelin χ^2 değerini, karşılaştırarak ölçülür ve NFI değeri bulunur. Yokluk hipotezine dayalı oluşturan model, değeri bulunan tüm değişkenlerin korelasyonsuz olduğunu belirttiği için en kötü durumları açıklayıp gösterdiği söylenebilir. Bu istatistik sonucu için çıkan değerler 0 ile 1 arasında yer alır. Bentler ve Bonnet (1980) .90'dan büyük değere sahip modellerin iyi uyum sergilediğini söylemişlerdir. Küçük örneklemelerin kullanıldığı durumlarda, diğer istatistikler iyi uyum değerleri gösterirken NNFI zayıf uyum değeri gösterebilir (Bentler, 1990; Kline, 2005; Tabachnick ve Fidell, 2007). Önemli araştırmacılar Bentler ve Hu (1999), NNFI için kesme noktası kriterinin .95 ve üzerinde olması gerektiği önerisinde bulunmuşlardır. Bu öneri genel olarak kabul görmüştür. Bu uyum indeksi örneklem büyüklüğüne karşı hassasiyet gösterir (Bennet, James, Lind, Mulaik, Stilwell, ve Van Alstine, 1989; Bentler, 1990) ve bu sebepten ötürü NFI indeksine analiz yorumlamalarında tek başına güvenilmemesi gerekmektedir (Kline, 2005). Bu sorun, daha basit modelleri tercih eden Normlaştırılmamış Uyum İndeksi (NNFI, aynı zamanda Tucker-Lewis indeksi olarak da bilinir) ile iyileştirilmiştir ve günümüzde çokça kullanılmaktadır. (Boehm, Lyubomirsk, Özer ve Tucker, 2006).

CFI: NFI'nın örneklem boyutu küçükken dahi iyi sonuç verdiği, örneklem boyutunu göz önünde bulundurulduğu Byrne (2010) tarafından literatüre kazandırılmıştır. Ayrıca Fidell ve Tabachnick (2013), NFI'nın revize edilmiş halinin CFI olduğunu da belirtmiştir. NFI gibi CFI istatistiği de örneklem kovaryansını, yokluk hipotezine dayalı olarak oluşturulan model ile kıyaslar ve tüm örtük değişkenlerin korelasyonsuz olduğunu kabul eder. NFI'da olduğu gibi bu istatistik için de değerler 0 ve 1 kesme değerleri arasında değişkenlik gösterir ve değerler 1'e yaklaştıkça model uyumu artar yani model benzerlik kazanır. Literatüre bakıldığında yapılan çalışmalarda göre mükemmel uyum değeri sayılabilmesi için .95'den büyük olması gerektiği kabul edilmektedir (Hu ve Bentler, 1999). Günümüzde bu indeks örneklem boyutundan en az etkilenen uyum indekslerinden biri olduğundan en çok rapor edilen uyum indekslerinden biri olmuştur ve böylece tüm YEM programlarına dâhil edilmiştir ve en sık kullanılan uyum indeksi sayılabilir (Fan, Thompson, ve Wang, 1999).

Uyum İndekslerini Raporlaştırma: Yapılan bir çalışmada alanyazında en çok rastlanan ve genel olarak kullanılan uyum indekslerinin SRMR, GFI, CFI, RMSEA, NNFI ve NFI olduğu bulunmuştur (Ho ve McDonald, 2000). Örneklem büyüklüğünden etkilendiği için Ki-kare istatistiğinden yola çıkarak yapılan değerlendirmeler bu istatistiğin, serbestlik derecesi ve ona bağlı p değeriyle birlikte her zaman kullanılması gerektiği söylenmektedir (Kline, 2005; Boadu, Boulianne, Cumming, Hayduk and Pazderka-Robinson, 2007). Bunun yanında, normal dağılımdan sapma gösteren ve örneklem büyüklüğünün çok olduğu koşullarda, Satorra-Bentler istatistiği ($S-B_{\chi^2}$) normallik gösteren ve küçük örneklemeye sahip durumlarda hesaplanan değerine yakın değerler üretir ve böylece bu bilgiler ışığında hesaplanır. (Satorra and Bentler, 2011). $S-B_{\chi^2}$ yani Satorra-Bentler düzeltmesi farklı puan dağılımına ve örneklem büyüklüğüne ve sahip veri setlerinde, örneklem büyüklüğünün üzerindeki etkisini azaltmak için kullanılan güvenilir bir istatistiktir (Byrne, 2013).

Test sonuçları Tablo 6'da verilmiştir.

Tablo 6.

<i>OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Kovaryans Matrislerinin Eşitliği Testi Sonuçları</i>							
Grup	$S-B_{\chi^2}(sd)$	P	χ^2/sd	RMSEA	GFI	CFI	SRMR
OECD Üyesi Ülkeler	18179.21(6960)	.00	2.61	.063	.94	.92	.063

Tablo 6'da χ^2/sd oranı 3'ün altında çıktığı, RMSEA değerinin .08'in altında, SRMR değerinin .08'in altında, CFI değerinin .90'ın üstünde ve GFI değerinin .90'ın üzerinde çıktığı görülmektedir. Uyum indeksleri literatür ile değerlendirildiğinde χ^2/sd manidarlığı oldukça iyi, RMSEA değerinin iyi

uyum, SRMR değerinin iyi uyum, CFI ve GFI değerlerinin oldukça iyi uyum gösterdiği görülmüştür. Ülkelere ait kovaryans matrislerinin oldukça iyi bir uyum gösterdiği ifade edilebilir. Ülkelere ait DFA sonuçlarının çok sayıda ülke olması sebebiyle ÇGDFA sonuçlarından ayrı tablo halinde sunulması uygun görülmüş ve sonuçlar Tablo 7’de verilmiştir.

Tablo 7.
OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Doğrulayıcı Faktör Analizi Sonuçları

Ülke	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	GFI	CFI	SRMR
Amerika	804.71(377)	.00	2.13	.039	.93	.97	.040
Avusturalya	818.38(377)	.00	2.17	.040	.93	.97	.039
Hong Kong	552.84(377)	.00	1.47	.040	.87	.98	.053
İngiltere	542.46(377)	.00	1.44	.036	.90	.97	.048
İrlanda	485.21(377)	.00	1.29	.029	.91	.98	.045
İsrail	595.53(377)	.00	1.58	.038	.90	.98	.042
İsveç	734.07(377)	.00	1.95	.056	.86	.94	.058
İtalya	707.66(377)	.00	1.88	.052	.87	.93	.057
Japonya	554.64(377)	.00	1.47	.037	.89	.98	.047
Kanada	689.57(377)	.00	1.83	.036	.93	.96	.040
Kore	583.17(377)	.00	1.55	.038	.89	.97	.050
Macaristan	556.80(377)	.00	1.48	.037	.90	.98	.046
Norveç	496.84(377)	.00	1.32	.031	.91	.97	.047
Şili	553.27(377)	.00	1.47	.037	.90	.96	.050
Slovenya	460.58(377)	.00	1.22	.027	.91	.98	.049
Türkiye	827.03(377)	.00	2.19	.052	.88	.96	.049
Yeni Zelanda	796.14(377)	.00	2.11	.042	.92	.96	.042

Tablo 7’deki sonuçlara göre 17 ülkenin χ^2 /sd oranının 3’ün altında olduğu görülmektedir. Bu durum da bize χ^2 /sd sonucunun iyi olduğunu göstermektedir. Uyum indekslerine bakıldığında tüm ülkelere ait CFI değerleri 0.90’nın üzerinde yani iyi uyum göstermektedir. GFI değerleri Hong Kong, İsveç, İtalya, Japonya, Kore ve Türkiye için iyi uyuma yakın, diğer ülkeler için iyi uyum göstermiştir. Hong Kong, İsveç, İtalya hariç diğer ülkelere ait SRMR değerlerinin .05’in altında olduğu, Hong Kong’a ait SRMR değerinin .053, İsveç’e ait SRMR değerinin .058, İtalya’ya ait SRMR değerinin .057 olduğu görülmektedir. Uyum indeksleri genel olarak literatür ile değerlendirildiğinde tüm ülkelere ait modellerin 17 OECD üyesi ülke için ayrı ayrı doğrulandığını kararı alınmıştır.

Tablo 8.
OECD Ülkelerinin Pearson Korelasyon Değer Aralığı

Ülke	N	Pearson Korelasyon Değer Aralığı	Çoklu Bağlantı İstatistiği	
			VIF Değeri	Tolerans Değeri
Amerika	741	.021 - 0.195	1.075 - 1.546	0.717 - 0.923
Avusturalya	732	.042 - 0.258	1.043 - 1.761	0.707 - 0.884
Hong Kong	288	.100 - 0.339	1.078 - 2.049	0.709 - 0.927
İngiltere	342	.068 - 0.323	1.085 - 1.776	0.725 - 0.922
İrlanda	339	.015 - 0.300	1.116 - 1.573	0.719 - 0.896
İsrail	393	.110 - 0.312	1.197 - 1.936	0.711 - 0.835
İsveç	303	.039 - 0.268	1.153 - 1.818	0.709 - 0.867
İtalya	321	.029 - 0.234	1.151 - 1.877	0.719 - 0.869
Japonya	345	.020 - 0.270	1.163 - 1.827	0.736 - 0.912
Kanada	631	.016 - 0.138	1.097 - 1.560	0.722 - 0.965
Kore	384	.086 - 0.397	1.136 - 1.758	0.771 - 0.872
Macaristan	354	.101 - 0.298	1.078 - 1.656	0.726 - 0.947
Norveç	342	.014 - 0.125	1.157 - 1.656	0.726 - 0.947
Şili	348	.052 - 0.232	1.111 - 1.580	0.736 - 0.955
Slovenya	301	.024 - 0.159	1.183 - 1.553	0.807 - 0.915
Türkiye	441	.017 - 0.138	1.182 - 1.838	0.732 - 0.914
Yeni Zelanda	618	.019 - .0205	1.100 - 1.786	0.774 - 0.909
Toplam	7.223	-		
Değer Aralığı	-	.014 - 0.397	1.043 - 2.049	0.707 - 0.965

Ölçme değişmezliğinde ülkeler bağımsız değişkenleri göstermektedir. Bağımsız değişkenler arasındaki korelasyonlara Tablo 8'deki değerler için bakılmış ve değişkenler arasındaki korelasyonların genel olarak .014 ile .397 arasındaki değiştiği görülmüştür. Tolerans değerleri .01'den büyük olduğu durumlarda çoklu bağlantı probleminin olmadığı söylenebilir (Fidell and Tabachnick, 2013). Tolerans değerleri .71 ile .97 arasında değişmektedir. Elde edilen değerler dikkate alındığında çoklu bağlantı problemi olmadığını söylemek mümkündür. Bu çerçevede, literatür bilgileri ışığında veri setine ilişkin tüm sayıtların karşılandığı söylenebilir.

Yapılan araştırmada veri setine ilişkin yapı geçerliliği kanıtlarını ortaya koymak amacıyla her bir grup için ayrı ayrı DFA yapılmıştır. DFA sonuçları bulgular bölümünde verilmiştir. Araştırmada, S-B χ^2 düzeltmesi farklı örneklem büyüklüğüne ve puan dağılımına sahip veri setlerinde, örneklem büyüklüğünün χ^2 üzerindeki etkisini azalttığı için, değeri olarak S-B χ^2 kullanılmış ve model uyumuna karar verilirken χ^2 , χ^2/sd , CFI ve SRMR uyum indeksleri dikkate alınmıştır.

Ölçme modelleri karşılaştırılırken ise uyum indekslerindeki değişimler incelenerek, bu indeksler arasındaki farka bakılmıştır. İlk olarak karşılaştırılan modellerin, χ^2 istatistiği ve serbestlik dereceleri arasındaki (sd) fark ($\Delta\chi^2$ ve Δsd) incelenir. Karşılaştırma sonucunda elde edilen farkın anlamlı olup olmadığı χ^2 tablosundaki değerler $p \leq .01$ ya da $p \leq .05$ düzeyinde karşılaştırılarak tespit edilir. Eğer χ^2 tablosundaki değer elde edilen sonuçtan büyükse modeller arasında anlamlı fark olduğu kabul edilir (Kline, 2005). Bu araştırmada χ^2 olarak S-B χ^2 değeri temel alındığı için fark derecesi belirlenirken Ts değerleri hesaplanmış ve modeller arasındaki farkın manidarlığı elde edilen Ts değeri ile χ^2 tablosundaki kritik değerler $p \leq .05$ düzeyinde karşılaştırılarak belirlenmiştir. Bununla birlikte

model karşılaştırmalarında ΔCFI , $\Delta SRMR$, ΔNFI , ΔGFI , $\Delta RMSEA$, ΔTLI gibi uyum indekslerinin değerlendirilmesi önerilmiştir (Cheng and Reinsvold, 2002). Araştırmada ΔCFI ve $\Delta SRMR$ değerleri temel alınarak karşılaştırmalar yapılmıştır. (Çelebi, Güner, Kaya ve Korumaz, 2016).

Bu karşılaştırmalar için kabul düzeyleri,

$n > 300$ olduğunda metrik değişmezlik için $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .015$, güçlü ve katı faktöriyel değişmezliğin karşılaştırılmasında $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .010$.

$n < 300$ olduğunda metrik değişmezlik için $\Delta CFI \leq -.005$ ve $\Delta SRMR \geq .025$, güçlü ve katı faktöriyel değişmezliğe ait karşılaştırmalarda $\Delta CFI \geq -.005$, $\Delta SRMR \geq 0.005$ olarak belirlenmiştir (Chen, 2007).

Tüm durumlara bakıldıktan sonra ölçme değişmezliğine karar verme aşamasında dikkat edilmesi gereken bazı hususlar bulunmaktadır.

Şekercioğlu'na (2018) göre bunlardan ilki χ^2 'de örnek sayısı arttıkça önem de artmaktadır, bu sebeple alternatif modeller, model çerçevesindeki faktör yapılarının veya kovaryans matrisleri arasındaki uyumun uygun değerlerde olup olmadığını araştırır. Bu değerler arasında ilk önce χ^2 değeri ve serbestlik derecesi karşılaştırılır Bu bağlamda, daha sınırlandırılmış modelden elde edilen χ^2 değeri, daha az sınırlandırılmış modelden elde edilen χ^2 değeri ve serbestlik derecesi arasındaki "delta" değeri (delta fark demektir ve sembolü Δ 'dır) hesaplanır. Bu hesaplama ile $\Delta\chi^2$ ve Δsd değerleri belirlenmiş olur. Bu belirleme sonucunda elde edilen χ^2 değerinin anlamlılık düzeyi, dağılım tablosundaki kritik değerler karşılaştırılarak $p \leq .01$ veya $p \leq .05$ düzeyinde kontrol edilir (Byrne, 2010; Jöreskog, 1971; Kline, 2005; Lee ve Leung, 1982; Steiger, 2007; Leung and Van de Vijver, 1997).

Bu durumda, H_0 ve H_1 hipotezleri aşağıdaki formatta geliştirilebilir:

H_0 : Daha sınırlandırılmış model ile daha az sınırlandırılmış model arasında uyum açısından anlamlı bir fark yoktur.

H_1 : Daha sınırlandırılmış model ile daha az sınırlandırılmış model arasında uyum açısından anlamlı bir fark vardır.

Bundan dolayı, belirli bir Δsd seviyesindeki serbestlik derecesi farkları temelinde hesaplanan $\Delta\chi^2$, kritik tablo değerlerinden düşükse, H_0 kabul edilir. Başka bir söyleyiş ile iki model arasında uyum açısından anlamlı bir fark yoktur, bu nedenle araştırmacı, $\Delta\chi^2$ 'ye dayanarak ölçme değişmezliği hakkında karar verebilir. (Büyüköztürk, Çokluk ve Şekercioğlu, 2010-2012).

3. BULGULAR

Test Eşitliği: Test eşitliği farklı ölçme işlemleri sonucu elde edilen puanların farklı gruplarda farklı zamanlarda karşılaştırılabilir olmasını sağlayan bir kavramdır. Yani test sonuçlarının farklı zamanlarda ve farklı kültürlerde olsa dahi aynı anlama gelmesi demektir. Van de Vijver ve Leung (1997); yapısal eşitlik, ölçme birimi eşitliği ve tam puan eşdeğerliği olmak üzere test eşitliği kavramını üç gruba ayırmışlardır. Yapısal eşitlik, ölçülen özelliğin kültürler arasındaki teorik eşitliğinin sağlanabilmesiyle ilgilidir. Ölçme birimi eşitliği, ölçülen özelliğin eşit aralıklı ölçek ya da oran ölçeği gibi aynı ölçek düzeyinde ölçülmesi demektir. Tam puan eşdeğerliğinde ise ölçme aracı aynı ölçme birimine ve başlangıç noktasına sahiptir. Bu eşitliğe, en üst düzey eşitlikte denebilir. Bu eşitlik sağlandığında kültürler arasında karşılaştırma ve birçok istatistiksel analiz yapılabilir (Matsumoto and Van de Vijver, 2012).

Ölçme değişmezliği, çoklu grup doğrulayıcı faktör analizi ile dört adet hiyerarşik modelin test edilmesi ile ortaya konmaktadır. Bu modeller sırasıyla "yapısal değişmezlik (configural invariance)" (Model 1), "metrik değişmezlik (metric invariance = weak factorial invariance)" (Model 2), "güçlü faktöriyel değişmezlik (strong factorial invariance = scalar invariance)" (Model 3) ve "katı faktöriyel

değişmezlik (strict factorial invariance)" (Model 4) olarak adlandırılmaktadır. Aşamalılık içeren bu dört modelin test edilmesi gereklidir (Meredith, 1993).

Tüm bunlar dikkate alınarak ölçmelerin kusursuz bir biçimde sağlanabilmesi yönündeki çalışmaların yapılması gerekir. Kurulan bir ölçme modeline ait yapının farklı gruplar için aynı olması, kullanılan ölçme aracındaki maddelerin faktör yüklerinin, faktör korelasyonlarının, hata varyanslarının aynı olması anlamına gelir (Bollen, 1989; Byrne, 1998; Jöreskog ve Sörbom, 1993). Ölçme değişmezliğinde deneysel olarak, gizil değişken puanları ile gözlenen puanları ilişkilendirerek bir ölçme modeli oluşturulur ve bu model üzerinden incelenebilir.

Bu bölümde araştırmanın amaçları doğrultusunda kullanılan veri seti için TIMSS 2015 matematik başarısının, OECD üyesi ülkelere göre değişmezliğine ilişkin bulgulara ve söz konusu gruplara ait test istatistiklerine, normallik testlerine ve güvenilirlik katsayılarına yer verilmiştir.

Araştırmanın problemi olan TIMSS 2015 Türkiye uygulaması 8. sınıf matematik başarı testi OECD ülkeleri için ölçme değişmezliğinin sağlanıp sağlanmadığı sorusuna cevap aranmıştır.

TIMSS 2015 matematik başarısının, ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını tespit etmek amacıyla yapılan çoklu grup doğrulayıcı faktör analizi (ÇGDFA) sonuçları Tablo 9'da verilmiştir.

Tablo 9.

OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S-B χ^2 (sd)	M.K.	$\Delta\chi^2$ (Δ sd)	χ^2 /sd	$\Delta\chi^2$ / (Δ sd)	CFI	Δ CFI	SRMR	Δ SRMR	Karar
Model 1^A	23382.36 (7337)	-	-	3.19	-	.86	-	.069	-	
Model 2^B	17923.45 (6873)	M1-M2	5458.91 (464)	2.61	.58	.91	-.05	.048	.021	H ₀ : Ret
Model 3^C	10735.69 (6409)	M2-M3	7187.76 (464)	1.68	.93	.96	-.05	.042	.006	H ₀ : Ret
Model 4^D	23030.71 (6873)	M3-M4	-12295.02 (-464)	3.35	-1.68	.86	.10	.061	-.019	H ₀ : Kabul

* p<.05

** Model karşılaştırmaları (M=Model)

AYapısal (Biçimsel) Değişmezlik (Temel Model) (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

BZayıf Faktöriyel (Metrik) Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

CGüçlü Faktöriyel (Ölçek / Scalar) Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

DKatı Faktöriyel (Tam) Değişmezlik (Artık Varyans Değişmezliği) (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

İlk olarak, matematik testi puanı için gerçekleştirilen analiz sonucunda elde edilen uygunluk indeksleri incelendiğinde, uygunluk indekslerinin genel olarak kabul seviyesini karşıladığı söylenebilir.

Ülkelerin sahip olduğu faktör yükleri, faktör korelasyonları ve hata varyanslarının sabit çıktığı Model 1 (yapısal değişmezlik)'de kovaryans matrislerinde faktör yüklerinin buna bağlı olarak faktör korelasyonlarının ve daha sonrasında hata varyanslarının tüm gruplar için eşit olup olmadığının test edilmesi sonucunda S-B χ^2 (sd) sonucunun 23382.36(7337) olduğu, S-B χ^2 ve serbestlik derecesi oranının χ^2 /sd=3.18695 çıktığı, 5'in altında ve 3'e yakın olması sebebi ile iyi, CFI değerinin .86 olduğu yani .90'a yakın olması sebebi ile kabul edilebilir düzeye yakın ve SRMR değerinin .069 olduğu yani .1'in altında

.05 ile .08 arasında olması sebebi ile iyi uyum gösterdiği görülmektedir. Uyum indekslerinin değerlerinin kabul edilebilir düzeylerde olduğu ve modelin doğrulandığı söylenebilir. Bu durumda yapısal değişmezliğin sağlandığını söylemek mümkündür.

Faktör yükleri, faktör korelasyonları ve hata varyanslarının sabit olduğu Model 1 (yapısal değişmezlik) ile faktör korelasyonları ve hata varyanslarının sabit, faktör yüklerinin serbest bırakıldığı Model 2 (zayıf faktöriyel değişmezlik) karşılaştırıldığında $S-B\chi^2(sd)$ sonucunun 17923.45(6873) olduğu, χ^2/sd oranının 2.6078'e düştüğü yani oranın 3'ün altına düşerek daha da iyileştiği görülmektedir. $\Delta\chi^2(\Delta sd)$ 'nin 5458.91(464) olduğu, böylece $\Delta\chi^2/\Delta sd$ oranının da .5791 olduğu görülmektedir. $S-B\chi^2$ fark derecesinin manidar olup olmadığını tespit etmek için T_s değeri hesaplanmış ve bu değer 7161.3943 olarak bulunmuştur. T_s değerinin χ^2 dağılımı tablosundaki kritik değerden büyük olduğu tespit edilmiştir, $\chi^2_{fark}(464) = 515.2183$, $p < .05$. Bu durumda modeller arasındaki farkın manidar olduğunu ifade etmek mümkündür. CFI değerinin .86'dan .91'e yani .90 ile 1 arasına yükselmesi değerin giderek daha da iyileştiği ve iyi uyum gösterdiği söylenebilir. SRMR değerinin .069'dan .048'e düşmesi değerin giderek daha da iyileştiği göstermekte iyi uyum gösterdiğini ifade etmektedir. Ayrıca ΔCFI değerinin $< -.01$ ve $\Delta SRMR \geq .025$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. ΔCFI değerinin $-.05$ ($< -.01$) olması dolayısı ile manidar bir fark vardır. $\Delta SRMR$ değerinin .021 ($< .025$) olması dolayısı ile farkın manidar olmadığı ifade edilebilir.

Model 1 ve Model 2 karşılaştırmasında Model 2'de genel olarak değerlerin iyileştiği, diğer bir ifadeyle Model 1'de sağlanan ölçme değişmezliğinin daha da iyileştiği ve Model 1 ile Model 2 yani yapısal değişmezlik ile zayıf faktöriyel (metrik) değişmezlik arasındaki değerler genel olarak incelendiğinde, manidar bir fark olduğu söylenebilir. Bu durumda Model 1 ve Model 2 arasında manidar bir fark olması sebebi ile H_0 ret edilmiştir.

Faktör yükleri serbest, faktör korelasyonları ve hata varyanslarının sabit olduğu Model 2 (zayıf faktöriyel değişmezlik) ile faktör yükleri ve hata varyanslarının serbest, faktör korelasyonlarının sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) karşılaştırıldığında, $S-B\chi^2(sd)$ sonucunun 10735.69(6409) olduğu, $\Delta\chi^2(\Delta sd)$ 'nin 7187.76(464) olduğu, böylece χ^2/sd oranının 1.6750'e düştüğü görülmektedir. $\Delta\chi^2/\Delta sd$ oranının 1.6750 olduğu ve oldukça iyileştiği tespit edilmiştir. $S-B\chi^2$ fark derecesinin manidar olup olmadığını tespit etmek için T_s değeri hesaplanmış ve bu değer 13212.7076 olarak bulunmuştur. T_s değerinin χ^2 dağılımı tablosundaki kritik değerden büyük olduğu tespit edilmiştir, $\chi^2_{fark}(464) = 515.2183$, $p < .05$. Bu durumda modeller arasındaki farkın manidar olduğu ifade etmek mümkündür. CFI değerinin .91'den .96'ya yani .90 ile 1 arasında 1'e daha yakın bir değere yükselmesi değerin giderek daha da iyileştiğini göstermekte iyi uyum gösterdiğini ifade etmektedir. SRMR değerinin .048'den .042'ye düşmesi değerin .00 ile .05 arasında bir değere gelmesi ve .00'a giderek yaklaşması sebebi ile değerin giderek daha da iyileştiği göstermekte iyi uyum gösterdiğini ifade etmektedir. Ayrıca ΔCFI değerinin $< -.01$ ve $\Delta SRMR > .01$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. ΔCFI değeri $-.05$ ($< -.01$) olması dolayısı ile manidar bir fark vardır. $\Delta SRMR$ değerinin .006 ($< .01$) olması dolayısı ile manidar bir fark yoktur denilebilir. Model 2 ve Model 3 karşılaştırmasında Model 3'te sağlanan ölçme değişmezliğinin daha da iyileştiği ve Model 2 ve Model 3 yani zayıf faktöriyel (metrik) değişmezlik ile güçlü faktöriyel değişmezlik arasındaki değerler ($S-B\chi^2(sd)$, χ^2/sd , $\Delta\chi^2(\Delta sd)$, CFI, SRMR, ΔCFI , $\Delta SRMR$, T_s değeri ve χ^2_{fark} kıyaslaması) genel olarak incelendiğinde, manidar bir fark olduğu ifade edilebilir. Bu durumda Model 2 ve Model 3 arasında manidar bir fark olması sebebi ile H_0 ret edilmiştir.

Faktör yükleri ve hata varyanslarının serbest, faktör korelasyonlarının sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) ve hata varyanslarının serbest, faktör yükleri ve faktör korelasyonlarının sabit olduğu Model 4 (katı faktöriyel değişmezlik) karşılaştırıldığında, $S-B\chi^2(sd)$ sonucunun 23030.71(6873) ve $\Delta\chi^2(\Delta sd)$ 'nin -12295.02(-464) olduğu, χ^2/sd oranının 3.3508'e yükseldiği yani oranın 3 ile 5 arasında bir değer olarak yeterli uyum sağladığı ancak Model 2 ve Model 3 kıyaslamasına göre

oldukça kötüleştiği görülmektedir. $\Delta\chi^2/\Delta sd$ oranının -1.6758 olduğu ve oldukça kötüleştiği tespit edilmiştir. S-B χ^2 fark derecesinin manidar olup olmadığını tespit etmek için T_s değerine bakılmış ve değerlerinin oldukça kötü olması dolayısı ile hesaplanamamış, ayrıca χ^2_{fark} değeri değer aralığı dışında çıkmış ve hesaplanamamıştır. Bu sebepten T_s değerinin χ^2 dağılımı tablosundaki değerle kıyaslaması yapılmamış ve sonuç vermemiştir. Bu durumda modeller arasındaki farkın manidar olmadığını ifade etmek mümkündür. CFI değerinin .96'dan .86'ya bir başka deyişle .90 değerinin altına düşmesi uyumun kötüleştiğini göstermektedir. SRMR değerinin .042'den .0461'e yükselmesi, değerin .00 ile .05 arasında olmasına rağmen Model 3'e göre değerin kötüleşmesi sebebi ile iyi uyum göstermediğini ifade etmektedir. Ayrıca ΔCFI değerinin $<-.01$ ve $\Delta SRMR >.01$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. ΔCFI değerinin .10 ($>-.01$) manidar bir fark olmadığı ve değerin biraz daha kötüleştiği tespit edilmiştir. $\Delta SRMR$ -.019 ($<.01$) manidar bir fark olmadığı yani SRMR değerlerinin de kötüleştiği, sonuçların genel olarak kötüleştiği ve manidar bir fark gözlenmediği tespit edilmiştir. Model 3 ve Model 4 karşılaştırmasında, Model 4'te genel olarak değerlerin kötüleştiği bu durumda Model 3 ve Model 4 yani güçlü faktöriyel değişmezlik ile katı faktöriyel değişmezlik arasında manidar bir farkın olmadığını söylemek mümkündür. Bu durumda Model 3 ve Model 4 arasında manidar bir fark olmaması sebebi ile H_0 kabul edilmiştir.

Tüm model karşılaştırmalarına ait bulgular değerlendirildiğinde karşılaştırılan dört model içerisinde en iyi çalışan model güçlü faktöriyel değişmezlik (Model 3) modelidir. Bu doğrultuda yapılan analizler sonucunda TIMSS 2015 matematik başarısının faktör yapısının OECD üyesi ülkeler için eşit olmadığı ve ölçme değişmezliğini sağlamadığı kararı alınmıştır.

4.TARTIŞMA VE SONUÇ

TIMSS 2015 matematik başarısı OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek için test istatistikleri, normallik ve güvenilirlik hesaplamaları yapılmıştır. Bu hesaplamalar sonucunda veri setinin normale yakın bir dağılım gösterdiği kabul edilmiştir. 17 ülke arasında kovaryans matrislerinin uyum düzeyini belirleyebilmek için yapılan kovaryans matrislerinin eşitliği testi sonucunda, ülkeler arasında kovaryans matrislerine ait modelin iyi uyum gösterdiği sonucuna varılmıştır. TIMSS 2015 matematik başarısının OECD üyesi ülkelere göre ölçme değişmezliğinin çoklu-grup doğrulayıcı faktör analizi ile dört model temel alınarak test edilmesi sonucunda temel model olarak kurulan Model 1'e ait uyum indekslerinin kabul edilebilir düzeyde olduğu ve modelin doğrulandığı tespit edilmiştir. Bu durumda OECD ülkelerine göre yapısal değişmezliğin sağlandığı ifade edilebilir. Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, Model 2 ve Model 3'e ait uyumun oldukça iyileştiği ancak manidar farklar olması sebebi ile ölçme değişmezliği olmadığı belirlenmiştir. Model 3 güçlü faktöriyel değişmezlik ve Model 4 katı faktöriyel değişmezlik arasında manidar bir farkın olmadığını söylemek mümkündür.

Tüm model karşılaştırmalarına ait bulgular değerlendirildiğinde karşılaştırılan dört model içerisinde en iyi çalışan model güçlü faktöriyel değişmezlik (Model 3) modelidir. Bu doğrultuda TIMSS 2015 matematik başarısının faktör yapısının OECD üyesi ülkeler için eşit olmadığı ve ölçme değişmezliğini sağlamadığı kararı alınmıştır. Ölçme değişmezliğinin en iyi güçlü faktöriyel değişmezlik düzeyinde sağlandığı sonucuna ulaşılmıştır. TIMSS 2015 Matematik başarı değişmezliğinin 11 no'lu kitapçık üzerinden değerlendirildiği bu araştırmanın bulguları, OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanamadığını göstermektedir.

Ülke değişkenine ilişkin sonuçlar değerlendirildiğinde, geniş ölçekli sınavlarda kullanılan ölçme araçlarının farklı dillere ve kültürlerle uyarmasının sınırlılık oluşturduğu söylenebilir.

1. Bazı araştırma sonuçları farklı ülkelerde uygulanan geniş ölçekli çalışmalarda farklı ülkeler arasında ölçme değişmezliğinin sağlandığını raporlaştırmışlardır (Akyıldız, 2009; Marsh ve diğerleri, 2006;).

2. Bazı çalışmalar farklı dil ve kültürle sahip bireyler arasında kullanılan ölçme araçlarında ölçme değişmezliğinin sağlanamadığını raporlamıştır (Ercikan ve Koh 2005; Karakoç Alatlı, 2016; Kıbrıslıoğlu, 2015; Marsh ve diğerleri, 2013; Öğretmen, 2006; Wu ve diğerleri 2007).

Bununla birlikte, geniş ölçekli sınav uygulamalarının sonuçları değerlendirilirken ölçme değişmezliğinin sağlanamıyor olması, özellikle ülkeler arası karşılaştırmalar ve yorumlamalar yapılırken göz önünde bulundurulmalıdır. Tüm sonuçlar değerlendirildiğinde, TIMSS 2015 Türkiye uygulamasında matematik başarısına ait elde edilen sonuçların farklı ülkeler yani, farklı kültür ve dillere sahip katılımcılardan elde edilen sonuçların karşılaştırılabilirliği ve buna bağlı olarak yapılan yorumlar ölçme değişmezliğinin sağlanamaması yani uygulanan ölçme aracının tüm katılımcılar için aynı anlamı ifade etmemesi nedeniyle tartışmaya açık hale gelmektedir.

Araştırma süreci ve elde edilen sonuçlar değerlendirildiğinde, uygulamaya yönelik ve araştırmacıya yönelik olmak üzere iki grupta öneriler verilebilir.

Uygulamaya yönelik öneriler:

1. Araştırma sadece OECD üyesi ülkelere ait veri seti üzerinden yürütülmüştür. Gelecek çalışmalarda bu araştırmada yer almayan diğer ülkelerin de bulunduğu ve demografik değişkenlerin dâhil edildiği araştırmalar yapılabilir.

2. Araştırmada çalışma yapılırken en güncel uluslararası sınav olan TIMSS 2015 ele alınmıştır. Sonraki araştırmalar TIMSS 2019 ya da PISA 2019 güncel data setinde yapılabilir.

3. Bu araştırma yalnızca TIMSS 2015 uygulaması 11 no'lu kitapçıkta bulunan 8. sınıf matematik başarısına ilişkin maddeler üzerinden yürütülmüştür. Sonuçların genellenebilmesi için gelecek araştırmalarda farklı kitapçıklarda ya da farklı başarı alanlarına (okuma becerileri, anketler ve fen başarısı) ilişkin ölçme değişmezliği incelenebilir.

4. Bu araştırmada ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi kullanılarak test edilmiştir. Gelecek araştırmalarda farklı yöntemler kullanılarak ölçme değişmezliği test edilirse, yöntemler arası karşılaştırmalar yapılması ve hangi yöntemin daha etkili olduğuna karar verilmesi mümkün olabilir.

5. Araştırmada ülkeler arasında ölçme değişmezliğinin sağlanamadığının belirlenmiş olması, farklı gruplar (cinsiyet, dil vb.) arasında karşılaştırmalar yapılmadan önce yapılan yorumların anlamlı ve daha sağlıklı olabilmesi için uygulanan ölçme aracına ilişkin ölçme değişmezliği çalışmalarının yapılmasının oldukça önem taşıdığını göstermiştir.

Uygulayıcıya yönelik öneriler:

1. Her ülkenin kendine has yaşanmışlığı, tarihi ve kültürel değer ve gelenekleri olduğu bir gerçektir. Bu sebeple madde yazımı ve çeviri komisyonlarında her bir ilgili kültürdeki temsilcilerin de yer alması gerekmektedir. Sadece birebir çevirinin yeterli olmadığı bilinmelidir.

2. Bu araştırmaya göre özellikle ortalamalar kullanılarak yapılan hipotez testlerinin bulgularının tartışmalı olduğu bilinmelidir. Bu bulgulara dayalı olarak verilecek eğitimsel ve politik kararların da tartışmalı olacağı gözden kaçırılmamalıdır.

3. Geniş ölçekli olarak yürütülen ölçme ve değerlendirme uygulamalarında, ölçülen özelliğin ya da yapılan uygulamanın türüne göre ölçme değişmezliğinin farklılık gösterdiğini, elde edilen sonuçlara göre sonucu bulma, karşılaştırma ve yorumlama yapılmadan önce ölçme değişmezliği çalışmalarının titizlikle yapılması gerektiği göz ardı edilmemelidir.

4. Araştırmalar sonucunda TIMSS'in PISA'ya nazaran daha az çalışıldığı görülmüştür. Araştırmacılar TIMSS üzerinde birçok alanda daha fazla çalışma yapabilirler.

5. Araştırmalar sonucunda TIMSS matematik başarı testinin fen başarı testine nazaran daha az çalışıldığı görülmüştür. Araştırmacılar matematik başarı testi üzerinde daha fazla çalışma yapabilirler.

KAYNAKÇA

- Akgün, Ö.E., Büyüköztürk, Ş., Çakmak, E. Demirel, F., ve Karadeniz, Ş. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Alma, Ö. ve Vupa, Ö. (2008). Regresyon analizinde kullanılan en küçük kareler ve en küçük medyan kareler yöntemlerinin karşılaştırılması. *Sdü Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi)*, 3(2), 2008, 219-229
- Alwin, D. F., Muthen, B., Summers, G. F. & Wheaton, B. (1977). Assessing reliability and stability in panel models. *Sociological methodology*, 8(1), 84-136.
- Arıkan, S., Sarı, M. H. ve Yıldızlı, H. (2017). 8. Sınıf Matematik Akademik Başarısını Yordayan Faktörler-TIMSS 2015. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 2017, 246-265.
- Arora, Foy, P., Martin, M.O. & Mullis, I. A. (2012). *TIMSS 2011 international results in mathematics*. 2012, 1.
- Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing*. Advance online publication. [Çevrim-içi: <https://www.tandfonline.com/doi/abs/10.1080/15305058.2015.1064431>] Erişim Tarihi: 14 Temmuz 2018.
- Atalay, Ö. Bayır, E. ve Çakıcı, Y. (2016). Fen bilimleri öğretmenlerinin bilimin doğasına ilişkin görüşleri: Bilişsel harita örneği. *Kastamonu Eğitim Dergisi*, 24(3), 1419-1436.
- Ayas, A., Çepni, S., Johnson, D. ve Turgut, F. (1997). *Fizik Öğretimi*. Ankara: YÖK/Dünya Bankası, Milli Eğitim Geliştirme Projesi
- Ayvallı, M. (2016). *PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğinin incelenmesi. An investigation into the measurement invariance of PISA 2012 mathematical literacy test*. Yayınlanmamış Yüksek Lisans Tezi, Akdeniz Üniversitesi, Antalya, 2016.
- Babakus, E. (1985). *The sensitivity of maximum likelihood factor analysis given violations of interval scale and multivariate normality*. Doctoral dissertation, The University of Alabama. [Çevrim-içi: <https://elibrary.ru/item.asp?id=7467190>] Erişim Tarihi: 14 Eylül 2018.
- Bahadır, E. (2012). *Uluslararası öğrenci değerlendirme programına (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, 2012.
- Bailey, K.D. (1982). *Methods of social research*. (Second Edition). New York: The Free Press.
- Balcı, A. (2013). *Sosyal bilimlerde araştırma yöntem, teknik ve ilkeler* (10. baskı). Ankara: Pegem Akademi
- Barrett (2007). *Personality and Individual Differences*, 42(5), 859-67.
- Başusta, N.B. ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. Examination of measurement invariance at groups' comparisons: a study on PISA student questionnaire. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Bennet, N., James, L.R., Lind, S., Mulaik, S.A., Stilwell, C.D. & Van Alstine, J., (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3), 430-45.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588.

- Bentler, P.M. & Hu, L.T. (1999). Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Bentler, P. M. & Satorra, A. (2011). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika: A Journal of Quantitative Psychology*, 66(4), 507-514.
- Boadu, K., Boulianne, S., Cummings, G., Hayduk, L. & Pazderka-Robinson, H. (2007). Testing! Testing! One, two, three—testing the theory in structural equation models!. *Personality and Individual Differences*, 42(5), 841-850.
- Boehm, J. K., Lyubomirsk, S., Özer, D. J. & Tucker, K. L. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research*, 78, 341–360.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley
- Bollen, K. A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological bulletin*, 107(2), 256.
- Bornstein, M.H. & Putnick, D.L. (2016). *Measurement invariance conventions and reporting: The state of the art and future directions for psychological research*. *Developmental Review* 41, 2015, 71-90.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Newyork: Guilford Publications, Inc.
- Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Çokluk, Ö. ve Şekercioğlu, G. (2010-2012). *Sosyal Bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi.
- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Byrne, B. M. (2010). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872-882.
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.
- Byrne, B. M., Muthen O. & Shavelson, J. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance May 1989 *Psychological Bulletin* 105(3):456-466 DOI: [10.1037/0033-2909.105.3.456](https://doi.org/10.1037/0033-2909.105.3.456)
- Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, 14(3), 464-504.
- Cheng, Y. & Reinsvold, R. B. (2002). *Factors Associated with foreign language writing anxiety*. Retrieved 25, 2002. doi: <https://doi.org/10.1111/j.1944-9720.2002.tb01903.x>
- Çelebi, N., Güner, H., Kaya G. T. ve Korumaz, M. (2016). Teachers loyalty to their supervisors and organizational commitment. *Educational Research and Reviews*, 11(12), 1161-1167.
- Çelebi, N., Güner, H., Kaya G. T. ve Korumaz, M. (2014). Neoliberal eğitim politikaları ve eğitimde fırsat eşitliği bağlamında uluslararası sınavların (PISA, TIMSS ve PIRLS) analizi. *Tarih Kültür ve Sanat Araştırmaları Dergisi*, 3(3), 33-75. doi: 10.5897/ERR2016.2808 (Yayın No: 2864277)
- Davidov & others, (2014). The Goals of Regulating Work: Between Universalism and Selectivity [Çevrim-ici: https://www.academia.edu/17915574/The_Goals_of_Regulating_Work_Between_Universalism_and_Selectivity] Erişim Tarihi: 22 Temmuz 2018

- Davidov & others, (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, 5, 982.
- Doğan, N. ve Uyar, Ş. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2, 30-43.
- EARGED (2014). *TIMSS 2011 ulusal matematik ve fen raporu: 8. sınıflar*. 01 Eylül 2018 tarihinde <http://timss.meb.gov.tr/wp-content/uploads/TIMSS-2011-8-Sinif.pdf> adresinden alınmıştır.
- Erdoğan Akan, O. ve Ertürk, Z. (2018). TIMSS 2015 matematik başarıları ile ilgili bazı değişkenlerin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Kuramsal Eğitimbilim Dergisi, UBK-2018*, 204-226, *Aralık 2018 Journal of Theoretical Educational Science, ICSE-2018*, 204-226, December 2018 [Çevrim-ici: <http://www.keg.aku.edu.tr>] Erişim Tarihi: 08 Temmuz 2018
- Ergül, H. (1999). Uzaktan öğretimde kalite verimlilik ve üretkenlik. *Anadolu Üniversitesi Açık Öğretim Fakültesi Kurgu Dergisi*, 16, 283-296.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural equation modeling: a multidisciplinary journal*, 6(1), 56-83.
- Fidell, L.S. & Tabachnick, B.G. (2013). *Using multivariate statistics* (6th ed.). New Jersey: Pearson.
- Gierl, M.J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education*, 25(4), 280-296.
- Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37, 1, GEFAD / GUJGEF 37, 1: 151-175 (2017) Yayına Kabul Tarihi: 27.12.2016
- Ho, M.-H.R. & McDonald, R.P. (2002). Principles and practice in reporting statistical equation analyses. *Psychological Methods*, 7(1), 64-82.
- Horn and McArdle, (1992). Life satisfaction across adulthood: different determinants at different ages. [Çevrim-ici: https://www.academia.edu/14360191/Life_satisfaction_across_adulthood_different_determinants_at_different_ages] Erişim Tarihi: 22 Temmuz 2018
- Hox, J., Lugtig, P. & Schoot, R., (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: structural equation modeling with the simplis command language*. USA: Scientific Software International, Inc.
- Karakoç Alatlı, B. (2016). *Uluslararası öğrenci değerlendirme programı (PISA - 2012) okuryazarlık testlerinin ölçme değişmezliğinin incelenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.
- Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye-Çin, (Şangay)-Endonezya örneği*. (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (Second Edition) New York: Guilford Press.
- 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 Journal of Measurement and Evaluation in Education and Psychology*, 6(2), 2015, 351-364

- Koğar, H. ve Yılmaz Koğar, E. (2017), Öğretmenlerin matematik konularına yönelik hazırlık düzeylerinin matematik başarıları ile ilişkisi: TIMSS 2015 Türkiye ve Singapur örneği. *Başkent Üniversitesi Eğitim Dergisi*, 4(2), 2017, 108-121,
- Korkmaz, H. (2004). *Fen ve teknoloji eğitiminde alternatif değerlendirme yaklaşımları*, Yeryüzü Yayınevi, Ankara.
- Köse, İ., Uysal Ö. ve Yandı, A. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi - Mersin University Journal of the Faculty of Education "Mersin Üniversitesi Eğitim Fakültesi Dergisi"*, 13(1), 2017, 243-253. doi: <http://dx.doi.org/10.17860/mersinefd.305952>.
- Lance, C. E. & Vandenberg, R. J. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69, 4-70. 24 Ekim 2005 tarihinde <http://orm.sagepub.com/cgi/reprint/3/1/4.pdf> adresinden alınmıştır.
- Lee, S. Y. & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, 47, 297-308.
- Leung, K. & Van de Vijver, F. J. R., (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage
- Li, Z., Wu, A. D. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data practical assessment, *Research and Evaluation*, 12(3), 1-26.
- Marcoulides, G. A. & Raykov, T. (2008). *An Introduction to Applied Multivariate Analysis* (First Edition). NY: Taylor & Francis Group.
- Matsumoto, D. & Van de Vijver, F. J. (Eds.). (2011). *Cross-cultural research methods in psychology*. NY:Cambridge University Press.
- Martin, M.O. & Mullis, I.V.S. (Eds.) (2012). *Methods and procedures in TIMSS and PIRLS 2011: Characteristics of national samples*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved, February, 16, 2015, from http://timss.bc.edu/methods/pdf/T11_Characteristics.pdf
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Miles, J. & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and individual differences*, 42(5), 869-874.
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge.
- Moraes, C.L. & Reichenheim, M.E. (2002). Cross-Cultural measurement equivalence of the revised conflict tactics scales (CTS2) portuguese version used to identify violence within couples. *Saúde Pública*, 18, 3. Retrieved October, 26, 2018 from http://www.scielo.br/scielo.php?script=sci_arttext&pid
- Öğretmen, T. ve Uzun, B. (2010). Fen başarıları ile ilgili bazı değişkenlerin TIMSS-R Türkiye örneğinde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *Eğitim ve Bilim*, 35, 155, 26-35.
- Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. 22 Şubat 2018 tarihinde <https://odsgm.meb.gov.tr/kurslar/> adresinden alınmıştır.
- Ölçülüoğlu, R. (2015). *TIMSS 2011 Türkiye sekizinci sınıf matematik başarılarını etkileyen değişkenlerin bölgelere göre incelenmesi*. Yayınlanmış yüksek lisans tezi. [Çevrim-ici: www.openaccess.hacettepe.edu.tr] Erişim Tarihi: 18 Ocak 2019

- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42, 893-898.
- Şekercioğlu, G. (2018). *Measurement invariance: Concept and implementation*. *International Online Journal of Education and Teaching (IOJET)*, 5(3), 609-634. 02 Temmuz 2018 tarihinde <http://iojet.org/index.php/IOJET/article/view/439/257> adresinden alınmıştır.
- TIMSS 2015 Ulusal Raporu. 17.03.2018 tarihinde http://timss.meb.gov.tr/?page_id=147 adresinden alınmıştır.
- TIMSS PIRLS web adresi. 22 Temmuz 2018 tarihinde <http://timssandpirls.bc.edu/> adresinden alınmıştır.
- Tyson, (2004). Theories Underlying Literary Criticism. [Çevrim-içi: https://www.academia.edu/7113249/THEORIES_UNDERLYING_LITERARY_CRITICISM] Erişim Tarihi: 22 Temmuz 2018

EXTENDED ABSTRACT

1. Introduction

In the globalized world, international studies as well as national measurement and evaluation activities, which are effective in determining and reorganizing educational policies, gain importance. In order to determine the position of countries at international level in terms of education system, their differences with other countries and lack of education, there are exams which are accepted by many countries as reference and applied at international level. In many countries, studies are carried out in order to identify the deficiencies of the education system and its components and to develop the required areas in consideration of student achievement (MEB, 2014). These are considered to be of high importance because evaluators use the results to compare student performance and evaluate the effectiveness of educational policies and practices in the participating countries (Gierl, 2000).

In this study, by using TIMSS 2015 data, the mathematical achievement of the eighth grade students who participated in the mathematics achievement test is analysed. It is tried to determined what kind of invariance is the best provided if this success is different in terms of countries, whether it provides invariance or not. Considering common and different aspects in the studies conducted previously in this field, the recommendations for the practitioner and the research are mentioned in important points.

2. Method

In first step, among cognitive test booklets implemented in TIMSS 2015, booklet 11 which was used commonly by all countries was selected for this correlational survey study and the study was conducted using dataset belonging to 301.472 students that took booklet 11 in the countries that attend TIMSS. Test statistics, normality tests and reliability coefficients were examined for the eighth grade mathematic achievement test scores of the countries and it was concluded that the measures of central tendency and distribution were similar. For simplicity, it is possible to say that the data set has a close distribution to normal. The reliability coefficients and internal consistency of test items were generally good. In the following stage, OECD countries were selected and 7.223 students attending the eighth grade in these countries were subjected to confirmatory factor analysis and the measurement models were validated for all groups. The χ^2 / sd ratio was found to be quite good. When considered to the compliance indices, it is seen that CFI datas are compatible with the patterns in all countries. It is determined GFI datas are delivering compatible results in Hong Kong, Swedish, Italy, Japan, Korea and Turkey , for the other countries it extremely shows the compatible results, SRMR datas are closed to harmony in Hong Kong, Swedish, Italy, and for the rest of the other countries it is estimated that they show the best compatible. When compatible index evaluated in general, it is decided that evaluation models of all countries are verified for all groups.

Then, for all groups, covariance matrices were tested for equality. The other step is testing of measurement invariance. Considering the results of the covariance matrices equality test for the mathematics achievement test scores of OECD member countries, it can be stated that the fit indexes are in good agreement. H_0 (there is no significant difference between the more restricted model and the less restricted model) and H_1 (there is a significant difference between the more restricted model and the less restricted model) hypotheses are established, and the measurement invariance is tested by multi-group confirmatory factor analysis. It can be said that the values of the fit indices for the structural invariance are acceptable and the model is confirmed. Then, H_0 has been rejected because of a significant difference between the models based on the nested model based on Model 2, Model 3 and Model 4, which were established as an alternative to Model 1, respectively the configural (structural) invariance (Model 1) and the weak factorial invariance (Model 2), configural (structural) invariance (Model 1) and strong factor invariance (Model 3), weak factorial invariance (Model 2) and strong

factorial invariance (Model 3). (H_1 was accepted.) H_0 was accepted because of a significant difference between the configural (structural) invariance (Model 1) and the strict factorial invariance (Model 4), the strong factorial invariance (Model 3) and the strict factorial invariance (Model 4) (H_1 was rejected). As a result, the results of the study showed that the measurement invariance was not achieved for OECD member countries, and the best model for measurement invariance was the strong factorial invariance model.

3. Findings, Discussion and Results

In the literature, some research results reported that the measurement invariance between different countries was ensured in large scale studies. On the contrary, some research results reported that the measurement invariance could not be achieved in the measurement tools used among individuals with different languages and different cultures. In the context of these results, the necessity of the presence of people from different cultures and nationalities in the article writing because each country has its own unique experience, historical, cultural values and traditions. From the preparation of the exam question, to the implementation of the exam and even to the examination and analysis of the exam result, the decision-makers and politicians in the education should not be too confident in the group fixtures because they have very important functions in these exams. The interpretation of the measurement invariance according to the measured characteristics and the type of application, and the efforts to be meticulously performed, create suggestions for the practitioner. Including different countries and demographic variables in future studies as suggestions for research, analysis of different booklet types and success areas, determining which method can be more effective by testing the measurement invariance with different methods, In order to make the interpretations made before different comparisons between different groups such as language, gender and culture, to be more meaningful and healthy, it is very important to carry out the measurement invariance of the measurement tool with great care.

When all the results are evaluated, the comparability of the results obtained from the results of mathematics achievement in the TIMSS 2015 application against the participants with different countries, languages, genders and cultures, and the comments made accordingly, cannot be ensured by the measurement invariance is becoming. However, measurement of the results of large-scale exam applications cannot be achieved in measuring the results, especially when making comparisons between countries and must be taken into consideration diligently.