

PISA Okuma Keyfi Ölçeğinin Farklı Başarı Gruplarında Yer Alan Ülkelere Göre Ölçme Değişmezliğinin İncelenmesi

Yusuf Kasap*, Nuri Doğan**, Mustafa Köroğlu***, Özlem Doğan****

Makale Geliş Tarihi:02/02/2025

Makale Kabul Tarihi:29/06/2025

DOI: 10.35675/befdergi.1630600

Öz

Bu çalışma, PISA 2018 öğrenci anketinde yer alan okuma keyfi ölçeğinin farklı başarı gruplarındaki ölçme değişmezliğini incelemeyi amaçlamaktadır. Ölçme değişmezliği, gizil değişkenler ile gözlenen değişkenler arasındaki ilişkinin farklı gruplarda aynı biçimde sürdürülüp sürdürülmediğini ortaya koyar. Araştırmanın örneklemini, üst başarı grubunu temsil eden Norveç (n=5659), orta başarı grubunu temsil eden Türkiye (n=6858) ve düşük başarı grubunu temsil eden Kolombiya'dan (n=7192) olmak üzere toplam 19.709 öğrenciden oluşmaktadır. Açıklayıcı faktör analizi sonucunda beş maddeli tek faktörlü model doğrulanmış, ardından çok gruplu doğrulayıcı faktör analiziyle ölçme değişmezliği test edilmiştir. Bulgular, modelin şekilsel, metrik, ölçek ve katı düzeylerde değişmezlik göstermediğini ortaya koymuştur. Bu sonuç, ülkeler arasında yapılan karşılaştırmaların geçerliliğini sınırlamakta ve ölçeğin farklı başarı gruplarında ortak bir ölçme yapısını sağlamadığını göstermektedir.

Anahtar Kelimeler: Okuma keyfi ölçeği, ölçme değişmezliği, PISA 2018

An Examination of the Measurement Invariance of the PISA Reading Enjoyment Scale across Countries with Different Achievement Levels

Abstract

This study aims to examine the measurement invariance of the reading enjoyment scale included in the PISA 2018 student questionnaire across different achievement groups. Measurement invariance reveals whether the relationship between latent variables and observed variables is maintained consistently across groups. The sample consists of a total of

*Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara, Türkiye, yusufkasap5806@hotmail.com, Orcid: [0000-0002-5114-1175](https://orcid.org/0000-0002-5114-1175) 

**Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara, Türkiye, nurid@hacettepe.edu.tr, Orcid: [0000-0001-6274-2016](https://orcid.org/0000-0001-6274-2016) 

***Erzincan Binali Yıldırım Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Erzincan, Türkiye, mustafa.koroglu@erzincan.edu.tr, Orcid: [0000-0001-9610-8523](https://orcid.org/0000-0001-9610-8523) 

****Millî Eğitim Bakanlığı, Ankara, Türkiye, sudoganozlem@gmail.com, Orcid: [0009-0004-5896-7895](https://orcid.org/0009-0004-5896-7895) 

Kaynak Gösterme: Kasap, Y., Doğan, N., Köroğlu, M. & Doğan, Ö. (2025). PISA okuma keyfi ölçeğinin farklı başarı gruplarında yer alan ülkelere göre ölçme değişmezliğinin incelenmesi. *Bayburt Eğitim Fakültesi Dergisi*, 20(47), 1190-1210.

19,709 students: Norway representing the high-achievement group (n=5659), Turkey representing the middle-achievement group (n=6858), and Colombia representing the low-achievement group (n=7192). Exploratory factor analysis confirmed a single-factor structure with five items, followed by multi-group confirmatory factor analysis to test measurement invariance. The findings indicated that the model did not demonstrate configural, metric, scalar, or strict invariance across groups. This result suggests that cross-country comparisons are limited in validity and that the scale does not provide a common measurement structure across different achievement levels.

Keywords: Reading enjoyment scale, measurement invariance, PISA 2018

Giriş

Eğitimde öğrencilerin bilgi düzeylerini belirlemek için yapılan ölçme ve değerlendirme süreci önem teşkil etmektedir. Ölçme ve değerlendirme sürecinde çeşitli sınav ve değerlendirme türleri uygulanmaktadır. Uygulanan sınavlar ve yapılan değerlendirme işlemleri ulusal ve uluslararası alanda yapılmaktadır. Bu çerçevede gerçekleştirilen sınavlardan biri, PISA, OECD'nin yürüttüğü Uluslararası Öğrenci Sınama Programı'dır. Türkiye'nin 2003 yılından itibaren katıldığı ve üç yılda bir belirli konu alanlarına ağırlık verilerek yapılan PISA'nın amacı, 15 yaşındaki öğrencilerin okuduğunu anlama becerisi, matematik okuryazarlığı ve fen okuryazarlığı olmak üzere üç alandaki başarı düzeylerini belirleyip bu alanlarda edindikleri becerileri günlük yaşama aktarıp aktaramadıklarını saptamaktır. Ayrıca PISA'dan elde edilen sonuçlar katılımcı ülkeler ve bu ülkelerde yer alan gruplar arasında ölçülen değişkenler bakımından karşılaştırma yapılmasına imkân sağlamaktadır.

PISA ve Okuryazarlık

PISA 2018, okuduğunu anlama becerisini, bireyin amaçlarına ulaşması, bilgi ve yeteneklerini geliştirmesi, ayrıca toplum içinde ticaret, üretim ve hizmet gibi alanlarda etkili bir şekilde yer alabilmesi için metinleri doğru bir şekilde kavrama, değerlendirme, eleştirel düşünme ve analiz etme yetisi olarak tanımlamaktadır (OECD, 2019). Bu çerçevede, PISA, okuduğunu anlama becerisini, bireyin metinlerden hem gerçek hem de mecazi anlamlar çıkarabilme yeteneği olarak görmektedir. Bu beceri, yalnızca kelimelerin anlamını çözmeyi değil, aynı zamanda öyküleyici metinlerdeki ana fikri belirlemeyi de içerir. Metinden elde edilen bilgilerin bir amaç doğrultusunda ya da belirli bir görüşü desteklemek için kullanılması da bu kapsamda değerlendirilmektedir. PISA'da yapılan okuma etkinliklerinin büyük bir bölümü bu tür bilgi kullanımını içermektedir. Birey, metinle ilgili düşüncelerini ifade ederken, okuduklarıyla kendi yaşam deneyimleri ve fikirleri arasında bağ kurar. Bu süreçte, kendi deneyimlerini farklı bir perspektiften değerlendirebilir veya metindeki bilgilerin doğruluğu ve geçerliliği hakkında bir sonuca ulaşabilir. Sonuç olarak, birey, metni okuyarak onun belirli bir amaca hizmet edip etmediğini analiz etme ihtiyacı hisseder (MEB, 2010).

Okuryazarlık, yalnızca okulda alınan örgün eğitimle sınırlı kalmayan, bireyin ailesi, arkadaş çevresi ve meslektaşları gibi sosyal gruplarla etkileşim içinde gelişimini sürdüren önemli bir unsurdur. Günümüzde bu kavram, eğitimin ilk aşamalarında edinilen bir beceri olmanın ötesine geçmekte ve bireyin yaşam boyu, sosyal çevresiyle kurduğu ilişkiler yoluyla geliştirdiği bir yetkinlik olarak değerlendirilmektedir (OECD, 2009). PISA'nın ele aldığı okuryazarlık, bireyin modern topluma üretken bir şekilde katılmasını sağlayan, gerçek yaşam koşullarında ihtiyaç duyulan bilgi ve becerilere dayalı bir yaklaşımı ifade etmektedir. Bu üst düzey okuryazarlık, bireyin yalnızca temel okuma yazma becerilerinden ziyade, ekonomik ve toplumsal hayata etkin katılımını mümkün kılan yetkinlikleri de kapsamaktadır. Bu bağlamda, bireyin okuryazarlık becerilerini yaşam boyu geliştirmesi ve bu becerileri ekonomik, toplumsal ve kültürel ortamlarda etkin biçimde kullanabilmesi, okuma tutumu ve istekliliğiyle doğrudan ilişkilidir. Bu nedenle, öğrencilerin okuma eylemine karşı duyduğu keyif ya da isteklilik düzeyinin, okuryazarlığın işlevsel kullanımına ve dolayısıyla bireysel ve toplumsal kalkınmaya etkisi olduğu söylenebilir. Dolayısıyla bu çalışmada, PISA 2018 uygulamasında yer alan okuma keyfi ölçeğinin farklı başarı düzeyine sahip ülkelerdeki ölçme geçerliliğinin incelenmesi, sadece eğitimsel değil, aynı zamanda sosyoekonomik açıdan da karşılaştırmalı analizler yapılabilmesi açısından önem taşımaktadır.

Ölçme Değişmezliği Kavramı

Uluslararası geniş ölçekli değerlendirmelerin tasarım süreci, kültürel, ekonomik ve siyasi bağlamlar açısından oldukça farklılık gösteren geniş bir katılımcı yelpazesi arasında karşılaştırmaların yapılabilmesini sağlamak için titiz standartlara bağlı kalmayı gerektirir. Bu değerlendirmelerin sonuçları karşılaştırılabilir olmalıdır ve bu tür çalışmalar farklı eğitim yaklaşımlarının başarı üzerindeki etkisini inceleme ve kendi eğitim sistemine dair ek içgörüler sağlama fırsatı sunar. Bu gereklilikleri karşılayabilmek için, ölçme araçları belirli bir özelliğe aynı düzeyde sahip olan katılımcıların testte aynı puanı almasını garanti etmelidir (Eryılmaz vd., 2020). PISA Okuma Keyfi Ölçeği, öğrencilerin okuma becerilerini değerlendirmek ve bu beceriyi kullanmaya yönelik isteklerini ölçmek amacıyla kullanılan bir ölçektir. Bu ölçeğin farklı başarı grupları içindeki ölçme değişmezliği incelenerek, ölçeğin geçerliği ve güvenilirliği hakkında önemli bilgiler elde edilebilir. Ölçme değişmezliği kavramı, bir ölçeğin ölçtüğü özelliklerin, farklı gruplar arasında ölçüm sırasında aynı şekilde değerlendirilmesini ifade eder (Emerson vd., 2017). Bu durumun sağlanması, ölçeğin güvenilir ve geçerli olduğunun bir göstergesidir (Ireland vd., 2019). Farklı başarı grupları arasındaki ölçme değişmezliğini incelemek için, öncelikle öğrenciler PISA sınavındaki performanslarına göre farklı başarı gruplarına ayrılır. Daha sonra, her başarı grubundan örneklem seçilerek ölçek uygulanır. Bu incelemeyi gerçekleştirmek için, ölçeğin iç tutarlılık katsayıları, faktör analizi ve ölçeğin madde analizi gibi yöntemler kullanılabilir (Jovanović vd., 2020). Bu yöntemler, ölçeğin farklı gruplar arasında aynı ölçüm sonuçları verdiğini gösterir.

PISA Okuma Keyfi ölçeği'nin farklı başarı grupları içindeki ölçme değişmezliğini incelemek, ölçeğin geçerliliği ve güvenilirliği hakkında önemli bilgiler sunabilir. Bu bilgiler, ölçeğin geliştirilme ve iyileştirilme süreçlerinde dikkate alınabilir. Yapılan çalışmalarda incelenen psikolojik özellikler farklı gruplar arasında da karşılaştırmalara tabi tutulmaktadır. Yapı geçerliği, bir psikolojik özelliğin veya yapıyı ölçme aracının ne derecede doğru bir şekilde yansıttığıyla ilgilidir (Anastasi, 1988). Bireyler ve sonuç olarak gruplar arasındaki farklılıklar, ölçme sonuçlarını etkileyebilir. Bu nedenle elde edilen sonuçlar farklılık gösterebilir. Grupların ve bireylerin özelliklerinin yanı sıra, ölçüm sonuçlarını etkileyen başka faktörler de mevcuttur. Bu yüzden, madde ve test istatistiklerinin geçerlilik ve güvenilirlik yönünden değerlendirilmesi, her grubun özellikleri doğrultusunda değişebilecek sonuçlar doğurabilir; çünkü bu istatistikler grup düzeyinde hesaplanmaktadır (Crocker & Algina, 1986). Bu faktörlerin yanı sıra, gruplar arası ölçme işlemleri sırasında cinsiyet ve bölge gibi farklılıklar da dikkate alınmalıdır. Bu tür gruplar arası karşılaştırmalar yapılırken, ölçme aracının, belirli bir psikolojik özelliği tutarlı bir şekilde ölçebilmesi gerekmektedir (Anastasi, 1988). Farklı gruplar arasındaki karşılaştırmalar öncesinde, her grup içinde aynı psikolojik yapı veya yapıların ölçüldüğünden emin olabilmek için ölçme değişmezliği analizi yapılmalıdır. Ölçme değişmezliği (measurement invariance), farklı grupların aynı ölçme aracını benzer şekilde algılayıp algılamadığını test etmek amacıyla kullanılan istatistiksel bir yaklaşımdır (Byrne, 2016). Benzer biçimde ölçme değişmezliği, örtük değişkenler ile gözlenen değişkenler arasındaki ilişkinin farklı gruplar arasında tutarlılık göstermesi durumu anlamındadır. (Widaman & Reise, 1997). Aynı zamanda, farklı gruplar arasında karşılaştırmalar yapılabilmesi için, ölçme aracının kullanıldığı ön ölçümlerin değişmezliği de göz önünde bulundurulmalıdır (Horn & McArdle, 1992). Örneğin, Amerikalı ve Çinli işçilerden oluşan iki farklı grup arasındaki iş memnuniyeti düzeyini incelemek için kullanılan ölçek, grupların farklı yorumlama şekilleri nedeniyle farklı sonuçlar verebilir (Cheung & Rensvold, 1999).

Sosyal bilimler alanında yapılan çalışmalarda gruplar arası karşılaştırmalar yapılırken, ölçme aracının farklılık yaratmayacağı varsayımı yanıltıcı olabilir. Bu nedenle, ölçme aracının değişmezliği analizlerle ortaya konulmalıdır. Davidov ve arkadaşları (2021) ise özellikle sosyal bilimler bağlamında ölçme değişmezliğinin metodolojik temellerini incelemiş ve bu alanda yapılacak çalışmalar için kapsamlı bir rehber sunmuştur. Psikometri ve sosyal bilimlerde oldukça önemli bir kavram olan ölçme değişmezliği, özellikle kültürlerarası veya demografik farklılıkların bulunduğu örneklerde yapılan karşılaştırmalı çalışmalarda geçerliliğin korunmasını sağlar (Byrne, 2016). Ölçme değişmezliği, farklı gruplar arasında geçerli karşılaştırmalar yapabilmek için gerekli bir koşuldur (Cheung & Rensvold, 2000) ve bu kavram, karşılaştırma yapılan gruplar için ölçülen özelliklerin geçerliğini sağlamak anlamına gelir (Tyson, 2004). Kullanılan ölçeklerin kültürlerarası geçerliliğinin sağlanması, bu ölçeklerin genelleştirilebilir olmasını ve elde edilen sonuçların güvenilirliğini artırır (Marsh vd., 2006). Literatür incelendiğinde, ölçme değişmezliği araştırmalarının çoğunlukla psikolojik yapıları kültürler arası genellenebilirlik ve bu yapıların

birbirleriyle karşılaştırılabilirliğini ortaya koymaya yönelik olduğu dikkat çekmektedir. Bu tür çalışmalar, örneklemdaki bireylerden ziyade ölçme aracının genelleştirilebilirliğini ve ölçme aracından elde edilen sonuçların aynı yapıyı ifade edip etmediğini belirlemeyi amaçlar. Ölçme değişmezliği üzerine yapılan çalışmaların (Akyıldız, 2009; Ayvalli, 2016; Bahadır, 2012; Başusta & Gelbal, 2015; Kıbrıslıoğlu, 2015; Uyar & Doğan, 2014) alanyazına sunduğu katkılar oldukça değerlidir. Ancak, PISA gibi geniş ölçekli sınavlarla ilgili çalışmaların yanı sıra, TIMSS gibi diğer büyük ölçekli sınavlar üzerinde yapılan ve bireysel ölçme araçlarıyla gerçekleştirilen ölçme değişmezliği ile ilgili araştırmalarda, yapılan ölçme işlemlerinin farklı cinsiyet, dil, kültür ve ırk grupları arasında tutarlı sonuçlar vermediği belirtilmiştir. (Ercikan & Koh, 2005; Önen, 2009; Steinmetz vd., 2009; Wu vd., 2007). Ölçme değişmezliği, genellikle doğrulayıcı faktör analizi (DFA) yoluyla test edilir ve sırasıyla yapı değişmezliği, metrik değişmezlik, skaler değişmezlik ve artan seviyelerde test edilir. Putnick ve Bornstein (2016), ölçme değişmezliğinin farklı düzeylerini (yapısal, metrik, skalar) sistematik olarak ele alarak, çok kültürlü ve gelişimsel araştırmalarda bu kavramın önemini vurgulamışlardır. Eğer bir ölçme aracı bu düzeylerde değişmezlik gösteriyorsa, farklı gruplar arasında karşılaştırma yapmak bilimsel açıdan daha güvenilir hale gelir (Meredith, 1993). Ölçme değişmezliğini incelemek için dört aşama ve hipotez testleri gereklidir. Bu aşamalar şunlardır:

Şekilsel değişmezlik (configural invariance), ölçme değişmezliğinin ilk safhasıdır. Bu safhada, ölçme aracının faktör yapısının gruplar arasında değişmediğine dair bir hipotez testi yapılır. Şekilsel değişmezlik için destekleyici kanıt bulunması, ölçme aracının gruplar arasında benzer bir psikolojik yapıyı ölçtüğü anlamına gelir.

Metrik değişmezlik (metric invariance), farklı grupların maddelere verdikleri yanıtların benzer olduğu durumlarda, bu gruplardan elde edilen puanların birbirleriyle karşılaştırılabilir olduğunu ortaya koyar (Steenkamp & Baumgartner, 1998). Bu safhada, ölçme aracını oluşturan maddelerin regresyon katsayıları, yani faktör yüklerinin, gruplar arasında tutarlı olup olmadığına dair bir hipotez testi yapılır.

Ölçek değişmezliği (scalar invariance), bu safhada ölçme aracında bulunan maddelerin yer aldığı regresyon denklemindeki sabit katsayının gruplar arasında tutarlı olup olmadığını test etme amaçlanır. Ölçek değişmezliğinin sağlanabilmesi için hem metrik değişmezlik hem de regresyon denkleminde aynı orijinlerin kullanılması gerekmektedir. Bu aşama, örtük yapıdaki değer gözlenen değerle eşit olduğunu ifade eder. Gruplar arasında anlamlı karşılaştırmalar yapılabilmesi için, bir yapının en azından ölçek değişmezlik düzeyine ulaşmış olması gerekir (Eryılmaz & Sandoval-Hernandez, 2024). Skaler değişmezlik düzeyine ulaşamadığında, modele bazı düzenlemeler yaparak değişmezlik düzeyini artıran kısmi değişmezlik yöntemi kullanılabilir. Bu yöntem, modelden tamamen vazgeçmek yerine tercih edilebilecek bir yaklaşımdır (Eryılmaz & Sandoval-Hernandez, 2023).

Katı değişmezlik (strict invariance) safhasında, ölçme aracında bulunan maddelerin hata terimlerinin gruplar arasında tutarlı olup olmadığına dair bir hipotez testi yapılır.

Her bir safha, bir önceki safhanın modeline dayalı olarak yapılandırılmaktadır. Bu nedenle, her düzeydeki ölçme değişmezliği, o düzeydeki modelin, bir önceki düzeydeki modelle veriye uyumunu incelemeye dayanır.

Genellikle, tüm değişmezlik aşamalarını tam olarak sağlamak mümkün olmayabilir. Bu durumlarda, kullanılan ölçeğin amacına bağlı olarak kısmi değişmezlik de kabul edilebilir bir ölçme değişmezlik derecesi olarak ifade edilir. Ölçeğin gruplar arasındaki ortalamalarının karşılaştırılabilmesi için ölçek değişmezliğine kadar değişmezliğin sağlanması gerekmektedir (Dimitrov, 2010). Yapısal eşitlik modeli çalışmaları genellikle teori geliştirmeye odaklanır. Bu tür çalışmalar, üzerinde çalışılan değişkenler arasındaki ilişkileri belirlemek amacıyla gerçekleştirilir. Ölçme değişmezliğini belirlemek için yapılan yapısal eşitlik modeli uygulamaları, incelenen teorinin genellemelerine ışık tutar. Bu sayede incelenen yapıların belirlenmesinde ölçme değişmezliği çalışmalarının önemli bir rol oynadığı söylenebilir. Değişmezlik, ölçme araçlarının yapı geçerliğiyle yakından ilişkilidir. Özellikle geniş ölçekli sınavlara yönelik gerçekleştirilen değişmezlik çalışmaları, gruplar arasında geçerli sonuçlar sunma kapasitesine sahiptir. Bu nedenle çeşitli ölçme araçlarının farklı değişkenleri ele alarak değişmezlik çalışmalarına daha fazla odaklanılmasının alanyazına önemli katkılar sağlayabileceği düşünülmektedir.

Okuma keyfi ölçeği, kişilerin metinleri anlama yeteneklerini değerlendirmek amacıyla kullanılan bir psikolojik araçtır. Bu ölçek, farklı başarı gruplarında ölçme değişmezliği ve okuma keyfi maddelerinin eşitlik durumunun incelenmesiyle daha sağlıklı sonuçlar sunabilir. Farklı başarı gruplarındaki ölçme değişmezliği analizi, okuma keyfi ölçeğinin farklı gruplar için geçerli olduğunu doğrulamayı amaçlar (Özgen & Bindak, 2008). Aynı şekilde, okuma keyfi maddelerinin eşitlik durumu analizi, ölçeğin her bir maddesinin farklı gruplarda aynı şekilde anlamlı olduğunu belirlemek hedeflenir. Bu nedenle, PISA 2018 okuma keyfi ölçeğinin her bir maddesi, farklı gruplar arasında aynı anlamı taşıdığından emin olmak amacıyla incelenmelidir. Okuma keyfi ölçeğinin geçerliliği ve güvenilirliği, farklı başarı gruplarında ölçme değişmezliği ve okuma keyfi maddelerinin eşitlik durumunun dikkatlice incelenmesiyle sağlanabilir. Bu tür analizler ve çalışmalar, ölçeğin farklı gruplar arasında kullanılabilirliği ve sonuçlarının genelleme potansiyelini aydınlatılabilir. Okuma keyfi ölçeğinin değişmezliğini ve maddelerinin eşitlik durumunu değerlendirmek için yapısal eşitlik modeli kullanmak mümkündür (Kaynak, 2012; Şimşek, 2020). Bu tür modellemeler, ölçüm araçlarının yapısal bütünlüğünü ve genel geçerliliğini anlamamızı sağlar.

Bu araştırmanın amacı, PISA 2018 sınavına katılan öğrencilerin okuma keyfi ölçeğinden aldıkları puanların Norveç, Kolombiya ve Türkiye gibi farklı başarı düzeyine sahip ülkeler için ölçme değişmezliğini ve okuma keyfi maddelerinin eşitlik

durumunu incelemektir. Bu doğrultuda, farklı başarı düzeyine sahip ülkelerdeki öğrencilerin okuma keyfi ölçeği puanlarının ölçme değişmezliği test edilecektir. Bu çalışmanın, PISA 2018 okuma keyfi maddelerinin farklı ülkeler arasındaki başarı düzeyine göre yapılan karşılaştırmaların istatistiksel olarak anlamlı olup olmadığını daha sağlam bir şekilde belirlemeye yardımcı olacağı düşünülmektedir. Bu bağlamda, araştırmada ölçme değişmezliğinin test edilmesi, farklı başarı düzeyine sahip ülkelerin gruplarının karşılaştırılmasını daha geçerli kılacaktır. Bu tür bir çalışma, PISA 2018 okuma keyfi ölçeği maddelerinin farklı başarı düzeyine sahip ülkelerde aynı ölçüm özelliklerine sahip olduğunu göstererek daha güvenilir sonuçlara ulaşmayı amaçlar. Bu çalışma ayrıca, literatürde daha çok farklı dil, kültür ve cinsiyet gruplarında yapılan ölçme değişmezliği çalışmalarının yanı sıra farklı başarı gruplarına yönelik bir yaklaşım sunarak literatüre katkıda bulunmayı hedefler. Araştırmanın temel problemi, PISA 2018 öğrenci anketinde yer alan okuma keyfi ölçeği verileri kullanılarak, farklı başarı düzeyine sahip ülkelerdeki öğrenciler arasında madde ölçme değişmezliği kanıtlarının var olup olmadığını belirlemektir. Bu çalışma, öğrenci başarısı düzeyine göre gruplandırılan ülkeler arasında yapılan karşılaştırmaların güvenilirliğini artırarak literatüre önemli bir katkı sunmayı amaçlar.

Yöntem

Bu bölümde, araştırma modeli, evren ve örneklem, veri toplama süreci, kullanılan veri toplama araçları ve verilerin analizi ile ilgili detaylar sunulmaktadır.

Araştırma Modeli

Bu araştırmanın problemi PISA 2018 okuma keyfi ölçeğinin Norveç, Kolombiya ve Türkiye gibi farklı başarı düzeyine sahip ülkeler için ölçme değişmezliğinin sağlanıp sağlanmadığının belirlenmesi olduğu için araştırmada PISA 2018 uygulamasında okuduğunu anlama becerisini ölçen maddelerin başarı bakımından sırasıyla alt, orta ve üst grup ülkeler olan Kolombiya, Türkiye ve Norveç açısından eşitlik durumu incelenmiştir. Bu doğrultuda bu çalışma ilişkisel tarama modelinde bir çalışmadır.

Evren-Örneklem

Çalışmanın amacı doğrultusunda, ülkeler PISA okuma puanları sıralamasına göre alt, orta ve üst başarı gruplarına ayrıldığı için amaçlı örnekleme yöntemi uygun görülmüştür. Amaçlı örnekleme, belirli bir amaç doğrultusunda, bilgi açısından zengin ve derinlemesine inceleme yapmaya olanak tanıyan örneklem seçilmesini sağlamaktadır (Büyüköztürk vd., 2018).

Bu araştırmanın amacına göre PISA 2018 sınavına katılan 79 ülkeden alt, orta ve üst grup ülkeler arasından her bir gruptan bir ülke olmak üzere, toplamda üç ülke seçilerek 79 ülkedeki farklı başarı düzeyine sahip ülkelerin temsil edilmesi hedeflenmiştir. Bu amaçla, 79 ülkenin sınav başarı puanları dikkate alınarak yapılan yüzdelik sıralamalara göre üç ülke seçilmiştir. Alt, orta ve üst başarı gruplarını

oluşturmak için, her grup için sırasıyla %25, %50 ve %75'lik dilimlere karşılık gelen ülkeler belirlenmiştir. Bu durumda, çalışma örneklemini alt grupta Kolombiya, orta grupta Türkiye ve üst grupta Norveç'ten katılan öğrenciler oluşmaktadır. Araştırmada Kolombiya (7192), Türkiye (6858) ve Norveç (5659) öğrenci verisinden yararlanılmıştır. Araştırmanın amacı doğrultusunda örneklem seçiminde farklı başarı grubu ülkelere odaklanılırken bu ülkelerdeki dil farklılıkları yapılan araştırmanın amacına yönelik bir sınırlılığa neden olabilir.

Veri Toplama Süreci

Araştırma verileri <http://www.oecd.org/pisa/data/2018database/> adresinden alınmıştır. Bu veriler 2019 yılında paylaşılmıştır. Araştırmada veriler PISA 2018 yılı uygulamasına ait ST160 kodlu öğrenci anketindeki okuma keyfine yönelik maddelerinden elde edilmiştir.

Veri Toplama Araçları

Bu çalışmada, PISA 2018'e ait ST160 kodlu öğrenci anketinin 5 maddelik okuma keyfi alt ölçeği kullanılmıştır. Söz konusu ölçek, "kesinlikle katılmıyorum" (1) ile "kesinlikle katılıyorum" (4) arasında derecelendirilen 4 dereceli Likert tipi bir yapıya sahiptir. ST160 kodlu öğrenci anketinde yer alan maddelerin kodları ve madde formları, Tablo 1'de sunulmuştur.

Tablo 1.

PISA 2018 ST160 Öğrenci Anketindeki Okuma Keyfi Maddeleri

Madde Kodu	Maddeler
ST160Q01IA	I only read when it's necessary (Sadece gerekli olduğunda okurum)
ST160Q02IA	Reading is one of my top hobbies (Okumak en büyük hobilerimden biridir)
ST160Q03IA	I enjoy discussing books with others (Başkalarıyla kitap tartışmaktan hoşlanıyorum)
ST160Q04IA	I view reading as a time-waster (Okumayı zaman kaybı olarak görüyorum)
ST160Q05IA	I only read to gather necessary information. (Sadece gerekli bilgileri toplamak için okuyorum)

Verilerin Analizi

Veri analizi için bu çalışmada SPSS23 ve Mplus7 for Windows yazılımları kullanılmıştır. Öncelikle kayıp veriler incelenip kayıp veri oranının % 1'in altında olduğu ve tesadüfi dağıldığı saptanmıştır. Madde puanlarının sıralama ölçeği düzeyinde olması nedeniyle kayıp veri yerine medyan değeri atanmıştır. Ayrıca, açımlayıcı faktör analizine geçmeden önce verilerin normal dağılıma uyup uymadığı ve ölçek maddeleri arasında çoklu bağlantıların bulunup bulunmadığı değerlendirilmiştir. Elde edilen sonuçlara göre verilerin normal dağıldığı ve ölçek maddeleri arasında çoklu bağlantı sorununun olmadığı belirlenmiştir. Daha sonra analiz öncesi verilerin açımlayıcı faktör analizine uygunluğunu belirlemek için KMO ve Bartlett testi uygulanmış ve sonuçlar, veri setinin bu analiz için uygun olduğunu ortaya koymuştur. Son olarak açımlayıcı faktör analizi yapıp bu analiz sonuçlarına göre bir ölçme modeli kurulmuştur.

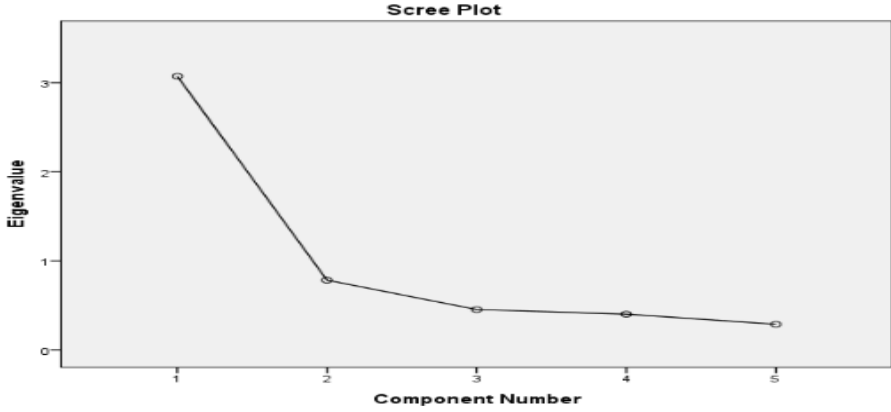
Modelin veriyle uyumunu belirlemek amacıyla çok gruplu doğrulayıcı faktör analizi (ÇG DFA) gerçekleştirilmiştir. Modelin uyumunu değerlendirmek için CFI, χ^2/df oranı ve RMSEA indeksleri incelenmiştir. Yapısal eşitlik modellemesi analizlerinde çoğu zaman ki-kare (χ^2/df) oranı değerlerinin büyük örneklemelerden kaynaklı olarak (5'ten büyük çıkarak) anlamlı çıktığı görülmüştür (Brown, 2006). Model-veri uyumunun iyi düzeyde olması için, örneklem büyüklüğünden kaynaklanabilecek istatistiksel yetersizliklere karşı daha güçlü olan ve büyük örneklemelerle uyumlu olan SRMR (Standartlaştırılmış ortalama hataların karekökü), RMSEA (Yaklaşık hataların ortalama karekökü) değerleri ve CFI değerlerinin tercih edilmesi gerektiği düşünülmüştür (Cheung & Rensvold, 2002). Bu çalışmada ülkelerden alınan örneklem hacimleri çok geniş olduğu için CFI, RMSEA ve SRMR değerleri dikkate alınmıştır.

Çoklu grup uygulamaları ile ölçülmeye çalışılan örtük özelliklerin gruplar arasında farklı olup olmadığı belirlenmeye çalışılmaktadır. Bu analizlerde, doğrulayıcı uyum indeksleri CFI arasındaki farklılıklar dikkate alınmaktadır. Uyumun belirlenmesinde örneklem hacminin büyük olmasının neden olduğu istatistiksel yetersizlikler olduğunda ki kareye alternatif olarak CFI uyum katsayısının daha tercih edilen ölçü olduğunu tavsiye eden araştırmalar vardır. Wu ve arkadaşları (2007), TIMSS verisiyle yaptıkları araştırmada, ki-kare farkı yerine CFI uyum indeksleri arasındaki farkların kullanılmasının gerektiğini şu şekilde açıklamışlardır: Ki-kare farkı, örneklem büyüklüğüne bağlı olarak değişir ve ölçme değişmezliğine karar verirken tek başına kullanılmasının uygun olmadığına dikkat çekmişlerdir.

Bu araştırmada kullanılan modelde değişmezlik testi olarak; şekilsel, metrik, ölçek ve katı değişmezlik uygulaması sonucunda hesaplanan CFI farkları (ΔCFI) karşılaştırılmıştır. ΔCFI değerlerini incelemadaki temel sebebi uyum katsayılarının gizil puanlarla gözlenen puanlar arasındaki ilişkiye yönelik bilgi vermesidir. Değişmelik şartının sağlanmasında ΔCFI değerlerinin $0.01 \geq \Delta CFI \geq -0.01$ aralığında olması dikkate alınarak değişmezlik incelemesi yapılmıştır. Ayrıca metrik ve katı değişmezlik için ΔCFI değerlerinin yanı sıra SRMR fark değerleri ($\Delta SRMR$) temel alınarak karşılaştırmalar yapılmıştır. Bu karşılaştırmalar için kabul edilebilir sınırlar, örneklem büyüklüğü $n > 300$ olduğunda, metrik değişmezlik için $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .015$, güçlü ve katı faktöriyel değişmezlik karşılaştırmaları için ise $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq 0.010$ olarak belirlenmiştir. Örneklem büyüklüğü $n < 300$ olduğunda ise, metrik değişmezlik için $\Delta CFI \leq -.005$ ve $\Delta SRMR \geq .025$, güçlü ve katı faktöriyel değişmezlik karşılaştırmaları için $\Delta CFI \geq -.005$ ve $\Delta SRMR \geq 0.005$ olarak kabul edilmiştir (Chen, 2007).

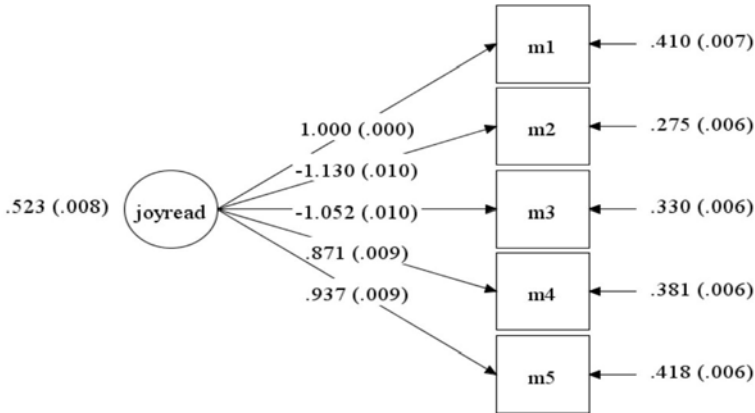
Bulgular

PISA okuma keyfi anketinde bulunan 5 maddeye yönelik açıklayıcı faktör analizi gerçekleştirilmiş ve madde gruplarının geliştirilen teori ve modele ne derece uyduğuna incelenmiştir. Faktörlerin özdeğerleri, aşağıda yer alan scree plot grafiğinde gösterilmiştir.



Şekil 1. Faktörlerin özdeğerleriyle oluşturulan scree plot grafiği

Şekil 1'de yer alan faktörlerin özdeğerlerine dayalı yamaç grafiğinde kırılmalar incelendiğinde bir faktörün diğerlerinden belirgin şekilde baskın olduğu gözlemlenmektedir. Açıklayıcı faktör analizi uygulanan 5 madde okuma keyfini tutarlı biçimde ölçtüğü için maddelerin tamamı birbiriyle ilişkili olduğundan 1 faktör çatısı altında toplanmıştır. Bir faktörün açıkladığı toplam varyans %61,413'tür. Açıklayıcı faktör analizi ile elde edilen yol (path) diagramı ile kurulan model şekil 2'de verilmiştir.



Şekil 2. Tüm ülkeler için tek faktörlü modelin yol diyagramı

Şekil 2’de teorisi kurulan modeli ölçme değişmezliği bakımından incelenmeden önce DFA ile Kolombiya, Türkiye, Norveç örneklemelerinden oluşan tüm veri, model-veri uyumu bakımından test edilmiştir. Elde edilen uyum istatistikleri makul düzeyde bulunmuştur. Örneklem hacmi çok geniş olduğu için CFI ve RMSEA değerleri dikkate alınmıştır. Verinin modele uyum kriteri olarak ($CFI > 0.90$) dikkate alındığında $CFI = 0,902$ yeterli ve $RMSEA = 0,114$ makul seviyesine yakın bulunduğu için kullanılan veri tek faktörlü model ile uyumlu çıkmıştır.

Test edilen değişmezlik düzeylerine dair bilgiler, Tablo 2’de yer almaktadır. Bu tabloda, şekil değişmezliği modeli; faktör yüklerinin, faktör korelasyonlarının ve hata varyanslarının serbest olduğu durumu, metrik değişmezlik modeli; faktör yüklerinin sabit, faktör korelasyonlarının ve hata varyanslarının serbest olduğu durumu, ölçek değişmezliği modeli; faktör yükleri ve faktör korelasyonlarının sabit, hata varyanslarının ise serbest olduğu durumu ve katı değişmezlik modeli; faktör yüklerinin, faktör korelasyonlarının ve hata varyanslarının sabit olduğu durumu ifade etmektedir. Şekil 2, faktör yükleri, faktör korelasyonları ve hata varyanslarının serbest bırakıldığı şekilsel değişmezlik modelini göstermektedir. Sınırlamalar eklenmeden oluşturulan modele sınırlamalar getirildiğinde elde edilen uyum istatistikleri ve bu istatistiklerin fark değerleri Tablo 2’de verilmektedir.

Tablo 2.

Ölçme Değişmezlik Düzeylerine Yönelik Uyum İstatistikleri

Aşamalar	χ^2	df	RMSEA	SRMR	Δ SRMR	CFI	Δ CFI
Şekil değişmezliği	8732.795	23	0.24	0.176	-	0.735	-
Metrik değişmezlik	3374.017	28	0.218	0.176	0	0.733	-0.002
Ölçek değişmezliği	6591.225	31	0.179	0.089	-0.087	0.8	0.065
Katı Değişmezlik	7360.486	43	0.161	0.157	-0.019	0.777	0.042

Tabloda verilen indeksler göre değişmezlik bakımından her bir düzey için yorum yapılacak olursa; şekilsel değişmezlik için uyum istatistikleri incelendiğinde ($CFI < 0.90$ ve $SRMR > 0.08$) olduğundan modelin şekilsel değişmezlik aşamasında veriye uyum sağlamadığı görülmektedir. Bu durumda, PISA 2018’de okuma keyfine yönelik uygulanan maddelerinin alt, orta ve üst grup ülkelerde benzer psikolojik yapıyı ölçmediği söylenebilir. Bu bulgu ölçülen yapının gruptan gruba benzer olmaması yani Kolombiya, Türkiye ve Norveçli öğrencilerin ölçek maddelerine cevap vermede benzer kavramsal bakış açısına sahip olmadıklarını göstermektedir. Şekilsel değişmezlik sağlanmadığı için diğer değişmezlik aşamaları olan metrik, ölçek ve katı değişmezliği incelemenin gereği yoktur. Çünkü herbir değişmezlik düzeyi bir öncekine bağlı ve hiyerarşiktir. Bu durumu doğrulamak için tabloda metrik değişmezlik için uyum istatistikleri incelendiğinde ($CFI < 0.90$ ve $SRMR > 0.08$) olduğundan modelin metrik değişmezlik aşamasında da veriye uyum sağlamadığı görülmektedir. Metrik değişmezliğe karar vermek için CFI ve SRMR fark testinden elde edilen (Δ CFI ve Δ SRMR) değerleri incelenip yorumlanmıştır. Elde edilen bulgu, okuma keyfi ölçeğinin maddelerinin ölçtüğü özellik ile okuma keyfi boyutu arasındaki ilişkinin, Kolombiya, Türkiye ve Norveçli öğrenciler için farklı olduğunu

göstermektedir. Diğer yandan ölçek değişmezliği için uyum istatistikleri incelendiğinde ($CFI < 0.90$ ve $SRMR > 0.08$) olduğundan modelin ölçek değişmezliği aşamasında da veriye uyum sağlamadığı görülmektedir. Ölçek değişmezliğini tespit etmek için çok gruplu DFA uygulamasından elde edilen CFI uyum indeksleri ile hesaplanan ΔCFI değeri incelendiğinde elde edilen değer değişmezlik koşulları sağlanmadığı için Kolombiya, Türkiye ve Norveçli öğrencilere yönelik maddeler bazında yanlılığın olduğu tespit edilmiştir. Bu bulguya yöre maddeler için kurulan regresyon denklemlerindeki sabit katsayıların Kolombiya, Türkiye ve Norveçli öğrenciler için değişmez olduğuna yönelik hipotez onaylanmamıştır. En son düzeyde ise katı değişmezlik için uyum istatistikleri incelendiğinde ($CFI < 0.90$ ve $SRMR > 0.08$) koşulunu sağlamadığından modelin katı değişmezlik aşamasında da veriye uyum sağlamadığı saptanmıştır. Buna ek olarak, ölçme aracındaki maddelere ilişkin hataların karşılaştırılan gruplar arasında değişmez olduğuna dair hipotez, yapılan çok gruplu DFA sonuçları ve ΔCFI değeri göz önünde bulundurularak doğrulanmamıştır. Böylece ölçme aracındaki maddelere yönelik hataların karşılaştırılan gruplar arasında farklılık gösterdiği sonucuna varılmıştır. Sonuç olarak şekil değişmezliği sağlamadığı için diğer değişmezliklerde sağlamamıştır. Böylece Kolombiya, Türkiye ve Norveç için okuma keyfi ölçeğinin ölçme değişmezliği sağlamamıştır. Tüm analizlerden elde edilen bulgular sonucunda AFA sonuçları dikkate alınarak oluşturulan modele ilişkin farklı başarı düzeyine sahip ülkeler için okuma keyfi ölçeğinin ortalamalarının karşılaştırılması anlamlı olmayacaktır.

Tartışma, Sonuç ve Öneriler

Bu araştırmanın temel amacı, PISA 2018 uygulamasında bulunan öğrenci anketinde bulunan okuma keyfi maddelerinin farklı okuduğunu anlama başarı gruplarında ölçme değişmezliğini incelemektir. Bu doğrultuda düşük başarı sınıfında yer alan Kolombiya, orta başarı sınıfında bulunan Türkiye, yüksek başarı sınıfında bulunan Norveç örneklemi analize alınmıştır. Ölçüm değişmezliğini sağlamak için modeller arası karşılaştırmalarda; $\Delta CFI < 0.004$, $\Delta RMSEA < 0.05$, $\Delta SRMR \leq 0.01$ ve anlamlı olmayan $\Delta \chi^2$ kriterlerinin karşılanması gerekir (Rutkowski & Svetina 2017). Araştırma kapsamında ilk olarak anket maddelerinin faktör yapısı incelenmiş ve maddelerin tek faktörlü bir yapı sergilediği belirlenmiştir. Daha sonra veri model uyumu incelenmiş ve verinin modele uyum kriteri olarak ($CFI > 0.90$) dikkate alındığında $CFI = 0.902$ ve $RMSEA = 0.114$ bulunduğu için CFI bakımından kullanılan veri tek faktörlü model ile uyumlu çıkmıştır. Burada RMSEA değeri bir miktar yüksek çıksa da makul uyum düzeyine yakındır.

Bu çalışmada, ülkelere yönelik ölçme değişmezliği incelenmiş ve okuma keyfi ile ilgili oluşturulan ölçme modeli kapsamında şekilsel, metrik, ölçek ve katı değişmezlik onaylanamamıştır. Araştırmada yapılan ölçme değişmezliği analizlerinin sonuçları, PISA öğrenci anketinde bulunan ve okuma keyfi üzerine hazırlanan maddelerden oluşan ölçme modelinin, Kolombiya, Türkiye ve Norveç gibi üç ülke için

genellenebilir bir psikometrik yapı sunmadığını ortaya koymaktadır. PISA gibi uygulamalarda farklı ana dilleri konuşan ve farklı kültürlerden gelen öğrencilerin cevaplarına ilişkin madde yanlılığı çalışmaları bulunmaktadır (Grisay vd., 2009; Gür, 2019; Lee, 2009; Uzun & Gelbal, 2017; Yıldırım & Berberoğlu, 2009). Bu sonuçlar, maddelerin farklı ana dillere çevrilmesinden kaynaklanan bir sorun olabileceğini göstermektedir. Başka bir deyişle, PISA gibi çok dilli uygulamalarda çeviri sorunlarının (kültürel farklılıklardan da kaynaklanan) meydana geldiği düşünülebilir. Kültür, bireylerin toplumsal çevreye uyum sağlama şekli olmanın yanı sıra, farklı toplumlar arasında çeşitlilik gösteren bir yaşam ve düşünme tarzıdır (Beals & Hoijer, 1972). Kültürler ve diller arasındaki maddelerin farklı işlevler göstermesinin olası sebepleri, uzman görüşlerinden çıkan ortak bir değerlendirmeye şu şekilde özetlenebilir: Uzmanlar, PISA maddelerindeki yanlılıkların temel sebeplerini, çeviri ve uyarlama hatalarından kaynaklanan anlam kaymaları ile sosyoekonomik düzey farklılıkları olarak belirtmişlerdir (Uzun & Gelbal, 2017).

Özellikle geniş katılımlı sınavlarda kullanılan anket ve testlerin farklı dillere çevrilmesi, çoğunlukla bir sınırlılık olarak değerlendirilmektedir. Bunun temel nedeni, bireylerin içinde bulundukları kültürün, maddeleri algılama şekilleri üzerinde etkili olabilmesidir. Bu alanda yapılan bazı araştırmalar, kültürel farklılıkların artmasıyla birlikte farklı diller konuşan bireyler arasında kullanılan test ve ölçeklerin ölçme değişmezliğini koruyamadığını ortaya koymuştur (Asil & Gelbal, 2012; Ercikan & Koh, 2005; Kıbrıslıoğlu, 2015; Marsh vd., 2013; Yandı vd., 2017). Örneğin, Kıbrıslıoğlu (2015), PISA 2012 kapsamında matematik öğrenme algısına yönelik modelin Türkiye, Çin-Şangay ve Endonezya örneklemelerinde ölçme değişmezliği gösterip göstermediğini incelemiş; araştırma sonucunda modelin ülkeler arasında yalnızca biçimsel düzeyde değişmezliği sağladığı belirlenmiştir. Öte yandan Yandı ve arkadaşları (2017) ise, kültürel ve dilsel farklılıkların sınırlı olduğu durumlarda dahi bazı maddelerde farklı işleyen madde (DMF) bulunduğu dikkat çekmiştir. Bu bulgular, çeviri süreçlerinde yaşanan zorluklar ve maddelerdeki kültürel unsurların yeterince uyarlanamamasının, sorun yarattığını göstermektedir.

Bu çalışmada hem seçilen ülkelerin dillerinin ve kültürlerinin farklı olması hemde alt, orta ve üst başarı grubu ülkeler sosyo ekonomik düzey bakımından farklılık göstermesi durumu anket maddelerinin ölçme değişmezliğinin sağlanamamasının nedenleri olarak düşünülebilir. Sosyoekonomik düzey (SED) farklılıkları, psikolojik ve eğitimsel ölçüm araçlarının farklı gruplarda aynı yapıyı ölçüp ölçmediğini değerlendiren ölçme değişmezliği üzerinde önemli etkiler yaratabilir. Bu durum, farklı SED gruplarındaki bireylerin aynı ölçek maddelerine farklı tepkiler verebileceği anlamına gelir ve bu da grup karşılaştırmalarının geçerliliğini sorgulatar (Putnick & Bornstein, 2016). Diğer bir neden ise; araştırmada grup sayısı ve örneklem hacminin büyük olması değişmezliğin olası ihlalini artırdığı düşünülmektedir. Benzer olarak Rutkowski ve Svetina (2014) değişmezlik çalışması yaparken örneklem hacminin büyük olması durumunda uygun olmadığını ya da özellikle grup sayısı çok olduğunda ayarlanması gerektiğini ifade etmiştir. Bu bilgiler ışığında bu araştırmada hem

örneklem hacminin büyük olması nedeniyle hem de farkları sıralamasında yer alan Kolombiya, Türkiye ve Norveç gibi ülkelerin seçilmesi nedeniyle değişmezliğin olası ihlalinin meydana geldiği düşünülmektedir.

Sonuç olarak ölçme değişmezliğinin araştırıldığı ölçme modelinin Kolombiya, Türkiye ve Norveçli öğrenciler için benzer biçimde ölçme yapamadığını ve böylece Ölçme aracından elde edilen verilere dayanarak, Kolombiya, Türkiye ve Norveçli öğrencilerin incelenen ölçme modeli çerçevesinde karşılaştırılmasının uygun olmadığı sonucuna varılabilir. Elde edilen bu sonuca göre Kasap (2022) tarafından yapılan PISA okuğunu anlama başarısını yordayan değişkenlerden biri olan okuma keyfi değişkeninin önem derecesinin farklı başarı düzeyine sahip ülkelerde farklılık gösterdiğine yönelik bulgu ve PISA 2018 raporunun 3. cildinde, öğrencilerin okuma keyfi düzeylerinin ülkeler arasında farklılık gösterdiğine yönelik bulgu olmasına rağmen ülkelerin okuma keyfi değişkeni bakımından karşılaştırılması anlamlı sonuç vermeyecektir. Çünkü bu çalışmada farklı başarı düzeyine sahip ülkelerin bu değişken bakımından ölçme değişmezliği sağlanamamıştır.

Bu çalışma, PISA 2018 okuma keyfi ölçeğinin farklı başarı düzeyine sahip ülkelerde ölçme değişmezliğini inceleyerek, başarı düzeyine dayalı karşılaştırmaların geçerliliğine dair literatürdeki sınırlı çalışmalara katkı sunmaktadır. Farklı başarı gruplarına odaklanması yönüyle, kültür, dil veya cinsiyet temelli önceki araştırmalardan ayrılmaktadır. Gelecekteki araştırmalar, daha fazla ülkeyi kapsayarak bulguların genellenebilirliğini artırabilir, farklı psikolojik yapıların ölçme değişmezliğini inceleyebilir, çeviri ve kültürel uyarlama süreçlerini daha derinlemesine analiz edebilir, zaman içindeki değişmezliği test ederek boylamsal karşılaştırmalar yapabilir ve eğitim politikalarını bu tür bulgular doğrultusunda şekillendirebilir. Bu araştırmanın sonuçlarından yola çıkarak, sadece başarı testlerinin değil, sonuçları önemli görülen PISA gibi uluslararası uygulamalarda yer alan anket maddelerinin de çeviri süreçlerindeki standartların gözden geçirilmesi, karşılaştırılan grupların sayısı ve her bir grupta seçilen örneklem büyüklüğünün uygun olarak ayarlanması önerilmektedir.

Yazarların Katkı Oranı

Araştırmanın yürütülmesi sürecinde yazarlar eşit oranda katkı sağlamışlardır.

Araştırma ve Yayın Etiği

Bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi"nde belirtilen kurallara uyulduğunu ve "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler"e dayalı hiçbir işlem yapmadığımızı beyan ederiz. Aynı zamanda tüm etik ihlallerde tüm sorumluluğun makale yazarlarına ait olduğunu beyan ederiz.

Çatışma Beyanı

Yazarlar çalışma kapsamında herhangi bir kurum ya da kişi ile çıkar çatışmasında bulunmadığını beyan etmektedir.

Etik Kurul İzni

Yazarlar için tüm etik kurallara uyulduğunu beyan ederiz. Bu çalışma yalnızca kamuya açık ikincil verilerle yürütüldüğünden etik kurul iznine tabi değildir.

Kaynakça

- Akyıldız, M. (2009). PIRLS 2001 testinin yapı geçerliliğinin ülkelerarası karşılaştırılması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 1, 18-47. <https://dergipark.org.tr/en/pub/yyuefd/issue/13711/165993>
- Asil, M. ve Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim*, 166, 236-249. <https://dergipark.org.tr/en/download/article-file/146293>
- Ayvallı, M. (2016) *PISA 2012 Matematik okur-yazarlığı testinin ölçme değişmezliğinin incelenmesi* (Tez No. 436748) [Yüksek Lisans Tezi, Akdeniz Üniversitesi-Antalya]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Anastasi, A. (1988). *Psychological Testing*. Macmillan Publishing Company.
- Bahadır, E. (2012). *Uluslararası Öğrenci Değerlendirme Programı'na (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi* (Tez No. 315014) [Yüksek Lisans Tezi, Hacettepe Üniversitesi-Ankara]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Başusta, N. B. & Gelbal, S. (2015). Gruplar arası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90. <https://www.researchgate.net/publication/330089813>
- Beals, R. L., & Hoişer, H. (1972). "The Macmillan Company", (çeviren: Gürbüz Erginer). The Nature and Scope of Anthropology.
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2018). *Bilimsel Araştırma Yöntemleri*. (25.Baskı). Pegem Yayıncılık.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Byrne, B. M. (2016). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). Routledge.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Toronto: Holt, Rinehart, and Winston, Inc.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>

- Cheung, G. W. & Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*, 25(1), 1–27. <https://doi.org/10.1177/014920639902500101>
- Cheung, G. W. & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of cross-cultural psychology*, 31(2), 187–212. <https://doi.org/10.1177/0022022100031002003>
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2021). Measurement invariance in cross-national research. *Annual Review of Sociology*, 47, 293–318. <https://doi.org/10.1146/annurev-soc-090120-012518>
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121–149. <https://doi.org/10.1177/0748175610373459>
- Emerson, S. D., Guhn, M., & Gadermann, A. M. (2017). Measurement invariance of the Satisfaction with Life Scale: Reviewing three decades of research. *Quality of Life Research*, 26, 2251–2264. <https://doi.org/10.1007/s11136-017-1552-2>
- Ercikan, K. & Koh, K. (2005). Examining the construct comparability of the English and French versionf of TIMSS. *International Journal of Testing*, 5, 23–35. https://doi.org/10.1207/s15327574ijt0501_3
- Eryilmaz, N., & Sandoval-Hernandez, A. (2024). Improving cross-cultural comparability: Does school leadership mean the same in different countries? *Educational Studies*, 50(5), 917–938. <https://doi.org/10.1080/03055698.2021.2013777>
- Eryilmaz, N., & Sandoval-Hernandez, A. (2023). Is distributed leadership universal? A cross-cultural, comparative approach across 40 countries: *An alignment optimisation approach. Education Sciences*, 13(2), 218. <https://doi.org/10.3390/educsci13020218>
- Eryilmaz, N., Rivera-Gutiérrez, M., & Sandoval-Hernández, A. (2020). Should different countries participating in PISA interpret socioeconomic background in the same way? A measurement invariance approach. *Revista Iberoamericana de Educación*, 84(1), 109–133. <https://rieoei.org/RIE/article/view/3981>
- Gür, E. (2019). *An investigation of the PISA 2015 in terms of differential item functioning based on culture*. (Thesis No. 582227) [Master's thesis, Hacettepe University-Ankara]. Council of Higher Education Thesis Center.
- Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. *IERI*, 2. <https://orbi.uliege.be/bitstream/2268/121044/1/IERI-Grisay-Gonzalez-Monseur.pdf>
- Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117–144. <https://doi.org/10.1080/03610739208253916>

- Ireland, M. J., Day, J. J., & Clough, B. A. (2019). Exploring scale validity and measurement invariance of the Toronto Mindfulness Scale across levels of meditation experience and proficiency. *Journal of clinical psychology*, 75(3), 445-461. <https://doi.org/10.1002/jclp.22709>
- Jovanović, V., Lazić, M., & Gavrilov-Jerković, V. (2020). Measuring life satisfaction among psychiatric patients: Measurement invariance and validity of the Satisfaction with Life Scale. *Clinical Psychology & Psychotherapy*, 27(3), 378-383. <https://doi.org/10.1002/cpp.2434>
- Kasap, Y. (2022). *Veri Madenciliği Yöntemleri ile Ülkelerin PISA Başarı Düzeylerini Etkileyen Değişkenlerin İncelenmesi* (Tez No. 721726) [Doktora Tezi, Hacettepe Üniversitesi-Ankara]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Kaynak, Z. N. (2012). *Yapısal eşitlik modelleri* (Tez No. 305950) [Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi-İstanbul]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürle ve cinsiyete göre ölçme değişmezliğinin incelenmesi* (Tez No. 394817) [Yüksek Lisans Tezi, Hacettepe Üniversitesi-Ankara]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy and math anxiety across 41 countries in PISA 2003 participating countries. *Learning and Individual Differences*, 19, 255-365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311-360. https://doi.org/10.1207/s15327574ijt0604_1
- Marsh, H. W., Abduljabbar, A. S., Ebu-Hilal, M. M. vd., (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: a comparison of arab and anglo-saxon countries. *Journal of Educational Psychology*, 105 (1), 108-128. <https://psycnet.apa.org/doi/10.1037/a0029907>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://doi.org/10.1007/BF02294825>
- Millî Eğitim Bakanlığı [MEB]. (2010). *Millî Eğitim İstatistikleri Örgün Eğitim 2010-2011*, <https://sgb.meb.gov.tr/www/milli-egitim-istatistikleri-orgun-egitim-2010-2011/icerik/69> adresinden 23 Aralık 2024 tarihinde alınmıştır.
- Organisation for Economic Co-Operation and Development [OECD] (2019). *Programme for International Student Assessment (PISA)*, <https://www.oecd.org/pisa/pisa-2019-results-in-focus.pdf>, 14.02.2019 adresinden 23 Aralık 2024 tarihinde alınmıştır.
- Organisation for Economic Co-Operation and Development [OECD] (2019). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. OECD Publishing, Paris. <https://doi.org/10.1787/acd78851-en>
- Organisation for Economic Co-Operation and Development [OECD]. (2009). *Country Statistical Profiles 2009*, <http://stats.oecd.org/WBOS/index.aspx> adresinden 23 Aralık 2024 tarihinde alınmıştır.

- Önen, E. (2009). *Ölçme değişmezliğinin yapısal eşitlik modellemesi teknikleri ile incelenmesi* (Tez No. 324837) [Doktora Tezi, Ankara Üniversitesi-Ankara]. Yükseköğretim Kurulu Başkanlığı Tez Merkezi.
- Özgen, K. & Bindak, R. (2008). Matematik okuryazarlığı öz-yeterlik ölçeğinin geliştirilmesi. *Kastamonu Eğitim Dergisi*, 16(2), 517-528. <https://dergipark.org.tr/en/pub/kefdergi/issue/49100/626538>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rutkowski, L. & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and psychological measurement*, 74(1), 31-57. <https://doi.org/10.1177/0013164413498>
- Rutkowski L, Svetina D. (2017) Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Appl. Meas. Educ.*;30(1): 39-51. <https://doi.org/10.1080/08957347.2016.1243540>
- Steenkamp E.M. & Baumgartner H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, Volume 25, Issue 1, June 1998, Pages 78–90, <https://doi.org/10.1086/209528>
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Qual Quant*, 43, 599-616. <https://doi.org/10.1007/s11135-007-9143-x>
- Şimşek, Ö. F. (2020). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları*. Ekinoks.
- Tyson, H.E. (2004). Ethnic differences using behavior rating scales to assess the mental health of children: a conceptual and psychometric critique. *Child Psychiatry and Human Development*, 34 (3), 167-201. 11. <https://doi.org/10.1023/B:CHUD.0000014996.26276.a5>
- Uyar, Ş. & Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2, 30-43. <https://dergipark.org.tr/en/pub/goputeb/issue/33496/380021>
- Uzun, N. B., & Gelbal, S. (2017). PISA fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446.
- Yandı, A.; Köse, İ. A.; Uysal, Ö. ve Oğul, G. (2017). *Küreselleşen dünyada eğitim. (PISA 2015 öğrenci anketinin (st094q01nast094q05na) ölçme değişmezliğinin farklı yöntemlerle incelenmesi)*. Pegem, Ankara
- Yıldırım, H. H. & Berberoglu, G. (2009). Judgemental and statistical analyses of the PISA 2003 mathematics literacy items. *International Journal of Testing*, 9(2), 108-
<https://doi.org/10.1080/15305050902880736>
- Wu, D. A., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: a demonstration with

TIMSS data. *Practical Assesment, Research & Evaluation*, 12(3), 1-26.
<https://doi.org/10.7275/mhqa-cd89>

Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.
<https://doi.org/10.1037/10222-009>

Extended Abstract

International large-scale assessments such as the Programme for International Student Assessment (PISA) have become a cornerstone for understanding student performance across nations. While PISA primarily assesses competencies in reading, mathematics, and science, it also collects extensive contextual data through student background questionnaires. Among these, the Reading Enjoyment Scale is particularly important, as it captures students' attitudes toward reading a factor strongly associated with reading performance and lifelong literacy development (OECD, 2019). However, valid international comparisons can only be made if the scale demonstrates measurement invariance across groups. Without invariance, observed differences may reflect translation problems, cultural variations, or socioeconomic disparities rather than genuine differences in reading enjoyment.

Measurement invariance refers to the degree to which the relationship between latent constructs and observed indicators remains stable across groups (Widaman & Reise, 1997). Establishing invariance ensures that the same psychological construct is measured consistently across populations. Previous studies have emphasized that cultural and linguistic differences, as well as demographic factors such as gender, can affect measurement equivalence (Byrne, 2016; Cheung & Rensvold, 2000). Yet, few studies have investigated invariance across achievement-level groups of countries, despite the policy relevance of such comparisons. Addressing this gap, the present study examines the invariance of the PISA Reading Enjoyment Scale across three countries representing different achievement levels in reading: Norway (high-achieving), Turkey (middle-achieving), and Colombia (low-achieving).

The study employed purposive sampling, selecting one country from each achievement tier out of the 79 PISA 2018 participants based on percentile rankings of reading scores (Rutkowski & Svetina, 2014). The sample consisted of 19,709 students: 5,659 from Norway, 6,858 from Turkey, and 7,192 from Colombia. Data were taken from the publicly available OECD PISA 2018 database, specifically the ST160 questionnaire, which includes five Likert-type items on reading enjoyment. Responses ranged from "strongly disagree" to "strongly agree." The items measured whether students engaged in reading for leisure or necessity, whether they viewed reading as enjoyable or a waste of time, and whether they enjoyed discussing books.

In the first stage, an exploratory factor analysis (EFA) was conducted to confirm the underlying structure of the scale. Results indicated a single-factor solution explaining over 60% of the variance, supporting the unidimensionality of the construct. In the second stage, multi-group confirmatory factor analysis (MGCFAs) was employed to test measurement invariance hierarchically across configural, metric, scalar, and strict levels (Meredith, 1993; Putnick & Bornstein, 2016). Analyses were carried out using SPSS 23 and Mplus 7. Fit indices such as CFI, RMSEA, and SRMR were prioritized due to the sensitivity of chi-square statistics to large sample sizes (Brown, 2006; Chen, 2007).

The results showed that configural invariance was not achieved ($CFI < 0.90$; $SRMR > 0.08$). This finding indicates that the basic factor structure was not equivalent across countries. Consequently, higher levels of invariance (metric, scalar, and strict) were also unsupported. The lack of configural invariance suggests that Colombian, Turkish, and Norwegian students may interpret or respond to the items differently. For example, the statement “*I only read when it is necessary*” might reflect pragmatic attitudes in one cultural setting but be perceived negatively in another. These discrepancies highlight the influence of translation, cultural interpretation, and socioeconomic context on item functioning (Beals & Hoiyer, 1972; Uzun & Gelbal, 2017).

The findings align with earlier research on international assessments, which documented item bias due to linguistic adaptation and cultural differences (Ercikan & Koh, 2005; Grisay, Gonzalez, & Monseur, 2009). Moreover, large sample sizes and multiple groups increase the likelihood of detecting non-invariance (Rutkowski & Svetina, 2017). Thus, while the scale may function adequately within individual countries, its cross-national comparability is compromised.

Implications for research and policy are substantial. Methodologically, the study underscores the necessity of conducting invariance testing before making cross-group comparisons in international surveys. Without such testing, observed differences in student motivation or reading enjoyment may reflect measurement artifacts rather than true disparities. From a policy perspective, the results caution against using PISA background data for straightforward cross-country comparisons. Policymakers should be aware that differences in reading enjoyment scores might be confounded by methodological issues, translation errors, or cultural factors.

This study makes two main contributions to the literature. First, it expands the focus of measurement invariance research beyond language, culture, and gender to achievement-level groups, an underexplored but highly relevant dimension. Second, it provides empirical evidence that the PISA Reading Enjoyment Scale does not demonstrate invariance across Norway, Turkey, and Colombia, thereby questioning the validity of direct comparisons among these countries.

Future studies should include more countries across achievement levels to increase generalizability and explore whether similar patterns of non-invariance persist. Longitudinal invariance analyses would also help determine whether reading enjoyment is stable across PISA cycles. Additionally, mixed-methods approaches combining quantitative invariance testing with qualitative interviews could provide deeper insights into how students interpret questionnaire items in different cultural and linguistic contexts.

In conclusion, the study reveals that the PISA 2018 Reading Enjoyment Scale does not exhibit measurement invariance across Norway, Turkey, and Colombia. As a result, direct comparisons of reading enjoyment scores across these achievement-level groups are not psychometrically justified. This finding calls for greater caution in interpreting PISA background data and emphasizes the importance of improving translation procedures, cultural adaptations, and sampling strategies in international large-scale assessments. By highlighting the limitations of current practices, the study contributes to the methodological refinement of cross-national research and offers valuable guidance for both scholars and policymakers in education.