





PISA 2018 Okuma Becerileri Testinin Bilişsel ve Sosyo-Ekonomik Gruplar Arasında Ölçme Değişmezliği¹

Measurement Invariance of the PISA 2018 Reading Literacy Test Across Cognitive and Socio-economic Groups

Sayfa | 777

Tuba YAZIKAN , Öğretmen, MEB, tubayazikan@gmail.com

Kübra ATALAY KABASAKAL , Doç. Dr., Hacettepe Üniversitesi, katalay@hacettepe.edu.tr

Geliş tarihi - Received: 8 Nisan 2025
Kabul tarihi - Accepted: 1 Nisan 2026
Yayın tarihi - Published: 28 Nisan 2026

¹ Makale, Tuba YAZIKAN'ın Doç. Dr. Kübra ATALAY KABASAKAL danışmanlığında hazırladığı yüksek lisans tezinden türetilmiştir.

Yazıkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi*, 17(1), 777-798.
DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

Sayfa | 778

Öz. PISA, uluslararası eğitim araştırmaları arasında önemli bir yere sahiptir ve elde edilen sonuçlar farklı alt gruplar arasında karşılaştırmalar yapmak için kullanılmaktadır. Ancak, bu karşılaştırmaların güvenilir olabilmesi için kullanılan ölçme aracının her alt grupta aynı yapıyı ölçtüğünün kanıtlanması gerekmektedir. Bu çalışmada, PISA 2018 uygulamasında okuma becerileri başarısını yordayan güvenilirliği değerlendirme, özetleme ve ekonomik-sosyal-kültürel durum değişkenlerine göre oluşturan alt gruplarda ölçme değişmezliği incelenmiştir. Araştırmanın örneklemini, PISA 2018'e Türkiye'den katılan 6890 öğrencinin öğrenci anketi ile okuma becerileri temel testinde yer alan 5 üniteye verdikleri yanıtlar üzerinden oluşturulmuştur. Verilerin analiz için gerekli varsayımları sağladığı belirlendikten sonra, R yazılımı kullanılarak çoklu grup doğrulayıcı faktör analizi (ÇGDFA) uygulanmış ve ölçme değişmezliği aşamalı olarak test edilmiştir. Analiz sürecinde her aşama için ölçeklenmiş uyum katsayıları değerlendirilmiş, ayrıca aşamalar arasındaki farklılıklar ölçeklenmiş CFI ve ölçeklenmiş TLI uyum indeksleri farkı üzerinden karşılaştırılmıştır. Çalışmada yapılan 15 analiz sonucunda, 4'ünde yapısal değişmezliğin sağlandığı, 10'unun metrik değişmezlik aşamasında kaldığı ve 1 ünite için modelin yakınsamaması nedeniyle uyum katsayılarının hesaplanamadığı belirlenmiştir.

Anahtar Kelimeler: Ölçme değişmezliği, Ekonomik-sosyal-kültürel durum, Güvenilirliği değerlendirme, Özetleme, Çoklu grup doğrulayıcı faktör analizi.

Abstract. PISA holds a significant place among international educational research studies, and its results are used to make comparisons across different subgroups. However, for these comparisons to be valid, it must be demonstrated that the measurement instrument assesses the same construct across all subgroups. This study examines measurement invariance in subgroups formed based on three key variables predicting reading proficiency in the PISA 2018 assessment: evaluating reliability, summarizing, and socioeconomic and cultural status. The study sample consists of responses from 6,890 students in Turkey who participated in PISA 2018, specifically their answers to the student questionnaire and five reading literacy test units. After verifying that the data met the necessary assumptions for analysis, multiple-group confirmatory factor analysis (MG-CFA) was conducted using R software, and measurement invariance was tested step by step. Throughout the analysis, scaled fit indices were evaluated for each stage, and differences between stages were assessed using the scaled CFI and scaled TLI fit index differences. The findings revealed that among the 15 analyses conducted, structural invariance was achieved in 4 cases, 10 remained at the metric invariance stage, and in one unit, model convergence was not achieved, preventing the calculation of fit indices.

Keywords: Measurement invariance, Economic-social,cultural situation, Reliability assessment, Summarizing, Multigroup confirmatory factor analysis.



Extended Abstract

Introduction. Measurement invariance refers to the extent to which a measurement instrument evaluates the same constructs consistently across different subgroups or time periods (van de Vijver, 2018). Without establishing invariance, observed groups differences may reflect measurement artifacts rather than true differences in the constructs. In large-scale assessments like PISA, which evaluate both cognitive and non-cognitive domains through tests and surveys, ensuring measurement invariance is essential to avoid misinterpretation due to differential item functioning across countries or cycles.

Beyond assessing performance, PISA aims to identify factors influencing academic achievement. Among these, metacognitive skills—key components of 21st-century competencies—are critical for understanding students' academic success. In PISA 2018, students' metacognitive reading skills were assessed in three areas: understanding and remembering (UNDREM), summarizing (METASUM), and reliability assessment (METASPAM). These skills reflect students' abilities to analyze, synthesize, and critically evaluate texts. Additionally, the economic, social, and cultural status (ESCS) variable is another important factor influencing educational outcomes by capturing students' environmental and familial contexts.

Studies on measurement invariance in Turkey have examined various subgroups, such as gender, school type, region, country, language, culture, test administration conditions, and teachers' educational background and experience. However, these studies have primarily focused on affective domain scales, with relatively limited attention to cognitive tests. While numerous studies explore factors influencing academic achievement, few address these factors within a measurement invariance framework.

A review of the literature reveals a lack of national or international studies on measurement invariance related to the "reliability assessment (METASPAM)" and "summarizing (METASUM)" variables. Although some studies have examined socioeconomic status (ESCS) in this context, they have mainly focused on affective scales rather than cognitive tests. This study addresses this gap by examining the measurement invariance of metacognitive skills and socioeconomic status using PISA 2018 cognitive item data, specifically focusing on METASPAM, METASUM, and ESCS variables to support the validity of academic achievement comparisons.

Method. Descriptive research aims to accurately and systematically define a population, situation, or phenomenon (Fraenkel, Wallen, and Hyun, 2012). This study employs multi-group confirmatory factor analysis (MGCFA) to examine measurement invariance across three predictor variables of reading proficiency. Prior to MGCFA, it is essential to conduct data screening, including missing data analysis, normality checks, outlier detection, and overall data cleaning to ensure accuracy and consistency (Tabachnick and Fidell, 2013). Accordingly, cognitive test items and survey items were separately analyzed for missing data. Subsequently, the METASPAM and METASUM variables, which were ordinal in nature, and the ESCS variable, which was continuous, were categorized. Given the narrow range of METASPAM and METASUM scores, they were divided into two categories: low and high. The ESCS variable was categorized into three levels: low, medium, and high, as commonly done in the literature. Additionally, complex multiple-choice items were dichotomized.



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

Following preliminary analyses, confirmatory factor analysis (CFA) was conducted using four items each from the "South Pole," "Machu Picchu," and "Nalini Nadkarni" units, as well as three items each from the "Sleep" and "Great Pacific Garbage Patch" units.

Results. Measurement invariance analyses were conducted by assessing model fit indices at different levels of invariance (configural, metric, and scalar) across the subgroups defined by METASPAM, METASUM, and ESCS variables.

For the reliability assessment (METASPAM) variable, metric invariance was achieved in the "South Pole" and "Nalini Nadkarni" units without deteriorating model fit, whereas scalar invariance constraints led to a decrease in model fit. In the "Sleep," "Machu Picchu," and "Great Pacific Garbage Patch" units, metric invariance constraints led to poorer model fit. Thus, metric invariance was achieved for "South Pole" and "Nalini Nadkarni," but scalar invariance was not established. In contrast, for the other three units, configural invariance was supported, but metric invariance was not achieved.

For the summarizing (METASUM) variable, model convergence issues prevented the computation of fit indices for the "Sleep" unit. For the remaining units, configural and metric invariance were supported based on RMSEA, SRMR, and CFI values. However, scalar invariance was not achieved, as indicated by poor model fit indices. The Δ CFI and Δ RMSEA values confirmed that metric invariance constraints did not significantly impact model fit. Therefore, metric invariance was supported for the "South Pole," "Machu Picchu," "Nalini Nadkarni," and "Great Pacific Garbage Patch" units, but scalar invariance was not established.

For the ESCS variable, configural and metric invariance were supported for all units, but scalar invariance was not achieved. Δ CFI and Δ RMSEA values indicated that metric invariance constraints did not degrade model fit, except in the "Great Pacific Garbage Patch" unit, where metric invariance constraints negatively impacted model fit. Consequently, for the "South Pole," "Sleep," "Machu Picchu," and "Nalini Nadkarni" units, metric invariance was supported, whereas scalar invariance was not achieved. In the "Great Pacific Garbage Patch" unit, configural invariance was supported, but metric invariance was not established.

Discussion and Conclusion. This study examined the measurement invariance of five PISA 2018 reading proficiency units across subgroups defined by the assessing credibility (METASPAM), summarizing (METASUM) and economic, social, and cultural status (ESCS) variables. Two groups were formed for METASPAM and METASUM, while three groups were formed for ESCS and MGCFA was employed.

The results indicate that for the METASPAM variable, configural invariance was achieved in three units, but metric invariance was not supported, suggesting potential item bias that precludes meaningful group comparisons. For the METASUM variable, one unit did not converge, while in the remaining units, metric invariance was supported, but scalar and strict invariance were not achieved. Regarding the ESCS variable, only one unit demonstrated configural invariance, while four units achieved metric invariance but did not reach scalar invariance.

Overall, due to the failure to establish scalar and strict invariance across all variables, reliable comparisons of reading proficiency across groups cannot be made. Previous research suggests that

Yazıkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi*, 17(1), 777-798.

DOI. 10.51460/baebd.1671573

Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper



socioeconomic factors may play a role in the lack of measurement invariance. Future studies should consider employing latent variable modeling or item response theory-based approaches to further investigate and address potential item biases.



Giriş

Uluslararası geniş ölçekli değerlendirmeler, farklı eğitim sistemlerine ve kültürel geçmişlere sahip ülkeler arasında karşılaştırmalar yapmayı amaçlamaktadır. Bu tür değerlendirmeler, ülkeler arasındaki eğitim farklılıklarını ortaya koymanın yanı sıra zaman içindeki değişimleri takip etme fırsatı sunar. Bu değerlendirmeler sadece ülkeler arasında değil, ülke içinde farklı alt gruplarda da karşılaştırmalar yapmak için kullanılmaktadır. Ancak, bu karşılaştırmaların geçerli ve güvenilir olabilmesi için ölçme değişmezliğinin sağlanması gerekmektedir. Ölçme değişmezliği, bir ölçme aracının farklı alt gruplarda veya zaman dilimlerinde aynı yapıların aynı şekilde ölçülmesini ifade eder (Van de Vijver, 2018). Eğer ölçme değişmezliği sağlanmazsa, farklı ülkelerde veya farklı alt gruplarda gözlemlenen değişikliklerin, ölçülen yapılar yerine ölçme aracından kaynaklanması riski ortaya çıkmaktadır.

Uluslararası Öğrenci Değerlendirme Programı (PISA) gibi geniş ölçekli eğitim araştırmaları, hem bilişsel hem de bilişsel olmayan yapıları değerlendirmek amacıyla test ve öğrenci anket verilerini kullanmaktadır. Bu yapılarda gözlenen farklılıkların, kullanılan maddelerin ülkeler ve döngüler arasında tutarsız işlemlerinden kaynaklanmaması için ölçme değişmezliği analizi kritik bir öneme sahiptir. Ölçme değişmezliği sağlandığında, ülkeler ve farklı alt gruplar arasında yapılan karşılaştırmalar daha güvenilir hale gelmektedir (Rutkowski ve Svetina, 2014; Vandenberg ve Lance, 2000). Örneğin, He ve diğerleri (2018), Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) ve Uluslararası Öğrenci Değerlendirme Programı'ndan (PISA) elde edilen bilişsel olmayan yapıların (enstrümantal motivasyon, bilimden zevk alma ve okula aidiyet duygusu) kültürler arası karşılaştırılabilirliğini incelemiş ve ölçme değişmezliğinin metrik değişmezlik aşamasının sağlandığını, ancak ölçek değişmezliği aşamasının sağlanmadığını tespit etmiştir. Bu nedenle, ölçek ortalamalarının doğrudan kültürler arası karşılaştırılmaması gerektiği belirtilmiştir. Ölçme değişmezliğinin sağlanmadığı ölçeklerde, gruplar arasında karşılaştırma amaçlı yapılan değerlendirmeler sağlıklı bilgi edinimini zorlaştırır. Cinsiyet, etnik köken, dil gibi değişkenler açısından farklılıklara sahip büyük gruplarda uygulanacak ölçme araçlarına ilişkin tüm ölçme özelliklerinin ilgili değişkenlere göre belirlenen alt gruplarda eşdeğerliğinin sağlanması gerekir (Öğretmen, 2006).

PISA, öğrenci performanslarını belirlemenin yanı sıra akademik başarıyı etkileyen faktörleri belirleme açısından da önemli bir araçtır. Akademik başarı, hem bireysel hem de toplumsal gelişim açısından büyük önem taşımaktadır. Başarıyı yordayan değişkenlerin belirlenmesi, eğitim politikalarının revize edilmesi ve eğitim sistemlerinin güncellenmesi gibi konularda önemli bir rol oynamaktadır. PISA sonuçlarına göre akademik başarıyı belirleyen değişkenler arasında üst bilişsel beceriler öne çıkmaktadır. Dunslosky ve Thiede (1998), üst bilişi, üst düzey zihinsel işlemlerden olan öğrenenin öğrenme sürecini planlama, problem çözme sırasında uygun strateji ve becerileri kullanma, kendi performansını değerlendirme ve öğrenme sürecini düzenlemeyi, öğrenme amacıyla kullanması şeklinde tanımlamıştır. 21. yüzyıl becerileri arasında değerlendirilen bu üst bilişsel beceriler, öğrencilerin akademik başarısını anlamada kritik bir role sahiptir. Üst biliş becerileri, bireyin kendi bilişsel süreçlerine ilişkin farkındalığını, bu süreçleri anlama ve yönetme kapasitesini ifade eder. Bu beceriler, öğrenme süreçlerini düzenleme,



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

kontrol etme ve görev performansını artırma açısından da oldukça önemlidir. PISA 2018 öğrenci anketleri, öğrencilerin okumaya yönelik üst bilişsel becerilerini üç temel başlık altında değerlendirmiştir: “anlama ve hatırlama (UNDREM)”, “özetleme (METASUM)” ve “güvenilirliği değerlendirme (METASPAM)”. Bu beceriler, öğrencilerin metinleri analiz etme, sentezleme ve eleştirel bir bakış açısıyla değerlendirme yeteneklerini ölçmeyi amaçlamaktadır.

Sayfa | 783

PISA 2018’in bilgisayar tabanlı gerçekleştirilmesi, güvenilirliği değerlendirme becerisini ön plana çıkaran yeni bir başlık olarak tanımlanmıştır. Bu beceri, metinlerin yazarının niyetini, yetkinliğini ve bilgisini analiz etmeyi; metindeki bilginin kaynağını belirlemeyi ve içeriğin geçerli, doğru ve objektif olup olmadığını değerlendirmeyi kapsar (OECD, 2019). Güvenilirliği değerlendirme, özellikle dijital ortamlarda bilgi kaynaklarının kalitesinin ve güvenilirliğinin analiz edilmesiyle doğrudan ilişkilidir. Özetleme becerisi ise, metinleri orijinalinden daha kısa ve öz bir şekilde açıklayarak iletme yeteneğini ifade eder (OECD, 2009). Bu beceri, öğrencilerin metinlerdeki ana fikirleri belirleme, önemli detayları ayırt etme ve bu bilgileri sentezleyerek etkili bir şekilde sunma kapasitesini ölçer. Özetleme, önceki PISA döngülerinde de ele alınan temel üstbilişsel becerilerden biridir.

Üstbilişsel becerilerin yanı sıra, PISA’daki akademik başarıyla doğrudan ilişkili bir diğer önemli değişken, “ekonomik-sosyal-kültürel durum (ESCS)” değişkenidir. Bu değişken, öğrencilerin eğitimsel başarılarını etkileyen sosyoekonomik faktörleri anlamak ve karşılaştırmak amacıyla kullanılmaktadır. ESCS, öğrencilerin eğitimsel çıktılarını etkileyen çevresel ve ailesel faktörlerin analizinde önemli bir rol oynar. Araştırmalar, öğrencilerin ailelerinin ekonomik, sosyal ve kültürel durumlarının, akademik dayanıklılık ve başarılarının güçlü bir belirleyicisi olduğunu ortaya koymaktadır (Cheung vd., 2013). Dezavantajlı sosyoekonomik geçmişe sahip öğrenciler, genellikle eğitim kaynaklarına sınırlı erişim, yetersiz beslenme ve artan aile stresi gibi bilişsel gelişimlerini ve öğrenme çıktılarını olumsuz etkileyebilecek bir dizi zorlukla karşı karşıya kalmaktadır (Francisca ve Mezoh, 2018). Bu faktörler, öğrencilerin eğitimsel performanslarını önemli ölçüde etkileyebilmekte ve eşitsizliklerin derinleşmesine neden olabilmektedir.

Türkiye içinde yapılan ölçme değişmezliği çalışmaları incelendiğinde cinsiyet (örn. Atılgan, 2022; Dinçer, 2023), okul türü ve bölge (örn. Ayvalli, 2016; Güngör, 2019), ülke (örn. Gönen, 2021; Kıbrıslıoğlu, 2015), dil ve kültür (örn. Bağdu Söyler vd., 2021; Gören vd., 2024; Uyumaz, 2021); testin uygulama ortamı (örn. Gündoğmuş 2017; Yalçınkaya, 2023); öğretmenlerin eğitim düzeyi ve deneyimleri (Ülkü, 2019) alt gruplarının ele alındığı görülmüştür. Bu çalışmalarda duyuşsal alan ölçeklerinin, bilişsel alan testlerine göre daha fazla olduğu gözlenmiştir. Ayrıca bilişsel alan testlerinde yapılan çalışmalarda da daha çok matematik okuryazarlığı testlerinin değişmezliğine odaklanıldığı görülmüştür. Literatürde akademik başarıyı belirleyen faktörler üzerine yapılan çalışmalar çeşitlilik göstermesine rağmen, bu faktörlerin ölçme değişmezliği bağlamında ele alınması nispeten sınırlıdır.

Ekonomik-sosyal-kültürel durum değişkeninin ise az sayıda çalışmada ele alındığı görülmüştür. Çakıcı Eser (2021), TIMSS 2015 verileriyle oluşturduğu matematik duyuşsal özellikler modelinin, evdeki kaynaklar değişkenine göre ölçme değişmezliğini incelemiş ve tüm alt boyutlarda değişmezliğin

Yazkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi, 17(1), 777-798.*
DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

sağlandığını bulmuştur. Stafford (2011), öğrencilerin bilişsel ve duyuşsal katılım ölçeğinin ESCS gruplarına göre ölçme değişmezliğini araştırmış ve düşük-yüksek gelir gruplarında yapısal ve metrik değişmezliğin sağlandığını belirlemiştir. Espinosa (2016), öğrenci katılım modelinin yüz yüze ve çevrim içi eğitim, etnik köken ve ESCS gruplarına göre ölçme değişmezliğini incelediğinde tüm gruplarda değişmezliğin sağlandığını ancak bunu beklediğini belirtmiştir. Bofah ve Hanula (2017), TIMMS 2011 verileriyle Gana'daki öğrencilerin ev eşyalarına sahip olma durumlarına göre üç SED (sosyo-ekonomik durum) profili belirlemiş ve bu profillerin cinsiyete göre ölçme değişmezliğini sağladığını göstermiştir. Lee (2019), PISA anketlerinden elde edilen evdeki kaynaklar ölçeğinin boylamsal ve ülkeler arası ölçme değişmezliğini inceleyerek kendi odasının olması, sessiz çalışma ortamı gibi maddelerin tüm PISA döngülerinde değişmezliği sağladığını tespit etmiştir.

PISA 2018 okuma becerileri temel alanı

PISA 2018'de okuma becerileri, bireyin yazılı metinleri anlayarak, kullanarak ve eleştirel bir şekilde değerlendirme yetisini geliştirmesi olarak tanımlanmıştır (MEB, 2019). Okuduğunu anlama, matematik ve fen okuryazarlığı gibi diğer alanlardaki başarıyı da doğrudan etkileyen bir faktördür (Coşguner, 2013). Bu kapsamda, PISA 2018'de okuma becerileri dört temel bilişsel süreç altında değerlendirilmiştir: akıcı okuma, bilgiye ulaşma, anlama ve değerlendirme/derinlemesine düşünme. Akıcı okuma, öğrencilerin metinleri etkili bir şekilde okuyabilmesini ifade ederken, bilgiye ulaşma süreci metindeki önemli bilgileri seçebilme yetisini içerir. Anlama süreci, metinler arasındaki ilişkileri kurabilmeyi kapsarken, değerlendirme ve derinlemesine düşünme süreci, metinlerin doğruluğunu ve güvenilirliğini analiz etmeyi gerektirir (OECD, 2019).

PISA 2018'de öğrencilerin okuma becerileri sekiz yeterlik düzeyinde değerlendirilmiştir. En düşük seviye olan 1c düzeyindeki öğrenciler kısa ve basit metinleri anlayabilirken, en yüksek olan 6. düzeyde öğrenciler soyut metinleri analiz edebilmekte, bilgileri karşılaştırarak değerlendirebilmekte ve metinler arası ilişkileri derinlemesine inceleyebilmektedir. Türkiye'de öğrencilerin büyük kısmının 2. ve 3. düzeylerde performans gösterdiği belirlenmiştir (MEB, 2019). Araştırma sonuçlarında, ekonomik-sosyal-kültürel durum değişkeniyle benzerlik gösteren değişkenlerin okuma başarılarını etkileyen faktörler olarak öne çıktığı görülmektedir (Arıcı ve Altıntaş, 2014; Kasap vd., 2021; Sarier, 2021). Ayrıca, üst bilişsel beceriler, özellikle de özetleme yetisi, okuma başarıyla ilişkilendirilen değişkenler arasında yer almaktadır (Eroğlu ve Aslan, 2020; Epçaçan, 2018).

Araştırmanın amacı

Başarı karşılaştırması yapılan PISA'da katılımcılar başarıyı yordayan değişkenlere göre gruplandığında, gruplar arasında ölçme sonuçlarının farklılık göstermemesi beklenir. Literatür taraması sonucunda özellikle "güvenilirliği değerlendirme (METASPAM)" ve "özetleme (METASUM)" değişkenlerinin başarıyı yordamada etkili olduğu görülmüş ve bu değişkenlerle ilgili literatürde bir ölçme değişmezliği çalışmasına rastlanmamıştır. Bunun yanı sıra, ölçme değişmezliği çalışmalarının bazılarında

Yazkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi, 17(1), 777-798.*

DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

Sayfa | 785

sosyoekonomik düzey değişkenine göre alt grupların oluşturulduğu görülmüş ancak bu çalışmaların genellikle duyuşsal ölçeklere odaklandığı tespit edilmiştir. Benzer şekilde bilişsel testlerin sosyoekonomik düzeye göre ölçme değişmezliğinin incelendiği çalışmaların oldukça sınırlı olduğu görülmektedir. Bu araştırma, üst bilişsel becerilerin ve sosyoekonomik durumun ölçme değişmezliğini inceleyerek, akademik başarı analizlerinde karşılaştırmaların ne ölçüde güvenilir olduğunu değerlendirmeye katkı sağlayacaktır. Bu çalışmada PISA 2018 bilişsel maddelerine ait veriler kullanılarak "güvenilirliği değerlendirme (METASPAM)", "özetleme (METASUM)" ve "ekonomik-sosyal-kültürel durum (ESCS)" değişkenleri açısından ölçme değişmezliğinin incelenmesi amaçlanmıştır.

Bu araştırma, PISA 2018 okuma okuryazarlığı testinin geçerliğini sadece genel bir çerçevede değil, başarı üzerinde kritik etkisi olduğu bilinen üstbilişsel stratejiler ve sosyoekonomik arka plan ekseninde derinlemesine inceleyerek literatüre iki temel açıdan özgün katkı sunmaktadır. Birincisi; literatürde METASPAM ve METASUM gibi üstbilişsel stratejilerin başarıyı yordama gücü sıklıkla çalışılsa da, bu stratejilerin farklı başarı seviyelerindeki öğrenciler için aynı psikometrik yapıyı temsil edip etmediği sorusu büyük oranda yanıtsız kalmıştır. Eğer bu stratejiler gruplar arasında farklı anlamlara geliyorsa, üstbilişsel becerilere dayalı başarı karşılaştırmaları metodolojik olarak hatalı sonuçlar verebilir. Bu çalışma, söz konusu değişkenlere dayalı ölçme değişmezliğini bilişsel maddeler düzeyinde kanıtlayarak, Türkiye örneklemindeki başarı farklarının gerçek yetenek farkından mı yoksa ölçme aracındaki yapısal bir farklılıktan mı kaynaklandığını ortaya koyması bakımından kritiktir. İkincisi; ESCS değişkeni PISA analizlerinde genellikle bir gruplandırma kriteri olarak kullanılmakta, ancak ESCS alt grupları arasında ölçme birimlerinin değişmezliği varsayımı çoğunlukla test edilmeden kabul edilmektedir. Mevcut çalışmaların büyük oranda duyuşsal ölçeklere odaklandığı göz önüne alındığında, bu araştırma doğrudan bilişsel maddelerin sosyoekonomik katmanlar arasındaki psikometrik eşdeğerliğini test ederek geniş ölçekli eğitim araştırmalarına metodolojik bir standart getirmeyi amaçlamaktadır. Bu yönüyle çalışma, politika yapıcıların ve eğitim araştırmacılarının PISA verileri üzerinden yapacağı karşılaştırmalı analizlerin adil ölçme ilkesine uygunluğunu değerlendiren öncü bir araştırma niteliğindedir.

Bu amaç doğrultusunda "PISA 2018 Türkiye örnekleminde okuma başarısını yordayan değişkenlerden güvenilirliği değerlendirme, özetleme ve ekonomik-sosyal-kültürel durum değişkenlerine göre oluşturulan alt gruplarda ölçme değişmezliği sağlanmakta mıdır?" sorusuna yanıt aranmıştır.

Yöntem

Araştırmanın türü

Bu çalışmada betimsel araştırma yöntemi benimsenmiştir. Betimsel araştırma bir olayı tam ve gerçek şekilde tanımlamak için kullanılan, çalışma grubunu betimleyecek istatistikleri içeren araştırma yöntemidir (Ocak, 2019).



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

Araştırmanın evreni ve örneklemi

Bu araştırmada PISA 2018 Türkiye verileri kullanıldığı için hedef evren, Türkiye'deki 15 yaş grubundaki öğrencilerdir. PISA'da 15 yaş grubu ile kastedilen 15 yaş 3 ay ile 16 yaş 2 ay aralığındaki, 7.sınıf ve üzeri örgün eğitime devam eden öğrencilerdir. PISA'da ülke örneklemleri, iki aşamalı tabakalı örnekleme yöntemiyle belirlenir.

Veri toplama araçları

PISA uygulaması, öğrencilerin bilişsel yeteneklerini değerlendiren testler ve geçmişleri, tutumları ile öğrenme deneyimlerini inceleyen anketlerden oluşur. Temel alandan daha fazla sorunun yer aldığı bilişsel testler, yaklaşık iki saatte tamamlanır. PISA 2018'de temel alan okuma olduğu için tüm öğrenciler okuma sorularını yanıtlamıştır. Anketler ise 35 dakika sürmekte ve öğrencilerin kendileri, aileleri, evleri, okulları ile öğrenme deneyimlerine dair bilgi toplamayı amaçlamaktadır. Araştırmada kullanılan tüm veriler OECD PISA uluslararası internet sitesinden (www.oecd.org/pisa/data/2018database/) alınmıştır. Araştırmanın verileri halka açık olarak bir web sitesinde bulunduğundan araştırma için etik kurul izni alınmasına gerek duyulmamıştır.

PISA 2018'de Türkiye dahil birçok ülke, bilgisayar tabanlı uygulamayı tercih etmiştir. Okuma testinde çok aşamalı uyarlanabilir model kullanılarak, sorular öğrencilerin yetenek düzeyine göre ayarlanmış ve beceriler daha hassas ölçülmüştür. Test, temel test, 1. aşama ve 2. aşama olmak üzere üç bölümden oluşmuş; temel test sabit kalırken, doğru cevap sayısına göre sonraki aşamalarda kolay veya zor sorular belirlenmiştir. Bu yöntem, her öğrencinin beceri düzeyine uygun sorularla karşılaşmasını sağlamıştır. Bu uyarlanabilir yaklaşım, öğrencilere beceri düzeylerine uygun sorular sunulduğu için öğrenci becerisinin daha hassas bir şekilde ölçülmesini sağlar (Pohl, 2013).

PISA 2018'de temel alan okuma olduğu için öğrencilerin tümü temel kitapçıktaki okuma becerileri maddeleriyle karşılaşmış ve en fazla veri bu kısımdan elde edilmiştir. Bu araştırmada okuma becerileri temel testinde en fazla cevaplanan beş ünite (Güney Kutbu, Uyku, Machu Picchu, Nalini Nadkarni, Büyük Pasifik Çöp Alanı) yer alan 19 maddeye verilen yanıtlar kullanılmıştır. Araştırma için seçilen 19 maddeden 17 tanesi basit çoktan seçmeli, 2 tanesi karmaşık çoktan seçmelidir. Araştırmada kullanılan değişkenlere (güvenilirliği değerlendirme, özetleme, ekonomik-sosyal-kültürel durum) ilişkin veriler öğrenci anketinden elde edilmiştir. Öğrenci anketinden elde edilen veriler direkt olarak kullanılıyorsa basit indeks, bir kaç maddeye verilen yanıt birleştirilerek yeni bir puan elde ediliyorsa bileşik indeks adını alır. Bu çalışmanın değişkenlerinden olan güvenilirliği değerlendirme (METASPAM) ve özetleme (METASUM) basit indekslerdir. METASPAM "Bir cep telefonu operatörünün yürüttüğü kampanya kapsamında bir akıllı telefon kazandığınızı bildiren ve kişisel bilgilerinizi girebilmeniz için bir link içeren bir e-postaya vereceğiniz tepki için aşağıda verilen yöntemlerin uygunluğunu 1 ile 6 arasında değerlendiriniz" sorusuna METASUM ise "Verilen metnin özetini yazarken aşağıda verilen her bir yöntemin yararlılığını 1 ve 6 arasında değerlendiriniz." sorusuna verilen cevaplardan elde edilmiştir. ESCS (ekonomik-sosyal-kültürel durum)

Yazkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi, 17(1), 777-798.*

DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

bileşik indeksi ise temelde anne ve babanın en yüksek eğitim düzeyi indeksi (PARED), anne ve babanın en yüksek mesleki durumu indeksi (HISEI), evdeki kaynaklar indeksi (HOMEPOS) olmak üzere üç indeksin birleşiminden oluşturulmaktadır.

Ölçme değişmezliği

Ölçme değişmezliği, bir testin farklı alt gruplarda (örn. cinsiyet, ırk) eşdeğer sonuçlar üretmesini ifade eder ve ölçme aracına bağlıdır (Kenny, 2006). Ölçme değişmezliğini test etmek için çoklu grup doğrulayıcı faktör analizi (ÇGDFA) kullanılır. ÇGDFA, doğrulayıcı faktör analizinin (DFA) özel bir türüdür ve yapısal eşitlik modellerinin (YEM) bir parçasıdır (Kline, 2016).

Kline (2016), YEM sürecini; model belirleme, model tanımlama, parametre kestirimi ve model veri uyumunun değerlendirilmesi aşamalarına ayırmaktadır. İlk aşama olan model belirlemede, araştırmacının hipotezine dayalı olarak değişkenler ve ilişkileri belirlenir, yol diyagramı oluşturulur. İkinci aşamada ise model tanımlama aşamasıdır. Model tanımlama aşamasında, belirlenen modeldeki tüm parametrelerin tanımlanabilmesi beklenir. Bir örtük değişken, en az üç göstergeyle ilişkilendirilmelidir; bu, model tanımlamada karşılaşılabilecek sorunları önlemek için önerilen bir kuraldır (Şen, 2020). Model tanımlama, veriye değil modele bağlı olduğundan örneklem büyüklüğünden bağımsızdır (Kline, 2016). Bu aşamada eksik, tam veya aşırı tanımlanmış modellerle karşılaşılabılır. Bu durumlar, serbestlik derecesinin (sd) 0'dan küçük, 0'a eşit veya 0'dan büyük olmasıyla ilişkilidir. Serbestlik derecesi, gösterge sayısı (p) ve gösterge sayısının 1 fazlasının çarpımının yarısı ile parametre sayısı arasındaki farkla hesaplanır ($sd = p.(p+1) / 2 -$ parametre sayısı). Eksik tanımlanmış modellerde sd 0'dan küçüktür ve modelin sonsuz çözümü olduğundan kestirim yapılmaz. Tam tanımlanmış modellerde sd 0'a eşittir ve modelin tek çözümü olduğundan kestirim yapılır ancak test edilmez. Aşırı tanımlanmış modellerde ise sd 0'dan büyüktür ve parametreler için tek çözüm kümesi bulunabildiğinden model kestirimi yapılarak test edilir. YEM'de genellikle aşırı tanımlanmış modeller test edilir (Şen, 2020), çünkü bu modellerde eşitlik sayısı parametre sayısını aşar ve bilinmeyenler için tek çözüm elde edilebilir. Eğer model tanımlanamıyorsa, ilk aşamaya geri dönülür. Üçüncü aşama ise parametre kestirimidir. Amos, LISREL, Mplus, R gibi yazılımlar kullanılarak model parametreleri tahmin edilir. Sürekli normal dağılımlı verilerde maksimum olabilirlik (ML), kategorik verilerde ağırlıklandırılmış en küçük kareler (WLS) yöntemi tercih edilebilir (Şen, 2020). Dördüncü aşama modelin veriyi ne kadar iyi yordadığının incelendiği aşamadır. Ki-kare (χ^2) testi temel değerlendirme ölçütüdür ancak büyük örneklerde yanıltıcı olabilir (Hoyle, 1995). Bu nedenle uyum indeksleriyle desteklenmesi önerilir. İyilik uyum indeksleri olarak geçen CFI, NFI, GFI, AGFI değerleri 0 ile 1 arasında değişir; 0.95 ve üzeri iyi uyumu, 0.90-0.95 arası kabul edilebilir uyumu gösterir. Mutlak uyum indeksleri olarak kullanılan RMSEA, SRMR'in 0.08'in altındaki RMSEA değerleri iyi uyumu gösterir. Bilgi kriteri olarak geçen AIC, BIC ise modellerin birbiri ile karşılaştırılmasını sağlar. Kline (2016), model uyum raporlarında özellikle χ^2 , RMSEA, CFI ve SRMR değerlerinin raporlanmasını önermektedir.



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

Çoklu grup doğrulayıcı faktör analizi (ÇGDFA)

Bazı araştırmalarda veriler farklı gruplardan (ör. farklı diller, etnik kökenler, uluslar) elde edilebilir. Bu tür çalışmalarda, belirli bir grupta geliştirilen ölçeğin faktör yapısının diğer gruplarda aynı olup olmadığını incelemek için çoklu grup doğrulayıcı faktör analizi (ÇGDFA) uygulanır (Tabachnick ve Fidell, 2013). ÇGDFA, doğrulayıcı faktör analizinin (DFA) özel bir türü olup, ölçme değişmezliğini değerlendirmek için kullanılır (Byrne ve Stewart, 2006). Ölçme değişmezliği, farklı grupların ölçek maddelerini aynı şekilde anladığını ve ölçek puanlarının anlamlı şekilde karşılaştırılabilir olduğunu gösterir. Ölçme değişmezliği sırasıyla dört aşamada test edilir (Kline, 2016).

Yapısal (Configural) Değişmezlik Aşaması: Faktör yapısının gruplar arasında aynı olup olmadığı test edildiği, en az kısıt içeren ilk aşamadır.

Metrik (Metric) Değişmezlik Aşaması: Ölçek maddelerinin faktör yüklerinin gruplar arasında eşit olup olmadığı test edildiği ikinci aşamadır. Bu aşamada faktör yüklerinde anlamlı farklılık olması, madde yanlılığına işaret edebilir (Byrne ve Watkins, 2003).

Ölçek (Scalar) Değişmezliği Aşaması: Ölçek maddelerine ilişkin regresyon sabitlerinin veya kategorik değişkenler için eşik değerlerin eşit olup olmadığı incelendiği üçüncü aşamadır. Bu aşamanın sağlanması, faktör ortalamalarının karşılaştırılmasını mümkün kılar (Bowen ve Masa, 2015).

Katı (Strict) Değişmezlik Aşaması: Ölçekteki hata terimlerinin gruplar arasında eşit olup olmadığı test edildiği son aşamadır. Literatürde genellikle sağlanamayan bir aşamadır (Schmitt ve Kuljanin, 2008).

Her bir aşamada, önceki aşamaya göre daha fazla kısıtlama getirilir ve aşamaların sağlanma durumu model veri uyum indeksleri ile değerlendirilir. Ölçek puanlarının güvenilir şekilde karşılaştırılabilmesi için genellikle ölçek değişmezliğinin sağlanması yeterli kabul edilir, ancak tüm aşamaların test edilmesi önerilmektedir (Meredith, 1993).

Normal dağılıma sahip sürekli değişkenler için parametre kestiriminde genellikle maksimum olabilirlik (ML) yöntemi kullanılmaktadır. Bu durumda ölçme değişmezliği analizinde incelenen temel parametreler faktör yükleri, kesişimler ve artık varyanslardır. Göstergeler sıralı olduğunda ise ML yerine en küçük kareler (WLS) yöntemi tercih edilmelidir. Bu durumda incelenen parametreler faktör yükleri, eşikler ve artık varyanslardır. Sürekli ve sıralı değişkenler arasındaki temel fark, ölçek değişmezliği aşamasında kısıtlanacak parametrenin farklı olmasıdır (Vandenberg ve Lance, 2000). Literatürde, faktör yükleri ve eşik değerlerin birlikte kısıtlanarak metrik değişmezlik aşamasının atlandığı üç adımlı model ile bu parametrelerin ayrı ayrı kısıtlandığı dört adımlı model olmak üzere iki farklı yaklaşım bulunmaktadır (Bowen ve Masa, 2015). Her iki durumda da ölçme değişmezliğinin sağlanabilmesi için faktör yüklerinin ve eşik değerlerin eşitliğinin istatistiksel olarak kanıtlanması gerekmektedir.



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

Çoklu gruplu doğrulayıcı faktör analizinde (ÇGDFA) ölçme değişmezliği aşamaları incelenirken her aşamada elde edilen ki-kare istatistiği, RMSEA, SRMR, CFI ve TLI gibi uyum indeksleri değerlendirilmektedir. Büyük örneklemelerde ki-kare indeksinin anlamlı çıkma eğiliminde olduğu göz önünde bulundurulmalıdır. Değişmezlik aşamaları iç içe kümelenmiş modeller şeklinde ilerlediğinden, yalnızca her bir aşamanın bağımsız değerlendirilmesi yeterli olmayıp, yapılan kısıtlamaların model veri uyumuna etkisinin de belirlenmesi gerekmektedir. Sürekli veya sürekli gibi davranan sıralı veriler için ML yöntemi kullanıldığında, ardışık aşamalar arasındaki karşılaştırmalar ki-kare fark testi ve uyum indekslerindeki değişimler (örneğin, CFI ve RMSEA farkları) üzerinden yapılmaktadır. Ancak ikili kategorik verilerde WLS yöntemi kullanıldığından, model karşılaştırmalarında doğrudan ki-kare farkı yerine Satorra-Bentler ölçeklenmiş ki-kare farkının hesaplanması gerekmektedir (Muthen ve Muthen, 1998, 2019; akt. Şen, 2020). Literatürde önerilen kriterlere göre, daha kısıtlı modelin uyum indekslerinde anlamlı bir kötüleşme göstermemesi durumunda, ölçme değişmezliğinin sağlandığı kabul edilmektedir. Cheung ve Rensvold'e (2002) göre karşılaştırma analizlerinde ΔCFI değerinin -0,010 ile 0,010 arasında olması, $\Delta RMSEA$ değerinin ise 0,015'ten küçük olması daha kısıtlı değişmezlik modelinin, uyumu kötüleştirmediği anlamına gelmektedir.

Verilerin çözümlenmesi

Bu araştırmada okuma becerileri başarısını yordayan 3 değişkene göre ÇGDFA yöntemiyle ölçme değişmezliğinin incelenmesi amaçlanmıştır. ÇGDFA analizi öncesi veri setinin taranarak kayıp veri, normallik, uç değer ve çoklu bağlantılılık açısından değerlendirilmesi, gerekli, doğru ve düzgün veri tarama ve temizliğinin yapılması önemlidir (Tabachnick ve Fidell, 2013). Bu amaçla ilk olarak veri setindeki bilişsel testten elde edilen maddeler ve anketten elde edilen maddeler ayrı ayrı kayıp veri açısından incelenmiş ve eksik veriler olduğu saptanmıştır.

Maddelerdeki kayıp veri oranları bir madde hariç %0,1 ile %1,6 arasında değişmektedir. Tabachnick ve Fidell (2013) büyük veri setleri için, %5'in altındaki kayıp verilerin, kayıp veriyle baş etme yöntemlerinden herhangi biriyle giderilmesinin sonuçların benzer şekilde etkileyeceğini belirtmiştir. Bunun nedeni, büyük bir örneklem büyüklüğünde, verilerin küçük bir yüzdesinin çıkarılmasının etkisinin muhtemelen minimum olmasıdır. Bu nedenle %5'in altındaki kayıp verilerin veri setinden silinmesine karar verilmiştir. Eksik verinin fazla olduğu Büyük Pasifik Çöp Alanı ünitesindeki CR560Q10S kodlu basit çoktan seçmeli madde ise analize dahil edilmemiştir.

Araştırmada kullanılan METASPAM, METASUM ve ESCS değişkenlerinde %5'in altında eksik verisi olan bireyler de analizden çıkarılmıştır. Son durumda araştırmada kullanılacak okuma okuryazarlığı testindeki soruları eksiksiz yanıtlayan ve öğrenci anketindeki araştırma değişkenleriyle ilgili maddeleri eksiksiz tamamlayan, analize dahil edilecek birey sayısı Güney Kutbu ünitesinde 2312 birey, Uyku ünitesinde 864 birey, Machu Picchu ünitesinde 3148 birey, Nalini Nadkarni ünitesinde 2345 birey, Büyük Pasifik Çöp Alanı ünitesinde 2345 bireydir.



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

Analiz öncesinde yapılan incelemelerde; tüm ünitelerde çarpıklık ve basıklık değerlerinin ± 1 aralığında kalarak normal dağılım sergilediği görülmüştür (Tabachnick ve Fidell, 2013). Çoklu bağlantılılık incelemesinde; Tolerans (0,86-0,99), VIF (1,01-1,15) ve CI (< 30) değerlerinin kritik eşiklerin oldukça içinde olduğu saptanmıştır.

Sayfa | 790

İkinci olarak veri setinde sıralı değişken olarak yer alan METASPAM ve METASUM ile sürekli değişken olarak yer alan ESCS değişkenleri kategorik hale getirilmiştir. METASPAM ve METASUM değişkenleri veri setindeki ranjin dar olması sebebiyle düşük ve yüksek olmak üzere ikişer kategoriye; ESCS değişkeni ise literatürde olduğu gibi düşük, orta ve yüksek olmak üzere 3 kategoriye ayrılmıştır.

Üçüncü olarak karmaşık çoktan seçmeli maddeler iki kategorili hale getirilmiştir. Orijinal veri setinde 3 kategoride olan bu maddelere yanlış cevap veren bireylere 0, kısmi ve tam doğru cevap veren bireylere 1 değeri atanarak kategori sayısı ikiye düşürülmüştür. Puan kategorilerinin birleştirilmesi, özellikle bazı kategorilerdeki gözlem sayısının az olduğu durumlarda modelin yakınsama sorunlarını önlemekte ve parametre tahminlerinin kararlılığını artırmaktadır (DiStefano ve Morgan, 2014). Bu çalışmada da analiz yöntemi olarak kategorik veriler için önerilen Robust WLS (WLSMV) tercih edildiğinden, maddelerin iki kategorili yapıda eşitlenmesi hem teorik hem de istatistiksel açıdan daha tutarlı modellerin kurulmasına olanak tanımıştır.

Ön incelemeler tamamlandıktan sonra, Güney Kutbu, Machu Pichu ve Nalini Nadkarni ünitelerinden 4'er soru, Uyku ve Büyük Pasifik Çöp Alanı ünitelerinden 3'er soru ile doğrulayıcı faktör analizinde model uyumu test edilmiştir. Uyku ve Büyük Pasifik Çöp Alanı ünitesi için kurulan model, ünitadaki soru sayısının 3 olması sebebiyle benzer şekildedir. Uyku ve Büyük Pasifik Çöp Alanı ünitelerinde gösterge sayısı 3, parametre sayısı 6 olduğundan serbestlik derecesi $(3.4 / 2) - 6 = 0$ olarak hesaplanır. Bu ünitelerde serbestlik derecesi 0'a eşit olduğundan kurulan modellerin tam tanımlanmış olduğu söylenebilir. Bu modellerde model veriye mükemmel uyum gösterir ve test edilmesine gerek yoktur. Güney Kutbu, Machu Picchu ve Nalini Nadkarni ünitelerinde ise gösterge sayısı 4, parametre sayısı 8 olduğundan serbestlik derecesi $(4.5 / 2) - 8 = 2$ olarak hesaplanır. Bu ünitelerde serbestlik derecesi 0'dan büyük olduğundan kurulan modellerin aşırı tanımlanmış olduğu söylenebilir. Bu modellerde model veri uyumu test edilir. Aşırı tanımlanmış modellerin doğruluğunu test etmek için yapılan DFA sonucu elde edilen Storra-Bentler ölçeklenmesine uygun ölçeklenmiş uyum katsayıları Tablo 1'de verilmiştir.

Tablo 1.

DFA ölçeklenmiş uyum katsayıları

	ölçeklenmiş χ^2	p	ölçeklenmiş RMSEA	SRMR	ölçeklenmiş CFI
Güney Kutbu	9,760	0,008*	0,0041	0,029	0,989
Machu-Picchu	0,357	0,837	0,000	0,006	1,000
Nalini-Nadkarni	1,775	0,412	0,000	0,016	1,000

* p değeri, 0,05 düzeyinde anlamlı.

Yazıkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi, 17(1), 777-798.*

DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

Tablo 1'e göre; Güney Kutbu ünitesinde ölçeklenmiş ki-kare istatistiğine bağlı anlamlılık (p) düzeyi manidardır. Bu durum istenen bir durum değildir ancak örneklemin büyük olmasına bağlı olarak böyle bir sonuç elde edilmiştir. Diğer uyum indeksleri birlikte değerlendirildiğinde Güney Kutbu, Machu Picchu ve Nalini Nadkarni ünitesinde model veri uyumunun sağlandığı, ünite içindeki maddelerin üniteye tek faktörlü yapıyı desteklediği sonucuna ulaşılabilir. Tek faktörlü yapı desteklendiğinden ÇGDFA ile ölçme değişmezliğinin her bir aşamasının kontrol edilmesi kısmına geçilmiştir. Bu analizin sonuçları ayrıntılı bir şekilde bulgular kısmında verilmiştir.

Sayfa | 791

Bulgular

Tablo 2'de güvenilirliği değerlendirme becerisine (METASPAM) göre oluşturulan alt gruplara ait beş üniteye ilişkin hiyerarşik ölçme değişmezliği aşamalarının uyum katsayıları ve ardışık değişmezlik aşamalarının karşılaştırıldığı analiz sonuçları verilmiştir. Uyum katsayıları SRMR hariç ölçeklenmiş uyum katsayılarıdır.

Tablo 2.

Güvenirliği değerlendirme becerisine göre oluşturulan alt gruplarda uyum katsayıları ve karşılaştırması

Üniteler	Aşama	χ^2	df	P	RMSEA	SRMR	CFI	ΔCFI	$\Delta RMSEA$
Güney Kutbu	Yapısal	11,951	4	0,018*	0,041	0,033	0,987		
	Metrik	15,062	8	0,058	0,028	0,043	0,989	-0,011	-0,021
	Ölçek	202,789	12	0,000*	0,117	0,043	0,691	-0,297	0,090
Uyku	Yapısal	0,000	0	-	0,000	0,000	1,000		
	Metrik	5,600	3	0,133	0,045	0,059	0,936	-0,050	0,065
	Ölçek	48,869	6	0,000*	0,129	0,059	0,000	-0,936	0,084
Machu Picchu	Yapısal	1,879	4	0,758	0,000	0,012	1,000		
	Metrik	10,618	8	0,224	0,014	0,031	0,996	-0,015	0,039
	Ölçek	324,425	12	0,000*	0,129	0,031	0,466	-0,529	0,114
Nalini Nadkarni	Yapısal	4,360	4	0,359	0,009	0,019	0,998		
	Metrik	16,517	8	0,036*	0,030	0,045	0,956	-0,026	0,011
	Ölçek	163,076	12	0,000*	0,104	0,045	0,222	-0,734	0,074
Büyük Pasifik Çöp Alanı	Yapısal	0,000	0	-	0,000	0,000	1,000		
	Metrik	4,104	3	0,251	0,018	0,030	0,995	-0,016	0,050
	Ölçek	140,911	6	0,000*	0,139	0,030	0,414	-0,581	0,121

* p değeri, 0,05 düzeyinde anlamlı.

Tablo 2'ye göre tüm ünitelerde yapısal ve metrik değişmezlik aşamalarında elde edilen RMSEA, SRMR ve CFI değerleri iyi uyuma işaret etmektedir. Ünitelerde ölçek değişmezliği aşamasındaki uyum katsayıları ise kabul edilebilir değerlerde değildir. ΔCFI ve $\Delta RMSEA$ değerlerine göre değişmezlik aşamaları karşılaştırıldığında Güney Kutbu ve Nalini Nadkarni ünitelerinde daha kısıtlı olan metrik değişmezliğin yapısal değişmezliğe göre uyumu kötüleştirmediği, ancak ölçek değişmezliğinde kısıtlanan eşik değerin metrik değişmezliğe göre uyumu kötüleştirdiği görülmektedir. Uyku, Machu Picchu ve Büyük Pasifik Çöp Alanı (Yazkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi*, 17(1), 777-798. DOI. 10.51460/baedb.1671573



Alanı ünitelerinde ise yapısal değişmezliğe göre daha kısıtlı olan metrik değişmezliğin uyumu kötüleştirdiği görülmektedir. Dolayısıyla Güney Kutbu ve Nalini Nadkarni ünitelerinde METASPAM alt gruplarına göre metrik değişmezlik sağlanmış, ölçek değişmezliği sağlanamamıştır. Uyku, Machu Picchu ve Büyük Pasifik Çöp Alanı ünitelerinde ise yapısal değişmezlik sağlanmış, metrik değişmezlik sağlanamamıştır. Bu durumda düşük ve yüksek güvenilirliği değerlendirme becerisine sahip bireylerin oluşturduğu alt gruplarda metrik değişmezliğin sağlandığı Güney Kutbu ve Nalini Nadkarni ünitelerine ilişkin soruların aynı psikolojik yapıyı temsil ettiği, aynı alt gruplardaki öğrencilerin aynı soruları aynı şekilde yorumladığı söylenebilir. Ancak ortalamalar arasındaki farklılığın nedeninin tam olarak bilinemez. Sadece yapısal değişmezlik aşamasının sağlandığı Uyku, Machu Picchu ve Büyük Pasifik Çöp Alanı ünitelerine ilişkin soruların ise aynı psikolojik yapıyı temsil ettiği, fakat aynı alt gruplardaki öğrencilerin aynı soruları farklı şekilde yorumladığı söylenebilir. Bu durum madde yanlılığının göstergesi olabilir.

Tablo 3'te özetleme (METASUM) becerisine göre oluşturulan alt gruplarda beş üniteye ait ölçme değişmezliği aşamalarının uyum katsayıları ve ardışık değişmezlik aşamalarının karşılaştırıldığı analiz sonuçları verilmiştir. Uyum katsayıları SRMR hariç ölçeklenmiş uyum katsayılarıdır.

Tablo 3.
Özetleme becerisine göre oluşturulan alt gruplarda uyum katsayıları ve karşılaştırması

Üniteler	Aşama	χ^2	df	p	RMSEA	SRMR	CFI	ΔCFI	$\Delta RMSEA$
Güney Kutbu	Yapısal	14,349	4	0,006	0,047	0,035	0,983		
	Metrik	13,487	8	0,096	0,024	0,040	0,991	0,006	-0,041
	Ölçek	187,643	12	0,000*	0,113	0,040	0,715	-0,276	0,088
Uyku	Yapısal	Model yakınsamadığı için değerler elde edilememiştir.							
	Metrik	Model yakınsamadığı için değerler elde edilememiştir.							
	Ölçek	Model yakınsamadığı için değerler elde edilememiştir.							
Machu Picchu	Yapısal	1,204	4	0,877	0,000	0,009	1,000		
	Metrik	4,841	8	0,774	0,000	0,020	1,000	0,000	0,000
	Ölçek	224,744	12	0,000*	0,106	0,020	0,660	-0,340	0,106
Nalini Nadkarni	Yapısal	2,983	4	0,561	0,000	0,015	1,000		
	Metrik	6,239	8	0,621	0,000	0,028	1,000	-0,004	0,000
	Ölçek	139,079	12	0,000	0,095	0,028	0,299	-0,701	0,095
Büyük Pasifik Çöp Alanı	Yapısal	0,000	0	-	0,000	0,000	1,000		
	Metrik	1,006	3	0,800	0,000	0,014	1,000	0,000	0,000
	Ölçek	139,620	6	0,000*	0,138	0,014	0,410	-0,590	0,138

* p değeri, 0,05 düzeyinde anlamlı.

Tablo 3'e göre Uyku ünitesinde model yakınsamadığı için uyum katsayıları elde edilememiştir. Uyku ünitesi dışındaki tüm ünitelerde yapısal ve metrik değişmezlik aşamalarında elde edilen RMSEA, SRMR ve CFI değerleri iyi uyuma işaret etmektedir. Ölçek değişmezliği aşamasındaki uyum katsayıları ise kabul edilebilir değerlerde değildir. Söz konusu 4 ünite içinde yapısal-metrik değişmezlik karşılaştırması sonucu elde edilen ΔCFI ve $\Delta RMSEA$ değerleri daha kısıtlı olan metrik değişmezlik modelinin uyumu



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

kötüleştirmedini göstermektedir. Ancak metrik-ölçek değişmezliği karşılaştırmasındaki ΔCFI ve $\Delta RMSEA$ değerleri, ölçek değişmezliğinin incelenmesi için kısıtlanan eşik değerlerin uyumun kötüleştirmesine sebep olduğu yönünde bilgi vermektedir. Bu bağlamda Güney Kutbu, Machu Picchu, Nalini Nadkarni ve Büyük Pasifik Çöp Alanı ünitelerinde özetleme becerisine göre oluşturulan alt gruplarda göre metrik değişmezliğin sağlandığı, ölçek değişmezliğinin sağlanmadığı görülmektedir. Bu durumda düşük ve yüksek özetleme becerisine sahip bireylerin oluşturduğu alt gruplarda bu dört üniteye ilişkin soruların aynı psikolojik yapıyı temsil ettiği, aynı alt gruplardaki öğrencilerin aynı soruları aynı şekilde yorumladığı söylenebilir. Ancak ortalamalar arasındaki farklılığın nedeni tam olarak bilinemez.

Tablo 4'te ekonomik – sosyal – kültürel durum (ESCS) değişkenine göre oluşturulan alt gruplara göre beş üniteye ilişkin hiyerarşik ölçme değişmezliği aşamalarının uyum katsayıları ve ardışık değişmezlik aşamalarının karşılaştırıldığı analiz sonuçları verilmiştir. Uyum katsayıları SRMR hariç ölçeklenmiş uyum katsayılarıdır.

Tablo 4.

Ekonomik – sosyal – kültürel durum (ESCS) değişkenine göre oluşturulan alt gruplarda uyum katsayıları ve karşılaştırması

Üniteler	Aşama	χ^2	df	p	RMSEA	SRMR	CFI	ΔCFI	$\Delta RMSEA$
Güney Kutbu	Yapısal	12,775	6	0,047*	0,038	0,031	0,990		
	Metrik	15,165	14	0,367	0,010	0,042	0,998	0,003	-0,036
	Ölçek	142,583	22	0,000*	0,084	0,042	0,813	-0,185	0,074
Uyku	Yapısal	0,000	0	-	0,000	0,000	1,000		
	Metrik	1.362	6	0,968	0,000	0,029	1,000	0,000	0,000
	Ölçek	69,305	12	0,000*	0.129	0,029	0,000	-1,000	0,129
Machu Picchu	Yapısal	10,412	6	0,108	0,026	0,028	0,993		
	Metrik	14,465	14	0,416	0,006	0,037	0,999	0,009	-0,033
	Ölçek	209,318	22	0,000	0,090	0,037	0,711	-0,289	0,084
Nalini Nadkarni	Yapısal	4,410	6	0,621	0,000	0,022	1,000		
	Metrik	18,437	14	0,188	0,020	0,045	0,979	0,000	0,000
	Ölçek	96,351	22	0,000*	0,060	0,045	0,653	-0,326	0,046
Büyük Pasifik Çöp Alanı	Yapısal	0,000	0	-	0,000	0,000	1,000		
	Metrik	21,151	6	0,002*	0,057	0,059	0,947	-0,059	0,090
	Ölçek	90,029	12	0,000*	0,091	0,059	0,729	-0,218	0,034

* p değeri, 0,05 düzeyinde anlamlı.

Tablo 4'e göre tüm ünitelerde yapısal ve metrik değişmezlik aşamalarında elde edilen değerler iyi uyuma işaret etmektedir. Ölçek değişmezliği aşamasındaki uyum katsayıları ise kabul edilebilir değerlerde değildir. ΔCFI ve $\Delta RMSEA$ değerlerine göre değişmezlik aşamaları karşılaştırıldığında Büyük Pasifik Çöp Alanı dışındaki ünitelerde daha kısıtlı olan metrik değişmezliğin yapısal değişmezliğe göre uyumu kötüleştirmede, ancak ölçek değişmezliğinde kısıtlanan eşik değerin metrik değişmezliğe göre uyumu kötüleştirdiği görülmektedir. Büyük Pasifik Çöp Alanı ünitesinde ise yapısal değişmezliğe göre daha kısıtlı



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

olan metrik değişmezliğin uyumu kötüleştirdiği görülmektedir. Dolayısıyla Güney Kutbu, Uyku, Machu Picchu ve Nalini Nadkarni ünitelerinde ESCS alt gruplarına göre metrik değişmezlik sağlanmış, ölçek değişmezliği sağlanamamıştır. Büyük Pasifik Çöp Alanı ünitesinde ise yapısal değişmezlik sağlanmış, metrik değişmezlik sağlanamamıştır. Bu durumda düşük, orta ve yüksek sosyoekonomik düzeye sahip bireylerin oluşturduğu alt gruplarda metrik değişmezliğin sağlandığı dört üniteye ilişkin soruların aynı psikolojik yapıyı temsil ettiği, aynı alt gruplardaki öğrencilerin aynı soruları aynı şekilde yorumladığı söylenebilir. Ancak ortalamalar arasındaki farklılığın nedeni tam olarak bilinemez. Sadece yapısal değişmezliğin sağlandığı Büyük Pasifik Çöp Alanı ünitesine ilişkin soruların ise aynı psikolojik yapıyı temsil ettiği, fakat aynı alt gruplardaki öğrencilerin aynı soruları farklı şekilde yorumladığı söylenebilir. Bu durum madde yanlılığının göstergesidir.

Özetle güvenilirliği değerlendirme değişkenine göre 3 ünitenin yapısal, 2 ünitenin metrik değişmezliği; özetleme değişkenine göre 4 ünitenin metrik değişmezliği; ESCS değişkenine göre 1 ünitenin yapısal değişmezliği, 4 ünitenin metrik değişmezliği sağladığı görülmüştür. Ölçek değişmezliği aşaması hiçbir üniteye sağlanamadığından katı değişmezlik aşaması ünitelerde test edilmemiştir.

Tartışma, Sonuç ve Öneriler

Bu çalışmada, PISA 2018 Okuma Becerileri Testinde yer alan beş üniteye yer alan maddelerin güvenilirliği değerlendirme becerisi, özetleme becerisi ve ekonomik – sosyal – kültürel durum (ESCS) açısından oluşturulan alt gruplara göre ölçme değişmezliği test edilmiştir. Değişkenlerin ranjı dikkate alınarak METASPAM ve METASUM değişkenlerine göre iki, ESCS değişkenine göre ise üç grup oluşturulmuştur. Ölçme değişmezliği testinden önce her bir üniteye yer alan maddelerin yapı geçerliğini oluşturulan alt gruplarda test etmek için DFA ayrı ayrı gerçekleştirilmiş ve model uyumu incelenerek genel faktör yapıları her bir grup için teyit edilmiştir. Bu aşamada Uyku ve Büyük Pasifik Çöp Alanı ünitelerinde kurulan modellerin tam tanımlanmış olmasından kaynaklı model veri uyumunun mükemmel olması sebebiyle bu iki üniteye DFA analizine gerek duyulmamıştır.

Güvenilirliği değerlendirme (METASPAM) değişkenine göre oluşturulan alt gruplarda Uyku, Machu-Picchu ve Büyük Pasifik Çöp Alanı ünitelerinde yapısal değişmezlik sağlanmış ancak metrik değişmezlik sağlanamamıştır. Bu durum madde yanlılığına işaret etmektedir ve gruplar arası anlamlı karşılaştırmalar yapılmasını engellemektedir (Şen, 2020). Güney Kutbu ve Nalini Nadkarni ünitelerinde metrik değişmezlik aşamasına kadar değişmezlik sağlanmış olsa da daha ileri seviyelerde sağlanamayan değişmezlik, gruplar arası farkların kaynağının belirsizliğini koruduğunu göstermektedir. Özetleme (METASUM) değişkenine göre oluşturulan alt gruplarda ise uyku ünitesinde model yakınsamadığı için sonuç alınamamış, bunun veri karmaşıklığı veya örneklem küçüklüğü gibi nedenlerden kaynaklanabileceği belirtilmiştir (Kline, 2016; Uyumaz ve Sırgancı, 2020). Güney Kutbu, Machu Picchu, Nalini Nadkarni ve Büyük Pasifik Çöp Alanı ünitelerinde metrik değişmezlik aşamasına kadar değişmezlik sağlanmış, ancak ölçek ve katı değişmezlik sağlanamadığından maddelerin grupları eşit şekilde ölçmediği ve gruplar arası



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.

Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.

Araştırma Makalesi / Research Paper

farkların karşılaştırılmasının anlamlı olmadığı sonucuna ulaşılmıştır. ESCS değişkenine göre Büyük Pasifik Çöp Alanı ünitesinde yapısal değişmezlik sağlanmış ancak metrik değişmezlik sağlanamamış; bu durum madde yanlılığına işaret ettiğinden alt grupların okuma becerilerinin karşılaştırılamayacağı sonucuna ulaşılmıştır. Uyku, Machu-Picchu, Nalini Nadkarni ve Büyük Pasifik Çöp Alanı ünitelerinde ölçek değişmezliği aşamasına kadar değişmezlik sağlanmış ancak sonraki aşamalarda sağlanamamıştır.

Sayfa | 795

Üç gruplama değişkenine göre oluşturulan alt gruplarda hiçbir üniteye ölçek ve katı değişmezlik sağlanmadığından, maddelerin öğrencileri eşit şekilde ölçmediği ve gruplar arası karşılaştırmaların anlamlı olmadığı sonucuna ulaşılmıştır. Vandenberg ve Lance (2000) ölçek değişmezliğinin gruplar arası karşılaştırmalar için ön koşul olduğunu söylemekte, Asil ve Brown (2015) ise genellenebilirlik açısından önemini vurgulamaktadır. Bu çalışmada hiçbir üniteye katı değişmezlik sağlanamamıştır. Bu durum bu çalışmada belirlenen alt gruplar arasında okuma becerilerinin karşılaştırılmasının uygun olmadığını göstermektedir. Ölçek değişmezliğinin sağlandığı durumlarda bazı alt grupların okuma puanları karşılaştırılabilirken, metrik değişmezlik aşamasına kadar ulaşan ünitelerde karşılaştırmaların yanlı olabileceği ve dikkatli olunması gerektiği ifade edilmiştir. ESCS'ye göre oluşturulan alt gruplarda sadece 1 üniteye yapısal, 4 üniteye metrik değişmezlik sağlanmış, diğer değişmezlik seviyelerine ulaşamamıştır. Asil ve Brown'ın (2015) PISA 2009 verilerini kullanarak okuma becerileri başarısının dil, kültür ve ekonomik kalkınmaya göre ülkeler arası ölçek değişmezliğini inceledikleri çalışmalarında 55 ülkeden sadece 3 tanesinin ölçek değişmezliği aşamasına kadar gelebildiğini ifade etmişlerdir. Ülkelerin ekonomik kalkınma düzeylerine PISA ESCS değişkenine göre karar verdikleri çalışmanın sonucunda sosyoekonomik kaynakların ölçek değişmezliğinde büyük bir rol oynadığını belirtmişlerdir. İmrol (2017) PISA 2012 verileriyle motivasyon ve öz-inanç psikolojik yapılarının, içinde SED de olan 3 değişkene göre ölçek değişmezliğini incelediği çalışmada sosyoekonomik düzeye göre motivasyon yapısının katı değişmezlik aşamasını, öz-inanç yapısının en çok metrik değişmezlik aşamasını sağladığını belirtmiştir. Stafford (2011) Bilişsel ve Duyuşsal Katılım Ölçeğinden oluşturduğu modelin SED'e göre ölçek değişmezliğini incelediği çalışmada en çok metrik değişmezliğin sağlandığı sonucunu bulmuştur. Espinosa (2016) öğrenci katılım modelinin SED'e göre ölçek değişmezliğini incelediği çalışmada değişmezliğin sağlandığını belirtmiştir. Bu çalışmalarda göstergeler duyuşsal maddelerden elde edilmiş olmalarına rağmen oluşturulan alt gruplar bu çalışmayla benzerlik göstermektedir. SED değişkenine göre bakılan ölçek değişmezliği seviyelerinde bir istikrar olmadığı görülmüştür. Bu çalışmada da ESCS alt gruplarında 1 üniteye yapısal değişmezlik, 4 üniteye metrik değişmezlik sağlandığı görülmüştür.

Bu çalışma, PISA 2018 okuma becerileri başarısını yordayan üç değişken üzerinden ölçek değişmezliğini incelemiştir. Ölçek değişmezliği, geçerlik kanıtı sunduğundan, farklı değişken ve yöntemlerle benzer çalışmalar yapılabilir. Geniş ölçekli testlerde kullanılan maddelerin geçerliğini daha kapsamlı bir şekilde değerlendirmek amacıyla DMF çalışmaları da gerçekleştirilebilir (Atalay Kabasakal ve Gören, 2025; Gök vd., 2014). Çoklu grup doğrulayıcı faktör analizi kullanılan çalışmada, örtük yapı analizi veya MTK'ya dayalı yöntemlerle yöntem karşılaştırmaları yapılması önerilmektedir. Ayrıca, ölçek değişmezliğinin sağlanmadığı durumlarda, sorunlu maddeleri tespit etmek için kısmi ölçek değişmezliği çalışmaları gerçekleştirilerek maddelerdeki yanlılık kaynaklarının giderilmesi önerilmektedir.

Yazkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçek değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi, 17(1), 777-798.*

DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

Kaynakça

- Arıcı, Ö. ve Altıntaş, Ö. (2014). PISA 2009 okuma becerileri yeterliklerinin sosyoekonomik alt yapı ve okul öncesi eğitime katılım açısından incelenmesi "Türkiye örneği". *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 47(1), 423-448. https://doi.org/10.1501/Egifak_0000001331
- Asil, M. and Brown, G. T. (2015). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing*, 16(1), 71-93. <http://doi.org/10.1080/15305058.2015.1064431>
- Atalay Kabasakal, K. ve Gören, S. (2025). Eğitimde madde takımlarının kullanımı: eTIMSS 2019 örneği. *Eğitim ve Bilim*, 50, 111-127. <http://dx.doi.org/10.15390/EB.2025.14104>
- Atılğan, M. (2022). *TIMMS 2019 matematik başarısına ilişkin duyuşsal özelliklerin cinsiyete göre ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Ankara Üniversitesi.
- Ayvallı, M. (2016). *PISA 2012 matematik okur-yazarlığı testinin ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Akdeniz Üniversitesi.
- Bağdu Söyler, P., Aydın, B. ve Atılğan, H. (2021). PISA 2015 reading test item parameters across language groups: a measurement invariance study with binary variables. *Eğitimde Ve Psikolojide Ölçme Ve Değerlendirme Dergisi*, 12(2), 112–128. <https://doi.org/10.21031/epod.800697>
- Bofah, E. ve Hannula, M. (2017). Home resources as a measure of socio-economic status in Ghana. *Large-scale Assessments in Education*, 5(1), 1-15.
- Bowen, N. K., and Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research*, 6(2), 229–249. <https://doi.org/10.1086/681607>
- Bryne, B. and Stewart, S.M. (2006). The MACS approach to testing for multigroup invariance of a second order structure: A walk through the process. *Structural Equation Modelling a Multidisciplinary Journal*, 287-321. <http://doi.org/10.1186/s40536-017-0039-5>
- Bryne, B. and Watkins, D. (2003). The issue of measurement invariance revisited, *Journal of Cross-cultural Psychology*, 34(2), 155-175. <https://doi.org/10.1177/0022022102250225>
- Cheung, G.W. and 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
- Coşguner, T. (2013). *Uluslararası öğrenci başarı değerlendirme programı (PISA) 2009 uygulaması okuma becerileri okuryazarlığını etkileyen faktörler*. (Yüksek lisans tezi). Akdeniz Üniversitesi.
- Çakıcı Eser, D. (2021). Investigation of measurement invariance according to home resources: TIMSS 2015 mathematical affective characteristics questionnaire. *International Journal of Assessment Tools in Education*, 8(3), 633-648. <https://doi.org/10.21449/ijate.817168>
- DiStefano, C., and Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling a Multidisciplinary Journal*, 21(3), 425–438. <https://doi.org/10.1080/10705511.2014.915373>
- Dinçer, L. (2023). *Sınıf içi değerlendirme yaklaşımları envanterinin Türkçe'ye uyarlanması ve ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Gaziantep Üniversitesi.
- Dunslosky, J. and Thiede, K.W. (1998). "What makes people study more? An evaluation of factors that affect self-paced study". *Acta Psychologica* 98 (1998), 3756.
- Epeçan, C. (2018). Okuduğunu anlama becerisinin gelişiminde özetleme tekniğinin etkisi üzerine bir değerlendirme. *Ekev Akademi Dergisi*, 74, 11-30.
- Eroğlu, B. ve Aslan, Y. (2020). Ortaokul öğrencilerinin özetleme tutumlarının bazı değişkenler açısından incelenmesi. *Dil Eğitimi ve Araştırmaları Dergisi*, 6(2), 281-194. <https://doi.org/10.31464/jlere.610085>
- Espinosa, J. E. (2016). *Assessing the factorial validity, measurement invariance, and latent mean differences of a second-order, multidimensional model of academic and social collage course engagement: A comparison accross course format, ethnic groups, and economic status*. (Doctoral thesis). California University.



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

- Fraenkel, J.R., Wallen, N.E. and Hyun, H.H. (2012). *How to design and evaluate research in education (8th ed.)*. New York: Mc Graw Hill.
- Francisca, E.E. and Mezoh, P.R. (2018). Home environment as a predictor of academic performance of pupils with learning disabilities in buea, south west region of cameroon. *International Journal of Trend in Scientific Research and Development*, 3(1), 711–729. <https://doi.org/10.31142/ijtsrd19067>
- Gök, B., Kabasakal, K. A., ve Kelecioğlu, H. (2014). PISA2009 öğrenci anketi tutum maddelerinin kültüre göre değişen madde fonksiyonu açısından incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 5(1), 72-87. <https://doi.org/10.21031/epod.64124>
- Gönen, E. (2021). *PISA 2018 okuma becerileri testinin ülkelere ve cinsiyete göre ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Hacettepe Üniversitesi.
- Gören, S., Sayın, A., ve Gelbal, S. (2024). An analysis of item bias in the PISA 2018 reading understanding and memorising strategies questionnaire. *Kastamonu Education Journal*, 32(2), 345-356. <https://doi.org/10.24106/kefdergi.1473662>
- Gündoğmuş, İ. (2017). *Kağıt-kalem, bilgisayar ve tablet ortamında gerçekleştirilen sınavlar için ölçme değişmezliğinin ve öğrenci görüşlerinin incelenmesi*. (Yüksek lisans tezi). Hacettepe Üniversitesi.
- Güngör, M. (2019). *Fen motivasyonu ve öz yeterliği modeli'nin ölçme değişmezliğinin incelenmesi: PISA 2015 Türkiye örneği*. (Yüksek lisans tezi). Hacettepe Üniversitesi.
- He, J., Barrera-Pedemonte, F. ve Buchholz, J. (2018). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education Principles Policy and Practice*, 26(4), 369–385. <https://doi.org/10.1080/0969594x.2018.1469467>
- Hoyle, R.H. (1995). *Structural Equation Modelling*. California: SAGE Publications.
- İmrol, F. (2017). *PISA 2012 Türkiye örneğinde matematiğe yönelik motivasyon ve öz-inanç yapılarının ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Ankara Üniversitesi.
- Kasap, Y., Doğan, N. ve Koçak, C. (2021). PISA 2018'de okuduğunu anlama başarısını yordayan değişkenlerin veri madenciliği ile belirlenmesi. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 19(4), 241-258. <https://doi.org/10.18026/cbayarsos.959609>
- Kenny, D. A. (Eds.). (2006). *Methodology in the social sciences*. The Guilford Press.
- Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye-Çin(Şangay)-Endonezya örneği*. (Yüksek lisans tezi). Hacettepe Üniversitesi.
- Kline, R. B. (2016). *Methodology in the social sciences: Principles and practice of structural equation modeling (4th ed.)*. The Guilford Press.
- Lee, S. S. (2019). *Longitudinal and cross-country measurement invariance of the PISA home possessions scale*. (Doctoral Thesis). The Pennsylvania State University.
- MEB (2019). *PISA 2018 Türkiye ön raporu*. Eğitim ve Analiz Değerlendirme Raporları Serisi No:10. T.C. Milli Eğitim Bakanlığı Basımevi.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543 (1993). <https://doi.org/10.1007/BF02294825>
- OECD (2009). *PISA 2009 Assessment framework: Key competencies in reading, mathematics and science*, OECD Publishing.
- OECD (2019). *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing Paris. <https://doi.org/10.1787/b25efab8-en>
- Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği*. (Doktora tezi). Hacettepe Üniversitesi.
- Pohl, S. (2013). Longitudinal multistage testing. *Wiley*, 50(4), 447-468. <https://doi.org/10.1111/jedm.12028>
- Rutkowski, L. and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>

Yazıkan, T. ve Atalay Kabasakal, K. (2026). PISA 2018 okuma becerileri testinin bilişsel ve sosyo-ekonomik gruplar arasında ölçme değişmezliği. *Batı Anadolu Eğitim Bilimleri Dergisi*, 17(1), 777-798.
DOI. 10.51460/baebd.1671573



Batı Anadolu Eğitim Bilimleri Dergisi, (2026), 17 (1), 777-798.
Western Anatolia Journal of Educational Sciences, (2026), 17 (1), 777-798.
Araştırma Makalesi / Research Paper

- Sarier, Y. (2021). PISA uygulamalarında Türkiye'nin performansı ve öğrenci başarısını yordayan değişkenler. *Türkiye Sosyal Araştırmalar Dergisi*, 25(3), 905-926. <https://dergipark.org.tr/tr/pub/tsadergisi/issue/66172/757533>
- Schmitt, N. and Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210–222. <https://doi.org/10.1016/j.hrmmr.2008.03.003>
- Stafford, R. E. (2011). *Evaluation of the student engagement instrument: Measurement invariance across economic status and association with academic achievement*. (Master Thesis). Southeastern Louisiana University.
- Şen, S. (2020). *Mplus ile yapısal eşitlik modellemesi uygulamaları. (1.baskı)*. Nobel Akademik Yayıncılık.
- Tabachnick, B.G. ve Fidell, L.S. (2013). *Using multivariate statistics*. 6th ed. Boston: Pearson Education.
- Uyumaz, G. (2021). *Yükseköğretimde Okul Terki Ölçeği'nin Türk kültürüne uyarlanması ve kültürlerarası ölçme değişmezliğinin belirlenmesi*. (Yüksek lisans tezi). Kocaeli Üniversitesi.
- Uyumaz, G. ve Sırgancı, G. (2020). Doğrulayıcı faktör analizi için gerekli örneklem büyüklüğü kaç kişidir: Bayes yaklaşımı ve maksimum olabilirlik kestirimi. *Uluslararası Toplum Araştırmaları Dergisi*, 16(32), 5304-5340.
- Ülkü, S. (2019). *ABİDE 2016 Türkçe ve fen bilimleri alt-testlerinin öğretmen özelliklerine göre ölçme değişmezliğinin incelenmesi*. (Yüksek lisans tezi). Hacettepe Üniversitesi.
- Van De Vijver, F. J. R. (2018). Towards an integrated framework of bias in noncognitive assessment in international large-scale studies: challenges and prospects. *Educational Measurement Issues and Practice*, 37(4), 49–56. <https://doi.org/10.1111/emip.12227>
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Yalçinkaya, M. (2023). *2019 TIMMS ve eTIMMS 4. ve 8.sınıf matematik ve fen bilimleri başarı testlerinin ölçme değişmezliği ve madde yanlılığının incelenmesi*. (Yüksek lisans tezi). Ege Üniversitesi.