

6. Sınıf Türkçe Ortak Sınavının Cinsiyete ve Okul Coğrafyasına Göre Ölçme Değişmezliğinin İncelenmesi: Niğde İli Örneği

ARAŞTIRMA MAKALESİ

Erdem BODUROĞLU¹, Oğuz KOÇ², Hüseyin EKİNAY³

¹ Dr, Müdür Yrd., Millî Eğitim Bakanlığı, Niğde Ölçme Değerlendirme Merkezi, erdemboduroglu@gmail.com, ORCID: 0000-0001-8318-4914

² Öğretmen, Millî Eğitim Bakanlığı, Polatlı Hatice-Mevlüt Yılmaz Bilim ve Sanat Merkezi, oguzkoc20@hotmail.com, ORCID: 0000-0002-8656-6069

³ Müdür, Millî Eğitim Bakanlığı, Niğde Ölçme Değerlendirme Merkezi, ekinay51@gmail.com, ORCID:0000-0001-8049-3682

Gönderilme Tarihi: 02.03.2023 Kabul Tarihi: 23.05.2023 DOI: 10.37669/milliegitim.1258944

Atf: "Boduroğlu, E., Koç, O., ve Ekinay, H. (2024). 6. Sınıf türkçe ortak sınavının cinsiyete ve okul coğrafyasına göre ölçme değişmezliğinin incelenmesi: Niğde ili örneği. *Millî Eğitim*, 53(244), 1773-1788. DOI: 10.37669/milliegitim.1258944"

Öz

İllerde kurulan Ölçme Değerlendirme Merkezleri aracılığıyla son dönemlerde farklı sınıf düzeylerinde bazı sınavlar il genelinde ortak sınav şeklinde yapılmaktadır. Bu sınavların güvenilirlik ve geçerliğinin artırılması amacıyla atılacak adımlar önem arz etmektedir. Bu araştırmada da Niğde ilinde 6.sınıf düzeyinde yürütülen Türkçe dersi ortak sınavının cinsiyet ve okul coğrafyasına göre ölçme değişmezliğinin incelenmesi amaçlanmıştır. 2019-2020 eğitim öğretim yılında 6.sınıf Türkçe sınavına katılan 5032 öğrenci araştırma grubunu oluşturmaktadır. Araştırma sonucunda Türkçe dersi ortak sınavının hem cinsiyete hem de coğrafyasına göre güçlü değişmezlik sağladığı bulgusuna ulaşılmıştır. Bu neticede gruplar arasındaki ortalama farklılıkların ölçülen örtük özellikteki gerçek farklılıktan kaynaklandığı ifade edilebilir.

Anahtar Kelimeler: ortak sınav, ölçme değerlendirme merkezi, ölçme değişmezliği, yapısal eşitlik modeli, çok gruplu doğrulayıcı faktör analizi

Examining the Measurement Invariance of the 6th Grade Turkish Common Exam by Gender and School Location: The Example of Niğde Province

Abstract

Through the Measurement and Evaluation Centers established in the provinces, some examinations at different grade levels have been held in the form of a common examination throughout the province in recent years. The steps to be taken in order to increase the reliability and validity of these exams are important. In this study, it was aimed to examine the measurement invariance of the Turkish lesson common exam conducted at the 6th grade level in Niğde province according to gender and school location. 5032 students who participated in the 6th grade Turkish exam in the 2019-2020 academic year constitute the research group. As a result of the research, it was found that the Turkish lesson common exam provides strong invariance according to both gender and school location. In this way, it can be stated that the mean differences between the groups are due to the real difference in the implicit trait measured.

Keywords: *common exam, measurement and evaluation centers, measurement invariance, structural equation modeling, multi-group confirmatory factor analysis*

Giriş

Psikolojik yapılar ancak dolaylı yoldan ölçülebildiği için bu tür değişkenleri ölçmek amacıyla çeşitli ölçme araçları geliştirilmiştir. Psikolojik ölçme süreçleri bazı sorunlar barındırmaktadır. Bu sorunlardan bazıları: herhangi bir yapının ölçümüne yönelik tek bir yaklaşımın evrensel olarak kabul edilememesi, davranış evreninin sınırlı sayıdaki örnekleme üzerinde çalışılabilmesi, elde edilen ölçme sonuçlarına birçok hata kaynağının karışabilmesi, kullanılan ölçeklerin biriminin iyi tanımlanamamış olması, psikolojik yapıların kavramsal ve operasyonel tanımlamalarında güçlüklerle karşılaşılması olarak ifade edilebilir. Bahsedilen bu sorunlar ölçme araçlarından elde edilen sonuçların geçerliğini de tartışmalı hâle getirmektedir (Crocker ve Algina, 1986). Literatürde geçerlik konusunda farklı tanımlamalar ve sınıflamalar yer almaktadır. Genel anlamda geçerlik, ölçme aracının ölçmek istediği özelliği ölçüp ölçmediğine ilişkin kanıt toplama süreci olarak ifade edilebilir. Yeteri kadar geçerlik kanıtı elde edilmesi ölçme aracından elde edilen sonuçların daha anlamlı olarak yorumlanmasını sağlayacaktır (Kane, 2006; Messick, 1995; Sireci, 2007). Özellikle farklı alt grupların ya da bireylerin karşılaştırıldığı çalışmalarda bir geçerlik kanıtı olarak ölçme değişmezliğinin incelenmesi gerekmektedir.

Ölçme değişmezliği, ölçülmek istenilen yapının farklı alt gruplarda (cinsiyet, sosyoekonomik düzey, kültür vb.) aynı şekilde ölçülüp ölçülmediğine dair istatistiksel kanıt sunar. Farklı gruplardaki bireyler ölçme aracında yer alan maddeleri aynı şekilde yorumlamıyorsa bu durum ölçme değişmezliğinin ihlal edildiği anlamına gelir ve gruplar arasında yapılan karşılaştırmaların hatalı yorumlanmasına yol açabilir (Vandenberg ve Lance, 2000). Literatürde ölçme değişmezliğinde kullanılan yöntemler genellikle iki temel başlık altında ele alınmaktadır. Bunlardan ilki Yapısal Eşitlik Modeli (YEM) çatısı altında ele alınan Doğrulamalı Faktör Analizi (DFA) yöntemi, diğeri ise Madde Tepki Kuramı altında ele alınan değişen madde ve test fonksiyonu belirleme yöntemidir. İki yöntem de çeşitli benzerlik ve farklılıklarına rağmen ölçme değişmezliğinin incelendiği araştırmalarda başvurulan yöntemlerdendir (Raju, Laffitte ve Byrne, 2002). YEM analizleri sosyal bilimlerde yaygın olarak kullanılmaktadır ve bu çerçevede Çok Gruplu Doğrulamalı Faktör Analizi (ÇGDFA) ile ölçme değişmezliği incelenmektedir (Cheung ve Rensvold, 2002). ÇGDFA, oluşturulan modelin alt gruplarda aynı olup olmadığının test edildiği ve DFA'nın iki ya da daha çok sayıdaki grupta eş zamanlı olarak uygulandığı bir yöntemdir (Tabachnick ve Fidell, 2013).

Ölçme değişmezliği dört aşamada ele alınır. Bunlar: 1-Yapısal/Biçimsel Değişmezlik, 2-Zayıf/Metrik Değişmezlik, 3-Güçlü/Skalar Değişmezlik, 4-Katı Değişmezlik modelleridir. Yapısal değişmezlik, ölçme değişmezliğinin ilk adımıdır ve bu adım sağlanmadığında diğer model karşılaştırma aşamalarına devam edilmesi anlamlı değildir. Değişmezlik sağlandığında maddelerin alt gruplarda aynı yapıyı ölçtüğü sonucuna varılır. Zayıf değişmezlikte yapısal değişmezliğe ilave olarak madde faktör yüklerinin gruplar arasında aynı olup olmadığı test edilir. Güçlü değişmezlikte ise diğerlerine ek olarak alt gruplarda regresyon sabitlerinin eşitliği sınanır. Bu değişmezliğin sağlandığı durumlarda gruplar arası ortalamaların karşılaştırması anlamlıdır. Son olarak katı değişmezlik aşamasında madde hata varyanslarının da eşit olup olmadığı test edilir. Bu aşamanın sağlanması pratikte pek mümkün olmamaktadır (Cheung ve Rensvold, 2002; Meredith, 1993; Wu, Li ve Zumbo, 2007).

Ülkemizde yürütülen araştırmalarda ölçme değişmezliğinin incelendiği sınırlı sayıda çalışma bulunmasına rağmen özellikle son dönemlerde PISA, PIRLS, TIMSS, TALIS gibi uluslararası araştırmalarda ve bazı ulusal araştırmalarda ölçme değişmezliği çalışmalarının daha yaygın

olarak yürütüldüğü söylenebilir. Seçme ve yerleştirme amacıyla yürütülen Liselere Giriş Sınavları (LGS) ve Yükseköğretim Kurumları Sınavlarında (YKS), Öğrenci Başarı İzleme ve Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) gibi eğitim araştırmalarında, iller bünyesinde yürütülen ortak sınav çalışmalarında ve diğer ulusal çalışmalarda ölçme değişmezliğinin incelenerek bir geçerlik kanıtı olarak sunulmasının önemli olduğu düşünülmektedir. Ülkemizde ulusal ölçekte yürütülen sınavlarda ölçme değişmezliği araştırmasının yapıldığı bazı çalışmalar aşağıda verilmiştir:

Kütük (2019) araştırmasında Akdeniz Üniversitesi Uluslararası Öğrenci Kabul Sınavı'nın cinsiyet, sınav dili ve ülke değişkenlerine göre ölçme değişmezliğini incelemiştir. Elde edilen bulgulara göre genel yetenek testinde cinsiyete göre yapısal değişmezliğin sağlandığı görülmüştür. Sınav dili ve ülke gruplarına göre ise ölçme değişmezliğinin sağlanmadığı ifade edilmiştir. Matematik alt testinde ise cinsiyet ve ülke değişkenleri için ölçme değişmezliği sağlanırken sınav dili değişkenine göre ölçme değişmezliğinin sağlanmadığı belirtilmiştir.

Ülkü (2019) ölçme değişmezliği incelemesini 2016 yılı ABİDE uygulaması üzerinde yürütmüştür. Araştırmada öğrenciler, öğretmenlerinin mesleki deneyimleri ve eğitim düzeylerine göre alt gruplara ayrılmıştır. Türkçe ve Fen testlerinin oluşturulan alt gruplarda ölçme değişmezliğinin aşamalarını sağladığı sonucuna ulaşılmıştır.

Altıntaş ve Kutlu (2020) araştırmasını Ankara Üniversitesi Yabancı Uyruklu Öğrenci Sınavı'na katılan 2134 birey arasından seçilen 1110 kişilik örneklem üzerinde yürütmüştür. Sınavda yer alan Temel Öğrenme Becerileri Testi'nin ülkelere ve cinsiyete göre ölçme değişmezliği incelenmiştir. Eş zamanlı Örtük Sınıf Analizi yöntemi ile ölçme değişmezliği sınanmış ve testin cinsiyet ve ülkelere göre oluşturulan alt gruplarda ölçme değişmezliğini sağladığı bulgusuna ulaşılmıştır.

Eriştiren (2021) çalışmasında, 2018 LGS Türkçe testi üzerinde cinsiyet ve okul türü değişkenlerine göre ölçme değişmezliği ve Değişen Madde Fonksiyonu (DMF) incelemesi yapmıştır. Sınava giren grup arasından rastgele seçilen 3000 örneklem üzerinde yürütülen bu çalışmada Türkçe testinin oluşturulan alt gruplarda ölçme değişmezliğinin tüm aşamalarını sağladığı görülmüştür.

Şenel (2022) araştırmasında ortaöğretime geçişte uygulanan merkezî sınavın, özel öğrenme güçlüğüne sahip olan ve olmayan öğrencilere göre ölçme değişmezliğini incelemiştir. Araştırmada 994 öğrenci özel öğrenme güçlüğüne sahip iken 1000 öğrencinin ise özel gereksinimi olmadığı belirtilmiştir. Araştırma bulgularına göre sınavın hiçbir alt testinin öğrencilerin özel öğrenme güçlüğü gösterme durumuna göre ölçme değişmezliğini sağlamadığı görülmüştür. Bu durumda özel öğrenme güçlüğü olan ve olmayan öğrencilerin merkezî sınav puanlarının kıyaslanmasının doğru olmayacağı belirtilmiştir. Büyük ölçekli ve yüksek riskli sınavlardaki test geliştirme süreçlerinde özel eğitim uzmanlarının görevlendirilmesinin uygun olacağı ifade edilmiştir.

Araştırmanın Amacı ve Önemi

Son dönemlerde 81 ilde kurulan Ölçme Değerlendirme Merkezleri sayesinde il genelinde bazı sınıf düzeylerinde ortak sınavlar yapılmaktadır. Yapılan bu sınavların sonuçları öğrencilerin karne notuna etki etmektedir. Test geliştirme sürecinde yer alan adımlara dikkat edilerek geliştirilen bu sınavlar için çeşitli güvenilirlik ve geçerlik kanıtları elde edilmektedir. Sınav sonrasında yayımlanan raporlar ile sahadaki paydaşlara dönüt verilerek öğrenme eksiklerinin giderilmesi amaçlanmaktadır. Ortaya konulan raporlarda çeşitli demografik özelliklere göre alt gruplar arası karşılaştırmalar yapılabilmektedir. Bu karşılaştırmaların daha anlamlı olarak yorumlanabilmesi için ölçme değişmezliğinin incelenmesi gerektiği bilinmektedir. Bu araştırma ile bundan sonra yapılacak ortak sınavlarda ve hazırlanacak raporlarda ölçme değişmezliğinin incelenmesi gerekliliği açısından bir farkındalık oluşturulması amaçlanmıştır. Bu amaçlarla araştırmada aşağıdaki sorulara cevap aranmıştır:

1. Türkçe ortak sınavında cinsiyet açısından ölçme değişmezliği sağlanmakta mıdır?
2. Türkçe ortak sınavında okul coğrafyası açısından ölçme değişmezliği sağlanmakta mıdır?

Yöntem

Araştırma Modeli

Bu araştırma, Niğde Ölçme Değerlendirme Merkezi koordinesinde 2019-2020 eğitim öğretim birinci yarıyılı 6. sınıflar Türkçe ortak sınavı kapsamında öğrencilerden elde edilen veriler üzerine gerçekleştirilmiştir.

Ortak sınav uygulamasının çalışma kapsamında, cinsiyet ve okul coğrafyası (kent-kırsal) değişkenlerine göre ölçme değişmezliği sağlayıp sağlamadığı incelenecektir. Bu açıdan çalışmanın ilişkisel modelde olduğu ifade edilebilir. İlişkisel araştırmalarda değişkenler arasındaki bağlantılar ortaya çıkarılmaya çalışılır. Bu sayede kavramların ve olguların daha iyi anlaşılabilmesi amaçlanır (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2014).

Evren ve Örneklem

Tablo 1

Araştırma Evreninin Cinsiyet ve Okul Coğrafyasına Göre Dağılımı

		Frekans(f)	Yüzde(%)
Cinsiyet	Kız	2417	48
	Erkek	2615	52
Bölge	Kent	3280	65
	Kırsal	1752	35

Tablo 1 incelendiğinde araştırma grubundaki 5032 öğrencinin %48'i kız, %52'si erkektir. Tüm öğrencilerin 3280'i (%65) kentsel bölge okullarında iken 1752'si (%35) kırsal bölge okullarında eğitim görmektedir.

Verilerin Toplanması

Bu araştırmanın verileri 2019-2020 eğitim öğretim yılında Niğde ili genelinde 6. sınıflara uygulanan Türkçe dersi ortak sınavından elde edilmiştir. Bu sınavda öğrencilere 20 maddeden oluşan bir başarı testi uygulanmıştır. Başarı testi, hazırlanan belirtke tablosuna uygun olarak 7 alan uzmanı ve 1 ölçme değerlendirme uzmanı tarafından geliştirilmiştir. Gerçekleştirilen pilot uygulama sonrasında test ve madde istatistiklerine göre nihai form oluşturulmuş ve bu form öğrencilere uygulanmıştır. Uygulanan başarı testinde yer alan 20 maddenin güçlük indeksinin 0,27 ile 0,90 arasında değiştiği ve madde güçlükleri ortalamasının 0,66 olduğu görülmüştür. Ayırt edicilik indeksleri nokta çift serili korelasyon katsayı ile hesaplanmıştır. Bu değerlerin 0,38 ile 0,63 arasında değiştiği ve ortalamasının 0,48 olduğu görülmüştür. Bu sonuçlara göre testin kolay ve yüksek ayırt edicilikte olduğu ifade edilebilir (Crocker ve Algina, 1986; Urbina, 2014). Testin güvenilirlik katsayısını hesaplamak için Kuder-Richardson (KR-20) formülü kullanılmış ve bu değer 0,81 olarak bulunmuştur.

Verilerin Analizi

Çalışmada öncelikle açımlayıcı ve doğrulayıcı faktör analizi gerçekleştirilmiştir. Ölçme değişmezliği analizi için ise ÇGDFA yaklaşımı tercih edilmiştir. Çalışma kapsamında gerçekleştirilen analizler Factor programı 11.05.01 (Ferrando ve Lorenzo-Seva, 2017) sürümü ve Mplus programı (Muthén ve Muthén, 2017) 8.3 sürümü kullanılmıştır.

Veri seti açımlayıcı ve doğrulayıcı faktör analizlerini gerçekleştirebilmek amacıyla tesadüfi olarak iki alt gruba ayrılmıştır. Veriler kategorik olduğu için açımlayıcı faktör analizi (AFA), tetrakorik korelasyon matrislerine dayalı olarak gerçekleştirilmiş ve faktörleştirme yöntemi olarak “uls” kullanılmıştır. AFA uygulamaları her bir değişken için 10-15 katı bireye ihtiyaç duyulmakta olup analiz için kullanılan 2516 öğrenci verisinin yeterli olduğu ifade edilebilir. KMO değerinin 0,93 olduğu ve Barlett Küresellik Testinin ise anlamlı olduğu $p < 0,001$ görülmüştür. Bu doğrultuda bulunan KMO ve Barlett değerleri, veri setinin AFA'ya uygun olduğuna işaret etmektedir (Büyüköztürk, 2017). Çoklu bağlantı problemi olup olmadığını incelemek için korelasyon matrisi incelenmiş ve 0,90'dan büyük ilişki düzeyi olmadığı görülmüştür.

Gerçekleştirilen AFA sonucunda özdeğeri 1'den büyük 3 boyut bulunmuştur. Ayrıca gerçekleştirilen Paralel Analiz sonucunda ise 2 adet boyut önerildiği görülmüştür. Bununla birlikte ilk boyutun tek başına varyansın %38'ini açıklaması, birinci boyut ve ikinci boyut arasında özdeğer açısından yaklaşık üç katlık bir fark bulunması, tek boyutlu yapı için madde faktör yüklerinin 0,30'dan büyük olması, uyum indekslerinin kabul edilebilir sınırlar içerisinde olması (RMSEA=0,08, TLI=0,95, CFI=0,95, GFI=0,94, AGFI=0,93) ve teorik olarak başarı testinin tek boyutlu yapıda olması gereklilikleri doğrultusunda Türkçe başarı testinin tek boyutlu olduğu ifade edilebilir.

AFA sonucundan sonra diğer veri seti ile DFA gerçekleştirilmiştir. Yapısal eşitlik modellemesinde kestirilen parametre sayısının en az 20 katı örneklem büyüklüğü önerilmektedir. Bununla birlikte parametre sayısının 10 katı ya da en az 200 örneklem olması gerekliliği de ifade edilmektedir (Kline, 2016). 2516 öğrenciye ait veri ile işlemler gerçekleştirildiği için örneklem büyüklüğünün yeterli olduğu ifade edilebilir. Çoklu bağlantı olup olmadığını

incelemek için ise tetrakorik korelasyon analizi gerçekleştirilmiştir. Çoklu bağlantı sorunu olmaması için maddeler arasında .90'dan daha alt düzeyde ilişkisi olması gerekmektedir (Tabachnick ve Fidell, 2013). Maddelere ait ilişki düzeylerinin .04 ile .72 arasında değiştiği bulunmuştur. Buradan hareketle çoklu bağlantı problemi olmadığı söylenebilir. Kestirim yöntemi olarak ikili puanlanan maddelerde kullanılması önerilen “WLSMV” kullanılmıştır (Gana ve Broc, 2019). Analiz sonucunda uyum indeksleri yeterli bulunmamış ve modifikasyon işlemi gerçekleştirilmiştir. Modifikasyon sonucunda uyum yakalanmasına karşın madde 1, madde 9 ve madde 11'in madde faktör yüklerinin 0,30'dan düşük olduğu görülmüştür. Maddelerin çıkarılmasından sonra DFA analizi tekrar gerçekleştirilmiş ve madde faktör yüklerinin 0,31 ile 0,82 arasında ve 0,05 düzeyinde anlamlı olduğu görülmüştür. Sonrasında ise model veri uyumu incelenmiştir. Uyum indekslerine yönelik bilgiler ise Tablo 2'de gösterilmiştir.

Tablo 2

Uyum İndeks Değerleri

Uyum İndeksi	Kabul Edilebilir Düzey	İyi Düzey	Model Değeri
χ^2	$p > 0,05$	$p > 0,05$	$p > 0,05$
χ^2/sd	$2 < \chi^2 / df \leq 8$	$0 \leq \chi^2 / df \leq 2$	2,16
RMSEA	$0,05 < RMSEA \leq 0,08$	$0 \leq RMSEA \leq 0,05$	0,021
TLI	$0,90 \leq TLI < 0,95$	$0,95 \leq TLI \leq 1,00$	0,989
CFI	$0,90 \leq CFI < 0,95$	$0,95 \leq CFI \leq 1,00$	0,990

(Hu ve Bentler, 1999; Schumacher ve Lomax, 2004; Kline, 2011)

Tablo 2'de yer alan uyum indeksleri incelendiğinde χ^2 değeri haricinde RMSEA değerinin kabul edilebilir, CFI ve TLI değerlerinin ise iyi düzeyde uyum gösterdiği görülmektedir. χ^2 değeri örneklemin büyük olmasında etkilenmesinden kaynaklı olarak anlamlı çıktığı ifade edilebilir (Kline, 2011). Buradan hareketle χ^2/sd değeri incelenmiş ve 0,06 olarak iyi düzeyde uyum gösterdiği bulgusuna ulaşılmıştır.

Brown (2015) ÇGDFA değişmezlik değerlendirilmesine her grup için ayrı ayrı modelin değerlendirilmesi ile başlanması gerektiğini belirtmiştir. Bu kapsamda çalışmada ölçme değişmezliği açısından ele alınacak kategoriler açısından ayrı ayrı DFA gerçekleştirilmiştir. Kız ve kent gruplarında

madde 6'nın faktör yükünün 0,30'dan küçük olduğu görülmüştür. Madde 6 analizlerden çıkarılarak tekrarlanmıştır. Kırsal grubu, madde 18 haricinde tüm gruplarda madde faktör yüklerinin 0,30'dan daha yüksek olduğu görülmüştür. Bu kapsamda cinsiyet açısından işlemler gerçekleştirilirken madde 18 dâhil edilmiş, okul coğrafyası açısından ise dâhil edilmemiştir. Uyum indekslerine yönelik bilgiler ise Tablo 3'te gösterilmiştir.

Tablo 3

Alt Gruplar İçin Uyum İndeks Değerleri

Uyum İndeksi	Kız	Erkek	Kent	Kırsal
²	P<0,05	P<0,05	P<0,05	P<0,05
² /sd	1,71	1,44	1,83	1,30
RMSEA	0,025	0,018	0,022	0,020
TLI	0,983	0,984	0,989	0,992
CFI	0,985	0,995	0,990	0,991

Tablo 3'te yer alan uyum indeksleri incelendiğinde uyum indekslerinin iyi düzeyde olduğu ifade edilebilir. Yukarıdaki bulgular sonucunda ÇGDFA ile ölçme değişmezliği analizine geçilmiştir. Muthén ve Muthén, (2017) ölçme değişmezliğinde ikili değişkenler ve ağırlıklı en küçük kareler tahmini için yalnızca biçimsel ve skaler modellerin dikkate alınması gerektiğini belirtmişlerdir. Bu sebeple ölçme değişmezliği için biçimsel ve skaler değişmezlik incelenmiş ve "Theta" parametreleştirmesi kullanılmıştır.

Ölçme değişmezliğinde ki-kare fark testi farklı seviyelerdeki modeller arasında karşılaştırma yapmak için kullanılmaktadır ancak büyük örneklem ile gerçekleştirilen çalışmalarda yanlış sonuçlar üretilmesine neden olabilmektedir (Şen, 2020). Cheung ve Rensvold (2002), Meade, Johnson ve Brady (2008) ve Chen (2007) büyük örneklemelerde uyum indeksleri arasındaki farka bakılmasını belirtmişlerdir. Cheung ve Resvold (2002) CFI değerinde 0,010 veya daha düşük değişikliklerin kısıtlı değişmezlik hipotezinin reddedilmemesi yönünde bir kanıt olduğunu belirtmişlerdir. Chen (2007) ise bazı durumlarda $\Delta CFI \leq 0.010$ ve $\Delta RMSEA \leq 0.015$ ölçütlerinin birlikte kullanıldığında doğru sonuç verdiğini ifade etmiştir. Bu çalışmada ΔCFI ve $\Delta RMSEA$ birlikte ele alınarak analiz edilmiştir.

Etik Konular

Araştırmanın etik kurul izni Niğde Ömer Halisdemir Üniversitesi Etik Kurulunun 04.10.2022 tarih ve 262357 sayılı onayı ile alınmıştır.

Bulgular

Bu çalışma Türkçe ortak sınavı uygulamasının cinsiyet ve okul coğrafyası gruplarına göre ölçme değişmezliğini araştırmak amacıyla gerçekleştirilmiştir. Bulgular araştırma problemleri çerçevesinde sırasıyla raporlanmıştır.

1. Türkçe ortak sınavının cinsiyete göre ölçme değişmezliği

Türkçe ortak sınavından elde edilen ikili veriler cinsiyet açısından ÇGDFA yöntemi ile ölçme değişmezliği analizi biçimsel ve skaler olarak aşamalı bir şekilde gerçekleştirilmiştir. Elde edilen bulgular Tablo 4'te gösterilmiştir.

Tablo 4

Cinsiyete Göre Ölçme Değişmezliği Sonuçları

Değişmezlik Türü	²	sd	RMSEA	CFI	TLI	Δ RMSEA	Δ CFI
Biçimsel	329,339	208	0,022	0,991	0,990	-	-
Skaler	407,928	222	0,026	0,986	0,985	0,004	0,005

Tablo 4 incelendiğinde hem biçimsel değişmezlik hem de skaler değişmezlik düzeyinde uyum indekslerinin iyi uyum düzeyinde olduğu görülmektedir (RMSEA<0,05; CFI>0,95; TLI>0,95). Bununla birlikte değişmezlik düzeyleri arasındaki uyum indeksi farklılıklarının da istenilen düzeyde olduğu ve skaler (güçlü) değişmezliğin sağlandığı ifade edilebilir (Δ CFI \leq 0,010; Δ RMSEA \leq 0,015). Buradan hareketle madde ortalamalarının cinsiyet grupları arasında aynı olduğu ve ölçeğin cinsiyet grupları için aynı şekilde kullandığı varsayılabilir.

2. Türkçe ortak sınavının okul coğrafyasına göre ölçme değişmezliği

Türkçe ortak sınavından elde edilen ikili veriler okul coğrafyası açısından ÇGDFA yöntemi ile ölçme değişmezliği analizi biçimsel ve skaler olarak aşamalı bir şekilde gerçekleştirilmiştir. Elde edilen bulgular Tablo 5'te gösterilmiştir.

Tablo 5*Okul Coğrafyasına Göre Ölçme Değişmezliği Sonuçları*

Değişmezlik Türü	²	sd	RMSEA	CFI	TLI	Δ RMSEA	Δ CFI
Biçimsel	279,653	180	0,021	0,991	0,990	-	-
Skaler	295,625	193	0,021	0,991	0,990	0,000	0,000

Tablo 5 incelendiğinde hem biçimsel değişmezlik hem de skaler değişmezlik düzeyinde uyum indekslerinin iyi uyum düzeyinde olduğu görülmektedir (RMSEA<0,05; CFI>0,95; TLI>0,95). Bununla birlikte değişmezlik düzeyleri arasındaki uyum indeksi farklılıklarının da istenilen düzeyde olduğu ve skaler (güçlü) değişmezliğin sağlandığı ifade edilebilir (Δ CFI \leq 0,010; Δ RMSEA \leq 0,015). Buradan hareketle madde ortalamalarının okul coğrafyası grupları arasında aynı olduğu ve ölçeğin okul coğrafyası grupları için aynı şekilde kullandığı varsayılabilir.

Tartışma, Sonuç ve Öneriler

81 ilde kurulan Ölçme Değerlendirme Merkezleri aracılığı ile il genelinde birçok dersten ortak sınav uygulamaları gerçekleştirilmektedir. Bu sınavların karne notu yerine geçtiği durumlarda test geliştirme sürecinin ve sonrasında gerçekleştirilen raporlama çalışmalarının titizlikle yürütülmesi gerekmektedir. Geliştirilen test maddelerinin herhangi bir gruba avantaj ya da dezavantaj sağlamaması, sürece dair güvenilirlik ve geçerlik kanıtlarının hazırlanan raporlarda sunulması ve paydaşlara dönüt verilmesi önem arz etmektedir. Yüksek riskli sınav kategorisinde olmayan bu uygulamalarda öğrencilere ve öğretmenlere verilecek geri bildirimler ile süreç içerisindeki güçlü ve zayıf yönler ortaya çıkarılabilecek, öğrenme eksiklikleri giderilebilecektir. Bu araştırmada Niğde Ölçme Değerlendirme Merkezi tarafından il genelinde gerçekleştirilen 6.sınıf Türkçe ortak sınavının cinsiyet ve okul coğrafyasına göre ölçme değişmezliği incelenmiştir. Elde edilen sonuçlara göre hem cinsiyete hem de okul coğrafyasına göre ölçme değişmezliğinin sağlandığı görülmüştür. Yani ölçülmek istenilen yapı, cinsiyet ve okul coğrafyası alt gruplarında aynı şekilde ölçülmektedir. Bu durumda gruplar arasında yapılacak karşılaştırmaların anlamlı olacağı ifade edilebilir.

Türkiye’de ulusal ve uluslararası düzeyde gerçekleştirilen bazı sınavlarda ölçme değişmezliği araştırmalarının yürütüldüğü görülmüştür. Bu sınavlardan bazıları ABİDE, LGS ve PISA gibi geniş ölçekli sınavlardır. Bu sınavlarda cinsiyet, ülke, sınav dili, eğitim düzeyi, öğrencilerin özel öğrenme güçlüğü olup olmama durumu vb. değişkenlerdeki alt gruplar açısından ölçme değişmezliği incelemeleri yapılmıştır. Başarı testleri haricinde geliştirilen ve uyarlanan birçok ölçek çalışmasında da ölçme değişmezliği araştırması yürütülmüştür. Özellikle cinsiyet değişkeni açısından yanlılık ve ölçme değişmezliği araştırmaları ile sıklıkla karşılaşılmaktadır. Bu araştırmaların bazılarında cinsiyete göre ölçme değişmezliğinin sağlanmadığı veya kısmi model düzeyinde sağlandığı görülmüştür (Coşkun, 2023; Gündeğer, 2023). Bazı araştırmalarda ise cinsiyete göre ölçme değişmezliğinin sağlandığı bulgusuna ulaşılmıştır (Altıntaş ve Kutlu, 2020; Eriştiren, 2021; Kütük, 2019).

Bölgesel çapta gerçekleştirilen ortak sınavlar ve eğitim araştırmalarında, test ve madde analizleri ile birlikte DMF, yanlılık ve ölçme değişmezliği çalışmalarının yapılmasının önemli olduğu düşünülmektedir. Özellikle esas uygulamalar öncesindeki pilot uygulamalarda bu çalışmaların gerçekleştirilmesi ile karşılaşılabilecek sorunların önüne geçilebileceği ifade edilebilir. Test geliştirme sürecinde görev alan öğretmenlerin ölçme değişmezliği ve yanlılık konularında uygulama öncesinde bilgilendirilmesinin faydalı olabileceği düşünülmektedir. Bu araştırmada cinsiyet ve okul coğrafyası değişkenlerine göre ölçme değişmezliği incelemesi yapılmıştır. Cinsiyet değişkeni, ölçme değişmezliği çalışmalarında yaygın olarak incelenmesi sebebiyle araştırma kapsamına dâhil edilmiştir. Okul coğrafyasının ve yaşanılan bölgenin ise Türkçe becerisi üzerinde etkili bir değişken olduğu bilinmektedir. Araştırmalarda kırsal ve kent öğrencilerinin Türkçe becerilerinin kentteki öğrenciler lehine farklılaştığını gösteren çalışmalar vardır (Coşkun, 2023; Deniz, 2003). Bu sebeple okul coğrafyası değişkeni araştırma kapsamında incelenmiştir. Başka araştırmalarda, farklı alt gruplar için (devlet-özel okul, sosyoekonomik düzey, anne-baba eğitim durumu vb.) ölçme değişmezliği ve yanlılık çalışmaları gerçekleştirilebilir. Türkçe ortak sınavının yapıldığı farklı illerdeki uygulamalarda benzer çalışmalar tekrarlanarak madde ve test hazırlama sürecinin daha nitelikli hâle getirilmesi sağlanabilir. Ölçme Değerlendirme Merkezleri aracılığıyla sadece bilişsel testlerin değil duyuşsal alana yönelik ölçeklerin de ölçme değişmezliği çalışmaları geniş örneklemeler üzerinde incelenebilir.

Kaynakça

- Abell, N., Springer, D. W., and Kamata, A. (2009). *Developing and validating rapid assessment instruments*. Oxford University Press.
- Aksu, G., Guzeller, C. O., and Eser, M. T. (2017). Analysis of maths literacy performances of students with Hierarchical Linear Modeling (HLM): The Case of PISA 2012 Turkey. *Education and Science*, Vol 42, 247-266.
- Altıntaş, Ö., ve Kutlu, Ö. (2020). Ankara üniversitesi yabancı uyruklu öğrenci seçme testinin ölçme değişmezliğinin örtük sınıf ve rasch modeline göre incelenmesi. *Eğitim ve Bilim*, 45(203).
- Arıkan, Ç. A., and Zorbaz, S. (2020). Measurement invariance of the satisfaction with life scale across gender and time. *Turkish Journal of Education*, 9(4), 260-272.
- Brown, T. A. (2015). Confirmatory factor analysis for applied research. *New York, NY: The Guilford Press*.
- Büyüköztürk, Ş. (2017). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum (23. baskı)*. Pegem Yayınları.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2014). *Bilimsel araştırma yöntemleri (16. Baskı)*. Pegem Akademi.
- 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.
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- Crocker, L., and Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
- Coşkun, F. (2023). PISA 2018 verisi üzerinden öğrencilerin kültürlerarası iletişim algısının örtük sınıf modelleri ile incelenmesi: Türkiye örnekleme. *Gazi Eğitim Bilimleri Dergisi*, 9(1), 1-20.

- Deniz, K. (2003). Yazılı anlatım becerileri bakımından köy ve kent beşinci sınıf öğrencilerinin durumu. *Türklük Bilimi Araştırmaları*, (13), 233.
- Durmaz, M. (2019). *Okul motivasyonu ölçeğinin cinsiyete ve sınıf düzeyine göre ölçme değişmezliğinin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi.
- Embretson, S. E., and Reise, S. P. (2013). *Item response theory*. Psychology Press.
- Eriştiren, İ. (2021). *Ortaöğretime geçiş sınavlarında ölçme değişmezliği ve dmf'nin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi.
- Ferrando, P.J., and Lorenzo-Seva, U. (2017). Program FACTOR at 10: origins, development and future directions. *Psicothema*, 29(2), 236-241. [https://doi: 10.7334/psicothema2016.304](https://doi.org/10.7334/psicothema2016.304)
- Gana, K., and Broc, G. (2019). *Structural equation modeling with lavaan*. John Wiley & Sons.
- Gündeğer, C. (2023). E-değerlendirmeye yönelik tutum ölçeğinin madde yanlılığının incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 23(1), 249-271.
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- İnan, E. (2018). *Kırsal kesimlerdeki ilkokullarda Türkçe öğretiminde karşılaşılan güçlüklerin incelenmesi* [Yayımlanmamış doktora tezi]. Necmettin Erbakan Üniversitesi.
- Kane, M. T. (2006). Validation. *Educational measurement*, 4(2), 17-64.
- Kline, R. B., (2011). *Principles and practices of structural equation modelling*. New York: The Guilford Press.
- Kline, R. B. (2016). *Practice of principles of structural equation modeling*. Guilford.

- Meade, A. W., Johnson, E. C., and Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of applied psychology*, 93(3), 568.
- Meredith, W. (1993). MI, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Muthén, L.K., and B.O. Muthén. (2017). *Mplus user's guide*, 8th ed. Muthén & Muthén.
- Raju, N. S., Laffitte, L. J., and Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517.
- Schumacker, R. E., and Lomax, R. G. (2004). A beginner's guide to structural equation modeling. *Psychology Press*.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher* 36, 477-481.
- Şen, S. (2020). *Mplus ile yapısal eşitlik modellemesi uygulamaları*. Nobel Akademik Yayıncılık.
- Şenel, S. (2022) Measurement invariance of Turkish "central exam for secondary education" by specific learning disability. *Journal of Advanced Education Studies*, 4(2), 88-105.
- Urbina, S. (2014). *Essentials of psychological testing (2nd Edition)*. John Wiley & Sons
- Ülkü, S. (2019). *ABİDE 2016 Türkçe ve fen bilimleri alt-testlerinin öğretmen özelliklerine göre ölçme değişmezliğinin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi.
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics (5. Baskı)*. Pearson Education.

- Van de Schoot, R., Lugtig, P., and Hox, J. (2012). A checklist for testing measurement invariance. *European journal of developmental psychology*, 9(4), 486-492.
- 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.
- Wu, A., D., Li, Z., and Zumbo, B., D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: a demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1-26.