

Sınıf İçi Değerlendirme Uygulamalarında Adalet Algısı: Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçeye Uyarlanması*

Perceptions of Fairness in Classroom Assessment Practices: Adaptation of the Fair Classroom Assessment Scale into Turkish

Abdulkadir Baygöl, Serkan Buldur

Yazar Bilgileri

Abdulkadir Baygöl
Doktora Öğrencisi, Sivas Cumhuriyet Üniversitesi, Eğitim Bilimleri Enstitüsü,
kadirbygl@gmail.com

Serkan Buldur
Prof. Dr., Sivas Cumhuriyet Üniversitesi, Fen Bilgisi Eğitimi,
serkan.buldur@gmail.com

ÖZ

Bu çalışmada, sınıf içi değerlendirme süreçlerinde adalet algısını ölçmeyi hedefleyen Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçeye uyarlanarak ortaokul öğretmenleri örnekleminde geçerlik ve güvenirlik özelliklerinin incelenmesi amaçlanmıştır. Adalet, çağdaş ölçme ve değerlendirme uygulamalarının temel ilkelerinden biri olarak öne çıkmakta; ancak alanyazında bu kavramı bütüncül şekilde ele alan ölçme araçlarının sayısı sınırlı kalmaktadır. Sosyal adalet psikolojisi ve örgütsel adalet kuramına dayanan bu ölçek, dağıtımsal, prosedürel ve etkileşimsel adaleti kapsayan beş faktörlü bir yapıya sahiptir. Araştırmanın çalışma grubunu, 2024-2025 eğitim öğretim yılında bir ilin merkez ilçesinde görev yapan 712 ortaokul branş öğretmeni oluşturmaktadır. Veri toplama aracı olarak Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçeye uyarlanmış formu kullanılmıştır. Ölçeğin Türkçe uyarlaması, çeviri-tersine çeviri, uzman görüşü ve dil geçerliği uygulamalarını içeren çok aşamalı bir süreç izlenerek gerçekleştirilmiştir. Verilerin analizinde Doğrulayıcı Faktör Analizi, Cronbach alfa iç tutarlılık katsayıları, madde-toplam korelasyonları ve %27'lik alt-üst grup karşılaştırmaları yapılmıştır. Doğrulayıcı Faktör Analizi sonuçları, ölçeğin özgün beş faktörlü yapısının Türkçe formda da geçerli olduğunu göstermiş ve modelin uyum indekslerinin yeterli düzeyde olduğu belirlenmiştir. Ölçekteki faktörlerin iç tutarlılığına ilişkin hesaplanan Cronbach alfa katsayıları .76 ile .84 arasında değişmiş ve tüm alt boyutlar açısından kabul edilebilir düzeyde olduğu belirlenmiştir. Sonuç olarak Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçe formu geçerli ve güvenilir bir ölçme aracı olarak değerlendirilmiştir. Ölçek, öğretmen eğitimi, değerlendirme farkındalığı ve okul temelli politika geliştirme süreçlerinde kullanılabilecek işlevsel bir araç niteliğindedir.

Makale Bilgileri

Anahtar Kelimeler
Adalet algısı
Sınıf içi değerlendirme
Adil sınıf değerlendirme
Ölçek uyarlama

Keywords
Perception of fairness
Classroom assessment
Fair classroom assessment
Scale adaptation

Makale Geçmişi
Geliş: 11.08.2025
Kabul: 21.01.2026

ABSTRACT

The aim of this study is to adapt the Classroom Assessment Fairness Inventory into Turkish and to examine its validity and reliability among a sample of middle school teachers. Fairness is emphasized as one of the core principles of contemporary assessment practices; however, the number of instruments that address this construct comprehensively in the literature remains limited. This scale, grounded in social justice psychology and organizational justice theory, consists of five factors encompassing distributive, procedural, and interactional fairness. The study group consisted of 712 subject-matter middle school teachers working in the central district of a province in Türkiye during the 2024-2025 academic year. The Turkish-adapted version of the Classroom Assessment Fairness Scale was used as the data collection instrument. The adaptation process was conducted using a multi-step procedure including translation, back-translation, expert review, and language validity assessment. Data were analyzed using Confirmatory Factor Analysis, Cronbach's alpha internal consistency coefficients, item-total correlations, and 27% upper-lower group comparisons. The Confirmatory Factor Analysis results confirmed the original five-factor structure of the scale for the Turkish version, and the model fit indices indicated excellent fit. Cronbach's alpha coefficients ranged from .76 to .84, demonstrating an acceptable level of internal consistency for all sub-dimensions. In conclusion, the Turkish version of the Classroom Assessment Fairness Inventory was found to be a valid, reliable, and contextually appropriate tool for assessing fairness in classroom assessment. The scale offers functional utility in teacher education, assessment awareness, and school-based policy development.

Bu çalışma ikinci yazar danışmanlığında, birinci yazar tarafından hazırlanan doktora tezinden üretilmiştir.

Makale Türü

Araştırma

Önerilen Atıf Baygöl, A. & Buldur, S. (2026). Sınıf içi değerlendirme uygulamalarında adalet algısı: Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçeye uyarlanması. *TEBD*, 24(1), 550-576.
<https://doi.org/10.37217/tebd.1762846>

Giriş

Eğitimde ölçme ve değerlendirme süreçleri hem öğrenci başarısını belirlemede hem de öğretim kalitesini artırmada kritik rol oynar (Brookhart, 2011; DeLuca ve Bellara, 2013). Doğru, geçerli ve güvenilir ölçme araçları hem öğrencilerin akademik gelişimlerini izleme hem de öğretmenlerin öğretim stratejilerini düzenlemeleri açısından önemli işleve sahiptir (Brookhart, 2011; DeLuca ve Klinger, 2010; McMillan, 2014). Bununla birlikte değerlendirme süreçlerinin sadece teknik doğrulukla değil, aynı zamanda adalet ilkeleri doğrultusunda uygulanması, çağdaş eğitim anlayışında sosyal adaletin sağlanması açısından kritik öneme sahiptir (DeLuca vd., 2016; Rasooli vd., 2019; Tierney, 2013). Bu doğrultuda değerlendirmede adalete ilişkin algılar ve bunun öğrenciler üzerindeki etkileri önemlidir zira adil bulunmayan değerlendirme uygulamaları, öğrenme motivasyonu ve güven duygusunu zedeleyerek eğitimde eşitliği tehdit etmektedir (Chory, 2007).

Değerlendirme süreçlerindeki adalet algıları üzerine yapılan çalışmalar giderek daha fazla önem kazanmaktadır. Sınıf içi değerlendirme uygulamalarında öğrencilerin adalet algıları; motivasyon, öğrenmeye katılım, akademik öz-yeterlik ve öğretmene duyulan güven gibi pek çok değişken üzerinde doğrudan etkiye sahiptir (Resh ve Sabbagh, 2014). Ancak ilgili alanyazın incelendiğinde bu kavram çoğunlukla sınırlı ve tutarsız biçimde ele alınmış; adalet algısını değerlendirmeye yönelik geliştirilen ölçeklerin çoğu yalnızca birkaç maddeyle bu algıyı ölçmeye çalışmıştır (Alkharusi, 2011; Dorman ve Knightley, 2006). Her ne kadar adalet, değerlendirme algısının önemli bir bileşeni olarak kabul edilse de (Dorman ve Knightley, 2006; McMillan, 2016) öğrencilerin değerlendirme süreçlerindeki adalet algılarına odaklanan çalışmalar oldukça sınırlıdır. Alanyazındaki bu sınırlılık, değerlendirme adaletini çok boyutlu bir yapı olarak ele alan daha kapsamlı kuramsal çerçevelere duyulan ihtiyacı açıkça ortaya koymaktadır.

Değerlendirmede adalete yönelik algıları incelemek birçok açıdan önemlidir çünkü bu algılar, öğrencilerin değerlendirmeye yönelik tutumlarını, sürece katılımlarını ve öğrenme motivasyonlarını etkiler (Brown, 2021; McMillan, 2016; Struyven vd., 2005). Ayrıca öğretmene duyulan güveni ve öğretmenin otoritesini nasıl algıladıklarını da şekillendirir (Resh ve Sabbagh, 2014). Adil olmayan değerlendirme algısı, öğrencilerde olumsuz tutumlara yol açabilir (Brown, 2021; Daniels vd., 2014). Adalet algıları, öğretmenlerin değerlendirme uygulamalarını anlamada önemli bir göstergedir. Ancak tüm bu etkilerine rağmen, bu konu alanyazında hâlâ yeterince kuramsallaştırılmamış ve sistemli biçimde ölçülmemiştir (Rasooli vd., 2019; Tierney, 2013). Bu eksiklikleri gidermek amacıyla, son yıllarda değerlendirme uygulamalarındaki adalet algılarını daha bütüncül biçimde ele alan kuramsal yaklaşımlar ön plana çıkmıştır (Rasooli, 2021; Rezai vd., 2022).

Son yıllarda değerlendirme uygulamalarında adalet algılarını daha kapsamlı biçimde anlamaya yönelik sosyal adalet psikolojisi ve örgütsel adalet teorisi gibi kuramsal yaklaşımlara dayanan

çalışmalar önem kazanmıştır (McMillan, 2014; Rasooli, 2021; Rasooli vd., 2019; Rasooli vd., 2023b; Rezai vd., 2022; Tierney, 2013). Bu kapsamda üç temel adalet boyutu ön plana çıkarmaktadır: Dağıtım sal adalet, prosedürel adalet ve etkileşimsel adalet (Kazemi ve Törnblom, 2008). Bu adalet türlerinin sınıf içi değerlendirme süreçlerindeki yansımaları Tablo 1’de görülmektedir.

Tablo 1. Sınıf İçi Değerlendirme Süreçlerinde Adaletin Boyutları

<i>Adalet Türü</i>	<i>Tanımı</i>	<i>Kaynak(lar)</i>
Dağıtım sal Adalet	Değerlendirme sonuçlarının (örneğin notların veya geri bildirimlerin), öğrencilerin çabaları, başarı düzeyleri veya ihtiyaçlarına göre adil olarak dağıtılması.	Colquitt vd., 2001
Prosedürel Adalet	Değerlendirme sürecinin tutarlılık, tarafsızlık, açıklık ve öğrenci katılımı gibi ilkeler doğrultusunda adil şekilde yürütülüp yürütülmemesi.	Greenberg, 1993; Leventhal, 1980
Etkileşimsel Adalet	Öğretmen ile öğrenci arasındaki iletişimde saygı, empati ve açıklığın sağlanması.	Bies ve Moag, 1986

Bu üç adalet boyutu, öğrencilerin sınıf içi değerlendirme uygulamalarını nasıl algıladıklarını anlamada güçlü bir kuramsal çerçeve sunmakta; öğretmenlerin uygulamalarına ilişkin farkındalıklarını artırma potansiyeli taşımaktadır (Rasooli vd., 2019). Ancak bu kuramsal temele dayalı olarak geliştirilen ölçme araçlarının sayısı oldukça sınırlı olup mevcut araçların çoğu değerlendirme adaletini yalnızca belirli bağlamlar üzerinden ele almaktadır. Örneğin Dorman ve Knightley (2006) tarafından geliştirilen araçta öğrencilerin sınıf içi değerlendirme deneyimleri yalnızca “genel adalet algısı” düzeyinde ve sınırlı sayıda maddeyle ele alınmış; geri bildirim süreçleri, değerlendirme bağlamlarının çeşitliliği ya da öğretmen-öğrenci etkileşimi gibi kritik bileşenler kapsam dışı bırakılmıştır. Benzer şekilde Alkharusi (2011) tarafından kullanılan ölçek daha çok ölçme araçlarının adillğine yönelik tutumları değerlendirmeye odaklanmış ve değerlendirme sürecinin sosyal boyutları özellikle etkileşimsel adalet göz ardı edilmiştir.

Bu bağlamda hâlihazırda geliştirilen ölçeklerin çoğu adaletin yalnızca belirli boyutlarına odaklanmakta; örneğin yalnızca not verme ya da sınav gibi sınırlı alanları kapsamaktadır (Sonnleitner ve Kovacs, 2020). Oysa değerlendirme adaletine ilişkin algılar çok daha geniş bir bağlamı içerir. Grup çalışmaları, kopya çekme durumlarının ele alınışı, geri bildirim süreçlerinin şeffaflığı ve genel sınıf içi değerlendirme politikaları, adalet algısını etkileyen faktörler arasında yer almaktadır (Rasooli, 2021).

Bu çalışmada uyarlanan ölçek, Rasooli (2021) tarafından özgün formunda öğrenci algılarına dayalı olarak geliştirilmiş olmakla birlikte, içerik bakımından öğretmenlerin not verme, geri bildirim sunma, ölçüt belirleme ve değerlendirme kararlarına ilişkin adalet uygulamalarını hedefleyen yargılar içermektedir. Sınıf içi değerlendirmede adaletin dağıtım sal, işlemsel ve etkileşimsel boyutlarının doğrudan öğretmenlerin mesleki karar ve uygulamalarıyla yürütüldüğü dikkate alındığında öğretmenlerin bu uygulamaları adalet açısından yargılaması kuramsal olarak temellendirilebilir bir yaklaşımdır (Alm ve Colnerud, 2015; Gipps ve Stobart, 2009). Bu doğrultuda maddeler senaryo temelli bir yapı içinde sunularak öğretmenlerin ortak bir değerlendirme bağlamı üzerinden başka bir

öğretmenin uygulamalarını (varsayımsal durumlar için) değerlendirmeleri sağlanmış; böylece öğretmenlerin yalnızca kişisel deneyimlerine değil, mesleki normlara dayalı adalet anlayışlarına dayalı yargılarının belirlenmesi hedeflenmiştir (Bazvand ve Rasooli, 2022; Herman ve Cook, 2019).

Araştırmanın Önemi ve Amacı

Eğitim ortamlarında adil değerlendirme uygulamaları, yalnızca akademik performansın ölçülmesiyle sınırlı kalmayıp aynı zamanda öğretmenlerin karar verme süreçlerini, sınıf yönetim anlayışlarını ve pedagojik yaklaşımlarını da yansıtmaktadır (DeLuca ve Bellara, 2013; Tierney, 2013). Bu nedenle, öğretmenlerin değerlendirme uygulamalarındaki adalet algılarını ölçen araçlar geliştirmek hem farkındalığı artırmak hem de değerlendirme sürecini iyileştirmek için önemlidir. Ancak mevcut alanyazın incelendiğinde değerlendirmede adaleti çok boyutlu bir yapıda ele alan ve öğretmen perspektifine odaklanan ölçeklerin oldukça sınırlı olduğu görülmektedir (Alkharusi, 2011; Dorman ve Knightley, 2006; Rezai vd., 2022; Sonnleitner ve Kovacs, 2020).

Örneğin Sonnleitner ve Kovacs (2020) tarafından geliştirilen ölçek, daha çok sınav ve not verme süreçlerinde algılanan adaleti ölçmekte; bu yönüyle öğretmen ve öğrenci arasındaki etkileşim, geri bildirim süreçleri ya da bağlamsal faktörleri göz ardı etmektedir. Benzer biçimde Rezai vd. (2022) tarafından geliştirilen ölçek, lisans öğrencilerinin değerlendirme uygulamalarına ilişkin algılarını ölçmekle birlikte katılımcıların bireysel deneyimleri olmasa bile belirli ifadelerle yanıt vermelerini zorunlu kılan doğrudan yargılara dayanmaktadır. Örneğin “Performansınız adil şekilde değerlendirildi mi?” gibi ifadeler, deneyim yaşamayan bireylerde ölçüm yanlılığına yol açabilmektedir. Oysa adalet gibi soyut bir kavramın, bireyin yaşadığı bağlamdan bağımsız olarak varsayılması, yanıtların geçerliliğini sınırlayabilir.

Önceki alanyazında sınıf içi değerlendirme süreçlerine ilişkin adalet algıları çoğunlukla sınav ve not verme gibi dar alanlarla sınırlı tutulmuş ve genellikle yalnızca birkaç maddeyle ölçülmüştür (Chory-Assad ve Paulsel, 2007; Dalbert ve Stoeber, 2002; Sonnleitner ve Kovacs, 2020). Bu yaklaşımlar, geri bildirim, grup çalışmaları ve kopya çekme gibi farklı değerlendirme bağlamlarında öğrencilerin adalet algılarını dikkate almamış ve adaletin çok boyutlu doğasını yansıtmada yetersiz kalmıştır (Rasooli, 2021; Rasooli vd., 2019). Bu durumu özetleyen bir derleme, değerlendirme algılarında adaletin nasıl kavramsallaştırıldığını ve mevcut ölçeklerde ne şekilde ölçüldüğünü göstermektedir. Tablo 2, çeşitli araştırmalarda kullanılan araçlarını, odaklandıkları adalet boyutlarını ve örnek madde içeriklerini karşılaştırmalı olarak sunmaktadır.

Tablo 2. Değerlendirme Algılarında Adaletin Kavramsallaştırılması ve Ölçülmesi

<i>Araştırmacılar</i>	<i>Kullanılan Ölçek</i>	<i>Boyut</i>	<i>Adalet Maddesi Sayısı ve Örneği</i>
Brown, 2021	Değerlendirme Anlayışları	Değerlendirme Önemsizdir	Tek madde (“Değerlendirme öğrenciler için adil değildir.”)
Alkharusi, 2011	Algılanan Sınıf Değerlendirme Ortamı	Performansa Yönelimli Sınıf İçi Değerlendirme Ortamı	İki madde (örn: “Bu derste değerlendirme sonuçları, harcanan çabayı adil biçimde yansıtmıyor.”)
Dorman ve Knightley, 2006	Değerlendirme Görevlerine İlişkin Algılar Envanteri	Planlanan Öğrenme ile Uyumlu Olma	Üç madde (örn: “Fen dersindeki değerlendirme, çalışmalarımın adil bir göstergesidir.”)
Flores vd., 2015	Adalet Algıları	Öğrenci Merkezli Yöntemlerin Adaletine Yönelik Algılar	İki madde (örn: “Portfolyo gibi öğrenci merkezli değerlendirme yöntemleri geleneksel yöntemlere göre daha adildir.”)

Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2023b). Classroom Assessment Fairness Inventory: A new instrument to support perceived fairness in classroom assessment. *Assessment in Education: Principles, Policy & Practice*, 30(5–6), 372–395. <https://doi.org/10.1080/0969594X.2023.2255936> kaynağından uyarlanmıştır.

Ayrıca mevcut araçlar çoğu zaman katılımcılardan doğrudan deneyimledikleri durumlara ilişkin yargılar bildirmelerini beklemekte, bu da ölçüm geçerliğini sınırlamaktadır. Bu noktada, varsayımsal durumlar içeren senaryo tabanlı ölçme yaklaşımları değerlendirme bağlamlarını daha gerçekçi ve işlevsel biçimde ortaya koyarak hem öğretmen hem öğrenci perspektiflerinden adalet algısının daha detaylı biçimde anlaşılmasına olanak tanımaktadır (Rasooli, 2021; Sonnleitner ve Kovacs, 2020). Tüm bu gerekçeler, güncel kuramsal temellere dayalı, psikometrik açıdan geçerli ve bağlamsal farklılıkları içeren yeni ölçeklere duyulan ihtiyacı açıkça ortaya koymaktadır (Rasooli vd., 2018).

Bu noktada Rasooli (2021) tarafından geliştirilen Adil Sınıf Değerlendirmesi Ölçeği (ASDÖ), senaryo tabanlı yapısıyla alanyazındaki bu sınırlılıklara önemli bir alternatif sunmaktadır. Katılımcılara gerçekçi sınıf içi durumları içeren kısa senaryolar sunularak bu senaryolardaki değerlendirme uygulamalarını ne derece adil buldukları sorulmaktadır. Bu yaklaşım, hem bağlam temelli yanıtlar alınmasını sağlamak hem de kişisel deneyim gerektirmediği için ölçümün genellenebilirliğini artırmaktadır. Özellikle sınıfta yaşanmamış bir olay hakkında bile öğretmenin muhtemel tutumu ölçülebilmekte, böylece değerlendirme ilkelerine dair bilişsel yapılar açığa çıkarılmaktadır. Bu da öğretmenlerin değerlendirme uygulamalarında hangi adalet ilkelerini benimsediklerini daha iyi analiz etmeyi sağlar (Rasooli, 2021).

Bu çalışmanın özgün katkılarından biri Rasooli (2021) tarafından geliştirilen senaryo tabanlı Adil Sınıf Değerlendirmesi Ölçeği'nin Türkçeye uyarlanarak ortaokul öğretmenleri örnekleminde geçerlik ve güvenilirlik analizlerinin ilk kez gerçekleştirilmiş olmasıdır. Türkiye’de sınıf içi değerlendirme süreçlerinde öğretmenlerin adalet algılarını ölçmeye yönelik kuramsal temelli ve çok boyutlu bir ölçme aracının bulunmaması, bu çalışmanın önemini daha da artırmaktadır. Böylece sadece alanyazına yeni bir araç kazandırılmakla kalmayıp aynı zamanda öğretmenlerin değerlendirme süreçlerine ilişkin adalet algılarının ulusal düzeyde sistemli biçimde incelenmesine imkân sağlayan

geçerli ve güvenilir bir araç ortaya konulmaktadır. Bu ölçme aracının, öğretmenlerin kendi uygulamalarını değerlendirmelerine, profesyonel gelişimlerine yön vermelerine ve eğitim ortamlarında daha adil değerlendirme iklimlerinin oluşturulmasına katkı sunma potansiyeline sahip olduğu düşünülmektedir. Yukarıda ifade edilen tüm gerekçeler ışığında bu çalışmanın amacı, Rasooli (2021) tarafından geliştirilen ASDÖ'nün ortaokul öğretmenleri örnekleminde Türkçe formunun bilimsel araştırmalar için gerekli geçerlik ve güvenirlik ölçütlerini sağlayıp sağlamadığını incelemektir.

Yöntem

Araştırmanın Deseni

Bir ölçme aracının Türkçeye uyarlanması ve psikometrik özelliklerinin incelenmesi amacıyla yürütülen bu araştırma bir ölçek uyarlama çalışmasıdır. Ölçek uyarlama çalışmaları, mevcut bir ölçme aracının farklı bir dile ve kültürel bağlama uygun hâle getirilerek, geçerlik ve güvenirlik özelliklerinin yeniden test edilmesini içeren sistematik araştırma süreçlerini ifade etmektedir (Hambleton ve Patsula, 1998). Bu yönüyle araştırma, veri toplama ve analiz süreci bakımından nicel bir araştırma olup, var olan durumu ortaya koymaya yönelik yapısıyla betimsel tarama yaklaşımı kapsamında ele alınmıştır.

Çalışma Grubu

Araştırmanın çalışma grubunu, 2024–2025 eğitim-öğretim yılı bahar döneminde bir ilin merkez ilçesinde görev yapan toplam 712 ortaokul öğretmeni oluşturmaktadır. Katılımcılar, erişilebilirlik ve gönüllülük esasına göre belirlenmiş olup çalışma grubunun oluşturulmasında kolay ulaşılabilir örnekleme yöntemi esas alınmıştır. Kolay ulaşılabilir örnekleme, araştırmacının erişebildiği, katılım sağlamaya istekli bireylerden oluşan ve uygulama kolaylığı sağlayan bir örnekleme türüdür (Büyüköztürk vd., 2018).

Katılımcıların cinsiyet, mesleki kıdem, eğitim durumu ve branşlarına ilişkin demografik dağılımları Tablo 3'te sunulmuştur.

Tablo 3. Katılımcıların Demografik Özellikleri

<i>Cinsiyet</i>	<i>f</i>	<i>%</i>
Kadın	423	59,4
Erkek	289	40,6
<i>Mesleki Kıdem</i>	<i>f</i>	<i>%</i>
1-5 Yıl	97	13,6
6-10 Yıl	108	15,2
11-15 Yıl	185	26,0
16-20 Yıl	145	20,4
20+ Yıl	177	24,8
<i>Eğitim Durumu</i>	<i>f</i>	<i>%</i>
Lisans	516	72,5
Tezsiz Yüksek Lisans	95	13,3
Tezli Yüksek Lisans	87	12,2
Doktora	14	2,0

<i>Branş</i>	<i>f</i>	<i>%</i>
Fen Bilimleri	102	14,3
Türkçe	108	15,2
Matematik	113	15,9
Sosyal Bilgiler	65	9,1
Din Kültürü ve Ahlak Bilgisi	76	10,7
İngilizce	70	9,8
Diğer	178	25,0
<i>Toplam</i>	<i>712</i>	<i>100</i>

Tablo 3 incelendiğinde öğretmenlerin %59,4'ünün kadın, %40,6'sının erkek olduğu görülmektedir. Kadın öğretmenlerin daha yoğunlukta olduğu dikkat çekmektedir. Mesleki kıdem açısından en büyük grubu %26 ile 11–15 yıl arası deneyime sahip öğretmenler oluşturmaktadır. Bu grubu %24,8 ile 20 yıl ve üzeri deneyime sahip öğretmenler takip etmektedir. Eğitim durumu incelendiğinde, öğretmenlerin büyük çoğunluğu (%72,5) lisans mezunudur. Yüksek lisans ve doktora mezunları daha sınırlı düzeydedir. Branşlara göre dağılımda en yoğun katılımın gözlemlendiği alanlar arasında Matematik (%15,9), Türkçe (%15,2) ve Fen Bilimleri (%14,3) öne çıkmaktadır.

Veri Toplama Araçları

Araştırmada, Rasooli (2021) tarafından geliştirilen ve sınıf içi değerlendirme süreçlerinde adalet algısını ölçmeye yönelik olarak tasarlanan “Classroom Assessment Fairness Inventory (CAFI)” adlı senaryo tabanlı ölçek kullanılmıştır. Ölçek, dağıtımsal, prosedürel ve etkileşimsel adalet olmak üzere üç temel kuramsal boyuta dayanmaktadır. Türkiye bağlamında ölçek, “Adil Sınıf Değerlendirmesi Ölçeği” (ASDÖ) adıyla kullanılmıştır.

ASDÖ, grup çalışması, sınav, not verme, kopya çekme ve geribildirim olmak üzere beş farklı senaryodan oluşmaktadır. Her senaryo, örgütsel ve sosyal adalet kuramlarına dayanan eşitlik, açıklık, şeffaflık gibi temel adalet ilkelerini yansıtan durumlar içermektedir. Katılımcılara beş senaryo doğrultusunda toplam 21 madde sunulmakta ve her madde, katılımcıların o durumu ne kadar adil bulduklarını belirtecekleri bir yargı içermektedir. Ölçekte yer alan maddeler, dağıtımsal, prosedürel ve etkileşimsel adalet olmak üzere üç temel alt boyutta yapılandırılmıştır. Katılımcılardan senaryoları okuyarak, altında yer alan maddeleri 7'li Likert tipi bir ölçek üzerinden değerlendirmeleri istenmiştir (1=Oldukça Adil Değil, 2=Adil Değil, 3=Biraz Adil Değil, 4=Ne Adil Ne de Adil Değil, 5=Biraz Adil, 6=Adil, 7=Oldukça Adil). Örneğin, “Kopya çekmede adalet” senaryosuna ait bir madde şu şekildedir: “Öğretmen kopya çeken herkesin cezalandırılması gerektiğine dikkat çekmek amacıyla öğrenciye puan olarak sıfır verdi.” Katılımcılar, bu ifadeyi senaryo bağlamında değerlendirerek adalet düzeyine ilişkin görüşlerini belirtmiştir. Ölçekte yer alan grup çalışmasında adaletsizlik ve dönüt vermede adaletsizlik alt boyutlarına ait maddeler, analiz öncesinde ters kodlanmıştır (recode). Bu işlem sonucunda, ölçekteki tüm maddeler için yüksek puanlar sınıf içi değerlendirme uygulamalarına ilişkin yüksek adalet algısını

temsil edecek biçimde düzenlenmiştir. Böylece adalet ve adaletsizlik yönlü alt boyutlar, aynı yönlü bir ölçüm mantığı içinde değerlendirilebilir hâle getirilmiştir. Ölçekten elde edilen puanlar, sınıf içi değerlendirme uygulamalarının nesnel ya da normatif olarak adil olup olmadığını belirlemeyi değil, öğretmenlerin bu uygulamaları nasıl algıladıklarını ortaya koymayı amaçlamaktadır (Rasooli, 2021). Bu nedenle bulgular, doğruluk-yanlışlık yargıları yerine algılanan adalet düzeyi çerçevesinde yorumlanmaktadır. Ölçekten alınabilecek toplam puanlar 7 ile 147 arasında değişmektedir. Buna göre yüksek toplam puan, öğretmenlerin sınıf içi değerlendirme uygulamalarını genel olarak daha adil algıladıklarını; düşük toplam puan ise değerlendirme uygulamalarına yönelik adalet algısının daha düşük düzeyde olduğunu göstermektedir. Alt boyut puanları ilgili değerlendirme alanına özgü adalet algılarını yansıtırken, toplam puan katılımcıların sınıf içi değerlendirme süreçlerine ilişkin genel adalet algısını ifade etmektedir.

Verilerin Analizi

Verilerin analizinde, ölçeğin dil geçerliliğini değerlendirmek amacıyla İngilizce eğitimi alan bir lisans programında öğrenim gören ve İngilizce düzeyleri B1 olan öğrencilere bir hafta arayla önce İngilizce sonra Türkçe formlar uygulanmış, elde edilen puanlar arasındaki korelasyon değerleri incelenmiştir. Ayrıca uzman görüşleri doğrultusunda maddelerin anlaşılabilirliği ve kültürel uygunluğu kontrol edilmiştir. Ardından, veri setine ilişkin betimsel istatistikler (madde ortalamaları, standart sapmalar, çarpıklık ve basıklık değerleri) hesaplanmıştır. Analizler, 2024–2025 eğitim-öğretim yılı bahar döneminde 712 ortaokul öğretmeninden elde edilen veriler üzerinden gerçekleştirilmiştir. Ölçeğin yapı geçerliliğini test etmek amacıyla LISREL programı kullanılarak Doğrulayıcı Faktör Analizi (DFA) uygulanmıştır. DFA, ölçeğin özgün formunda yer alan beş faktörlü yapının Türkçe formda da doğrulanabilirliğini test etmek amacıyla yürütülmüştür. Modelin uyumuna ilişkin değerlendirmelerde; χ^2/sd , RMSEA, CFI, GFI, SRMR gibi uyum indekslerinden yararlanılmıştır. Ayrıca faktör düzeyinde güvenilirlik ve yakınsak geçerlik kanıtları sunmak amacıyla bileşik güvenilirlik (CR) ve ortalama açıklanan varyans (AVE) değerleri hesaplanmıştır.

Cronbach alfa iç tutarlılık katsayıları ve madde-toplam korelasyonları, ölçeğin güvenilirliğini ve madde ayırt ediciliğini belirlemek amacıyla SPSS 25 programı kullanılarak hesaplanmıştır. Güvenirlik analizleri hem genel ölçek hem de alt faktörler için ayrı ayrı yapılmıştır. Maddelere ait .30 ve üzeri madde-toplam korelasyon katsayıları ayırt edici kabul edilmiştir. Buna ek olarak ölçek maddelerinin cinsiyete göre diferansiyel madde işleyişi (DIF) gösterip göstermediğini belirlemek amacıyla lojistik regresyon temelli DIF analizleri gerçekleştirilmiş ve bu analizler SPSS programı kullanılarak yürütülmüştür. DIF değerlendirmesinde, istatistiksel anlamlılıkla birlikte etki büyüklüğü ölçütleri (ΔR^2) dikkate alınmıştır.

Ölçeğin farklı gruplar arasında eşdeğer biçimde çalışıp çalışmadığını incelemek amacıyla ölçme değişmezliği (MG-DFA) analizleri RStudio ortamında gerçekleştirilmiştir. Analizler, çok gruplu doğrulayıcı faktör analizi yaklaşımı kullanılarak yürütülmüş ve ordinal veri yapısına uygun olarak WLSMV kestirim yöntemi tercih edilmiştir. Bu kapsamda cinsiyet değişkeni için sırasıyla konfigüral, metrik ve skaler değişmezlik modelleri test edilmiştir. Modeller arası karşılaştırmalarda, örneklem büyüklüğüne duyarlılığı daha düşük ve daha güvenilir olduğu kabul edilen ΔCFI ve $\Delta RMSEA$ ölçütleri esas alınmıştır.

Mesleki kıdem (5 grup) ve branş (7 grup) değişkenleri ise çok kategorili yapıları ve bazı alt gruplarda sınırlı örneklem büyüklüklerinin bulunması nedeniyle MG-DFA kapsamı dışında tutulmuştur. Ölçme değişmezliği alanyazınında, grup sayısının artması ve küçük alt grup örneklemelerinin MG-DFA sonuçlarının kararlılığını ve yorumlanabilirliğini olumsuz etkileyebileceği vurgulanmaktadır (Putnick ve Bornstein, 2016). Bu doğrultuda, ölçme değişmezliği analizleri metodolojik açıdan daha sağlam olan ikili grup yapısı (cinsiyet) ile sınırlandırılmıştır.

Geçerlik ve Güvenirlik

Bu çalışmada, ölçeğin geçerlik ve güvenirlik düzeylerini belirlemek amacıyla çeşitli istatistiksel analizler gerçekleştirilmiştir. Yapı geçerliği, ölçeğin faktör yapısının Türkçe formda da korunup korunmadığını belirlemek amacıyla DFA ile incelenmiştir. DFA, özgün ölçekte yer alan beş faktörlü yapının Türkçe örnekleme doğrulanabilirliğini test etmeye yönelik olarak uygulanmıştır.

Güvenirlik çalışmaları kapsamında ise ölçeğin iç tutarlılığını belirlemek amacıyla Cronbach alfa katsayısı hesaplanmış, ayrıca her bir maddenin ölçeğe katkısını değerlendirmek üzere madde-toplam korelasyonları incelenmiştir. Bu analizler aracılığıyla ölçeğin alt boyutlarının ve genel yapısının istatistiksel açıdan tutarlı ve güvenilir olup olmadığı değerlendirilmiştir.

Ayrıca veri toplama sürecinde, ölçeğin geçerli ve güvenilir sonuçlar üretmesini sağlamak amacıyla bazı yöntemsel önlemler alınmıştır. Veriler, araştırmacı tarafından yüz yüze yöntemle toplanmıştır. Ölçek uygulaması, öğretmenlerin ders yükünün bulunmadığı ve daha rahat zaman ayırabildikleri seminer haftasında gerçekleştirilmiştir. Bu sayede katılımcıların dikkatle yanıt verebileceği uygun bir uygulama ortamı sağlanmıştır. Bununla birlikte ölçek maddeleri doğrudan kişisel yargular yerine senaryo temelli olarak hazırlanmıştır. Bu sayede öğretmenlerin değerlendirme uygulamalarına yönelik tutum ve algıları, daha doğal ve gerçekçi bir bağlamda ölçülmüştür. Senaryo kullanımı, sosyal beğenilirlik etkisini azaltarak katılımcıların daha dürüst yanıtlar vermesine yardımcı olmuştur. Bu tercih, özellikle adalet algısı gibi hassas konularda daha güvenilir veri elde edilmesini sağlamış ve klasik öz bildirim (self-report) formlarının sınırlılıklarını azaltmıştır.

Bununla birlikte yalnızca dönem içinde ders kapsamında aktif olarak ölçme ve değerlendirme uygulamaları yürüten öğretmenler çalışmaya dâhil edilmiştir. Bu kapsamda öğrencilerine yazılı sınav,

kısa sınav, proje, ödev gibi akademik başarıya yönelik değerlendirme araçlarını uygulayan branş öğretmenlerine odaklanılmıştır. Performansa dayalı değerlendirmelerin ağırlıklı olduğu (beden eğitimi, müzik, resim vb.) dersleri yürüten öğretmenler çalışma dışında bırakılmıştır. Formları yanıtlamadan önce katılımcılara bu kapsama ilişkin gerekli bilgilendirmeler yapılmış; böylece ölçek maddeleriyle doğrudan deneyim sahibi öğretmenlerden veri toplanması sağlanarak veri kalitesinin artırılması hedeflenmiştir.

Bulgular

Bu bölümde ASDÖ'nün geçerlik ve güvenilirliğine ilişkin analiz bulguları sunulmuştur.

Türkçe Formun Hazırlanması ve Dil Geçerliliği

Bu çalışmada kullanılan ASDÖ, Rasooli (2021) tarafından geliştirilmiş ve orijinal formu İngilizce olarak yayımlanmıştır. Ölçeğin Türkçeye uyarlanması süreci, çok aşamalı bir yöntem izlenerek gerçekleştirilmiştir.

İlk aşamada ölçek, ileri düzey İngilizce bilgisine sahip, ölçme ve değerlendirme alanında çalışmalar yapan bir akademisyen, bir İngilizce öğretmeni ve araştırmacılar tarafından bağımsız olarak (ayrı ayrı) Türkçeye çevrilmiştir. Ardından çeviri metinleri karşılaştırılarak anlam bütünlüğü ve kavramsal tutarlılık esas alınmış ve uzlaşılı yoluyla tek bir ortak Türkçe form oluşturulmuştur. Anlam eşdeğerliğinin sağlanması amacıyla, Türkçeye çevirisi yapılan form İngiliz Dili ve Edebiyatı Ana Bilim Dalında görevli bir akademisyenin katkılarıyla tekrar İngilizceye çevrilmiş (tersine çeviri) ve orijinal form ile karşılaştırılmıştır. Yapılan karşılaştırmalar sonucunda, ölçeğin kavramsal ve anlamsal düzeyde yüksek ölçüde örtüştüğü tespit edilmiştir.

Çeviri süreçlerinin hemen ardından, ölçeğin Türkçe formu kapsam geçerliliği ve dil uygunluğu açısından uzman görüşüne sunulmuştur. Bu kapsamda, ölçme ve değerlendirme alanında uzman bir akademisyen, yüksek lisans derecesine sahip üç öğretmen ve alanında deneyimli bir Türkçe öğretmeni tarafından içerik ve dil açısından değerlendirilen form, uzman görüşleri doğrultusunda gerekli düzenlemeler yapılarak dil eşdeğerliliği uygulamasına geçilmeden önce son hâline getirilmiştir.

İkinci aşamada, dil geçerliliğini test etmek amacıyla B1 düzeyinde İngilizce yeterliliğe sahip 35 lisans öğrencisine, ilk hafta ölçeğin orijinal (İngilizce) formu, ikinci hafta ise Türkçe formu uygulanmıştır. Uygulamalardan elde edilen veriler arasında yapılan korelasyon analizi sonucunda, Grup Çalışmasında Adaletsizlik boyutu için .84, Kopya Çekmede Adalet boyutu için .81, Not Vermede Adalet boyutu için .83, Dönüt Vermede Adalet boyutu için .80 ve Dönüt Vermede Adaletsizlik boyutu için .82 değerleri elde edilmiştir. Bu yüksek ve anlamlı korelasyon katsayıları, Türkçe formun anlam bütünlüğünü ve ölçüm yeterliliğini koruduğunu göstermektedir.

Pilot Uygulamaya İlişkin Betimsel Bulgular

Öğretmen örnekleminde elde edilen verilere ilişkin olarak ölçeğin geneli ve alt boyutlarına ait betimleyici istatistikler Tablo 4'te sunulmuştur.

Tablo 4. Ölçek Geneli ve Alt Boyutlara İlişkin Betimleyici İstatistikler

<i>Değişken</i>	<i>n</i>	<i>ort.</i>	<i>ss</i>
Grup Çalışmasında Adaletsizlik	712	5,63	1,42
Kopya Çekmede Adalet	712	5,39	1,21
Not Vermede Adalet	712	5,80	0,99
Dönüt Vermede Adalet	712	5,99	1,04
Dönüt Vermede Adaletsizlik	712	5,14	1,33
Ölçek Geneli	712	5,58	0,76

Tablo 4 incelendiğinde öğretmenlerin ölçek geneli ve tüm alt boyutlarda ortalama puanlarının orta noktanın üzerinde olduğu, bu durumun sınıf içi değerlendirme uygulamalarına ilişkin adalet algısının orta–yüksek düzeyde olduğunu gösterdiği görülmektedir.

Yapı Geçerliliğine İlişkin Bulgular

Uyarlama sürecinde, ölçeğin kuramsal temellere dayanan ve daha önce ampirik olarak test edilmiş bir faktör yapısına sahip olması nedeniyle DFA tercih edilmiştir. DFA, önceden belirlenmiş kuramsal yapının farklı kültür ve örneklerde geçerliliğini sınamak amacıyla kullanılan bir tekniktir. Bu çalışmada da mevcut ölçeğin faktör yapısının Türk örnekleminde doğruluğunu test etmek hedeflenmiş ve DFA bu amaçla uygun analiz yöntemi olarak değerlendirilmiştir (Brown, 2017; Tabachnick ve Fidell, 2013).

DFA, LISREL yazılımı aracılığıyla yürütülmüştür. Modelin yapısal geçerliliğini değerlendirmek amacıyla çeşitli uyum iyiliği indeksleri dikkate alınmıştır. Bu kapsamda, modelin veriyle olan uyum düzeyini değerlendirmek için genel kabul gören bazı istatistiksel göstergelerden yararlanılmıştır. Analiz bulguları, elde edilen yapının örneklem verileriyle anlamlı bir uyum sergilediğini göstermektedir. DFA'ya ilişkin ayrıntılı bulgular Tablo 5'te sunulmaktadır.

Tablo 5. ASDÖ'ye İlişkin DFA ile Hesaplanan Uyum İndisleri

<i>Uyum indisleri</i>	<i>Kabul Edilebilir Model Ölçütleri</i>	<i>Mükemmel Model Ölçütleri</i>	<i>1. DFA Sonuçları</i>	<i>2. DFA Sonuçları (Modifikasyon Sonrası)</i>
χ^2/sd	< 5	< 2	889/179=4,96	612/176=3,47
GFI	≥ .85	≥ .90	.89	.93
AGFI	≥ .85	≥ .90	.86	.90
NNFI	≥ .90	≥ .95	.95	.97
CFI	≥ .90	≥ .97	.96	.98
TLI	>.90	>.95	.95	.97
SRMR	≤ .08	≤ .05	.051	.047
RMSEA	≤ .08	≤ .05	.075	.058

DFA sürecinde, ölçeğin orijinalinde yer alan beş faktörlü yapı esas alınarak analizler gerçekleştirilmiştir. Faktör yapısının veriyle uyumunu değerlendirmek amacıyla çeşitli model uyum

indekslerinden yararlanılmıştır. İlk olarak, herhangi bir modifikasyon yapılmadan yürütülen analiz sonucunda elde edilen uyum indeksleri şu şekildedir: $\chi^2/sd=4,96$, GFI=.89, AGFI=.86, NNFI=.95, CFI=.96, SRMR=.051 ve RMSEA=.075. Bu değerler, alanyazında belirtilen kabul edilebilir sınırlar içinde yer almakta olup modelin genel olarak uygun bir yapıya sahip olduğunu göstermektedir (Çokluk vd., 2010; Seçer, 2015). Özellikle CFI ve NNFI değerlerinin .95 ve üzerinde olması, modelin yapısal geçerliliğini desteklemektedir.

Modelin uyumunu daha da iyileştirmek amacıyla analizde önerilen bazı modifikasyonlar değerlendirilmiştir. Bu doğrultuda, 9. ve 10. maddeler ile 18. ve 21. maddeler arasında hata kovaryansları tanımlanmıştır. Modifikasyon yapılan maddelerin her biri aynı alt boyut içinde yer almakta olup, içerik açısından birbirine benzer konuları ölçmektedir. Örneğin, 9. ve 10. maddeler etik değerlendirme ile ilgili öğrenciye sıfır verilmesi durumunu ele alırken 18. ve 21. maddeler öğretmenin dönüt verme süreciyle ilişkilidir. ([S_3_M9] "Öğretmenin sıfır verme nedeni kopya çekmenin etik açıdan yanlış olmasıdır."; [S_3_M10] "Öğretmen, kopya çeken herkesin cezalandırılması gerektiğine dikkat çekmek amacıyla öğrenciye sıfır puan vermiştir."; [S_5_M18] "Öğretmen, düşük performans gösteren öğrencilere görece daha fazla dönüt sağlamamıştır."; [S_6_M21] "Öğretmen, dönüt verme sürecinin her öğrencinin gösterdiği çabaya bağlı olduğunu açıklamıştır."). Bu modifikasyonlar, yalnızca istatistiksel değil aynı zamanda kavramsal gerekçelerle yapılmıştır. Bu sayede maddeler arasındaki ortak varyanslar modele yansıtılmış ve ölçüm hatalarının etkisi azaltılmıştır. Özellikle sosyal bilimlerde kullanılan ölçeklerde, benzer içerikli maddelerin belirli düzeyde ilişkili olması beklenmekte ve bu durum kabul edilebilir bir uygulama olarak değerlendirilmektedir (Byrne, 2010).

Modifikasyonlar sonrası yapılan ikinci DFA analizinde model uyumu daha da iyileşmiştir. Uyum indeksleri şu şekilde hesaplanmıştır: $\chi^2/sd=3,47$, GFI=.93, AGFI=.90, NNFI=.97, CFI=.98, SRMR=.047 ve RMSEA=.058. Bu sonuçlar, modelin iyi düzeyde uyum sağladığını göstermektedir.

Ölçeğin yapı geçerliğine ilişkin ek kanıtlar sunmak amacıyla, doğrulayıcı faktör analizi sonuçlarına dayalı olarak maddelere ait R² (Squared Multiple Correlations) değerleri ile faktör düzeyinde hesaplanan Bileşik Güvenirlik (CR) ve Ortalama Açıklanan Varyans (AVE) değerleri hesaplanmış ve Tablo 6'da sunulmuştur.

Tablo 6. ASDÖ'ye İlişkin Madde Düzeyi R² ve Faktör Düzeyi CR–AVE Değerleri

<i>Faktör</i>	<i>Madde</i>	<i>R²</i>	<i>CR</i>	<i>AVE</i>
Grup Çalışmasında Adaletsizlik	M_1	.35	.84	.64
	M_2	.22		
	M_3	.50		
Kopya Çekmede Adalet	M_4	.32	.88	.59
	M_5	.34		
	M_6	.29		
	M_7	.60		
	M_8	.50		

	M_9	.55		
	M_10	.56		
Not Vermede Adalet	M_11	.40	.85	.49
	M_12	.48		
	M_13	.62		
	M_14	.44		
	M_15	.36		
Dönüt Vermede Adalet	M_16	.44	.85	.65
	M_17	.26		
	M_18	.51		
Dönüt Vermede Adaletsizlik	M_19	.26	.82	.53
	M_20	.37		
	M_21	.73		

Tablo 6 incelendiğinde maddelere ait R^2 değerlerinin büyük bölümünün .30'un üzerinde olduğu görülmektedir. Bu durum, her bir maddenin ilgili faktör tarafından yeterli düzeyde açıklandığını göstermektedir (Hair vd., 2019). Faktör düzeyinde hesaplanan CR değerlerinin tüm boyutlarda .70'in üzerinde olması, ölçeğin yüksek iç tutarlılığa sahip olduğunu ortaya koymaktadır (Raykov, 1997). AVE değerlerinin çoğu faktörde .50 ve üzerinde olması, faktörlerin yakınsak geçerliğini desteklemekte; AVE'nin .50'nin biraz altında kaldığı boyutta ise CR değerinin .70'in üzerinde olması, bu boyut için de yakınsak geçerliğin kabul edilebilir olduğunu göstermektedir (Fornell ve Larcker, 1981).

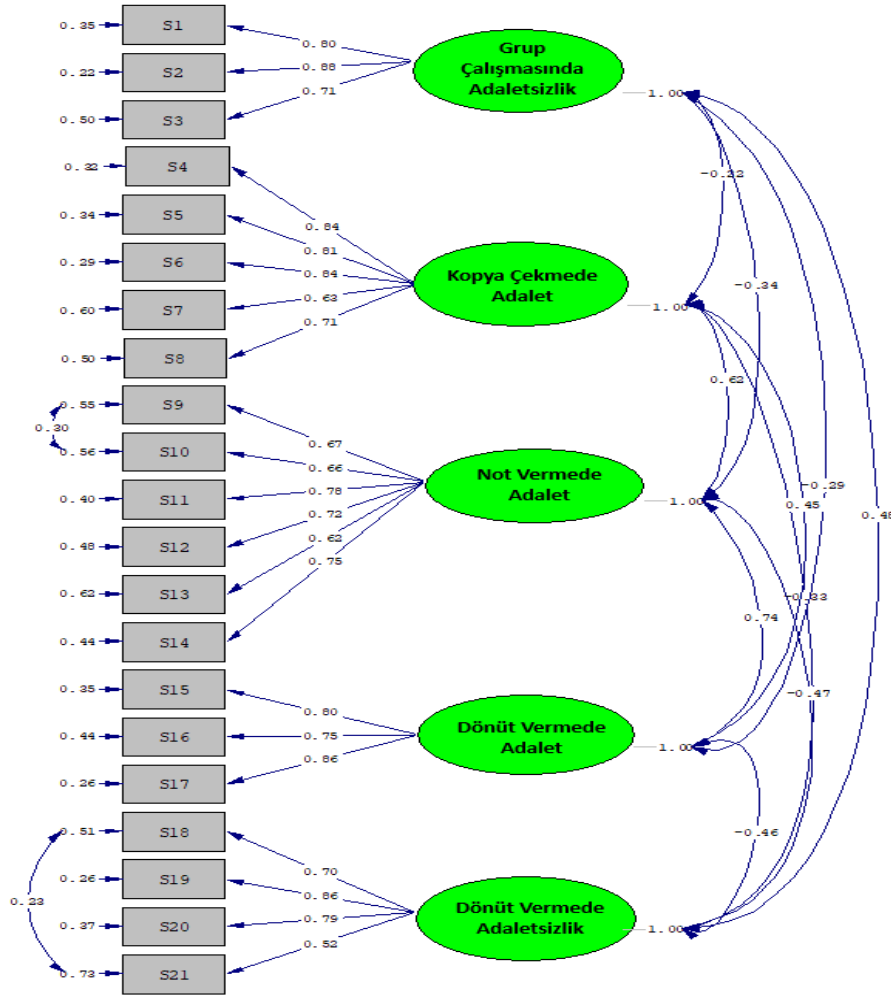
Sonuç olarak modifikasyon öncesi modelin de yeterli düzeyde geçerli olduğu, ancak yapılan teorik temelli düzenlemelerle modelin uyumunun belirgin şekilde iyileştiği görülmüştür. Ölçeğin Türkçe formu, orijinal beş faktörlü yapıyı korumuş ve yapı geçerliği sağlamıştır. Ölçeğe ilişkin standardize edilmiş faktör yükleri ve faktörler arası ilişkiler Şekil 1'de sunulmuştur.

Güvenirliliğe İlişkin Bulgular

ASDÖ'nün Türkçe formunun güvenirliliğini belirlemek amacıyla her bir alt boyuta ait Cronbach alfa (α) iç tutarlılık katsayıları hesaplanmıştır. Cronbach alfa katsayısı, maddelerin birbirleriyle olan tutarlılığını ve ölçülen yapıya olan güvenirliliğini ortaya koyar. Elde edilen sonuçlar Tablo 7'de sunulmuştur.

Tablo 7. ASDÖ Alt Boyutlarına Ait Cronbach Alfa (α) Katsayıları

<i>Alt Boyut Adı</i>	<i>Cronbach Alfa (α)</i>
Grup Çalışmasında Adaletsizlik	.81
Kopya Çekmede Adalet	.84
Not Vermede Adalet	.80
Dönüt Vermede Adalet	.76
Dönüt Vermede Adaletsizlik	.81
Ölçeğin Geneli	.85



Şekil 1. ASDÖ'ye ilişkin doğrulayıcı faktör analizine ait yol diyagramı

Tablo 7'de sunulan bulgulara göre ASDÖ'nün tüm alt boyutlarının Cronbach alfa (α) katsayılarının .70'in üzerinde olduğu görülmektedir. Bu durum, ölçeğin alt boyutlarının her birinin yeterli düzeyde iç tutarlılığa sahip olduğunu göstermektedir.

Bununla birlikte ölçeğin tüm maddeleri dikkate alınarak hesaplanan genel Cronbach alfa katsayısı .85 olarak bulunmuştur. Bu değer, ölçeğin genelinin de yüksek düzeyde iç tutarlılığa sahip olduğunu göstermekte; ölçeğin yalnızca alt boyut düzeyinde değil, bütünsel olarak da güvenilir bir yapı sunduğunu ortaya koymaktadır.

Sonuç olarak elde edilen bulgular ASDÖ'nün Türkçe formunun hem alt boyutlar hem de bütün ölçek düzeyinde tutarlı, güvenilir ve araştırmalarda kullanılabilir psikometrik özelliklere sahip bir ölçme aracı olduğunu göstermektedir.

Maddelerin Ayırt Edicilik Özelliğine İlişkin Bulgular

ASDÖ'nün ölçme yeterliğini belirlemek amacıyla her bir maddenin, ait olduğu boyuttaki bireyleri ne ölçüde ayırt edebildiğini ortaya koymak için madde-toplam korelasyon analizleri gerçekleştirilmiştir. Ayrıca ölçeğin ayırt ediciliğini desteklemek üzere, üst %27'lik ve alt %27'lik gruplar

(n=193) arasındaki puan farklarının istatistiksel olarak anlamlı olup olmadığı bağımsız örneklem için t-testi ile incelenmiştir. Bu analizlere ilişkin sonuçlar Tablo 8’de ayrıntılı biçimde sunulmuştur.

Tablo 8. ASDÖ’ye Ait Madde-Toplam Korelasyonları ve Üst %27–Alt %27 Grup t-Testi Sonuçları

<i>Boyut 1 Grup Çalışmasında Adaletsizlik</i>		
<i>Madde No</i>	<i>Madde-Toplam Korelasyonu</i>	<i>Maddeler için t Değeri (Üst %27-Alt %27)</i>
S1	.29	-7,93*
S2	.36	-6,93*
S3	.37	-4,72*
<i>Boyut 2 Kopya Çekmede Adalet</i>		
S4	.48	-14,68*
S5	.52	-13,10*
S6	.46	-16,90*
S7	.44	-12,46*
S8	.42	13,71*
<i>Boyut 3 Not Vermede Adalet</i>		
S9	.48	-11,53*
S10	.48	-13,84*
S11	.54	-10,12*
S12	.48	-9,88*
S13	.38	-10,62*
S14	.58	-9,16*
<i>Boyut 4 Dönüt Vermede Adalet</i>		
S15	.49	-7,44*
S16	.36	-8,80*
S17	.48	-8,39*
<i>Boyut 5 Dönüt Vermede Adaletsizlik</i>		
S18	.43	-5,54*
S19	.47	-3,54*
S20	.43	-4,53*
S21	.34	-7,07*

n=712 n₁=n₂=193 *p<.05

Tablo 8 incelendiğinde ASDÖ’ye ait tüm maddelerin madde-toplam korelasyon katsayılarının .29 ile .58 arasında değiştiği görülmektedir. Alanyazında madde-toplam korelasyonunun .30 ve üzeri olması, maddenin ölçtüğü kavrama katkı sağladığını ve ayırt ediciliğinin yeterli düzeyde olduğunu göstermektedir (Çokluk vd., 2010). Bu bağlamda, madde-toplam korelasyon değerleri ölçeğin genel olarak yüksek ayırt ediciliğe sahip olduğunu ortaya koymaktadır.

Bunun yanı sıra her bir madde için üst %27 ve alt %27’lik gruplar arasında yapılan bağımsız örneklem t-testi sonuçları incelendiğinde tüm maddeler için istatistiksel olarak anlamlı farklar olduğu görülmüştür (p<.05). Bu bulgular, ölçek maddelerinin her birinin bireyleri ayırt etmede istatistiksel olarak anlamlı düzeyde katkı sağladığını desteklemektedir. Sonuç olarak hem madde-toplam korelasyonları hem de gruplar arası t-testi sonuçları, ölçeğin maddelerinin yeterli düzeyde ayırt edici özellik taşıdığını ve geçerli ölçümler sunduğunu göstermektedir.

ASDÖ'nün Cinsiyete Göre Ölçme Değişmezliği

ASDÖ'nün kadın ve erkek öğretmen grupları arasında aynı yapıyı ölçüp ölçmediğini belirlemek amacıyla çoklu grup doğrulayıcı faktör analizi (MG-DFA) uygulanmıştır. Ölçme değişmezliği testleri, sırasıyla konfigüral, metrik ve skaler düzeylerde gerçekleştirilmiştir. Analizlerde ordinal veri yapısına uygun olarak WLSMV kestirim yöntemi kullanılmıştır. Model uyumunun değerlendirilmesinde CFI, TLI, RMSEA ve SRMR indeksleri raporlanmış; modeller arası karşılaştırmalarda ΔCFI ve $\Delta RMSEA$ değerleri esas alınmıştır. Ölçme değişmezliği için $|\Delta CFI| \leq .010$ ve $|\Delta RMSEA| \leq .015$ ölçütleri kabul edilmiştir (Chen, 2007). Cinsiyete göre ölçme değişmezliği (MG-DFA) sonuçları Tablo 9'da verilmiştir.

Tablo 9. Cinsiyete Göre ASDÖ'nün Ölçme Değişmezliği (MG-DFA) Sonuçları

<i>Model</i>	χ^2	<i>sd</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>	ΔCFI	$\Delta RMSEA$
Konfigüral	961,952	358	.989	.987	.069	.065	—	—
Metrik	1042,895	374	.988	.986	.071	.066	-.001	.002
Skaler	1025,533	454	.990	.990	.060	.065	.002	-.011
Katı	1025,533	454	.990	.990	.060	.065	.000	.000

Tablo 9'da sunulan bulgular incelendiğinde öncelikle konfigüral modelin her iki cinsiyet grubunda da kabul edilebilir uyum indeksleri sunduğu görülmektedir (CFI=.989, TLI=.987, RMSEA=.069). Bu sonuç, ölçeğin faktör yapısının kadın ve erkek öğretmen gruplarında aynı biçimde çalıştığını göstermektedir.

Konfigüral modele kıyasla faktör yüklerinin eşitlendiği metrik modelde uyum indekslerindeki değişimlerin önerilen eşik değerlerin altında kaldığı belirlenmiştir ($\Delta CFI = -.001$, $\Delta RMSEA = .002$). Bu bulgu, maddelerin ilgili faktörleri her iki grupta benzer biçimde temsil ettiğini ve faktör yüklerinin cinsiyete göre değişmez olduğunu göstermektedir.

Son olarak eşik değerlerin de eşitlendiği skaler modelin metrik modele kıyasla uyum indekslerinde anlamlı bir bozulma göstermediği görülmüştür ($\Delta CFI = .002$, $\Delta RMSEA = -.011$). Ayrıca katı (strict) modelde de uyum indekslerinin korunması, hata varyanslarının da cinsiyet grupları arasında değişmez olduğunu göstermektedir. Bu bulgular birlikte değerlendirildiğinde ASDÖ'nün cinsiyete göre konfigüral, metrik, skaler ve katı düzeylerde ölçme değişmezliğini sağladığı sonucuna ulaşılmıştır. Bu sonuçlar doğrultusunda ASDÖ'den elde edilen puanların kadın ve erkek öğretmen grupları arasında anlamlı ve adil karşılaştırmalara olanak tanıdığı söylenebilir.

Cinsiyete Göre Diferansiyel Madde İşleyişi (DIF) Analizleri

Ölçek maddelerinin cinsiyet değişkenine göre diferansiyel madde işleyişi (DIF) gösterip göstermediğini belirlemek amacıyla lojistik regresyon temelli DIF analizi yapılmıştır. Bu yaklaşım kapsamında uniform DIF toplam ölçek puanı kontrol edilerek grup değişkeninin modele eklenmesiyle, non-uniform DIF ise grup değişkeni ile toplam puan arasındaki etkileşim dikkate alınarak

değerlendirilmiştir. Analiz sonuçları, bazı maddelerde istatistiksel anlamlılık görülmesine karşın, etki büyüklüğü ölçütleri ($\Delta R^2 < .02$) dikkate alındığında hiçbir maddede pratik açıdan anlamlı bir uniform ya da non-uniform DIF bulunmadığını göstermiştir (Zumbo, 1999). Bu bulgular, ölçek maddelerinin cinsiyet grupları arasında eşdeğer biçimde işlediğini ve ölçekte DIF yokluğu bulunduğunu ortaya koymaktadır.

Sonuç ve Tartışma

Bu çalışmada, sınıf içi değerlendirme uygulamalarında adalet algılarını ölçmeye yönelik olarak geliştirilen ASDÖ ortaokul öğretmenleri örnekleminde Türkçeye uyarlanmış ve psikometrik özellikleri incelenmiştir. Elde edilen bulgular, ölçeğin geçerli ve güvenilir bir ölçme aracı olduğunu ortaya koymuştur. Ölçeğin beş faktörlü yapısı, kuramsal temele uygun biçimde yapılandırılmış ve DFA ile doğrulanmıştır. Modifikasyon öncesinde dahi kabul edilebilir düzeyde uyum indeksleri sunan model, kavramsal temele dayalı yapılan modifikasyonlar sonucunda daha güçlü bir yapıya ulaşmıştır. Ölçek, toplamda 21 madde ve beş alt boyuttan oluşmaktadır: "Grup Çalışmasında Adaletsizlik" (3 madde), "Kopya Çekmede Adalet" (5 madde), "Not Vermede Adalet" (6 madde), "Dönüt Vermede Adaletsizlik" (3 madde) ve "Dönüt Vermede Adalet" (4 madde). Ölçekten alınabilecek puan 21 ile 147 arasında değişmektedir. Güvenirlik analizleri sonucunda elde edilen Cronbach alfa katsayıları, hem genel ölçek (.85) hem de faktörler (.76-.84 arası) açısından iyi düzeydedir. Ayrıca hesaplanan madde-toplam korelasyonları ve üst-alt %27'lik gruplar arasında yapılan t-testi sonuçlarının anlamlı çıkması, ölçeğin madde düzeyinde ayırt edici ve işlevsel olduğunu göstermektedir. Ayrıca ölçekte cinsiyete göre anlamlı bir DIF bulgusunun elde edilmemesi, adalet algısının bireysel özelliklerden ziyade ortak normlar ve mesleki uygulamalara dayandığını savunan prosedürel ve etkileşimsel adalet kuramlarıyla tutarlıdır (Colquitt, 2001; Greenberg, 1993; Leventhal, 1980). Özellikle prosedürel ve etkileşimsel adalet yaklaşımları, değerlendirme süreçlerinin şeffaflık, tutarlılık ve saygı gibi evrensel ilkelere dayanması durumunda, farklı gruplar tarafından benzer biçimde algılanacağını öngörmektedir (Colquitt, 2001; Rasooli vd., 2023a). Bu bağlamda elde edilen DIF yokluğu bulgusu, ölçeğin adaletin çok boyutlu ve bağlamsal doğasını gruplar arası yanlılık oluşturmadan yansıtılabildiğini göstermektedir. Tüm bulgular doğrultusunda ASDÖ'nün Türkçe formu, eğitim araştırmalarında kullanılacak geçerli ve güvenilir bir ölçme aracı olarak değerlendirilmektedir.

ASDÖ'nün kuramsal temeli ve yapısal özellikleri değerlendirmede adaletle ilgili yürütülen benzer alanyazınla da uyumludur. Örneğin Murillo ve Hidalgo (2017, 2020) çalışmalarında değerlendirme süreçlerinin sadece teknik yeterlikle değil aynı zamanda bağlamsal ve sosyal adalet ilkeleriyle yapılandırılması gerektiğini vurgulamaktadırlar. Özellikle öğretmen-öğrenci etkileşimi, sürece katılım ve karar alma şeffaflığı gibi unsurların değerlendirmede adaletin temel bileşenleri olduğunu belirtmişlerdir. ASDÖ'nün temelli yapısı, Murillo ve Hidalgo'nun (2020) savunduğu gibi

bağlamı temsil eden, gerçekçi ve çok boyutlu değerlendirme durumlarını kapsamakta; böylece katılımcıların sadece bilişsel değil, etik ve pedagojik değerlendirme yargılarını da ortaya koymalarına olanak tanımaktadır. Ayrıca Murillo ve Hidalgo'nun (2017, 2020) öğretmen eğitimi programlarında adalet temelli ölçme yaklaşımlarının sistematik olarak yer alması gerektiğine dair önerileri, bu ölçeğin mesleki gelişim süreçlerinde bir farkındalık ve eğitim aracı olarak kullanılma potansiyelini desteklemektedir.

Ayrıca bu çalışmada elde edilen bulgular, öğretmenlerin değerlendirme süreçlerindeki adalet farkındalıklarını artırmaya yönelik geliştirilen diğer güncel yaklaşımlarla da örtüşmektedir. Özellikle Sonleitner ve Kovacs (2020) tarafından geliştirilen "Fairness Barometer (Adalet Barometresi)" adlı öz değerlendirme aracı, öğretmenlerin değerlendirme uygulamalarındaki adalet anlayışlarını öğrencilerin algılarıyla karşılaştırarak pedagojik öz farkındalıklarını geliştirmeyi amaçlamaktadır. Onların çalışmasında öğretmenlerin kendilerini çoğunlukla daha adil algıladıkları, ancak özellikle geri bildirim süreçleri ve sınav adaleti konularında öğrencilerle önemli görüş ayrılıkları yaşandığı ortaya konmuştur. Bu bulgular, ASDÖ'nün bağlamsal ve senaryo temelli yapısının, öğretmenlerin değerlendirme davranışlarını sadece teknik doğruluk açısından değil, aynı zamanda sosyal etkileşim ve süreç şeffaflığı açısından da değerlendirme fırsatı sunduğunu göstermektedir. Her iki çalışmada da amaç, öğretmenlerin değerlendirme süreçlerinde adalet ilkelerine dayalı daha bilinçli ve yansıtıcı bir yaklaşım geliştirmelerine katkı sağlamaktır. Bu bağlamda, söz konusu ölçek yalnızca bir ölçüm aracı değil, aynı zamanda öğretmen eğitimi ve mesleki gelişim programlarında kullanılacak işlevsel bir öğrenme aracıdır.

Çalışma bulguları, ölçme aracının özgün olarak geliştirildiği Rasooli'nin (2021) çalışmasıyla karşılaştırıldığında dikkat çekici benzerlik ve bazı üstünlükler barındırmaktadır. Rasooli (2021), özgün ölçeği Kanada'da 217 katılımcıya uygulamış ve beş faktörlü modelin geçerliliğini göstermiştir. Ancak bu çalışmada elde edilen bazı uyum indisleri (örneğin: CFI=.98 ve RMSEA=.058) değerleri, orijinal çalışmadaki değerlerden (CFI=.91; RMSEA=.06) daha yüksek model uyumu sergilemektedir. Bu farklılığın başlıca nedenlerinden birisi, örneklem büyüklüğü ve niteliği olabilir. Rasooli (2021) çalışmasını pandemi koşullarında, çevrim içi yöntemle ve sınırlı sayıda katılımcıyla gerçekleştirmiştir. Bu araştırma ise yüz yüze veri toplama yöntemiyle, doğrudan sınıf içi değerlendirme süreçlerini yürüten 712 ortaokul öğretmeni ile gerçekleştirilmiştir. Ölçek uygulaması, öğretmenlerin seminer haftasında, yani ders yüklerinin olmadığı ve zaman açısından daha esnek oldukları bir dönemde yürütülmüştür. Bu sayede, katılımcıların ölçek maddelerini dikkatle ve kendi deneyimlerine dayanarak yanıtlamaları sağlanmış; böylece hem veri kalitesi yükseltilmiş hem de ölçüm geçerliliği artırılmıştır. Sonuç olarak bu çalışmanın örneklem niteliği ve veri toplama koşulları, elde edilen psikometrik göstergelerin güvenilirliğini ve yorumlanabilirliğini önemli ölçüde desteklemektedir.

Çalışmada uyarlanan ASDÖ ile öğretmenlerin değerlendirmede adalet algılarının belirlenebilmesi de bu çalışmanın önemli avantajlarından birisidir. Çünkü öğretmenler, sınıf içi ölçme ve değerlendirme süreçlerini planlayan ve uygulayan kişilerdir. Bu nedenle değerlendirme uygulamalarına ilişkin adalet algılarının öğretmenler üzerinden incelenmesi anlamlıdır (Brookhart, 2011; Tierney, 2013). Bu yaklaşım, değerlendirme ortamlarında öğretmen algılarının belirleyici rolü olduğunu vurgulayan alanyazınla da uyumludur (Guskey, 2003; McMillan, 2014).

Sınırlılıklar ve Öneriler

ASDÖ'nün Türkçe formu hem kuramsal tutarlılığı hem de ampirik geçerliliği ile öğretmen örnekleminde kullanılabilir, güvenilir bir ölçme aracı olarak değerlendirilmiştir. Bu çalışmada, ASDÖ'nün Türkçe formunun geçerlik ve güvenilirlik analizleri yalnızca bir ilin merkez ilçesinde görev yapan ortaokul öğretmenlerinden elde edilen verilerle gerçekleştirilmiştir. Bu durum, ölçeğin farklı bölgeler, okul türleri ve sosyo-kültürel bağlamlarda nasıl işlediğine dair genellenebilirliği sınırlamaktadır. Ayrıca çalışma grubunun sadece ortaokul kademesinde görev yapan öğretmenlerden oluşması da bir diğer sınırlılıktır. Bu sınırlılıkların aşılması adına ölçeğin farklı öğretim kademelerindeki ve bölgelerdeki öğretmenlerle yürütülecek yeni çalışmalarda kullanılması önerilmektedir.

Farklı olarak bu çalışmada dil geçerliği, hedef kitle dışındaki ancak İngilizce yeterliliği yüksek lisans öğrencileriyle yürütülmüştür. Gelecek çalışmalarda, yeterli düzeyde İngilizce bilen öğretmenlerle yapılacak ek uygulamalar, dil eşdeğerliğine ilişkin bulguları güçlendirebilir. Ayrıca bu çalışmada, değerlendirme adaletiyle ilişkili olduğu düşünülen etik algı ya da ölçme okuryazarlığı gibi dış ölçütlerle kriter/örtüşen geçerlik analizleri gerçekleştirilememiştir. Bu durum, çalışmanın bir sınırlılığı olarak değerlendirilmekte olup gelecekte yapılacak araştırmalarda ilgili değişkenlerle yakınsak ve ayrışan geçerlik kanıtlarının sunulması önerilmektedir.

Ayrıca mesleki kıdem (5 grup) ve branş (7 grup) değişkenleri için MG-DFA uygulanamamıştır. Ölçme değişmezliği alanyazınında grup sayısının artması ve bazı alt gruplarda sınırlı örneklem büyüklüklerinin bulunması durumunda MG-DFA sonuçlarının kararsızlaşabileceği ve model karmaşıklığının önemli bir sorun hâline geldiği belirtilmektedir (Putnick ve Bornstein, 2016). Bu nedenle, ölçme değişmezliği analizleri metodolojik açıdan daha sağlam olan ikili grup yapısına sahip cinsiyet değişkeni ile sınırlandırılmıştır.

Diğer yandan çalışma sadece nicel verilerle sınırlandırılmış ve bu nedenle öğretmenlerin senaryo temelli maddelere verdikleri yanıtların altında yatan düşünme süreçleri detaylı biçimde analiz edilmemiştir. Nitel veri toplama yöntemleri (örneğin açık uçlu sorular, odak grup görüşmeleri veya bireysel mülakatlar) kullanılarak öğretmenlerin adalet algılarının daha derinlemesine incelenmesi, senaryo temelli yaklaşımın açıklayıcılığını artırabilir.

Kaynaklar

- Alkharusi, H. A. (2011). An analysis of the internal and external structure of the teacher assessment literacy questionnaire. *The International Journal of Learning*, 18(1), 515–528. <https://doi.org/10.18848/1447-9494/CGP/v18i01/47461>
- Alm, F. & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment*, 20(2), 132–150. <https://doi.org/10.1080/10627197.2015.1028620>
- Bazvand, A. D. & Rasooli, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation*, 72, 101118. <https://doi.org/10.1016/j.stueduc.2021.101118>
- Bies, R. J. & Moag, J. F. (1986). Interactional justice: Communication criteria of fairness. R. Lewicki, B. Sheppard, & M. Bazerman (Ed.), *Research on negotiation in organizations* (c. 1, s. 43–55).
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2018). *Eğitimde bilimsel araştırma yöntemleri* (25. b.). Pegem.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, G. T. L. (2017). *Assessment of student achievement*. Routledge.
- Brown, G. T. L. (2021). *Classroom assessment: Concepts and applications*. Routledge.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2. b.). Routledge.
- Chory, R. M. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education*, 56(1), 89–105. <https://doi.org/10.1080/03634520600994300>
- Chory-Assad, R. M. & Paulsel, M. L. (2007). Classroom justice: Student aggression and resistance as reactions to perceived unfairness. *Communication Education*, 53(3), 253–273. <https://doi.org/10.1080/0363452042000265189>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86(3), 425–445. <https://doi.org/10.1037/0021-9010.86.3.425>

- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Pegem.
- Dalbert, C., & Stoeber, J. (2002). Gerechtes Schulklima. J. Stoeber (Ed.), *Skalendokumentation "Persönliche Ziele von SchülerInnen"* (Hallesche Berichte zur Pädagogischen Psychologie(3), s. 32–34). Martin Luther Universität.
- Daniels, L. M., Poth, C., Papile, C., & Hutchison, M. (2014). Validating the conceptions of assessment-iii scale in canadian preservice teachers. *Educational Assessment*, 19(2), 139–158. <https://doi.org/10.1080/10627197.2014.903654>
- DeLuca, C. & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4), 356–372. <https://doi.org/10.1177/0022487113488144>
- DeLuca, C. & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. <https://doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- Dorman, J. P. & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, 32(1), 47–58. <https://doi.org/10.1080/03055690500415951>
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: A study in higher education. *Studies in Higher Education*, 40(9), 1523–1534. <https://doi.org/10.1080/03075079.2014.881348>
- Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* içinde (s. 79–103). Erlbaum.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6–11.
- Gipps, C. & Stobart, G. (2009). Fairness in assessment. C. Wyatt-Smith & J. J. Cumming (Ed.), *Educational assessment in the 21st century* içinde (s. 105–118). Springer.

- Hambleton, R. K. & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*(45), 153–171. <https://doi.org/10.1023/A:1006941729637>
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate data analysis* (8. b.). Cengage Learning.
- Herman, J. & Cook, L. (2019). Fairness in classroom assessment. S. M. Brookhart & J. H. McMillan (Ed.), *Classroom assessment and educational measurement* içinde (s. 243–264). Routledge/Taylor & Francis. <https://doi.org/10.4324/9780429507533-14>
- Kazemi, A. & Törnblom, K. Y. (2008). Social psychological perspectives on distributive and procedural justice theories. T. Törnblom & R. Vermunt (Ed.), *Distributive and procedural justice: Research and social applications* içinde (s. 15–50). Routledge.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. K. Gergen, M. Greenberg, & R. Willis (Ed.), *Social exchange: Advances in theory and research* içinde (s. 27–55). Springer. https://doi.org/10.1007/978-1-4613-3087-5_2
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6. b.). Pearson.
- McMillan, J. H. (2016). Section discussion: Students' perceptions of assessment. G. T. L. Brown & L. R. Harris (Ed.), *Handbook of human and social conditions in assessment* içinde (s. 237–260). Routledge.
- Murillo, F. J. & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation*, 53, 10–16. <https://doi.org/10.1016/j.stueduc.2017.01.001>
- Murillo, F. J. & Hidalgo, N. (2020). Fair student assessment: A phenomenographic study on teachers' conceptions. *Studies in Educational Evaluation*, 65, 100860. <https://doi.org/10.1016/j.stueduc.2020.100860>
- Putnick, D. L. & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rasooli, A. (2021). *Fairness in classroom assessment: Conceptual and empirical investigations*. (Doktora Tezi). <http://hdl.handle.net/1974/29060> sayfasından erişilmiştir.
- Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation*, 56, 164–181. <https://doi.org/10.1016/j.stueduc.2017.12.008>

- Rasooli, A., Zandi, H., & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: Exploring the value of organisational justice theory. *Assessment in Education: Principles, Policy & Practice*, 26(5), 584–611. <https://doi.org/10.1080/0969594x.2019.1593105>
- Rasooli, A., Zandi, H., & DeLuca, C. (2023a). Measuring fairness and justice in the classroom: a systematic review of instruments' validity evidence. *School Psychology Review*, 52(5), 639–664. <https://doi.org/10.1080/2372966X.2021.2000843>
- Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2023b). Classroom assessment fairness inventory: A new instrument to support perceived fairness in classroom assessment. *Assessment in Education: Principles, Policy & Practice*, 30(5–6), 372–395. <https://doi.org/10.1080/0969594X.2023.2255936>
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173–184. <https://doi.org/10.1177/01466216970212006>
- Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: Insights from Iranian university teachers. *Language Testing in Asia*, 12(8), 1–20. <https://doi.org/10.1186/s40468-022-00157-6>
- Resh, N. & Sabbagh, C. (2014). Justice, belonging and trust among Israeli middle school students. *British Educational Research Journal*, 40(6), 1036–1056. <https://doi.org/10.1002/berj.3129>
- Seçer, İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci (SPSS ve LISREL uygulamaları)*. Anı.
- Sonnleitner, P. & Kovacs, C. (2020). Differences between students' and teachers' fairness perceptions: Exploring the potential of a self-administered questionnaire to improve teachers' assessment practices. *Frontiers in Education*, 5, 17. <https://doi.org/10.3389/feduc.2020.00017>
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325–341. <https://doi.org/10.1080/02602930500099102>
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6. b.). Pearson.
- Tierney, R. (2013). Fairness in classroom assessment. J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment içinde* (s. 125–144). Sage.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.

Extended Summary

This study aimed to adapt the Classroom Assessment Fairness Inventory (CAFI), originally developed by Rasooli (2021), into Turkish and examine its psychometric validity and reliability in a sample of middle school teachers. In educational settings, assessment not only serves to measure

student achievement but also guides pedagogical practices, informs classroom management, and shapes the nature of teacher–student interaction. Therefore, structuring assessment procedures on the basis of fairness principles, rather than merely technical adequacy, is considered an essential component of contemporary teaching (DeLuca and Bellara, 2013; Tierney, 2013).

From a social justice perspective, fair assessment practices by teachers can directly influence students' motivation to learn, their trust in the assessment process, and their overall engagement in classroom learning (McMillan, 2016). Despite its importance, the literature reveals a limited number of comprehensive, valid, and reliable instruments that capture fairness in classroom assessment through a multidimensional lens (Sonnleitner and Kovacs, 2020). Existing tools primarily focus on narrow domains such as grading or test fairness, excluding key components like group work, feedback processes, ethical dilemmas, and teacher–student interactions (Alkharusi, 2011; Dorman and Knightley, 2006).

The CAFI, adapted in this study, is grounded in the theories of social justice psychology (Kazemi and Törnblom, 2008) and organizational justice (Colquitt et al., 2001). It integrates three dimensions of fairness: distributive, procedural, and interactional. The Turkish version of CAFI is based on Rasooli's (2021) scenario-based instrument, which aims to measure students' perceptions of fairness in classroom assessment contexts. The instrument includes 21 items organized around five contextual scenarios (e.g., group work, cheating, grading, feedback, and ethical dilemmas). Participants were asked to rate each scenario on a 7-point Likert scale ranging from 1 (Very Unfair) to 7 (Very Fair), thereby assessing fairness perceptions independent of personal experience (Rasooli, 2021).

The study sample consisted of 712 middle school teachers working in the central district of a province in Türkiye during the spring semester of the 2024–2025 academic year. Among the participants, 59.4% were female and 72.5% held a bachelor's degree. The most frequently represented subject areas were mathematics (15.9%), Turkish (15.2%), and science (14.3%). Participants were selected using convenience sampling, and the CAFI was administered face-to-face during the in-service week when teachers were free from teaching duties, enhancing attention to the items and overall data quality.

The translation and adaptation of the CAFI followed the linguistic and cultural guidelines proposed by Hambleton and Patsula (1998). Initially, expert translators fluent in English and Turkish prepared the forward translation. Subsequently, a reverse translation was conducted by an academic in English language and literature. Following this, the Turkish version was reviewed by experts in educational measurement, Turkish language education, and teacher training. Revisions were made based on their feedback. To assess linguistic validity, the instrument was administered to 35 preservice teachers with a B1 level of English—first in English and one week later in Turkish—yielding a strong correlation and confirming successful linguistic adaptation.

To assess construct validity, Confirmatory Factor Analysis (CFA) was performed using LISREL. The initial model yielded acceptable fit indices ($\chi^2/df=4.96$, CFI=.96, RMSEA=.075). After incorporating theoretically justified error covariances, model fit improved substantially ($\chi^2/df=3.47$, CFI=.98, RMSEA=.058). These results confirmed the five-factor structure of the original scale was retained in the Turkish sample, demonstrating satisfactory construct validity (Byrne, 2010; Çokluk et al., 2010).

The reliability of the instrument was evaluated using Cronbach's alpha coefficients, which ranged between .76 and .84 for the five subscales. The coefficient for the entire scale was calculated as .85, indicating strong internal consistency (Çokluk et al., 2010). Item-total correlations ranged from .29 to .58, and independent samples t-tests conducted between the upper and lower 27% groups for each item showed statistically significant differences ($p<.05$), confirming item discrimination and reliability.

The results indicated that most item-level R^2 values exceeded .30, suggesting that the items were adequately explained by their corresponding latent factors (Hair et al., 2019). Composite reliability (CR) values calculated at the factor level were above the recommended threshold of .70 for all dimensions, demonstrating high internal consistency of the scale (Raykov, 1997). In addition, average variance extracted (AVE) values were .50 or higher for most factors, supporting convergent validity. For the factor in which AVE was slightly below .50, a CR value exceeding .70 indicated that convergent validity could still be considered acceptable (Fornell and Larcker, 1981).

Measurement invariance across gender was examined using multi-group confirmatory factor analysis. The findings supported configural, metric, scalar, and strict invariance, as changes in model fit indices across successive models remained within recommended cutoff values. These results indicated that the factor structure, factor loadings, item intercepts, and error variances were equivalent for female and male teachers, allowing for meaningful group comparisons.

Furthermore, differential item functioning (DIF) was investigated using logistic regression. Although some items showed statistically significant effects, the associated effect sizes ($\Delta R^2<.02$) indicated no practically meaningful uniform or non-uniform DIF (Zumbo, 1999). Overall, these findings suggested that the scale operated equivalently across gender groups and provided valid, reliable, and fair measurement for comparative analyses.

Compared to the original study by Rasooli (2021), the present research utilized a larger and more diverse sample and implemented a face-to-face data collection procedure, both of which contributed to greater generalizability and data quality. The model fit indices in the Turkish adaptation even exceeded those in the original study, reflecting the robustness of the data and contextual applicability of the adapted instrument.

The CAFI also provides insights into which principles of justice teachers prioritize in assessment, and which pedagogical and ethical stances they adopt in different evaluative contexts. The

results align with Murillo and Hidalgo's (2020) emphasis on the contextual dimension of assessment fairness. When used in combination with tools like the Fairness Barometer (Sonnleitner and Kovacs, 2020), the CAFI serves as a strategic instrument for improving teacher awareness of justice-related practices.

In conclusion, the Turkish adaptation of the Classroom Assessment Fairness Inventory (CAFI) presents a theoretically grounded, psychometrically sound, and practically applicable tool. It enables the objective measurement of fairness perceptions in assessment practices and supports professional development, reflective teaching, and policy formulation. Future implementations of the instrument across different regions, educational levels, and sociocultural contexts, along with its integration into qualitative research, will further deepen the understanding of fairness in classroom assessment.

Furthermore, the importance of assessing fairness in educational contexts has been emphasized in recent studies focusing on both student and teacher perspectives. While most instruments are designed to capture fairness from the student's point of view, there remains a significant gap in assessing how teachers themselves perceive and implement fairness in classroom assessments. The CAFI bridges this gap by providing a structured, scenario-based tool that reflects real-world classroom situations. Its contextual richness enhances the interpretability and relevance of fairness judgments among educational professionals.

In adapting CAFI for Turkish educators, care has been taken to ensure that each scenario remains culturally meaningful and pedagogically appropriate, whilst preserving the integrity of the original items. Fair assessment is not merely a technical endeavor, but also an ethical commitment. Tools such as CAFI encourage equitable practices by enabling educators to critically evaluate their classroom decisions and strengthen trust among students. Such tools can play a significant role in teacher training programmes; policy design and the evaluation of educational reforms aimed at promoting justice and inclusivity.

Future research could examine the long-term impact of CAFI on teaching practices. For example, incorporating CAFI into ongoing professional development training could enhance teachers' self-awareness and consistency in applying fair assessment standards. Furthermore, applying this tool across different cultural and institutional contexts will strengthen its generalizability and reveal cross-national patterns in perceptions of equity.

Arařtırmacıların Katkı Oranı Beyanı

Bu arařtırmanın planlanması, yürütülmesi ve yazılı hale getirilmesinde arařtırmacılar eşit oranda katkı sağlamıştır.

Destek ve Teşekkür Beyanı

Bu çalışma, TÜBİTAK tarafından desteklenen 125K218 numaralı proje tarafından finanse edilmiştir. TÜBİTAK'ın sağladığı bu destek, araştırmanın planlanması, veri toplama ve analiz süreçlerinin yürütülmesinde önemli katkı sağlamıştır.

Çatışma Beyanı

Araştırmacıların araştırma ile ilgili diğer kişi ve kurumlarla herhangi bir kişisel ve finansal çıkar çatışması yoktur.

Etik Kurul Beyanı

Bu araştırma Sivas Cumhuriyet Üniversitesi Eğitim Bilimleri Araştırma Önerisi Etik Değerlendirme Kurulunun 30.05.2025 tarih ve E-50704946-050.04-569395 sayılı onayı ile yürütülmüştür.