

Öğrenci ve Okul Özelliklerinin Ortak Değişken Olduğu Test Eşitleme ile Ortak Maddeli Test Eşitlemenin Karşılaştırılması*

A Comparison of Test Equating with Student and School Characteristics as Covariates and Test Equating by Common Item

Elif Sezer Başaran¹, Mehtap Çakan²

¹Sorumlu Yazar, Dr. Öğr. Üyesi, Bursa Uludağ Üniversitesi, elifszr@gmail.com, (<https://orcid.org/0000-0002-7302-2724>)

²Prof. Dr., Gazi Üniversitesi, cakanmehtap@hotmail.com, (<https://orcid.org/0000-0001-6602-6180>)

Geliş Tarihi: 27.12.2025

Kabul Tarihi: 04.09.2026

ÖZ

Bu araştırmada, TIMSS 2019 sekizinci sınıf matematik testi verileri kullanılarak farklı test eşitleme desenlerinin [denk olmayan gruplarda ortak değişkenli (NEC) ve ortak maddeli (NEAT)] performansları, Kernel eşitleme yöntemleri (lineer ve eşit yüzdelikli) aracılığıyla karşılaştırılmıştır. NEC deseninde öğrenci ve okul değişkenleri arasından fen başarısı, evdeki eğitim kaynakları, matematiğe karşı özgüven, matematik öğrenmeyi sevme, okulun akademik başarıya verdiği önem ve sosyoekonomik düzey ortak değişken olarak seçilmiştir. Bulgular, Kernel lineer ve Kernel eşit yüzdelikli yöntemlerin benzer sonuçlar verdiğini ve ortak değişkenlerin matematik başarısıyla olan ilişki düzeyine göre eşitleme hatalarının farklılaştığını göstermiştir. Matematik başarısıyla yüksek düzeyde korelasyona sahip fen başarısının kullanıldığı NEC deseninin, NEAT desenine (son tabakalama ve zincirleme eşitleme yöntemleri) daha yakın olduğu tespit edilmiştir. Zayıf korelasyonlu değişkenlerin (sosyoekonomik düzey ve okulun başarıya verdiği önem) kullanıldığı NEC desenlerinin NEAT deseninden farklılaştığı ve gruplar arasındaki yetenek farklılığını açıklamada yetersiz kaldığı bulunmuştur. Orta düzeyde ilişkili değişkenlerin kullanıldığı NEC desenlerinde ise sonuçların gruplara göre değişkenlik gösterdiği görülmüştür. Bu bulgular, test eşitlemede ortak değişken seçiminin eşitleme doğruluğunu önemli ölçüde etkilediğini göstermektedir.

Anahtar Kelimeler: Test eşitleme, Kernel eşitleme, ortak maddeli desen, ortak değişkenli desen, TIMSS 2019 matematik testi.

ABSTRACT

This study aimed to compare the performance of different test equating designs—nonequivalent groups with covariates (NEC), and nonequivalent groups with anchor test (NEAT)—through Kernel equating methods (linear and equipercentile) using the data from the TIMSS 2019 eighth-grade mathematics test. In the NEC design, student- and school-level variables such as science achievement, home educational resources, students' confidence in mathematics, liking learning mathematics, school's emphasis on academic success, and socioeconomic level were selected as common covariates. The findings revealed that Kernel linear and Kernel equipercentile methods produced similar results, and that equating errors

*Bu çalışma Prof. Dr. Mehtap Çakan danışmanlığında Dr. Öğr. Üyesi Elif Sezer Başaran tarafından yazılan "Farklı Ortak Değişkenlerle Test Eşitlemenin Ortak Maddeli Test Eşitlemeyle Karşılaştırılması" başlıklı doktora tezinden hareketle hazırlanmıştır.

varied depending on the strength of the correlation between the covariates and mathematics achievement. The NEC design using science achievement, which was highly correlated with mathematics achievement, produced results closer to the NEAT design (post-stratification and chained equating methods). Conversely, NEC designs employing weakly correlated variables (socioeconomic status and school emphasis on success) diverged from the NEAT design, proving insufficient in accounting for ability differences between groups. In NEC designs using moderately correlated variables, results were observed to vary across groups. These findings indicate that the selection of covariates plays a critical role in determining the accuracy of test equating results.

Keywords: Test equating, Kernel equating, anchor test design, covariates design, TIMSS 2019 mathematics test.

GİRİŞ

Eğitim sisteminde bireylerin yeterliklerini ve eksikliklerini belirlemek amacıyla uygulanan testler, öğrencilerin başarılarını ölçmek açısından önemli araçlardır. Bu puanlar yalnızca öğretim sürecine geribildirim sağlamakla kalmaz; aynı zamanda liseye, üniversiteye veya iş hayatına geçiş gibi yüksek riskli kararlarda da belirleyici rol oynar (Büyüköztürk, 2016; Özer-Özkan & Turan, 2021). Ancak farklı test formları veya uygulama zamanları söz konusu olduğunda, puanların doğrudan karşılaştırılması adil olmayabilir. Farklı testlerin içerik, güçlük ve psikometrik özelliklerinden doğan farklılıkların giderilmesi için test eşitleme yöntemleri kullanılmaktadır (Crocker & Algina, 1986; Kolen & Brennan, 2014; Livingston, 2014).

Test eşitleme, benzer içerik ve güçlükteki testler arasında puanların karşılaştırılabilirliğini sağlayan istatistiksel bir süreçtir (Kolen & Brennan, 2014). Bu amaçla klasik test kuramına dayalı yöntemler, madde tepki kuramı ve Bayes yöntemi gibi yöntemler geliştirilmiştir. Son yıllarda özellikle Kernel eşitleme, kesikli puan dağılımlarını sürekli dağılımlara dönüştürerek uygulanan esnek bir eşitleme yöntemi olarak öne çıkmaktadır (von Davier vd., 2004). Araştırmalar Kernel eşitlemenin bazı durumlarda klasik yöntemlere göre daha düşük eşitleme hatası ürettiğini göstermektedir (Akın-Arık, 2018; Mao vd., 2006; Sezer Başaran vd., 2023; von Davier vd., 2006).

Eşitleme sonuçlarını yalnızca eşitleme yöntemi değil, aynı zamanda kullanılan eşitleme deseni de etkilemektedir. Eşdeğer gruplar (equivalent groups design, EG), ortak maddeli denk olmayan gruplar (nonequivalent groups with anchor test design, NEAT) ve ortak değişkenli denk olmayan gruplar (nonequivalent groups with covariates design, NEC) desenleri bu alandaki başlıca desenler arasında yer almaktadır (Kolen & Brennan, 2014; von Davier et al., 2004). Eşitleme sürecinde NEAT deseninde gruplara verilen testlerdeki aynı maddelerden (ortak test), NEC deseninde gruplar arası farklılıkları gidermek için test dışı değişkenlerden (ortak değişken) yararlanılır. Bu değişkenler genellikle cinsiyet, yaş, sosyoekonomik düzey gibi bireysel ve çevresel faktörlerden oluşur (Gonzalez & Wiberg, 2017; Wiberg & Branberg, 2015). Özellikle yüksek riskli sınavlarda test güvenliği gibi nedenlerle ortak madde kullanılamayan durumlarda, ortak değişkenlerin kullanıldığı NEC deseni alternatif bir çözüm olarak tercih edilmektedir (Gonzalez & Wiberg, 2017; Wiberg & Branberg, 2015).

NEC deseninde, gruplar arası yetenek farklılıkları ortak değişkenler aracılığıyla kontrol edilmeye çalışılır. Ancak bu desenin başarısı, seçilen değişkenlerin test başarısıyla olan ilişkisinin gücüne bağlıdır. Alan yazında, NEC deseninde kullanılan ortak değişkenlerin seçimine ilişkin çok sayıda çalışma bulunmamakla birlikte, yapılan araştırmalarda cinsiyet, yaş, sosyoekonomik düzey, farklı bir derse ait başarı (okuma veya matematik), matematik özyeterlik puanları, eğitim düzeyi, bilgisayara/tablete sahip olup olmama durumu, evdeki eğitim kaynakları gibi farklı değişkenlerin ortak değişken olarak daha çok kullanıldığı görülmektedir (Akın Arık, 2020; Albano & Wiberg, 2019; Branberg & Wiberg, 2011; Özsoy, 2021; Sezer Başaran vd., 2023; Wiberg & Branberg, 2015; Yıldırım Seheryeli, 2022; Yurtçu vd., 2021). Ayrıca, bu değişkenlerin

test başarısı ile olan korelasyon düzeyinin, eşitleme hatasının büyüklüğünü etkileyebileceği vurgulanmaktadır (Sezer Başaran vd., 2023; Wiberg & Branberg, 2015; Yıldırım Seheryeli, 2022).

Bununla birlikte alan yazında NEAT ve NEC desenlerinin karşılaştırıldığı araştırmalarda farklı sonuçlara ulaşılmıştır. Bazı çalışmalar NEAT deseninin daha küçük eşitleme hatası verdiğini belirtirken (Branberg & Wiberg, 2011; Özsoy, 2021; Sezer Başaran vd., 2023; Wallin & Wiberg, 2016), bazı araştırmalar NEC deseninin NEAT desenine göre daha tutarlı sonuçlar sunduğunu göstermektedir (Akın-Arkan, 2020; Yurtçu vd., 2021; Wallin & Wiberg, 2019). Bazı araştırmalarda ise her iki desende de birbirine benzer hatalar elde edilmiştir (Akın-Arkan, 2020; Yurtçu vd., 2021). Bu iki deseni de içeren NEATNEC deseninde en doğru sonuçların elde edildiğini belirten araştırmalar da vardır (Albano & Wiberg, 2019; Lu & Guo, 2018; Wiberg & Branberg, 2015). Bu tutarsızlıkların gruplar arası yetenek farkı, ortak testin yetersizliği ve ortak değişkenlerin özelliklerinden kaynaklandığı belirtilmektedir (Albano & Wiberg, 2019; Laukaityte vd., 2026; Lu & Guo, 2018).

Gruplar arası yetenek farkı büyüdükçe NEC ve NEAT desenlerinin performansı farklılaşmaktadır. Lu ve Guo (2018), ortak testin yetersiz olduğu durumlarda (madde sayısının az olması) NEC deseninin daha doğru sonuçlar verdiğini; yetenek farkı büyük olduğunda NEAT deseninin avantaj sağladığını; yetenek farklılığı küçük olduğunda NEC ve NEAT sonuçlarının benzer olduğunu ortaya koymuştur. Albano ve Wiberg (2019) ise gruplar arası yetenek farkı arttığında ortak değişkenlere olan ihtiyacın arttığını vurgulamaktadır. Laukaityte ve arkadaşları (2026) da eğilim puanlarını kullandığı çalışmada gruplar arası yetenek farkı büyük olduğunda NEAT deseninin NEC desenine göre daha doğru sonuç verdiğini, ancak iki desen birlikte kullanıldığında daha düşük yanlılık ve standart hata elde edildiğini tespit etmiştir. Bu durum NEAT ve NEC desenlerinin birbirinin alternatifi değil, tamamlayıcısı olarak ele alınması gerektiğine işaret etmektedir. Ancak gerçek yaşamda eşitlenecek testlerde her zaman ortak madde bulunmayabilir; bu durumlarda NEC deseninde hangi ortak değişkenlerin kullanılabileceğinin incelenmesi önem kazanmaktadır.

Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) ve Uluslararası Öğrenci Değerlendirme Programı (PISA) gibi uluslararası sınavlar, zengin öğrenci ve okul değişkenleri sayesinde NEC deseninin kapsamlı biçimde incelenmesine olanak tanımaktadır. Ancak TIMSS verisiyle yapılan NEC temelli çalışmaların sınırlı olduğu ve özellikle ortak değişkenlerin korelasyon düzeyine göre sistematik karşılaştırmaların yeterince ele alınmadığı görülmektedir (Özsoy, 2021; Yıldırım Seheryeli, 2022). Bu araştırma, söz konusu boşluğu doldurmayı ve farklı korelasyon düzeyindeki değişken gruplarının NEC desenindeki eşitleme performansını NEAT deseniyle karşılaştırarak analiz etmeyi amaçlamaktadır. Araştırma sonuçlarının, eşitleme desenlerinin seçiminde ve NEC deseninde ortak değişken belirleme sürecinde yol gösterici olması beklenmektedir.

1.1. Araştırmanın Amacı

Bu araştırmanın amacı, TIMSS 2019 sekizinci sınıf matematik testlerinde Kernel eşitleme yöntemi kullanılarak gerçekleştirilen NEC ve NEAT desenlerindeki eşitleme hatalarını karşılaştırmaktır. NEC deseninde ortak değişken olarak öğrenci ve okul özelliklerinden yararlanılmıştır. Seçilen ortak değişkenler şunlardır: (i) fen bilgisi başarısı, (ii) evdeki eğitim kaynakları, (iii) matematiğe karşı özgüven, (iv) matematik öğrenmeyi sevmek, (v) okulun akademik başarıya verdiği önem ve (vi) sosyoekonomik düzey.

1.2. Araştırma Soruları

Araştırmada şu sorulara yanıt aranmıştır:

1. Matematik başarıyla yüksek düzeyde ilişkili değişkenler (fen1 ve fen5 puanları) kullanıldığında NEC desenindeki Kernel eşitleme hataları, NEAT desenine göre nasıl farklılaşmaktadır?
2. Matematik başarıyla orta düzeyde ilişkili değişkenler (matematiğe karşı özgüven, evdeki eğitim kaynakları) kullanıldığında NEC desenindeki Kernel eşitleme hataları, NEAT desenine kıyasla ne düzeyde değişmektedir?
3. Matematik başarıyla düşük düzeyde ilişkili değişkenler (matematik öğrenmeyi sevmek, okulun akademik başarıya verdiği önem, sosyoekonomik düzey) kullanıldığında NEC desenindeki Kernel eşitleme hataları, NEAT deseniyle karşılaştırıldığında nasıl bir farklılık göstermektedir?

YÖNTEM

2.1. Araştırmanın Modeli

Bu temel araştırmada, sekizinci sınıf düzeyinde öğrenci ve okul özelliklerinin matematik testlerinin Kernel eşitlemesindeki rolü incelenmiştir. En az hata veren eşitleme yöntemi, veri toplama deseni ve ortak değişkenler araştırılmıştır; bu yönüyle çalışmanın kuramsal bilgiye katkı sağlayacağı öngörülmüştür. Temel araştırmalarda mevcut kuramın altında yatan süreçlerin açıklığa kavuşturulması beklenir (Fraenkel vd., 2011), yani temel özellikler incelenerek mevcut kuram geliştirilir. Bu nedenle bu araştırma, temel araştırma desenine göre planlanmıştır.

2.2. Çalışma Grubu

TIMSS 2019 araştırması, hem kâğıt-kalem formunda hem de bilgisayar ortamında (e-TIMSS) 14 kitapçık kullanılarak gerçekleştirilmiştir. Bu araştırmanın çalışma grubunu e-TIMSS uygulamasına katılarak matematik dersine ait 1, 2, 13 ve 14. kitapçıkları ile ortak değişkenlere yanıt veren 6973 sekizinci sınıf öğrencisi oluşturmaktadır. Örneklem OECD üyesi 13 ülkeden seçilmiştir. Ortalama matematik puanı TIMSS ortalamasının üstünde olan ülkeler Amerika Birleşik Devletleri, İngiltere, Norveç, Kore, Finlandiya, Litvanya, Macaristan ve İsveç'tir. Ortalama puanı TIMSS ortalamasının altında olan ülkeler Türkiye, Şili, İtalya ve Fransa'dır. Ortalaması TIMSS ortalamasına eşit olan ülke ise Portekiz'dir. Her bir ülkenin öğrenci sayısı Tablo 1'de sunulmuştur.

Tablo 1

Kitapçıkları Yanıtlayan Öğrencilerin Ülkelere Göre Dağılımı

Ülke	Kitapçık 1	Kitapçık 2	Kitapçık 13	Kitapçık 14	Toplam Öğrenci Sayısı
ABD	349	328	390	390	1457
Finlandiya	159	154	195	158	666
Fransa	59	47	88	62	256
İngiltere	79	61	81	82	303
İsveç	84	53	130	101	368
İtalya	83	83	115	92	373

Kore	217	194	245	214	870
Litvanya	117	81	146	130	474
Macaristan	165	151	193	159	668
Norveç	57	65	96	78	296
Portekiz	95	82	137	117	431
Şili	78	52	96	66	292
Türkiye	113	108	154	144	519
Toplam	1655	1459	2066	1793	6973

2.3. Veri Toplama Araçları

Bu araştırmada TIMSS’in resmi internet sitesinde yayımlanan TIMSS 2019 araştırmasına ait ikincil veri kullanılmıştır (International Association for the Evaluation of Educational Achievement, 2021). Bu araştırmanın verisi, e-TIMSS uygulamasında sekizinci sınıf düzeyinde uygulanan matematik ve fen bilgisi başarı testleri, öğrenci ve okul anketlerinden elde edilmiştir. Bu araştırmada matematik başarı puanı olarak TIMSS ekibi tarafından hesaplanan plausible values (BSMMAT01–BSMMAT05) kullanılmıştır. Ortak değişkenlerde ise TIMSS tarafından hesaplanan kategorik değerler esas alınmıştır. Araştırmada kullanılan ortak değişkenler fen bilgisi başarı, evdeki eğitim kaynakları, matematiğe karşı özgüven, matematik öğrenmeyi sevme, okulun akademik başarıya verdiği önem ve sosyoekonomik düzeydir. Bu değişkenlerin özellikleri Ek 1’de sunulmuştur.

Eşitlemenin yapıldığı matematik başarı testi, hem konu alanı hem de bilişsel düzey dikkate alınarak hazırlanmıştır (Lindquist vd., 2017). Konu alanı bakımından öğrenci puanlarının %30’u sayılar, %30’u cebir, %20’si geometri, %20’si veri ve olasılıktan oluşmaktadır. Bilişsel düzeyler ise %35 bilme, %40 uygulama ve %25 akıl yürütme şeklinde dağıtılmıştır. Çoktan seçmeli, kısa yanıt gerektiren ve uzun yanıt gerektiren maddelerden oluşan bu testler toplam 14 kitapçık halinde hazırlanmıştır; bu çalışmada B1, B2, B13 ve B14 kitapçıkları kullanılmıştır. Kitapçık seçiminde uzun yanıtlı soru sayısının az ve eşitlenecek kitapçıklarda benzer dağılım göstermesi dikkate alınmıştır.

Veri analizinde çoktan seçmeli ve kısa yanıt gerektiren maddeler doğru yanıtlar için “1” ve yanlış yanıtlar için “0” olacak şekilde kodlanmıştır. Uzun yanıt gerektiren maddelerde kısmi puanlama uygulanmasına rağmen bu çalışmada ikili puanlama modeli benimsenmiştir. Bu nedenle yalnızca tamamen doğru yanıtlar “1”, diğer tüm yanıtlar “0” olarak değerlendirilmiştir. Kitapçıkları faktör yapısı MPLUS 7.4 programında WLSMV (weighted least squares mean and variance) parametre kestirim yöntemi ile doğrulayıcı faktör analizi (DFA) yapılarak incelenmiştir. DFA her kitapçık için ayrı ayrı yapılmış ve tüm ülkelerdeki öğrencilerin bu kitapçığa verdikleri yanıtlar tek bir veri setinde birleştirilerek bütünleşik bir örneklem oluşturulmuştur. Tek faktörlü modele ait uyum indekslerine bakıldığında B1 ve B14 kitapçıkları $RMSEA < 0,05$; $TLI > 0,95$ ve $CFI > 0,95$ değerleriyle veriye iyi uyum göstermektedir. B2 ve B13 kitapçıkları ise $RMSEA < 0,08$; $TLI > 0,90$ ve $CFI > 0,90$ değerleriyle veriye kabul edilebilir uyum göstermektedir. Kategorik veride kullanılan bir uyum iyiliği indeksi olan WRMR değeri tüm kitapçıklarda birin üzerindedir; ancak diğer indeksler kabul edilebilir veya iyi uyum aralığında olduğu için model uyumunun değerlendirilmesinde WRMR sorun oluşturmamaktadır.

2.4. Veri Analizi ve Test Eşitleme Süreci

Bu araştırmada TIMSS 2019 matematik testi puanları NEC ve NEAT desenlerinde Kernel eşitleme yöntemleriyle (lineer ve eşit yüzdelikli) eşitlenerek karşılaştırılmıştır. Eşitleme sürecinde yeni formlar (X testi) B1 ve B14 kitapçıkları, eski formlar (Y testi) ise B2 ve B13 kitapçıkları olarak belirlenmiştir. Bu doğrultuda Kernel eşitleme yöntemleri kullanılarak B14 (X1) testinden

alınan puanlar B13 (Y1) testi puanlarına, B1 (X2) testinden alınan puanlar B2 (Y2) testi puanlarına eşitlenmiştir.

Kernel eşitleme, gözlenen puan eşitleme yöntemleri arasında yer alır ve bu yöntemde kesikli puan dağılımları sürekli dağılımlara dönüştürüldükten sonra eşitleme yapılır (von Davier vd., 2004). Eşitleme fonksiyonunun kestiriminde sonlu parametreler ve dağılım fonksiyonları yer aldığı için Kernel eşitlemenin yarı-parametrik eşitleme kestirimine sahip olduğu düşünülebilir (Gonzalez & Wiberg, 2017). Kernel eşitleme süreci (i) ön düzgünleştirme, (ii) puan olasılıklarının kestirilmesi, (iii) süreklileştirme, (iv) eşitleme dönüşümünün hesaplanması ve (v) ölçüm doğruluğunun hesaplanması olmak üzere beş adım ile özetlenir (von Davier vd., 2004). Bu adımlar, eşitleme sürecinin sistematik bir biçimde uygulanmasına olanak sağlar.

Bu araştırmada ilk aşama olan ön düzgünleştirmede, gözlenen puan dağılımlarındaki düzensizlikleri gidermek amacıyla log-lineer modeller kullanılmıştır. Bu doğrultuda NEC ve NEAT desenleri için iki değişkenli polinomial modeller aracılığıyla düzgünleştirme yapılmıştır. Model uyumu; AIC, moment değerleri (aritmetik ortalama, standart sapma, çarpıklık, basıklık), Freeman–Tukey artık istatistikleri ve ki-kare olasılık oranı testi ile değerlendirilmiştir. NEAT deseninde, yapısal sıfır (structural zero) probleminden dolayı iç ortak test yerine dış ortak test kullanılmıştır. İç ortak testte, ortak madde puanları toplam test puanının bir alt kümesi olduğundan, ortak test puanının toplam puandan büyük olduğu hücreler mantıksal olarak imkânsızdır; bu durum log-lineer modellerin kestirdiği puan olasılıklarında yanlışlığa yol açabilmektedir (Holland & Thayer, 2000; von Davier vd., 2004). Örneğin bir X testinde toplam test puanı X, ortak test puanı A ve testteki diğer maddelerden alınan puan X^* olsun. Bu durumda $X = X^* + A$ olur. X^* en az sıfır olabileceği için ortak testten 5 puan ($A=5$) alınırsa toplam test puanı 5 ve üzerinde olmalıdır ($X \geq 5$). Yani iç ortak test kullanıldığında toplam test puanı hiçbir zaman ortak test puanından daha küçük olamayacaktır. Bu durum yapısal sıfır durumunun ortaya çıkmasına sebep olur. İç ortak testle eşitlemede ön düzgünleştirme aşamasında küçük de olsa bazı pozitif olasılıklar veya $X < A$ durumuna yol açan kestirimler elde edilebilir (Kim, 2014; von Davier vd., 2004). Ancak dış ortak testle eşitlemede yapısal sıfır durumu ortaya çıkmaz. İç ortak ve dış ortak testlerle yapılan eşitlemelerde sadece ön düzgünleştirmedeki hesaplamalar farklılaşmaktadır; diğer tüm aşamalardaki hesaplamalar aynıdır. Bu çalışmada da bu problemi önlemek amacıyla dış ortak test kullanılmıştır.

İkinci aşamada, bu düzgünleştirilmiş dağılımlar üzerinden marjinal puan olasılıkları (r ve s) hesaplanmıştır. NEAT desenindeki son tabakalama yönteminde sentetik evren oluşturulurken P ve Q evrenleri eşit ağırlıklandırılmıştır ($w = 0,5$). Üçüncü aşama olan süreklileştirmede Gauss Kernel fonksiyonu kullanılmıştır. Bant genişliği (h) parametresi eşitleme fonksiyonunun yapısını belirlediği için kritik bir rol oynamaktadır: ideal h değerleri ile eşit yüzdellikli, büyük h değerleri ile lineer fonksiyonlar elde edilir. Bu çalışmadaki h parametreleri, R yazılımındaki kequate paketi tarafından seçilmiş olup kullanılan değerler Ek 2’de verilmiştir.

Dördüncü aşamada eşitleme işlemleri yapılmıştır. Bu kapsamda NEC ve NEAT desenlerinde Kernel lineer ve Kernel eşit yüzdellikli yöntemler uygulanmıştır. NEAT deseninde dış ortak test aracılığıyla hem son tabakalama (post-stratification equating) hem de zincirleme (chained equating) eşitleme yöntemleri kullanılmıştır. Son tabakalama yönteminde X ve Y’nin ortak test puanları üzerindeki koşullu puan dağılımları kullanılarak hedef popülasyona (sentetik evren) ait marjinal puan dağılımları dolaylı olarak kestirilir ve eşitleme bu sentetik evren üzerinden gerçekleştirilir. Zincirleme eşitleme yönteminde ise sentetik bir evren tanımlanmaksızın, X testi önce ortak teste, ardından ortak test Y testine sırayla eşitlenerek iki aşamalı bir dönüşüm zinciri oluşturulur.

Eşitlemenin son aşamasında ise eşitleme doğruluğu incelenmiştir. Değerlendirme kriteri olarak NEAT (son tabakalama yöntemi) deseni belirlenmiştir. Ayrıca eşitleme sonuçlarının değerlendirilmesinde PRE (percent relative error; yüzde görelî hata), SEE (standard error of

equating; eşitlemenin standart hatası), RMSD (root mean squared difference) (Kim & Lu, 2018; Suh vd., 2009), WMSE (weighted mean square error) (Skaggs & Lissitz, 1986), RMSE (root mean squared error) (von Davier vd., 2006) ve fark grafikleri kullanılmıştır. Puan farklarının pratik anlamlılığı Suh ve arkadaşları (2009) tarafından önerilen “önemli fark” (difference that matters, DTM) eşiğine göre yorumlanmıştır. Bu çalışmada birim puan bir olduğu için DTM sınırı 0.5 olarak belirlenmiş ve yorumlamalar şu kriterlere dayandırılmıştır:

- *Tam anlamlılık*: RMSD veya fark, 0,50’ye eşit veya büyük (DTM vardır)
- *Marjinal anlamlılık*: RMSD veya fark, 0,05 ile 0,49 aralığında (DTM yoktur. Bu durum farkın ihmal edilebilir sınırlar dahilinde olduğunu göstermektedir.)
- *Anlamsızlık*: RMSD veya fark, 0,05’ten küçük (DTM yoktur. Bu durum eşitlenmiş puanların anlamlı bir şekilde farklılaşmadığını göstermektedir.)

Hesaplanan eşitleme hatalarından SEE, örneklem büyüklüğünden kaynaklanmakta ve eşitleme sonuçlarına karışan tesadüfi hatayı temsil etmektedir. RMSE ve RMSD, tesadüfi ve sistematik hatayı bir arada ele alarak eşitleme sürecindeki toplam hatayı, WMSE ise puan dağılımındaki öğrenci frekanslarını dikkate alarak hesaplanan ağırlıklandırılmış toplam hatayı ifade etmektedir. PRE ise eşitleme fonksiyonunun puan dağılımına ait momentlerin ne derecede korunduğunu değerlendiren göreceli hata indeksidir. Eşitleme hatalarına ait denklemler aşağıdaki gibidir. Bu denklemlerde yeni formdaki puanların göreceli dağılımı (w_i), X testindeki madde sayısı (k), X testindeki ham puanların varyansı (S_X^2), X testinde i. puana karşılık gelen ham puanı (X_{crit}), eşitlenmiş puanlar (X_E), X testindeki puanların frekansı (f_i), farkların ortalaması ($\bar{d}$) ve farkların standart sapması (sd) değerleri bulunmaktadır. SEE eşitliği ise eşitleme fonksiyonunun Jacobian matrisi (J_{eY}), desen fonksiyonunun Jacobian matrisi (J_{DF}) ve ön düzgünleştirmedeki puan dağılımlarının asimptotik kovaryans matrisi (C) olmak üzere üç parçadan oluşmaktadır (von Davier vd., 2004).

$$SEE = \|J_{\varphi} J_{DF} C\|$$

$$PRE(p) = 100 \frac{\mu_p(\varphi(X)) - \mu_p(Y)}{\mu_p(Y)}$$

$$RMSD = \sqrt{\sum_{i=0}^k w_i [\hat{e}_i(x_i) - e_i(x_i)]^2}$$

$$RMSE = \sqrt{\bar{d}^2 + sd_d^2}$$

$$WMSE = \frac{\sum_{i=1}^{k-1} f_i (X_E - X_{crit})^2}{\sum_{i=1}^k f_i S_X^2}$$

Analizler öncesinde kayıp veri incelenmiş, en az bir değişkende eksik yanıt olan öğrenciler listwise yöntemiyle veri setinden çıkarılmıştır. Kernel eşitleme analizleri R (4.2.2, RStudio 2022.07.2) ortamında kequate paketi (Andersson vd., 2013, 2022) ile yapılmıştır. Ortak değişkenler ile matematik başarısı arasındaki korelasyon hesabı için psych (Revelle, 2022) paketi; betimsel istatistikler ve eşitleme hatalarının hesaplanması ile fark grafiklerinin çizilmesinde SPSS 25.0 ve Microsoft Excel programlarından yararlanılmıştır. Matematik başarısı ile ortak değişkenler arasındaki korelasyonun yorumlanmasında Guilford (1956) tarafından önerilen ölçütler esas alınmıştır. Bu ölçütlere göre korelasyonun mutlak değeri 0,20’den düşükse çok zayıf; 0.20 ve 0.39 arasındaysa zayıf; 0.40 ve 0.69 arasındaysa orta; 0.70 ve 0,89 arasındaysa yüksek; 0.90 ve üstündeyse çok yüksek ilişki bulunmaktadır.

BULGULAR

Bu bölümde öncelikle iki öğrenci grubu için testlere ait betimsel istatistikler ve eşitleme için incelenen varsayımlar sunulmuştur. Daha sonra NEC ve NEAT desenlerinde matematik testlerinin Kernel eşitleme (lineer ve eşit yüzdelikli) sonuçları karşılaştırılmıştır.

3.1. Betimsel İstatistikler

Eşitleme uygulamalarında kullanılan X ve Y testleri ile ortak testlere (B1, B2, B13 ve B14) ait betimsel istatistikler Tablo 2’de sunulmuştur. Testlerdeki madde sayıları birbirinden farklı olduğu için ham puanlar, başarı yüzdesine dönüştürülerek analiz edilmiştir. Puan dağılımlarının çarpıklık ve basıklık değerleri incelendiğinde normal dağılım varsayımına uygun olmadığı görüldüğü için Mann Whitney U testi ile test puanlarının sıra ortalamaları arasındaki farklılık incelenmiştir (Örn. X1 ve Y1). Sonuçlar hem birinci hem de ikinci gruptaki testlere ait sıra ortalamaları arasındaki farklarının istatistiksel olarak anlamlı olduğunu ortaya koymuştur ($p < 0,05$). Ancak hesaplanan etki büyüklüğü (effect size) değerlerinin 0,10’un altında kalması, bu farkların pratik açıdan ihmal edilebilir düzeyde olduğunu kanıtlamaktadır. Ancak mevcut olan bu farkın, öğrencilerin ham puanlarının doğrudan karşılaştırılmasını engelleyerek bir test eşitleme uygulamasını gerekli kıldığını ortaya koymaktadır.

Tablo 2

Test Puanlarına Ait Betimsel İstatistikler

<i>İstatistik</i>	<i>Yeni Form (B14)</i>		<i>Eski Form (B13)</i>		<i>Yeni Form (B1)</i>		<i>Eski Form (B2)</i>	
	<i>X1</i>	<i>Ortak</i>	<i>Y1</i>	<i>Ortak</i>	<i>X2</i>	<i>Ortak</i>	<i>Y2</i>	<i>Ortak Test</i>
	<i>Testi</i>	<i>Test</i>	<i>Testi</i>	<i>Test</i>	<i>Testi</i>	<i>Test</i>	<i>Testi</i>	
N	1793	1793	2066	2066	1655	1655	1459	1459
Madde Sayısı	16	27	15	27	16	28	17	28
Aritmetik Ort.	8.36	16.45	7.43	15.58	8.41	16.23	8.96	16.87
Standart Sapma	4.03	5.90	3.68	5.84	4.03	6.52	4.29	6.52
Çarpıklık	0.11	-0.03	0.12	0.16	0.04	-0.03	0.01	-0.14
Basıklık	-0.93	-0.90	-0.88	-0.81	-0.92	-0.92	-1.04	-1.03
Aritmetik Ort.*	52.24	60.94	49.51	57.71	52.54	57.96	52.71	60.24
Standart	25.19	21.85	24.52	21.65	25.20	23.29	25.23	23.29
Sapma*								
Güvenirlik**	0.83	0.88	0.80	0.88	0.83	0.89	0.84	0.89
Ortalama	0.52	0.61	0.50	0.58	0.52	0.58	0.53	0.60
Madde Güçlüğü								

Not. * Başarı yüzdesine göre. **Cronbach alfa.

Testlerin yapısal özelliklerini gösteren madde istatistikleri incelendiğinde tüm testlerin ortalama madde güçlüğü 0.50 civarında olduğu ve dolayısıyla testlerin "orta güçlük" düzeyinde olduğu saptanmıştır. X ve Y testlerinin madde güçlükleri ve ayırt edicilik indeksleri arasındaki farklar bağımsız örneklem t-testi ile analiz edilmiş; hem birinci hem ikinci grupta istatistiksel olarak manidar bir farklılık elde edilmemiştir. Benzer şekilde, X testi ve ortak test; Y testi ve ortak test arasında yapılan karşılaştırmalarda da madde güçlüğü ve ayırt edicilik indeksi açısından anlamlı farklılık olmadığı görülmüştür ($p > 0.05$). Güvenirlik analizleri ise, testlerin iç tutarlık katsayılarının 0,80 ile 0,89 arasında değiştiğini ve Fisher Z dönüşümü testlerine göre bu katsayılar arasındaki farkların istatistiksel olarak anlamsız olduğunu göstermiştir ($p > 0.05$). Özetle, madde istatistikleri ve güvenirlik analizlerinden elde edilen bu bütüncül bulgular; eşitlenecek testlerin istatistiksel özellikler açısından birbirine benzer olduğunu, hedef kitle için benzer ölçme hassasiyeti sunduğunu ve dolayısıyla eşitleme uygulamaları için gerekli aynı yapıyı ölçme ve eşit güvenirlik varsayımlarını karşıladığını doğrulamaktadır.

Son olarak, NEC deseninde kullanılan ortak değişkenlerin testler ile farklı düzeylerde ilişki sergilediğini ortaya koymuştur (Tablo 3). Analiz sonuçlarına göre fen başarısı matematik başarısıyla "yüksek" düzeyde korelasyon gösterirken; evdeki eğitim kaynakları ve sosyoekonomik düzey "orta", diğerlerinin ise "zayıf" düzeyde ilişki gösterdiği olduğu saptanmıştır.

Tablo 3

Eşitleme Uygulamalarındaki Matematik Testleri ile Ortak Değişkenler Arasındaki Korelasyon Katsayıları

Ortak Değişken	Birinci Grup		İkinci Grup		Korelasyon Yorumu	NEC Deseninin Adı
	X1	Y1	X2	Y2		
Fen başarısı (Fen1)	0.76	0.75	0.77	0.75	Yüksek	Fen1
Fen başarısı (Fen5)	0.76	0.76	0.76	0.75	Yüksek	Fen5
Matematiğe karşı özgüven	0.46	0.47	0.49	0.44	Orta	Özg
Evdeki eğitim kaynakları	0.37	0.44	0.42	0.42	Orta	EvKay
Okulun akademik başarıya verdiği önem	0.32	0.25	0.31	0.32	Zayıf	Önem
Matematik öğrenmeyi sevme	0.29	0.24	0.28	0.28	Zayıf	Sevme
Sosyoekonomik düzey	0.23	0.21	0.26	0.28	Zayıf	SED

3.2. Kernel Eşitleme Yöntemlerinin Karşılaştırılması: PRE ve Eşitlenmiş Puanlar

Kernel eşitleme uygulamaları, PRE değerleri ve eşitlenmiş puanlar üzerinden değerlendirilmiştir. PRE değerleri, gözlenen ve eşitlenmiş puan dağılımlarının ilk on momenti arasındaki farkı temel alır. X puanlarının Y formuna dönüşümün kalitesini simgeler. von Davier ve ark. (2004) tarafından önerildiği üzere, PRE değerlerinin ve aralıklarının düşük olması eşitleme işleminin başarısına dair bir kanıt olarak kabul edilir.

Bulgular, iki desende de (NEAT ve NEC) Kernel eşit yüzdelikli yöntemin, lineer yöntemle kıyasla daha düşük mutlak PRE değerlerine sahip olduğunu göstermektedir. Bu durum, eşit yüzdelikli eşitlemenin daha iyi uyum sağladığını; lineer yöntemin ise görece daha yüksek hata içerdiğini ortaya koymaktadır. Ancak ilk üç momentteki farkların %1'in altında kalması, her iki yöntemin de genel olarak kabul edilebilir düzeyde olduğunu göstermektedir (Kolen & Brennan, 2014; von Davier vd., 2004).

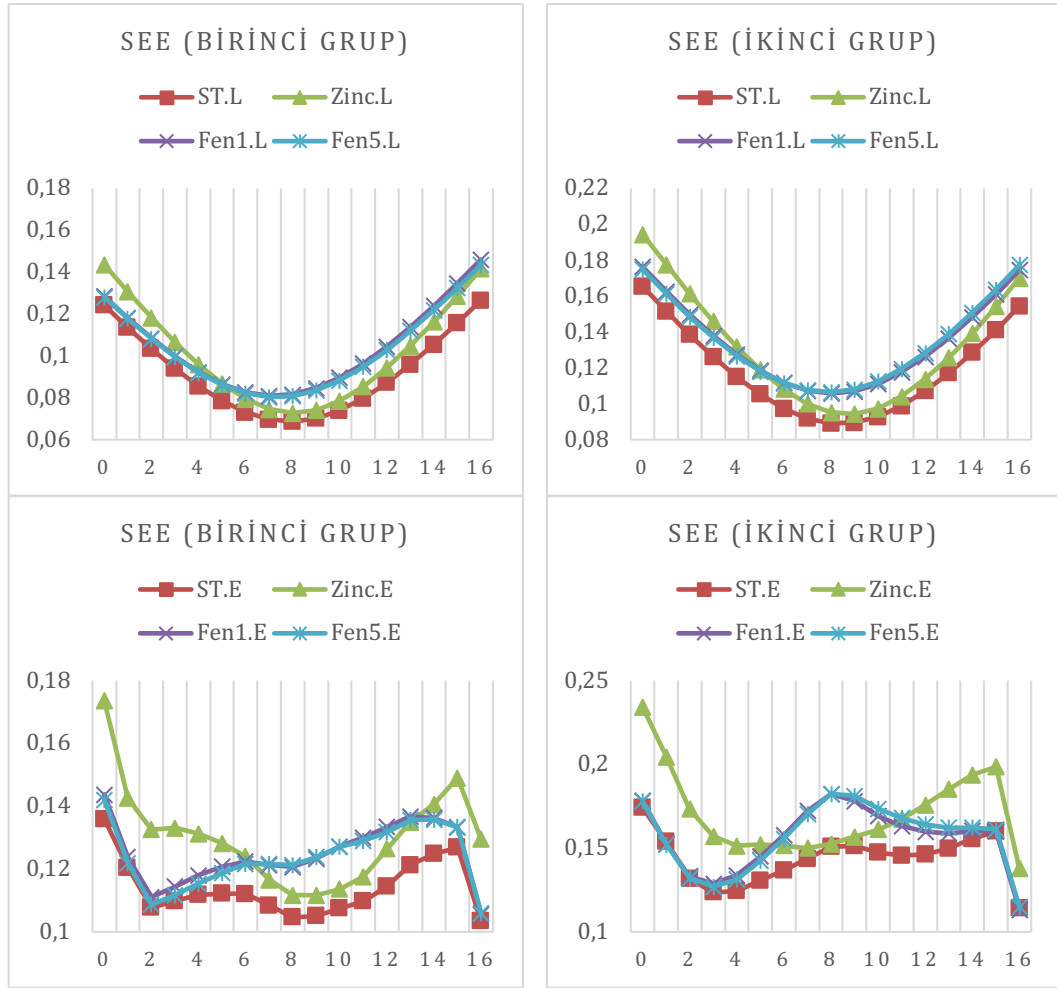
Ayrıca eşitleme sürecinde kesme işlemi uygulanmıştır (Kolen & Brennan, 2014; Livingston, 2004). Bu kapsamda negatif çıkan puanlar sıfır, Y testinin en yüksek ham puanını aşan değerler ise testin üst sınırı olarak kabul edilmiştir. Bu işlem, eşitlenmiş puanların ölçekte geçerli bir aralıkta kalmasını sağlamaktadır (Livingston, 2004). Bazı eşitlenmiş puanlar Ek 3'te sunulmuştur.

3.3. Yüksek Korelasyonlu Ortak Değişkenlerle Yapılan Kernel Eşitleme Bulguları

Matematik başarısıyla yüksek düzeyde ilişkili fen başarısı değişkenlerinin (Fen1 ve Fen5) ortak değişken olarak kullanıldığı NEC ve NEAT desenlerindeki Kernel eşitlemelere ait SEE değerleri Şekil 1'de verilmiştir. Bulgulara göre hem lineer hem de eşit yüzdelikli eşitlemelerde her bir puan için en düşük eşitleme hatası NEAT deseninde son tabakalama yönteminde elde edilmiştir. Ayrıca NEAT deseninde zincirleme ve son tabakalama yöntemleri arasında büyük bir fark olmadığı görülmektedir. NEC desenlerinde ise SEE değerleri birbirine oldukça yakın olup genel olarak NEAT desenine benzemektedir.

Şekil 1

Kernel Eşitlemelere ait SEE (Yüksek Düzeyde Korelasyona Sahip Ortak Değişkenler)



Kernel lineer eşitlemelerde grupların uç puanlarında (en düşük ve en yüksek puanlar) SEE değerlerinin daha yüksek olduğu saptanmıştır. Bu durumun, uç düzeylerde öğrenci sayısının azlığı ve uç puanlardaki kestirimlerin daha kararsız olmasıyla ilişkili olabileceği düşünülmektedir (Akın Arıkan, 2020; Wiberg & Branberg, 2015). Eşit yüzdelikli eşitlemelerde ise NEC desenlerindeki hataların birinci grupta $X=2$ ve $X=16$ iken; ikinci grupta $X<4$ ve $X>14$ iken son tabakalama yöntemiyle yakın olduğu belirlenmiştir.

Kernel eşitlemeler değerlendirme kriteri olarak NEAT deseniyle karşılaştırıldığında, hesaplanan RMSD, WMSE ve RMSE değerleri Tablo 4'te verilmiştir. Tüm incelemelerde RMSD, WMSE, RMSE ve fark grafikleriyle benzer yorumlara ulaşıldığı görülmüştür. Bu sebeple sadece RMSD değerleri yorumlanarak her bir NEC deseni ayrıntılı bir şekilde incelenmiştir. Bulgular, NEC desenleri ile NEAT desenleri arasındaki farkların tüm gruplarda marjinal düzeyde olduğunu ($0.05 < \text{RMSD} < 0.5$) ve ihmal edilebilir sınırlar içinde kaldığını göstermektedir.

Tablo 4

Yüksek Korelasyonlu Ortak Değişkenlerin Kullanıldığı NEC Desenlerine Ait Kernel Eşitleme Hataları

<i>Karşılaştırma</i>		<i>NEC İLE ST</i>			<i>NEC İLE ZİNCİRLEME</i>		
		<i>RMSE</i>	<i>RMSD</i>	<i>WMSE</i>	<i>RMSE</i>	<i>RMSD</i>	<i>WMSE</i>
<i>Grup 1</i>	Fen1.L	0.17	0.16*	0.00	0.11	0.09*	0.59
	Fen1.E	0.16	0.17*	0.00	0.13	0.13*	0.56
	Fen5.L	0.11	0.09*	0.00	0.09	0.07*	0.51
	Fen5.E	0.10	0.10*	0.00	0.13	0.13*	0.49
<i>Grup 2</i>	Fen1.L	0.21	0.22*	0.00	0.28	0.29*	0.12
	Fen1.E	0.21	0.24*	0.00	0.32	0.35*	0.12
	Fen5.L	0.16	0.16*	0.00	0.23	0.24*	0.18
	Fen5.E	0.16	0.18*	0.00	0.28	0.30*	0.18

Not: *Marjinal fark vardır. **DTM vardır.

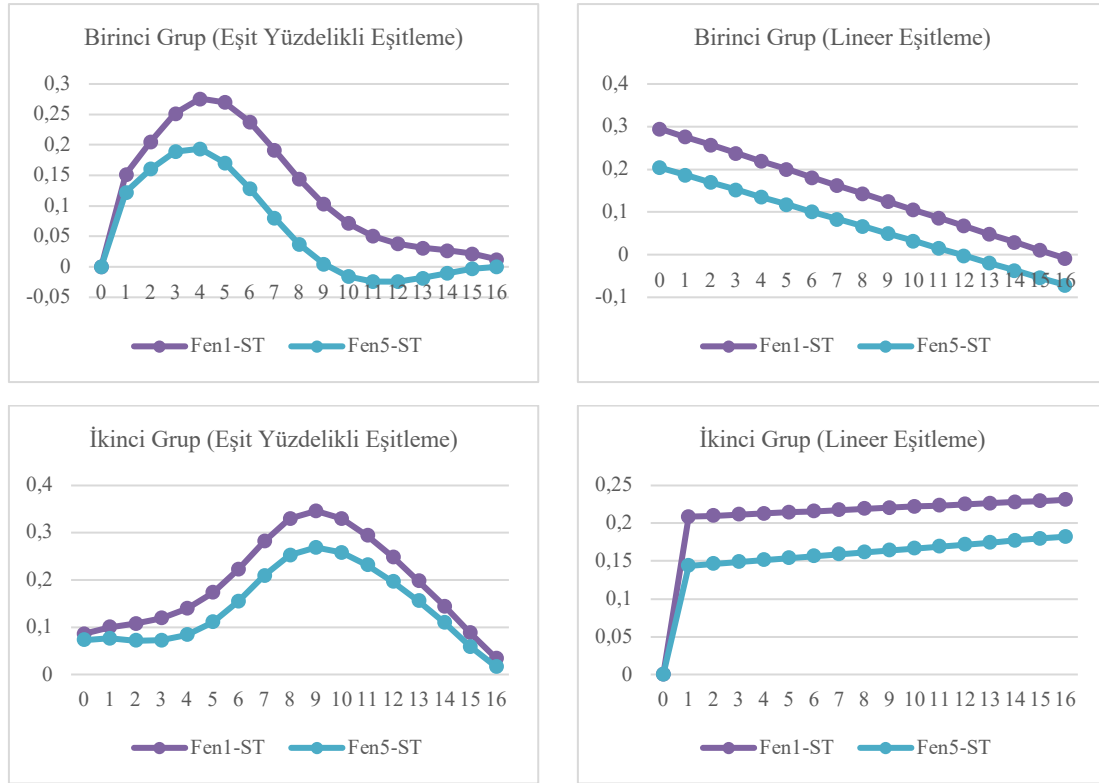
***Fen1.L ve Fen5.L: Fen başarısının (Fen1 ve Fen5) ortak değişken olduğu desenlerde Kernel lineer eşitlemeler.

****Fen1.E ve Fen5.E: Fen başarısının (Fen1 ve Fen5) ortak değişken olduğu desenlerde Kernel eşit yüzdelikli eşitlemeler.

NEC desenlerinde eşitlenmiş puanlar ile NEAT (son tabakalama yöntemi) deseninde eşitlenmiş puanlar arasındaki farkı gösteren fark grafikleri Şekil 2’de sunulmuştur. Bulgular, iki grupta da Fen5’in çoğu durumda son tabakalama yöntemine daha yakın sonuçlar verdiğini ortaya koymuştur. Lineer eşitlemede birinci grupta özellikle orta–yüksek puan aralıklarında ($12 < X$ için Fen1; $8 < X < 15$ için Fen5); ikinci grupta ise yalnızca $X=0$ puanında marjinal farkın anlamlı olmadığı belirlenmiştir. Eşit yüzdelikli eşitlemede de benzer biçimde, birinci grupta üst puanlarda ($11 < X$ iken Fen1 ve $7 < X$ iken Fen5), ikinci grupta ise yalnızca $X=16$ iken (Fen1 ve Fen5 için) marjinal farkların anlamsız olduğu görülmüştür.

Şekil 2

NEC Desenlerinin NEAT Deseninden Farkına İlişkin Fark Grafikleri (Fen1 ve Fen5)



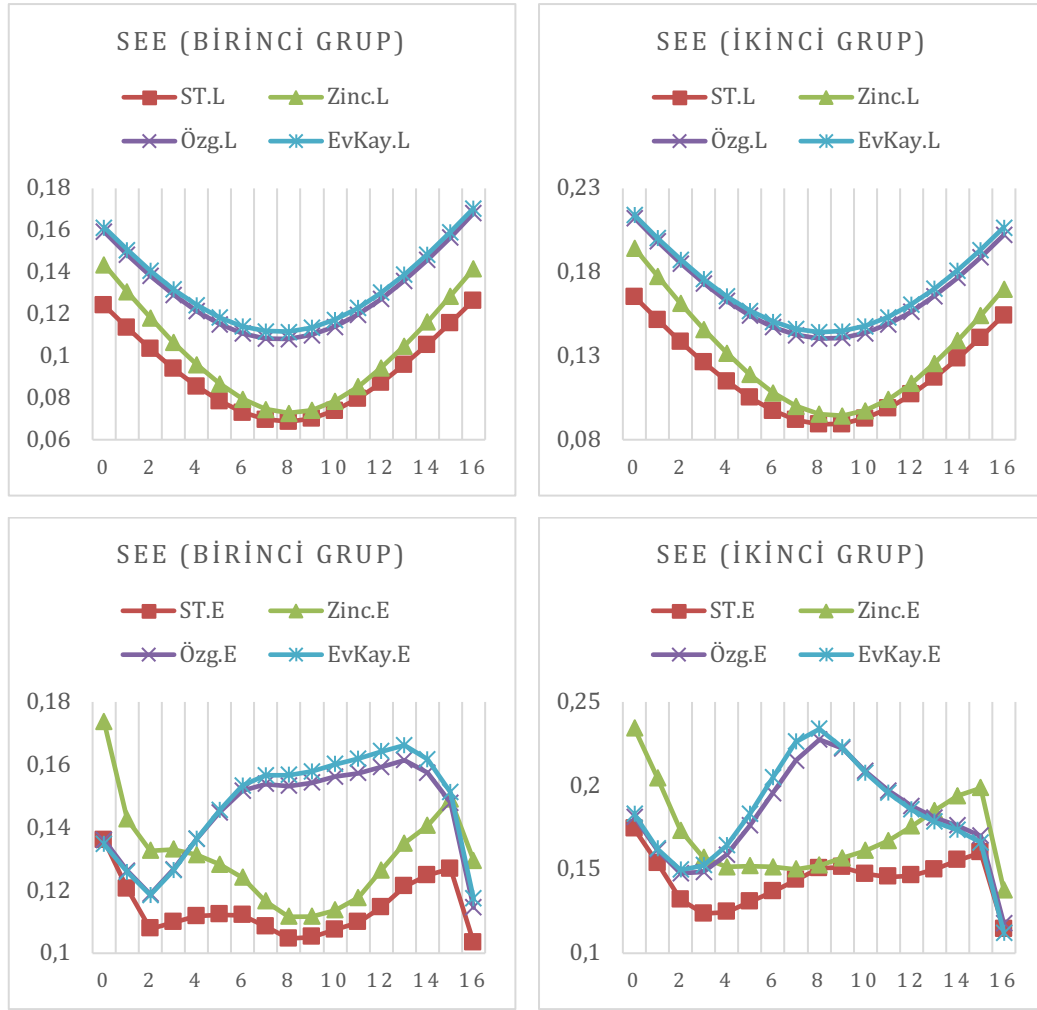
Özetle bu alt araştırma sorusunda SEE'ye göre tüm gruplarda Fen1 ve Fen5'e ait hataların birbirine ve son tabakalama yöntemine yakın olduğu saptanmıştır. Ayrıca RMSD değerlerine ve fark grafiklerine göre ilgili NEC desenleri ile son tabakalama yönteminde eşitlenmiş puanlar arasındaki fark, ihmal edilebilir sınırlar dâhilindedir.

3.4. Orta Korelasyonlu Ortak Değişkenlerle Yapılan Kernel Eşitleme Bulguları

Matematik başarısıyla orta düzeyde korelasyona sahip değişkenlerin (Özg ve EvKay) ortak değişken olarak kullanıldığı NEC ve NEAT desenlerindeki Kernel eşitlemelere ait SEE değerleri Şekil 3'te sunulmuştur. Grafikler incelendiğinde, iki grupta da NEC desenlerindeki eşitleme hatalarının birbirine oldukça yakın olduğu ve genel olarak NEAT deseninden farklılaştığı görülmektedir.

Şekil 3

Kernel Eşitlemelere ait SEE (Orta Düzeyde Korelasyona Sahip Ortak Değişkenler)



NEAT deseninin değerlendirme kriteri olarak alındığı durumlarda hesaplanan RMSD, WMSE ve RMSE değerleri Tablo 5'te verilmiştir. Buna göre NEC desenleri ile NEAT desenindeki son tabakalama yöntemi arasındaki RMSD değerleri 0.05 ile 0.5 aralığındadır. Bu sonuç, Özg ve EvKay desenleri ile son tabakalama yöntemi arasında marjinal farklılık bulunduğunu, ancak bu farkın ihmal edilebilir sınırlar dahilinde olduğunu göstermektedir. Zincirleme eşitleme yöntemi kriter olarak alındığında da benzer bulgular elde edilmiştir; sadece birinci gruptaki EvKay.E deseninde RMSD değeri 0.5'in üzerinde olup, ihmal edilebilir sınırların üzerinde bir farklılık göstermektedir.

Tablo 5

Orta Korelasyonlu Ortak Değişkenlerin Kullanıldığı NEC Desenlerine Ait Kernel Eşitleme Hatalar

		NEC İLE ST			NEC İLE ZİNCİRLEME		
	Karşılaştırma	RMSE	RMSD	WMSE	RMSE	RMSD	WMSE
Grup 1	Özg.L	0.27	0.28*	0.00	0.36	0.37*	0.01
	Özg.E	0.26	0.30*	0.01	0.36	0.41*	0.01
	EvKay.L	0.36	0.37*	0.01	0.45	0.46*	0.01
	EvKay.E	0.35	0.39*	0.01	0.45	0.51**	0.02
Grup 2	Özg.L	0.17	0.18*	0.00	0.25	0.25*	0.00
	Özg.E	0.18	0.20*	0.00	0.29	0.31*	0.01
	EvKay.L	0.34	0.35*	0.01	0.41	0.42*	0.01
	EvKay.E	0.34	0.38*	0.01	0.44	0.48*	0.01

Not: *Marjinal fark vardır. **DTM vardır.

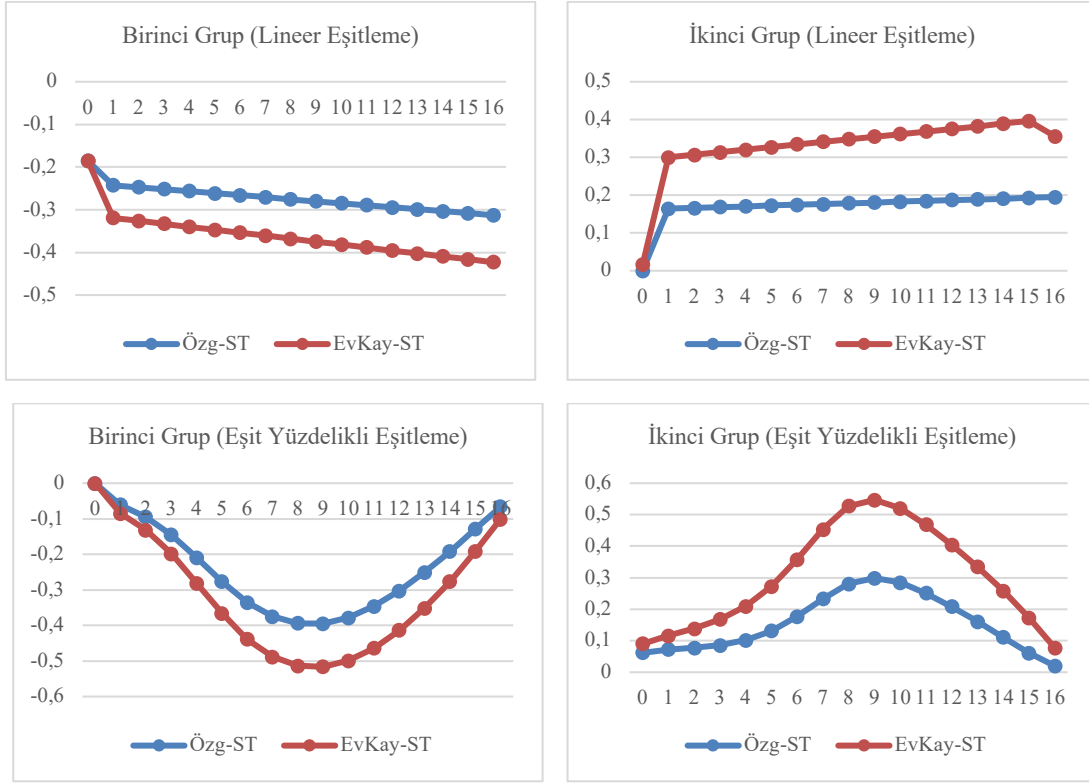
***Özg.L ve EvKay.L: Matematiğe karşı özgüvenin (Özg) ve evdeki eğitim kaynaklarının (EvKay) ortak değişken olduğu desenlerde Kernel lineer eşitlemeler.

**** Özg.E ve EvKay.E: Matematiğe karşı özgüvenin (Özg) ve evdeki eğitim kaynaklarının (EvKay) ortak değişken olduğu desenlerde Kernel eşit yüzdelikli eşitlemeler.

Fark grafikleri (Şekil 4) incelendiğinde, NEC desenleri ile NEAT desenindeki son tabakalama yöntemi arasındaki farkların çoğu puanda marjinal düzeyde olduğu ($0.05 < \text{RMSD} < 0.5$) belirlenmiştir. Özetle bu alt araştırma sorusunda tüm gruplarda Özg ve EvKay desenlerine ait SEE değerlerinin birbirine çok yakın olduğu ve son tabakalama yönteminden farklılaştığı görülmüştür. Tüm gruplardaki RMSD değerlerine ve fark grafiklerine göre ilgili NEC desenleri ile son tabakalama yönteminde eşitlenmiş puanlar arasındaki fark, ihmal edilebilir sınırlar dâhilindedir. Son tabakalama yöntemine daha yakın olan NEC deseni tüm gruplarda Özg'dir.

Şekil 4

NEC Desenlerinin NEAT Deseninden Farkına İlişkin Fark Grafikleri (Orta Korelasyonlu Değişkenler)

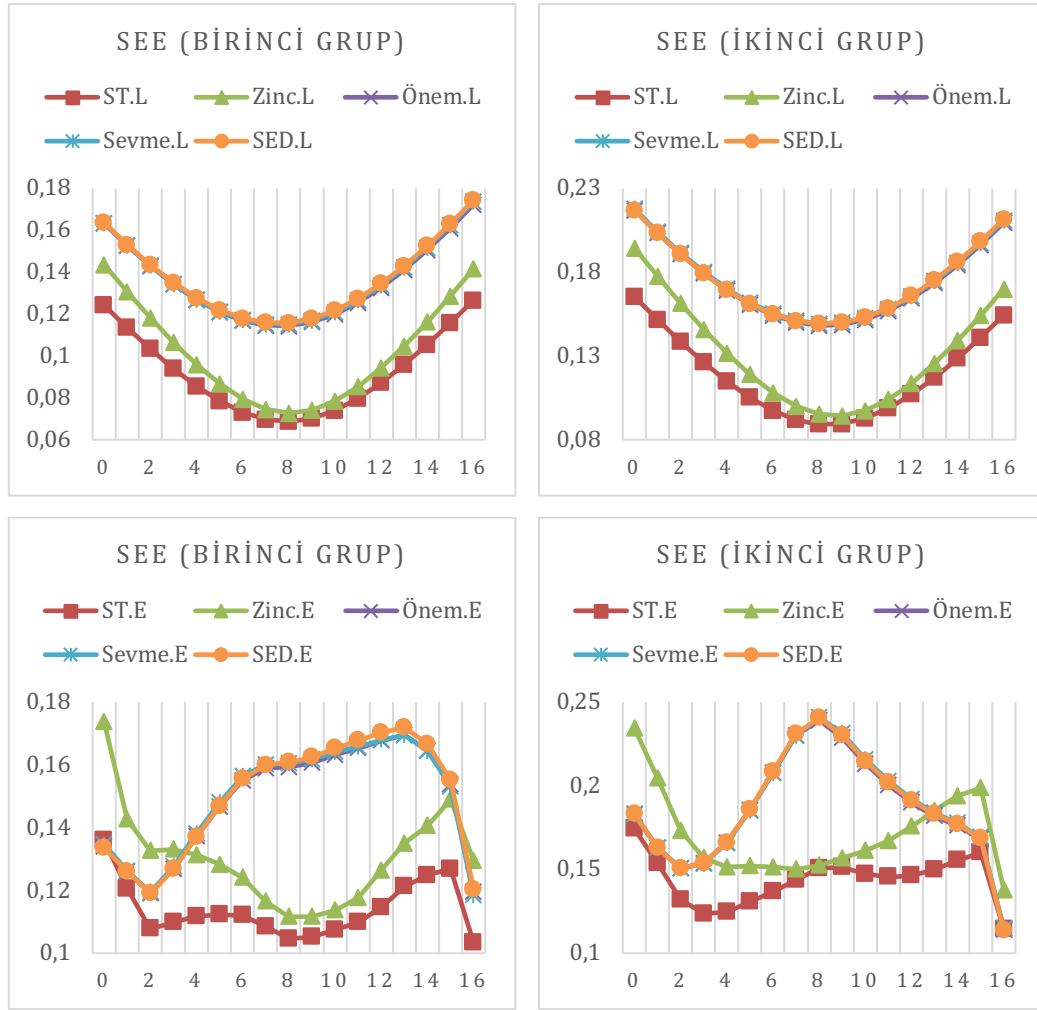


3.5. Zayıf Korelasyonlu Ortak Değişkenlerle Yapılan Kernel Eşitleme Bulguları

Matematik başarısıyla zayıf düzeyde korelasyona sahip değişkenlerin (Sevme, Önem ve SED) ortak değişken olarak kullanıldığı NEC ve NEAT desenlerindeki Kernel eşitlemelere ait SEE değerleri Şekil5'te sunulmuştur. Grafiklere genel olarak bakıldığında tüm NEC desenlerinde SEE değerlerinin birbirine çok yakın olduğu ve bu desenler arasındaki farklılıkların belirgin olmadığı görülmüştür. Ayrıca NEC desenlerindeki eşitleme hatalarının NEAT desenindeki yöntemlerden oldukça farklılaştığı belirlenmiştir.

Şekil 5

Kernel Eşitlemelere ait SEE (Zayıf Düzeyde Korelasyona Sahip Ortak Değişkenler)



NEAT desenindeki Kernel eşitlemeler (lineer ve eşit yüzdelikli) değerlendirme kriteri olarak kullanıldığında hesaplanan RMSD, WMSE ve RMSE değerleri Tablo 6'da verilmiştir. NEC desenleri ile NEAT desenindeki son tabakalama yöntemi arasındaki RMSD değerleri iki grupta da 0.05 ile 0.5 aralığında bulunmuştur. Bu bulgu, tüm desenlerde marjinal fark olduğunu ancak farkların ihmal edilebilir düzeyde kaldığını göstermektedir. Zincirleme eşitleme yöntemi değerlendirme kriteri olduğunda ise tüm desenlerde marjinal farkın veya DTM'nin olduğu görülmektedir.

Tablo 6

Zayıf Korelasyonlu Ortak Değişkenlerin Kullanıldığı NEC Desenlerine Ait Kernel Eşitleme Hataları

<i>Karşılaştırma</i>	<i>NEC İLE ST</i>			<i>NEC İLE ZİNCİRLEME</i>		
	<i>RMSE</i>	<i>RMSD</i>	<i>WMSE</i>	<i>RMSE</i>	<i>RMSD</i>	<i>WMSE</i>
<i>Grup 1</i>	Önem.L	0.42	0.43*	0.01	0.51	0.52**
	Önem.E	0.41	0.45*	0.01	0.50	0.56**
	Sevme.L	0.38	0.39*	0.01	0.47	0.48*
	Sevme.E	0.37	0.41*	0.02	0.46	0.52**
	SED.L	0.44	0.45*	0.01	0.53	0.54**
	SED.E	0.43	0.48*	0.01	0.52	0.59**
<i>Grup 2</i>	Önem.L	0.31	0.32*	0.01	0.38	0.39*
	Önem.E	0.31	0.34*	0.01	0.41	0.45*
	Sevme.L	0.28	0.29*	0.00	0.35	0.36*
	Sevme.E	0.28	0.31*	0.01	0.38	0.42*
	SED.L	0.32	0.33*	0.01	0.39	0.40*
	SED.E	0.32	0.36*	0.01	0.42	0.46*

Not: *Marjinal fark vardır. **DTM vardır.

***Önem.L, Sevme.L ve SED.L: Okulun akademik başarıya verdiği önemin (Önem), matematik öğrenmeyi sevmenin (Sevme) ve sosyoekonomik düzeyin (SED) ortak değişken olduğu desenlerde Kernel lineer eşitlemeler.

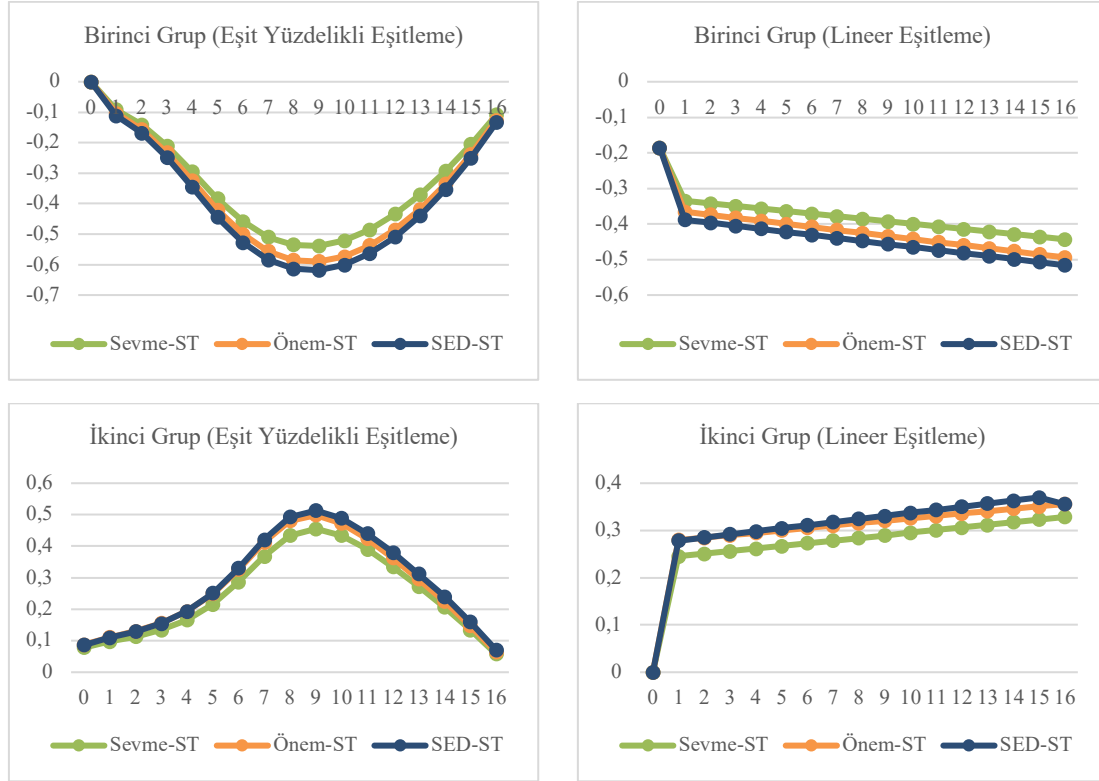
**** Önem.E, Sevme.E ve SED.E: Okulun akademik başarıya verdiği önemin (Önem), matematik öğrenmeyi sevmenin (Sevme) ve sosyoekonomik düzeyin (SED) ortak değişken olduğu desenlerde Kernel eşit yüzdelikli eşitlemeler.

NEC desenlerinde eşitlenmiş puanlar ile NEAT (son tabakalama yöntemi) deseninde eşitlenmiş puanlar arasındaki farkı gösteren fark grafikleri Şekil 6'da sunulmuştur. İlgili grafiklere göre tüm gruplardaki tüm karşılaştırmalara ait farkların birbirine yakın olduğu görülmüştür. Buna göre son tabakalama yöntemiyle yapılan karşılaştırmada bazı puanlardaki farkların 0.5'ten büyük, yani ihmal edilebilir sınırların üzerinde olduğu saptanmıştır. Diğer karşılaştırmalardaki farklar 0.5'ten küçük olduğu için ihmal edilebilir sınırlar dâhilindedir.

Özetle bu alt araştırma sorusunda SEE'ye göre tüm gruplarda tüm NEC desenlerine ait hataların birbirine çok yakın olduğu ve bu eşitleme hatalarının NEAT desenindeki son tabakalama yönteminden farklılaştığı saptanmıştır. Tüm gruplardaki RMSD değerlerine ve fark grafiklerine göre ilgili NEC desenleri ile son tabakalama yönteminde eşitlenmiş puanlar arasındaki fark, çoğu puan düzeyinde ihmal edilebilir sınırlar dâhilindedir.

Şekil 6

NEC Desenlerinin NEAT Deseninden Farkına İlişkin Fark Grafikleri (Zayıf Korelasyonlu Değişkenler)



TARTIŞMA, SONUÇ VE ÖNERİLER

Bu araştırmada TIMSS 2019 sekizinci sınıf matematik testleri üzerinden Kernel eşitleme yöntemlerinin (lineer ve eşit yüzdelikli) performansı, NEC ve NEAT desenleri çerçevesinde karşılaştırmalı olarak incelenmiştir. Farklı kitapçıkları yanıtlayan iki farklı öğrenci grubu seçilmiş ve sonuçlar hem gruplara göre hem de desenlere göre karşılaştırılmıştır. Eşitleme sonuçları WMSE, RMSD, RMSE ve SEE ile incelenmiş, fark grafikleri ile görselleştirilmiş ve bulguların pratik anlamlılığı DTM ve marjinal farklılığa göre yorumlanmıştır. Özellikle NEC deseninde kullanılan fen başarısı, sosyoekonomik düzey ve matematiğe karşı özgüven gibi ortak değişkenlerin seçiminin, eşitleme sonuçları üzerindeki belirleyici rolü mercek altına alınmıştır. Elde edilen sonuçlar, Kernel lineer ve Kernel eşit yüzdelikli eşitleme sonuçlarının birbirine benzer olduğunu, ancak ortak değişkenlerin matematik başarısıyla kurdukları ilişkinin düzeyine göre eşitleme performansının doğrudan değiştiğini ortaya koymuştur.

Bu araştırmanın bulguları fen başarısı gibi matematik puanlarıyla yüksek düzeyde korelasyona sahip değişkenlerin dahil edildiği NEC desenlerinin, NEAT desenine daha yakın performans sergilediğini göstermiştir. Öte yandan, düşük korelasyonlu değişkenlerin kullanıldığı desenlerde NEAT deseninden farklı sonuçların elde edildiği belirlenmiştir. Bu sonuç, test eşitlemede ortak değişkenlerin gücünün, değişkenlerin ölçülmek istenen test puanıyla olan ilişkisine bağlı olduğunu ortaya koyan alanyazınla tutarlıdır (Albano & Wiberg, 2019; Akın Arıkan, 2020; Sezer Başaran vd., 2023; Wiberg & Branberg, 2015; Yıldırım Seheryeli, 2022). Özellikle fen başarısının baskın etkisi, Tekeli Akdemir'in (2021) TIMSS 2019 Türkiye verisinde bu değişkeni en güçlü ayırt edici değişken olarak tanımlamasıyla desteklenmektedir. Alanyazında

fen başarısının ortak değişken olarak kullanıldığı özgün eşitleme çalışmalarına rastlanmamış olması ise mevcut çalışmanın alana sağladığı katkıyı güçlendirmektedir.

Matematik başarısıyla zayıf düzeyde ilişkili değişkenlerin (sosyoekonomik düzey, okulun akademik başarıya verdiği önem ve matematik öğrenmeyi sevme) kullanıldığı NEC desenlerinde ise sonuçlar NEAT deseninden farklılık göstermektedir. Bulgulara göre özellikle sosyoekonomik düzeyin ve okulun akademik başarıya verdiği önemin ortak değişken olduğu desenlerde eşitlenmiş puanların birbirine çok benzediği, gruplar arasında anlamlı fark oluşmadığı görülmüştür. Bu durum, bu değişkenlerin öğrencilerin bireysel düzeydeki başarı farklarını yansıtmada ve dolayısıyla gruplar arasındaki yetenek farkını açıklamada yetersiz kaldığını göstermektedir. Alanyazında da benzer biçimde, başarıyla zayıf ilişkiye sahip değişkenler ortak değişken olarak kullanıldığında bu değişkenlerin eşitleme doğruluğunu artırmada etkili olmadığı vurgulanmaktadır (Wiberg & Branberg, 2015; Yıldırım Seheryeli, 2022). Buna karşılık matematik öğrenmeyi sevme deseninde, özellikle bazı puan aralıklarında marjinal düzeyde farkların ortaya çıkması, bu değişkenin gruplar arasındaki yetenek farklılığını sınırlı ölçüde yansıtabildiğini göstermektedir. Bu sonuç, ortak değişken seçiminde yalnızca korelasyon katsayılarının değil, dağılım özelliklerinin (çarpıklık, basıklık, frekans farklılıkları) de dikkate alınması gerektiğini ortaya koymaktadır (Wiberg & Branberg, 2015; Yıldırım Seheryeli, 2022). Ancak genel olarak düşük korelasyonlu değişkenlerin kullanıldığı NEC desenlerinde eşitleme hatalarının NEAT desenine göre belirgin bir üstünlük sağlamadığı söylenebilir.

Bu çalışmada kullanılan sosyoekonomik düzey değişkeni, okul müdürü beyanına dayanmaktadır ve okulun sosyoekonomik düzeyini göstermektedir. Bu durum bireysel düzeydeki sosyoekonomik düzey tanımlamalarının kullanıldığı (örn. ABİDE 2016) eşitleme sonuçlarıyla olan farklılığın sebebini açıklayabilir. Bu çalışmanın bulgusuna ters olarak Akın Arıkan (2020) ABİDE 2016 uygulamasındaki matematik testi puanlarını eşitlediği çalışmasında sosyoekonomik düzeyin ortak değişken olarak kullanıldığı NEC desenindeki eşitleme hatalarının NEAT desenden daha küçük olduğunu saptamıştır. Bu düşük eşitleme hatası bulgusuyla mevcut çalışmanın çelişmesi, bu değişkenin ölçülme ve tanımlanma biçiminin eşitleme modelleri üzerindeki kritik etkisini açıkça göstermektedir. ABİDE 2016 uygulamasında sosyoekonomik düzey ailenin aylık geliri, anne-babanın eğitim düzeyi, öğrenciye ait kendi odasının olup olmaması gibi değişkenler dikkate alınarak hesaplanmıştır (Milli Eğitim Bakanlığı, 2017).

Bununla birlikte, matematik başarısıyla orta düzeyde ilişkili değişkenlerin (matematiğe karşı özgüven ve evdeki eğitim kaynakları) kullanıldığı NEC desenlerinde daha karmaşık bir tablo ortaya çıkmıştır. Bulgular, matematiğe karşı özgüven değişkeninin NEAT desenine daha yakın sonuçlar verdiğini göstermektedir. Bu durum, ortak değişkenlerin yalnızca korelasyon katsayılarına göre değil, gruplar arasında yetenek dağılımını ne ölçüde açıklayabildiğine göre de farklı işlevler üstlenebileceğini ortaya koymaktadır. Alanyazında da matematiğe karşı özgüvenin matematik başarısıyla anlamlı fakat sınırlı düzeyde ilişkili olduğu (Yıldırım Seheryeli, 2022; Yurtçu vd., 2021), evdeki eğitim kaynaklarının ise öğrenci başarısına farklı örneklemelerde değişen etkiler gösterebildiği belirtilmektedir (Berger vd., 2020; Sarı vd., 2017). Bu bağlamda, orta düzeyde korelasyonlu değişkenlerin ortak değişken olarak kullanılması, test eşitlemede yüksek korelasyonlu değişkenler kadar güçlü bir katkı sağlamamakta; ancak düşük korelasyonlu değişkenlere kıyasla daha güvenilir ve tutarlı sonuçlar üretmektedir.

Genel olarak bu çalışmanın bulguları, Kernel eşitlemede ortak değişken seçiminin kritik önemini bir kez daha vurgulamaktadır. Özellikle yüksek korelasyonlu değişkenlerin (fen başarısı) kullanıldığı NEC desenleri, NEAT desenine daha yakın sonuçlar vererek eşitleme doğruluğunu arttırmakta; buna karşılık düşük korelasyonlu değişkenler (sosyoekonomik düzey, okulun başarıya verdiği önem) gruplar arası yetenek farklılığını açıklamada yetersiz kalmaktadır. Orta korelasyonlu değişkenler ise bulgulara göre değişkenlik göstermekle birlikte, eşitleme sonuçlarını kısmen iyileştirebilmekte, ancak tutarlılıkları gruptan gruba farklılaşabilmektedir. Bu sonuç, alanyazında vurgulanan “ortak değişkenin gücünün yalnızca korelasyon düzeyine değil, dağılım

özelliklerine, açıklanan varyansa ve gruplar arasındaki tutarlılığına bağlı olduğu” görüşüyle örtüşmektedir (Wiberg & Branberg, 2015; Wallin & Wiberg, 2019; Yıldırım Seheryeli, 2022).

Bu araştırmada NEC desenlerinin performansı değerlendirilirken NEAT deseni kriter olarak kullanılmıştır. Bulgular, NEAT deseninde özellikle son tabakalama yönteminin zincirleme yöntemine kıyasla daha düşük hata verdiğini göstermektedir. Bu durum, daha önceki çalışmalarda da vurgulanmış olup (Akın Arıkan, 2017, 2020; Wiberg, 2021), son tabakalama yönteminin uç puanlardaki örneklem küçüklüğünden kaynaklanan hataları sınırlı düzeyde azalttığını ortaya koymaktadır.

Alanyazında farklı test eşitleme çalışmalarında EG deseninin değerlendirme ölçütü olarak tercih edildiği görülmektedir (von Davier vd., 2006; Wiberg & Branberg, 2015). Özellikle gerçek verilerle yapılan araştırmalarda, tüm öğrencilerin aynı testi yanıtladığı EG deseni ya da tek grup deseni sıklıkla karşılaştırma ölçütü olarak kullanılmaktadır (Albano & Wiberg, 2019; von Davier vd., 2006; Wiberg & Branberg, 2015). Bu tür çalışmalarda eşitleme işlemleri, evrenden seçilen öğrenci puanları ve test maddeleri üzerinden yürütülmektedir. Bununla birlikte, bazı araştırmalarda açık bir kriter belirtilmemekte, daha çok eşitleme yöntemlerinden elde edilen değerler üzerinden yorum yapılmaktadır. Örneğin Kernel eşitleme bağlamında, SEE değerleri desenler arasında karşılaştırma amacıyla kullanılmıştır (Akın Arıkan, 2020; Altıntaş & Wallin, 2021; Özsoy, 2021). Benzer biçimde Bayes yöntemine dayalı uygulamalarda ise eşitlenmiş puanların hedef teste uzaklığı incelenmiştir (Yurtçu vd., 2021). Ayrıca kimi çalışmalarda, eşitleme hatalarıyla birlikte özdeş yöntemlerin sonuçları da karşılaştırma ölçütü olarak alınmakta ve fark grafikleri değerlendirme sürecine dahil edilmektedir (Wallin & Wiberg, 2019; Özsoy, 2021).

Bu çalışmada NEC desenlerindeki Kernel eşitleme sonuçları SEE, RMSD, WMSE, RMSE ve fark grafikleri kullanılarak NEAT (son tabakalama ve zincirleme) deseniyle karşılaştırılmıştır. Özellikle RMSD değerleri 0,05 ve 0.5 sınırlarına göre değerlendirilmiştir. Bulgular, bazı desenlerde elde edilen RMSD değerlerinin bu eşiklere oldukça yakın olduğunu göstermektedir. Örneğin evdeki eğitim kaynakları deseninde zincirleme eşitleme değerlendirme kriteri iken ikinci grupta 0,48 hesaplanmıştır. Bu değer marjinal fark olarak yorumlanmasına rağmen aslında DTM şeklinde de yorumlanabilir. Benzer şekilde matematik öğrenmeyi sevme deseninde birinci grupta 0,07 şeklinde hesaplanmıştır. Bu değer marjinal fark olarak sınıflansa da, pratik açıdan oldukça düşük farklılıklardır. Bu durum, eşitleme hatalarının yalnızca sayısal sınıflandırmalarla değil, bağlamsal olarak da değerlendirilmesi gerektiğine işaret etmektedir (Kolen & Brennan, 2014).

Genel olarak incelendiğinde RMSD, WMSE ve RMSE değerleri farklı büyüklükler göstermesine rağmen desenlerin sıralamasının değişmediği görülmektedir. Bu da farklı hata ölçütlerinin çoğunlukla benzer eğilimler sunduğunu, ancak tek bir ölçüte bağlı kalmanın yanıltıcı olabileceğini ortaya koymaktadır. Literatürde de benzer biçimde, Kernel eşitleme hatalarının farklı boyutları yansıtmaları nedeniyle birden fazla ölçütün birlikte değerlendirilmesinin daha sağlıklı olduğu belirtilmektedir (Altıntaş & Wallin, 2021; Wiberg & Branberg, 2015). Ayrıca SEE’nin veri toplama desenini dikkate aldığı, buna karşın RMSD ve RMSE’nin bu özelliğe sahip olmadığı; dolayısıyla eşitleme sonuçlarının yorumlanmasında farklı hata ölçütlerinin avantaj ve sınırlılıklarının göz önünde bulundurulması gerektiği vurgulanmaktadır (Kolen & Brennan, 2014; von Davier vd., 2004). Bu çalışmada da görüldüğü üzere, hata ölçütleri arasındaki küçük farklar dahi eşitleme yorumlarını değiştirebilmektedir. Bu nedenle gelecekte yapılacak eşitleme çalışmalarında, farklı hata ölçütlerinin birlikte raporlanması ve sonuçların bu bütünlük içinde ele alınması önerilmektedir.

Test eşitleme uygulamalarında rastgele ve sistematik hataların ortaya çıkması olasıdır (Kolen & Brennan, 2014). Örneklem büyüklüğü ve özellikleri genellikle rastgele hatalara yol açarken, varsayımların karşılanmaması veya düzgünleştirme işlemleri sistematik hataları

artırabilir. Kernel eşitlemede kullanılan ön düzgünleştirme adımı, rastgele hatayı azaltabilse de sistematik hatanın büyümesine neden olabilir.

Bu araştırmanın sonuçları uygulayıcılara yönelik bazı önemli çıkarımlar sunmaktadır. Bulgular, fen başarısının ortak değişken olarak kullanıldığı NEC deseninin, sekizinci sınıf öğrencilerinde en düşük eşitleme hatalarıyla sonuçlandığını ve NEAT desenine daha yakın puanlar ürettiğini göstermiştir. Bu nedenle, matematik testlerinde ortak maddelerin bulunmadığı ya da kullanılmadığı durumlarda, fen başarısı ortak değişken olarak NEC deseninde tercih edilebilir. Bununla birlikte, matematiğe karşı özgüven, evdeki eğitim kaynakları ve matematik öğrenmeyi sevmeye gibi matematik başarısıyla orta veya zayıf düzeyde korelasyona sahip ortak değişkenlerin eşitleme sonuçlarının farklı gruplarda değişkenlik göstermesi, ortak değişken seçiminde yalnızca korelasyon katsayılarının dikkate alınmaması gerektiğini ortaya koymaktadır. Ortak değişkenlerin dağılım özellikleri, betimsel istatistikleri ve frekans farklılıkları da göz önünde bulundurulmalıdır. Ayrıca, TIMSS verilerinde kullanılan sosyoekonomik düzey değişkeni okul düzeyinde ölçüldüğü için bireysel başarı farklılıklarını yeterince yansıtmamıştır. Uygulamada aile gelir düzeyi, ebeveyn eğitimi veya öğrencinin yaşam koşulları gibi bireysel verilere dayalı sosyoekonomik göstergelerin tercih edilmesi, daha doğru eşitleme sonuçları sağlayabilir.

Araştırmacılar açısından ise, elde edilen bulgular bazı okul temelli değişkenlerin (örneğin okulun sosyoekonomik düzeyi veya akademik başarıya verilen önem) gruplar arasındaki yetenek farklarını açıklamada sınırlı kaldığını göstermektedir. Bu durum, gelecekte bireysel düzeydeki değişkenlerin daha kapsamlı biçimde ele alınması gerektiğini düşündürmektedir. Ayrıca NEC desenlerinin değerlendirilmesinde yalnızca NEAT desenine değil, EG desenine de başvurulması önemlidir; çünkü bazı durumlarda NEC desenleri NEAT desenine yakın görünse de, EG ile neredeyse aynı sonuçlar üretebilmektedir. Bunun yanında, bu çalışmada kullanılan eşitleme hatalarının (WMSE, RMSD, RMSE, SEE) büyük ölçüde benzer yorumlar sunduğu görülmüştür. Bu nedenle, eşitleme hatalarının değerlendirilmesinde farklı yöntemlerin geliştirilmesi ve alternatif ölçütlerin denenmesi, ileriki araştırmalar için yararlı olacaktır. Son olarak, bu çalışmanın TIMSS 2019 sekizinci sınıf matematik testi verisiyle sınırlı olduğu dikkate alındığında, benzer çalışmaların dördüncü sınıf uygulamaları veya fen testleri üzerinde gerçekleştirilmesi, hatta daha geniş örneklemelere sahip ulusal ve uluslararası veri setlerinin kullanılması, ortak değişken seçimine ilişkin daha güçlü ve genellenebilir sonuçlar elde edilmesine katkı sağlayacaktır.

Etik Kurul Onay Bilgileri:

Bu çalışmada TIMSS'in resmi internet sitesinde yayımlanan TIMSS 2019 araştırmasına ait ikincil veri kullanılmıştır. Bu veri açık erişime sahip olduğu için etik kurul iznine gerek yoktur.

Çıkar Çatışması Beyanı:

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir.

Yapay Zekâ Kullanımı Beyanı:

Bu çalışmanın hazırlanma sürecinde yazarlar, ChatGPT, Gemini ve Grammarly'yi metin iyileştirme ve Türkçe'den İngilizce'ye çeviri amacıyla kullanmıştır. Yapay zekâ araçları yalnızca destekleyici bir araç olarak kullanılmış olup çalışmanın tüm bilimsel içeriği, yorumlanması ve nihai sorumluluğu yazarlara aittir. Yazarlar, kullanılan yapay zekâ araçlarının etik kullanımına ilişkin akademik dürüstlük ve yayın etiği ilkelerine uygun hareket etmiştir.

KAYNAKÇA

- Akın Arıkan, Ç. (2017). *Kernel eşitleme ve madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*. [Yayınlanmamış doktora tezi] . Hacettepe Üniversitesi.
- Akın Arıkan, Ç. (2020). The impact of covariate variables on kernel equating under the non-equivalent groups. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 362-373. <http://dx.doi.org/10.21031/epod.706835>
- Akın-Arıkan, Ç., & Gelbal, S. (2018). A comparison of traditional and kernel equating methods. *International Journal of Assessment Tools in Education*, 5(3), 417-427. <http://dx.doi.org/10.21449/ijate.409826>
- Albano, A. D. & Wiberg, M. (2019). Linking with external covariates: examining accuracy by anchor type, test length, ability difference, and sample size. *Applied Psychological Measurement*, 43(8), 597-610. <http://dx.doi.org/10.1177/0146621618824855>
- Altıntaş, O. & Wallin, G. (2021). Equality of admission tests using Kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, 8(4), 729-743. <https://doi.org/10.21449/ijate.976660>
- Andersson, B., Branberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating using the R package kequate. *Journal of Statistical Software*, 55, 1-25.
- Andersson, B., Branberg, K., & Wiberg, M. (2022). *Package 'kequate'*. <https://cran.r-project.org/web/packages/kequate/kequate.pdf>
- Berger, N., Mackenzie, E., & Holmes, K. (2020). Positive attitudes towards mathematics and science are mutually beneficial for student achievement: a latent profile analysis of TIMSS 2015. *The Australian Educational Researcher*, 47, 409-444. <https://doi.org/10.1007/s13384-020-00379-8>
- Branberg, K. & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement*, 48(4), 419-440. <https://doi.org/10.1111/j.1745-3984.2011.00153.x>
- Büyüköztürk, Ş. (2016). Sınavlar üzerine düşünceler. *Kalem Eğitim ve İnsan Bilimleri Dergisi*, 6(2), 345-356. <https://doi.org/10.23863/kalem.2017.64>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education* (8th ed.). McGraw-Hill Humanities/Social Sciences/Languages.
- Gonzalez, J., & Wiberg, M. (2017). *Applying test equating methods using R*. Springer.
- Guilford, J. P. (1956). *Fundamental statistics in psychology and education* (3th ed.). McGraw-Hill.
- International Association for the Evaluation of Educational Achievement. (2021). *TIMSS 2019 international database*. [Data set]. https://timss2019.org/international-database/?_gl=1*bf0qid*_ga*NDQyNzY0MjI0LjE2NDI2MjMxOTU.*_ga_L2FMXN42HR*MTY0MjYyMzE5NS4xLjAuMTY0MjYyMzE5NS4w
- Kim, S. & Lu, R. (2018). *The pseudo-equivalent groups approach as an alternative to common-item equating* (Research Report No. ETS RR-18-02). <https://doi.org/10.1002/ets2.12195>

- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking* (3rd ed.). Springer.
- Laukityte, I., Wallin, G., & Wiberg, M. (2026). Combining propensity scores and common items for test score equating. *Applied Psychological Measurement*, 50(3), 115-144. <https://doi.org/10.1177/01466216251363240>
- Livingston, S. A. (2004). *Equating test scores (without IRT)* (2nd. ed.). Educational testing service.
- Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd. ed.). Educational testing service.
- Lu, R. & Guo, H. (2018). *A simulation study to compare nonequivalent groups with anchor test equating and pseudo-equivalent group linking*. (Research Report No. ETS RR-18-08). <https://doi.org/10.1002/ets2.12196>
- Mao, X., von Davier, A. A., & Rupp, S. (2006). *Comparisons of the Kernel equating method with the traditional equating methods on Praxis™ Data*. ETS Research Report Series, 2006(2). <https://files.eric.ed.gov/fulltext/EJ1111483.pdf>
- Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. (2017). *Akademik becerilerin izlenmesi ve değerlendirilmesi 8. sınıflar raporu*. Milli Eğitim Bakanlığı. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_11/30114819_iY-web-v6.pdf
- Özer-Özkan, Y. & Turan, S. (2021). Düşündürdükleri ve ötesiyle yüksek riskli sınav gerçeği. *Alanyazın Eğitim Bilimleri Eleştirel İnceleme Dergisi*, 2(1), 59-63. <http://dx.doi.org/10.22596/cresjournal.0201.59.63>
- Özsoy, Ş. N. (2021). *Kernel eşitleme yöntemlerinin karşılaştırılması: TIMSS 2019 fen testi örneği*. [Yayınlanmamış yüksek lisans tezi]. Bolu Abant İzzet Baysal Üniversitesi.
- Revelle, W. (2022). *Package 'psych'*. <https://cran.r-project.org/web/packages/psych/psych.pdf>
- Sarı, M. H., Arıkan, S., & Yıldızlı, H. (2017). 8. sınıf matematik akademik başarısını yordayan faktörler TIMSS 2015. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(3), 246-265. <https://doi.org/10.21031/epod.303689>
- Sezer Başaran, E., Mutluer, C., & Çakan, M. (2023). A comparison of covariates, equating designs, and methods in equating TIMSS 2019 science tests. *Participatory Educational Research*, 10(5), 41-63. <https://doi.org/10.17275/per.23.74.10.5>
- Skaggs, G., & Lissitz, R. W. (1986). An exploration of the robustness of four test equating models. *Applied Psychological Measurement*, 10(3), 303-317. <https://doi.org/10.1177/014662168601000308>
- Suh, Y., Mroch, A. A., Kane, M. T., & Ripkey, D. R. (2009). An empirical comparison of five linear equating methods for the NEAT design. *Measurement*, 7, 147-173. <http://dx.doi.org/10.1080/15366360903418048>
- Tekeli Akdemir, H. (2021). TIMSS 2019 Türkiye örnekleme göre 4. ve 8. sınıf öğrencilerin matematik başarı durumlarının çeşitli değişkenlere göre chaid analizi ile incelenmesi. [Yayınlanmamış yüksek lisans tezi]. Akdeniz Üniversitesi.
- von Davier, A. A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). *An evaluation of the kernel equating method: A special study with pseudotests constructed from real test data*. ETS Research Report Series, 2006(1), i-31. <https://doi.org/10.1002/j.2333-8504.2006.tb02008.x>

- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. Springer.
- Wallin, G. & Wiberg, M. (2016). Nonequivalent groups with covariates design using propensity scores for Kernel equating. Paper presented in *Quantitative Psychology The 81st Annual Meeting of the Psychometric Society*, Asheville, North Carolina.
- Wallin, G. & Wiberg, M. (2019). Kernel equating using propensity scores for nonequivalent groups. *Journal of Educational and Behavioral Statistics*, 44(4), 390-414. <https://doi.org/10.3102/1076998619838226>
- Wiberg, M. (2021) On the use of different linkage plans with different observed-score equipercentile equating methods. *Practical Assessment, Research, and Evaluation*, 26(23), 1-16. <https://doi.org/10.7275/21481778>
- Wiberg, M. & Branberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, 39(5), 349-361. <https://doi.org/10.1177/0146621614567939>
- Yıldırım Seheryeli, M. (2022). *Ortak değişkenlerin özelliklerine göre Kernel eşitleme sonuçlarının karşılaştırılması*. [Yayınlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Yurtçu, M., Kelecioğlu, H., & Boone, E. L. (2021). The comparison of the equated tests scores by various covariates using Bayesian nonparametric model. *Journal of Measurement and Evaluation in Education and Psychology*, 12(2), 192-211. <http://dx.doi.org/10.21031/epod.864744>

EXTENDED ABSTRACT

Introduction

When direct comparison of different test forms is not feasible, test equating is conducted (Crocker & Algina, 1986; Kolen & Brennan, 2014). Test equating is a statistical procedure that ensures the comparability of scores across tests with similar content and difficulty levels (Kolen & Brennan, 2014). The kernel equating method stands out due to its flexibility and low error rate (Akın-Arkan & Gelbal, 2018; von Davier et al., 2004). Test equating designs represent another factor that directly influences the equating results (Kolen & Brennan, 2014). The equivalent groups (EG), nonequivalent groups with anchor test (NEAT), and nonequivalent groups with covariates (NEC) designs are among the principal designs used in this field (Kolen & Brennan, 2014; von Davier et al., 2004). Particularly in NEC, the relationship between the common covariates and test performance determines the accuracy of the equating (Wiberg & Branberg, 2015). However, systematic analyses based on the correlation levels of these covariates remain limited (Özsoy, 2021; Yıldırım Seheryeli, 2022). This study aims to fill this gap.

The purpose of this study is to compare the equating errors obtained from the NEC, and NEAT designs using the kernel equating method in the TIMSS 2019 eighth-grade mathematics tests. In the NEC design, student- and school-level characteristics were utilized as common covariates. The selected covariates include: (i) science achievement, (ii) home educational resources, (iii) confidence in learning mathematics, (iv) like learning mathematics, (v) the school's emphasis on academic success, and (vi) socioeconomic level. The research questions are as follows:

Method

Within the framework of a basic research model, the participants of this study consisted of 6,973 eighth-grade students who participated in the TIMSS 2019 e-TIMSS assessment and

responded to booklets B1, B2, B13, and B14. In the study, mathematics achievement scores were represented by the plausible values (BSMMAT01–BSMMAT05) computed by the TIMSS research team. For the common covariates, the categorical values provided by TIMSS were used.

Kernel equating was conducted in five stages: (i) pre-smoothing, (ii) estimation of score probabilities, (iii) continuization, (iv) computation of the equating function, and (v) evaluation of equating accuracy. The obtained equating results were evaluated using criteria such as RMSD, RMSE, WMSE, and SEE. During the equating process, the new forms (test X) were designated as booklets B1 and B14, while the old forms (test Y) were designated as booklets B2 and B13.

Results and Discussion

An examination of the item statistics for the X and Y tests administered in the study revealed that the X and Y tests, as well as the common tests, exhibited similar levels of difficulty and reliability across all groups. This finding indicates that the X and Y tests to be equated are statistically comparable. The kernel equatings were evaluated based on PRE values and equated scores. The results demonstrated that, in both groups (NEAT, and NEC), the kernel equipercentile and linear equating methods yielded generally acceptable levels of accuracy (Kolen & Brennan, 2014; von Davier et al., 2004). When examining the correlations between mathematics achievement and the common covariates, strong relationships were found for science achievement ($r \geq .70$), moderate relationships for confidence in learning mathematics and home educational resources ($r = .40-.69$), and weak relationships for the remaining variables ($r = .20-.39$).

In the NEC designs using highly correlated variables, the lowest SEE value was obtained in the NEAT design. Differences between the NEC and NEAT designs were mostly negligible. These findings indicate that the NEC designs more closely resemble the NEAT design. For moderately correlated variables, the NEC designs generally diverged slightly more from the NEAT design, with differences remaining within negligible limits. In the NEC designs employing weakly correlated variables, all SEE and RMSD values were found to be further from those of the NEAT design. Furthermore, similar interpretations were reached across all analyses when examining the RMSD, WMSE, RMSE, and difference plots.

The findings of the study revealed that the relationships between the covariates and mathematics achievement significantly influenced the equating results. In particular, NEC designs employing highly correlated variables (science achievement) produced results more similar to NEAT. Conversely, when weakly correlated variables were used, the NEC design yielded results further to the NEAT. These findings are consistent with the literature suggesting that the effectiveness of common covariates in equating depends on their strength of association with the target score (Albano & Wiberg, 2019; Sezer Başaran et al., 2023; Wiberg & Branberg, 2015).

For moderately correlated variables (confidence in learning mathematics and home educational resources), the differences among the designs varied across groups. Confidence in learning mathematics yielded results more similar to the NEAT, whereas home resources produced results further to the NEAT. This finding indicates that not only the correlation but also the explanatory power of the variable in accounting for ability differences between groups plays a determining role. The literature likewise emphasizes that the effects of such variables vary contextually (Berger et al., 2020; Sari et al., 2017; Sezer Başaran et al., 2023).

Overall, NEC designs incorporating highly correlated variables enhance equating accuracy, while moderately correlated variables provide only limited improvement. Weakly correlated variables, on the other hand, fail to adequately capture differences between groups. This suggests that not only the level of correlation but also the distribution and explanatory power of the common covariate in accounting for variance should be taken into consideration.

EKLER

Ek 1. Ortak Değişkenler ve Özellikleri

<i>Değişken</i>	<i>TIMSS 2019'daki Adı</i>	<i>Kodu</i>	<i>Kategori Sayısı*</i>	<i>Veri Kaynağı</i>
1. Fen bilgisi başarısı	Plausible value science	BSSIBM01-BSSIBM05	5	Fen dersine ait başarı testi
2. Evdeki eğitim kaynakları	Home educational resources	BSDGHER	3*	Öğrenci anketi
3. Matematiğe karşı özgüven	Student confident in mathematics	BSDGSCM	3*	Öğrenci anketi
4. Matematik öğrenmeyi sevme	Students like learning mathematics	BSDGSLM	3*	Öğrenci anketi
5. Okulun akademik başarıya verdiği önem	School Emph on Acad Success-Prncpl	BCDGEAS	3*	Okul anketi
6. Sosyoekonomik düzey	School composition by socioeconomic background	BCDGSBC	3*	Okul anketi

Not. *Veri analizinde kategoriler için ters kodlama yapılmıştır.

Ek 2. Kernel Eşitlemelerdeki Bant Genişlikleri

<i>No</i>	<i>Desen</i>	<i>Birinci Grup</i>		<i>İkinci Grup</i>	
		<i>Eşit Yüzdelikli</i>	<i>Lineer</i>	<i>Eşit Yüzdelikli</i>	<i>Lineer</i>
1	ST	$h_X = 0,5215767$ $h_Y = 0,5434823$	$h_X = 4013,735$ $h_Y = 3697,019$	$h_X = 0,5226339$ $h_Y = 0,5333531$	$h_X = 4045,544$ $h_Y = 4278,42$
2	Zincirleme	$h_{XP} = 0,5192624$ $h_{XQ} = 0,543403$ $h_{aP} = 0,5224511$ $h_{aQ} = 0,5177481$	$h_{XP} = 4030,623$ $h_{XQ} = 3677,911$ $h_{aP} = 5898,526$ $h_{aQ} = 5845,012$	$h_{XP} = 0,5240862$ $h_{XQ} = 0,5320392$ $h_{aP} = 0,5357077$ $h_{aQ} = 0,5355315$	$h_{XP} = 4032,302$ $h_{XQ} = 4288,762$ $h_{aP} = 6522,411$ $h_{aQ} = 6519,925$
3	Fen1	$h_X = 0,5203193$ $h_Y = 0,5461513$	$h_X = 4058,926$ $h_Y = 3661,612$	$h_X = 0,5229702$ $h_Y = 0,5332937$	$h_X = 4042,069$ $h_Y = 4280,838$
4	Fen5	$h_X = 0,5204086$ $h_Y = 0,5457565$	$h_X = 4053,94$ $h_Y = 3664,42$	$h_X = 0,5226919$ $h_Y = 0,5336214$	$h_X = 4040,902$ $h_Y = 4283,866$
5	Ozguven	$h_X = 0,5210516$ $h_Y = 0,5435744$	$h_X = 4021,937$ $h_Y = 3685,617$	$h_X = 0,5225907$ $h_Y = 0,5334412$	$h_X = 4041,357$ $h_Y = 4282,268$
6	Evkayn	$h_X = 0,5202312$ $h_Y = 0,5435967$	$h_X = 4026,902$ $h_Y = 3681,228$	$h_X = 0,5242436$ $h_Y = 0,5326158$	$h_X = 4030,914$ $h_Y = 4290,731$
7	Onem	$h_X = 0,5196315$ $h_Y = 0,5436392$	$h_X = 4030,768$ $h_Y = 3678,236$	$h_X = 0,5238645$ $h_Y = 0,5329418$	$h_X = 4033,856$ $h_Y = 4286,633$
8	Sevme	$h_X = 0,5201845$ $h_Y = 0,5435973$	$h_X = 4027,089$ $h_Y = 3680,297$	$h_X = 0,5237462$ $h_Y = 0,532919$	$h_X = 4033,252$ $h_Y = 4287,894$
9	SED	$h_X = 0,5195342$ $h_Y = 0,5435709$	$h_X = 4030,417$ $h_Y = 3678,086$	$h_X = 0,5240941$ $h_Y = 0,5327079$	$h_X = 4031,537$ $h_Y = 4289,853$

Ek 3. Eşitleme Uygulamalarında Kernel Lineer ve Kernel Eşit Yüzdelikli Eşitleme Yöntemleriyle Eşitlenmiş Puanlar

<i>Desen</i>	<i>BİRİNCİ EŞİTLEME</i>						<i>İKİNCİ EŞİTLEME</i>					
	<i>Lineer</i>			<i>Eşit Yüzdelikli</i>			<i>Lineer</i>			<i>Eşit Yüzdelikli</i>		
	<i>0</i>	<i>8</i>	<i>16</i>	<i>0</i>	<i>8</i>	<i>16</i>	<i>0</i>	<i>8</i>	<i>16</i>	<i>0</i>	<i>8</i>	<i>16</i>
<i>ST</i>	0.19	7.55	14.92	0	7.75	14.79	0	8.18	16.64	0.04	8.10	16.81
<i>Zinc</i>	0.28	7.64	15	0.06	7.92	14.70	0	8.11	16.62	0	8.00	16.87
<i>Özg</i>	0	7.28	14.61	0	7.35	14.72	0	8.36	16.84	0.10	8.38	16.83
<i>EvKay</i>	0	7.19	14.50	0	7.23	14.69	0.02	8.53	17	0.13	8.62	16.89
<i>Sevme</i>	0	7.17	14.48	0	7.21	14.68	0	8.47	16.97	0.12	8.53	16.87
<i>Fen1</i>	0.48	7.70	14.91	0	7.89	14.80	0	8.40	16.88	0.13	8.43	16.84
<i>Fen5</i>	0.39	7.62	14.85	0	7.78	14.79	0	8.35	16.83	0.12	8.35	16.82
<i>Önem</i>	0	7.13	14.43	0	7.16	14.66	0	8.50	17	0.13	8.58	16.87
<i>SED</i>	0	7.11	14.41	0	7.13	14.66	0	8.51	17	0.13	8.59	16.88

Not: Birinci eşitleme uygulamasındaki Y testi 15 maddeden; ikinci eşitleme uygulamasındaki Y testi 17 maddeden oluşmaktadır.