



Amasya Üniversitesi
Eğitim Fakültesi Dergisi
8(1), 24-61, 2019
Özgün araştırma makalesi

<http://dergi.amasya.edu.tr>

Eğilim Puanı Eşleştirme Analizinin Eğitim Araştırmalarında Kullanılması**

Esin Yılmaz Koğar* 

Niğde Ömer Halisdemir Üniversitesi, Türkiye

Alındı: 16.10.2018 - Düzeltildi: 07.04.2019 - Kabul Edildi: 24.04.2019

Atf: Yılmaz Koğar, E. (2019). Eğilim puanı eşleştirme analizinin eğitim araştırmalarında kullanılması. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 8(1), 24-61.

Öz

Eğitim araştırmalarında önemli sorunları incelemek için rastgele atama yöntemlerinin kullanıldığı deneysel araştırmalar çoğu zaman mümkün olmamakta veya etik sorunlar nedeniyle tercih edilmemektedir. Bu nedenle eğitim araştırmalarında seçilen karşılaştırma grupları arasında rastgele atama işleminin gerçekleştirilememesinin ortaya çıkaracağı olumsuzlukları gidermek için eğilim puanı eşleştirilmesi gibi farklı yöntemler kullanılmaktadır. Eğilim puanı, rastgele atanmanın olası olmadığı durumlarda kullanılan veya gözlemsel verilerden nedensel sonuçlar elde etmek için geliştirilmiş bir koşullu olasılıktır. Eğilim puanı eşleştirilmesi,

* Sorumlu yazar: Tel.: 0 388 2254422, E-posta: esinyilmaz@ohu.edu.tr

**Bu makalenin bir kısmı 5-8 Eylül 2018 Prizren/KOSOVA'da gerçekleştirilen 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sunulmuştur. ISSN: 2146-7811, ©2019

karşılaştırma analizindeki grupları istatistiksel olarak daha fazla eşdeğer hale getirmek için kullanılmaktadır. Bu analiz ile çok sayıda işlem öncesi ortak değişken tek bir skaler fonksiyona indirgenmekte ve araştırmacılara işlem etkisini belirlemede daha karşılaştırılabilir gruplar sağlamaktadır. Bu makalede (a) eğilim puanı eşleştirme analizinin metodolojisi anlatılmış ve (b) PISA 2015 Türkiye örnekleme üzerinden elde edilen veriler yardımıyla eğilim puanı eşleştirmesinin eğitim araştırmalarında kullanımını gösteren bir örnek sunulmuştur. Araştırmanın sonucunda eğilim puanı eşleştirmesi ile grupların daha karşılaştırılabilir bir duruma getirildiği belirlenmiştir. Eğilim puanı eşleştirmesi öncesinde karşılaştırılan gruplar (özel ve devlet okulunda öğrenim görme) okuryazarlık puanları bakımından anlamlı farklılıklar gösterirken eşleştirme sonrasında gruplar arasında okuryazarlık puanları arasında anlamlı farklılık bulunmamıştır. Eğilim puanı eşleştirmesi, regresyon temelli yöntemlerin varsayımlarının karşılanmadığı durumlarda da kullanılabilirdiğinden eğitim araştırmalarında bu analiz yöntemine daha fazla yer verilmesi önerilmektedir.

Anahtar Kelimeler: Eğilim Puanı, Ortak Değişken, Eşleştirme

Giriş

Deneysel çalışmalarda, bağımlı değişken üzerinde gözlemlenen değişimlerin gerçekten işlem etkisinden kaynaklı olup olmadığını belirlemek önemlidir. Bu nedenle çalışmalarını güçlendirmek isteyen araştırmacılara katılımcıların rastgele atamalarla gruplandırıldığı gerçek deneysel desen kullanmaları önerilmektedir (Büyüköztürk, 2007). Charness, Gneezy ve Kuhn (2012) da rastgele atamanın, kullanılabilir sonuçlar üretmek için güvenilmesi gereken güçlü bir yöntem olduğunu ifade etmiştir. Çünkü rastgele atama yapıldığında sonucu etkileyebilecek tüm potansiyel değişkenlerin eşdeğer olduğu ve grupların dengelendiği varsayılır (Fan ve Nowell, 2011). Ancak lojistik, etik, politik ve ekonomik nedenlerden dolayı çoğu sosyal bilimler ve eğitim bilimleri çalışmalarında rastgele atamanın yapılması mümkün olmayabilir (Bostian, 2008; Luellen, Shadish ve Clark, 2005; Titus, 2007). Bu durumda da rastgele olmayan atama süreci, gruplar arasında farklılıklara

yol açabilir ve sonucu etkileyebilir. Çünkü rastgele oluşturulmayan gruplar, herhangi bir sayıda ortak (kovaryant) değişkene dayanarak sistematik bir şekilde farklılık gösterebilir (Rosenbaum ve Rubin, 1983) ve gruplar arasındaki bu farklılıklar araştırma tasarımında göz önünde bulundurulmadığında araştırma sonuçları üzerinde yanlış bir işlem etkisine ulaşılabilir (Grunwald ve Mayhew, 2008). Bu nedenle rastgele olmayan atama yönteminden kaynaklı problemlere veya seçim yanlılığı sorunlarına karşı önlemler alınmalıdır. Ayrıca deneysel olmayan çalışmalarda da grup karşılaştırmalarını daha etkili hale getirmek oldukça önemlidir. Tüm bu durumların göz önünde bulundurulmasıyla araştırmacıların karşılaştıkları bu gibi sorunlarla başa çıkabilmeleri için eğilim puanı eşleştirme-EPE (Propensity Score Matching-PSM) analizi geliştirilmiştir.

EPE, karşılaştırma analizindeki grupları istatistiksel olarak daha fazla eşdeğer hale getirmek için kullanılacak bir analizdir ve bu sayede nedensel çıkarımların temelini daha güçlü kılar (Fan ve Nowell, 2011). Heckman (1978) çalışmasında, atama işleminin rastgele olmadığı zaman bağımlı değişkendeki işlem etkisi konusuna odaklansa da eğilim puanı eşleştirmesi terimi ilk olarak Rosenbaum ve Rubin'in (1983) gözlemsel verilerle yürüttüğü çalışmada kullanılmıştır. EPE, gruplar arasında karşılaştırma yaparken tahmin yanlılığının azaltılması ve gözlemsel verilerden nedensel çıkarım yapmak amacıyla geliştirilmiştir (Rosenbaum ve Rubin, 1983, 1984). Bir eğilim puanı, birtakım ortak değişkenler göz önüne alındığında kişinin durumunun (örn. kontrol grubunda mı, işlem grubunda mı) tahmin edilmesinde kullanılan bir koşullu olasılıktır ve bu nedenle bütün olasılıklar gibi 0 ile 1 arasında değer alır. Eğilim puanı için eşitlik şu şekildedir (Rosenbaum ve Rubin, 1983):

$$e(x) = \text{prob}(z = 1|x)$$

$e(x)$ verilen değişkenlerden kişinin işlem grubunda olmasına ilişkin koşullu olasılığı, $z=1$ işlem grubu ve $z=0$ kontrol grubu olacak şekilde grup göstergesini ve x , gözlenen ortak değişkenler kümesini göstermektedir. Bu yaklaşımla, ilgili

ve uygun ortak değişkenler tanımlandığında istatistiksel olarak bu ortak değişkenler üzerinde eşdeğer gruplar oluşturulur ve daha sonra alt gruplar üzerinde istatistiksel testler uygulanır (Fan ve Nowell, 2011). Bir başka ifadeyle gözleme dayalı çalışmalarda gruplar arasındaki ortak değişkenlerin dengesizliği giderilmeye çalışılır. Özetle eğilim puanı, dengelenmiş bir puandır.

Eğilim puanlarını belirlemede hangi ortak değişkenlerin kullanılacağını belirlemek oldukça önemlidir (Shadish, Clark ve Steiner, 2008). Araştırmacı kendi çalışmasındaki ortak değişkenleri teorik bir yaklaşımla literatürdeki çalışmaları göz önünde bulundurarak seçebilir. Ancak genellikle işlem ve kontrol grupları arasında istatistiksel olarak anlamlı farkların bulunduğu değişkenler, ortak değişkenler olarak kullanılmaktadır (Luellen ve diğ., 2005). Baser (2006) de grup üyeliği ve sonuç değişkeni ile ilişkileri olan değişkenlerin, ortak değişken olarak seçilebileceğini ancak ilgili tüm değişkenlerin tanımlandığına %100 emin olunamayacağını belirtmiştir. Literatürde ortak değişkenlerin sayısı bakımından herhangi bir sınırlama belirtilmemiştir. Ancak ortak değişken sayısı arttığı zaman aynı veya benzer eşleştirilmiş çiftler bulmak zorlaşacaktır (Altunok, Bekiroğlu ve Genceli, 2005). Yine de ortak değişken sayısını çok düşük tutmaya gerek yoktur. Çünkü eğilim puanı işlem etkisini tahmin ederken bir skaler fonksiyon olarak hareket eder (Adelson, 2013). Bu teknikte her ne kadar zengin bir ortak değişken kümesi isteniyorsa da araştırmacılar eldeki verilerin niteliğine ve varsayımların ihlâl edilme olasılığına karşı duyarlı olmalıdırlar (Guo ve Fraser, 2010).

Ortak değişkenler belirlendikten sonra grup değişkeninin bağımlı, ortak değişkenlerin ise bağımsız değişkenler olarak ele alındığı bir model kurulur. Model seçiminde logit ya da probit modeller kullanılabilir. Ancak literatür incelendiğinde model olarak logit modellerin daha çok kullanıldığı görülmektedir (Fan ve Nowell, 2011; Jiang ve McComas, 2015). Thommes ve Kim (2001) çalışmalarında eğilim puanlarını kullanan

çalışmaları incelemişler ve bu çalışmaların %77.9'unun logit, %11.6'sının probit modelleri kullandığını belirlemişlerdir.

Lojistik regresyon analizi kullanılarak eğilim puanı kestirilirken grup değişkeni iki kategorili ise ikili (binary), ikiden çok kategorili ise çok kategorili (multinomial) lojistik regresyon kullanılır. Eğilim puanının avantajı, çok boyutlu ortak değişkenlerinin tek boyutlu bir puana indirgenmesidir (Barth, Guo ve McCrae, 2008).

Eğilim puanları hesaplandıktan sonra bir eşleştirme işlemi gerçekleştirilmelidir. Eğilim puanları eşleştirmede birçok farklı yöntem bulunmaktadır ve günümüzde yeni eşleştirme yöntemleri hızla gelişmektedir (Ho, Imai, King ve Stuart, 2007). Araştırmacılar, denk olamayan grupları eşleştirmek (dengelemek) için eşleştirme (matching), tabakalara ayırma (stratification or subclassification), regresyon düzeltmesi (regression adjustment) veya eğilim puanı üzerinde ağırlıklandırma (weighting) kullanarak eğilim puanları hesaplayabilirler (Hair, 2015). Eşleştirme işleminde Rosenbaum ve Rubin (1985) kestirilen eğilim puanına bağlı olarak mümkün olan en yakın eşleştirme, eğilim puanını içeren mahalnobis metrik eşleştirmesi, eğilim puanları tarafından tanımlanan aralıklar arasında mümkün olan en yakın mahalnobis metrik eşleştirmesi şeklinde üç farklı teknik önermişlerdir. Bu teknikler hakkında ayrıntılı bilgiler için Hirano ve Imbens (2001), Lunceford ve Davidian (2004), Rosenbaum ve Rubin (1983, 1984, 1985), Caliendo ve Kopeinig (2008) çalışmalarına bakılabilir. Bu çalışmada en yakın eşleştirme yöntemi üzerine odaklanılmaktadır. Çünkü farklı yöntemleri karşılaştıran Austin (2011), eğilim puanı analizlerinde eşleştirmenin tabakalandırma ve regresyon düzeltmesinden daha iyi olduğunu belirtmiştir.

Araştırmacı eşleştirme işleminde gruplar arasında dengenin gerçekten gerçekleşip gerçekleşmediğinin kontrol etmek için çeşitli istatistikler gerçekleştirebilir. Eşleştirmeden önce ve sonra, çoğu zaman standartlaştırılmış ortalama farklılıklar ve varyans oranı ile işlem ve kontrol grubunun birkaç istatistiğinin karşılaştırılması yapılabilir. Örneğin

eşleştirme işleminden sonra ortak değişkenlerin standartlaştırılmış ortalama farkı 0'a varyans oranı ise 1'e yakın olmalıdır (Thoemmes, 2012). Eşleştirme ile gruplar arasında istatistiksel olarak sistematik farklılıklar olmayan yeni bir örneklem oluşturularak artık birbiriyle karşılaştırılabilir ve dengeli gruplar elde edilmiş olur (Lane, To, Shelley ve Henson, 2012). Bu durumda son olarak işlem etkisini test etmek için t testi, varyans analizi gibi çalışmanın amacına uygun olan herhangi bir hipotez testi uygulanır. Böylece gruplar arasında işlem etkisinin olup olmadığı test edilmiş olur. Biri işlem diğeri kontrol grubunda olan kişilerin eğer aynı veya benzer bir eğilim puanı varsa bu kişilerin her bir gruba rastgele olarak atandığı (Baser, 2006) ve hem gözlenen hem de gözlenmeyen önceden mevcut olan özellikler bakımından dengelendikleri düşünülebilir (Lee, 2010). Böylece yalnızca benzer veya eşit eğilim puanlarına sahip olan kişiler karşılaştırılarak, işlem ve kontrol grubu arasındaki ortak değişkenlerin her birinin dağılımı arasında bir denge oluşturulur ve bu durum deneysel olmayan çalışmalarda etkili sonuçlara ulaşılmasını sağlar.

Grupların ortak değişkenler bakımından dengesiz olması durumuyla baş etmek için regresyon tabanlı teknikler de (ör. kovaryans analizi-ANCOVA, yapısal eşitlik modellemesi) kullanılmaktadır. Lee (2010) eğilim eşleştirme yaklaşımının, regresyon tabanlı modellerin tüm ön koşul varsayımlarının karşılaması ve iyi desteklenmiş verilere sahip olunması halinde bu modellerle benzer tahminler üreteceğini ancak bu koşullar sağlanamadığı zaman, eğilim puanı eşleştirme yönteminin parametrik yaklaşımlara göre belirgin avantajları olduğunu ifade etmiştir. Örneğin EPE yalnızca benzer durumları eşleştirerek ciddi uyumsuzluklardan kaçınırken standart regresyon yöntemleri, bu uyumsuzlukları tutacak ve gerçekçi olmayan bir ortalama işlem etkisi yaratacaktır. ANCOVA kullanarak nedensel etkileri tanımlamak teorik olarak mümkün olsa da grupların ortak değişkenlerin dağılımı bakımından çok farklı olduğu durumlarda eğilim puanlarının kullanılması daha doğrudur (Dehejia ve Wahba, 1999; Rubin, 1997). Çünkü ortak değişkenlerinin dağılımında büyük ölçüde farklılık gösteren

işlem ve kontrol grupları, regresyon modellerinin temel varsayımını – ortak değişkenlerin ve sonucun doğrusal bir ilişki içerdiğini – bozar ve bu nedenle güvenilir sonuçlara yol açabilir (Newgard, Hedges, Arthur ve Mullins, 2004). Buna karşın eğilim puanı, her gruptaki sonuç ve işlem öncesi ortak değişkenler arasındaki ilişki için doğrusallık gibi herhangi bir özel forma dayanmamaktadır (Rubin, 1997). Bu avantajlarına karşın eğilim puanları eşleştirmesi sosyal bilimlerde yaygın olarak kullanılan bir işlem olmasa da günümüzde farklı alanlarda kullanımı artmaktadır (Thoemmes, 2012). Özellikle ekonomi ve sağlık alanında oldukça sık kullanılmaya başlanan EPE (Barth ve diğ., 2008) eğitim araştırmalarında benimsenmesi ve kullanılması ise oldukça yavaş olan bir analiz olmuştur. Bu durum kısmen yaklaşımın anlaşılmasından ve eğitim araştırmacıları arasındaki uygulamalarının az olmasından kaynaklı olabilir (Fan ve Nowell, 2011).

Eğitim ortamlarında çoğu zaman önemli yapıları ve soruları incelemek için rastgele atama yöntemlerinin kullanıldığı deneysel çalışmalar yapmak olanaksızdır veya etik değildir ve bu nedenle eğitim araştırmalarında genelde gözleme dayalı yöntemlerden elde edilen veriler kullanılır (Adelson, 2013). EPE, eğitim alanındaki gözlemsel çalışmalarda, sistematik hatanın azaltılmasını ve belirli ortak değişkenlerin etkilerini ortaya koymaya yardımcı olur. Ayrıca bulgulardaki nedensel çıkarımın geçerliğini de daha fazla arttırılabilir. Amerika Birleşik Devletleri Eğitim Bakanlığı (2003) iyi eşleştirilmiş gruplar üzerinden yürütülen çalışmaların eğitimde değerli bir rolü olduğunu belirtmiştir.

Birçok sanayileşmiş ülkede eğitim politikaları hakkında tartışmaların ve uluslararası araştırmaların bir parçası olarak öğrencilere Uluslararası Öğrenci Değerlendirme Programı (Program for International Student Assessment-PISA), Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study-TIMSS) gibi deneysel olmayan uygulamalar yapılmaktadır. Bu çalışmada, uluslararası platformda gerçekleştirilen bu uygulamalardan PISA üzerine odaklanılmıştır. 15 yaşındaki öğrencilerin okuma,

fen ve matematik alanındaki okuryazarlıklarının uluslararası olarak değerlendirilmesini kapsayan PISA, ilk olarak 2000 yılında yürütülmüş olup her üç yılda bir tekrarlanmaktadır. Her PISA uygulamada bir alan daha kapsamlı olmak üzere üç alan da değerlendirilmektedir. PISA uygulamalarına dayalı olarak yapılan puan hesaplamalarında birçok farklı değişkenin etkisi olan söz konusu olabilmektedir. Ancak eğitimsel araştırmalarda bu değişkenleri belirlemek kolay değildir. Çünkü eğitimsel durumları uygun bir şekilde ayarlama işi oldukça karmaşıktır (Graesser, 2009).

PISA uygulamalarında öğrenci hakkında birçok bilgiye ulaşılmasını sağlayan anketler uygulanmaktadır. Araştırmacılar da bu anketlerde yer alan değişkenleri çalışmalarında kullanarak öğrenci okuryazarlıkları üzerinde yorumlar yapmaktadırlar. Okul türü değişkeni bu yorumlamalarda kullanılan değişkenlerden biridir. Aileler kendi çocukları için seçecekleri okulun birçok farklı faktörünü göz önünde bulundurlar ve bu faktörlerden biri de o okulda öğrenim gören öğrencilerin göstermiş oldukları akademik performanstır. Bu nedenle PISA uygulamalarında, öğrenci performanslarını okul türü değişkenini (devlet-özel) göz önüne alarak inceleyen çalışmalar bulunmaktadır (Dronkers ve Avram, 2009; Dronkers ve Robert, 2008a; 2008b; Vandenberghe ve Robin, 2003). Dronkers ve Robert (2008a; 2008b) PISA 2000 veri seti ile eğilim puanı analizleri yaparak devlet, yarı-devlet ve özel okul türü kıyaslamaları ile ilgili bir çalışma yürütmüşler, okuma ve matematik test puanlarının bu okul türlerine göre anlamlı farklılık gösterdiğini belirlemişlerdir. Dronkers ve Avram (2009) farklı ülkelerin PISA sonuçlarını alarak eğilim puanı eşleştirmesi ile okuma becerilerini incelemişlerdir. Fransa, İngiltere ve Amerika'daki özel bağımsız ve devlet okulları arasında eğilim puanı eşleştirmesinden önce okuma başarısında fark bulunmuşken eğilim puanı eşleştirme sonrasında önemli bir fark olmadığını ortaya koymuşlardır. PISA dışında veriler kullanılarak öğrencilerin akademik başarılarını okul türüne göre inceleyen ve eğilim puanı analizleri kullanan başka

araştırmalar da alanyazında bulunmaktadır (Bridgewater, 2013; Fan ve Nowell, 2011; Hahs-Vaughn ve Onwuegbuzie, 2006).

Adelson (2013) okumada üstün yetenekli programı olan (480) ve olmayan (370) okulları okuma becerileri bakımından karşılaştırmıştır. Bölge, ortalama sosyoekonomik düzey, okul amaçları gibi farklı değişkenleri eşleştirme değişkeni olarak kullanarak bir eğilim puanı eşleştirilmesi yapmıştır. Başlangıçta gruplar arasında anlamlı farklılıklar belirlemişken eşleştirme puanlarına göre oluşturduğu tabakalarda farklı üstün yetenekli okuma programı politikaları olan okullar arasında istatistiksel olarak anlamlı bir farklılık bulamamıştır. Vandenberghe ve Robin (2003) akademik başarının, aile ve okul kaynaklarının bir işlevi olduğunu varsayarak yürüttükleri çalışmalarında okul türlerinin başarı üzerindeki etkisini kestirmeyi amaçlamışlardır. Özel veya devlet okulunun başarı üzerindeki etkisini aile kaynakları ya da sosyo-ekonomik düzey gibi diğer belirleyici faktörlerden ayırmak için farklı ülkelerde eğilim puanları eşleştirmesi yapmışlardır. Ortalama yanlılığının eşleştirme işlemi yapıldıktan sonra birçok ülke için azaldığını belirlemişlerdir. Fan ve Nowell (2011) özel okulda (227) ve devlet okulunda (500) öğrenim gören öğrencilerin matematik akademik performanslarını incelemek için yürüttükleri çalışmada sosyoekonomik düzey ve aile beklentisi değişkenlerini ortak değişken olarak belirlemiş ve bir eğilim puanı eşleştirmesi yapmışlardır. Araştırmanın sonucunda bu ortak değişkenlere göre dengelenen iki grup arasındaki matematik performansındaki farklılığın hâlâ özel okul lehine olduğunu ve özel okullardaki öğrencilerin daha iyi performans göstermesinin ailelerin sosyoekonomik düzeylerine ve ebeveynlerinin eğitim kazanımlarına ilişkin beklentilerine bağlanamayacağını belirlemişlerdir. Fan ve Nowell'in (2011) çalışmasında Adelson'ın (2013), Vandenberghe ve Robin'in (2003), Dronkers ve Avram'ın (2009) bulgularından farklı olarak eşleştirme işlemi sonrasında da ele alınan bağımlı değişken bakımından gruplar arasında anlamlı farklılıklar bulunmuştur.

Eğitim alanında yapılan deneysel çalışmalarda rastgele atamalar genellikle öğrenci düzeyinde değil grup düzeyinde

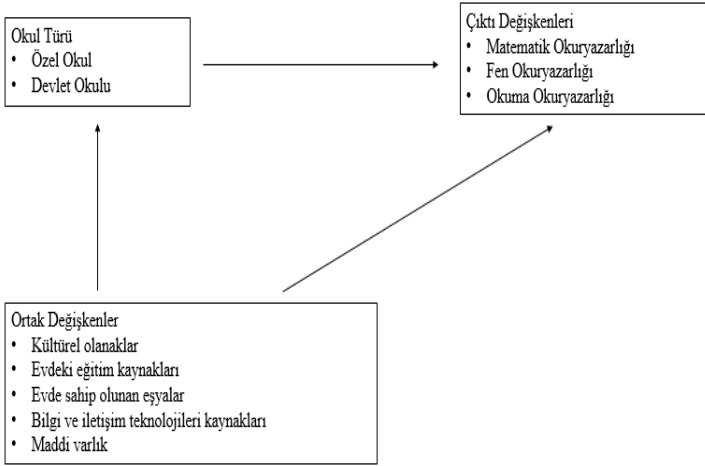
(okul ya da sınıf) gerçekleştirilmektedir (Schochet, 2008). Bu nedenle bu çalışmada da grup düzeyinde bir inceleme yapmak amaçlanarak yukarıda belirtildiği gibi başka çalışmalarda da değişken olarak alınan okul türü değişkeni seçilmiştir. Okul türü değişkeninin çalışma için bağımsız değişken olarak seçilmesinin bir diğer nedeni ise bu okullarda öğrenim gören öğrencilerin benzer sosyoekonomik ve sosyokültürel olanaklarından kaynaklı oluşan korelasyon nedeniyle oluşabilecek durumların etkisi giderildikten sonra inceleme yapabilmektir. Böylece gerçekten okul türünün literatürdeki çalışmalarda belirtildiği gibi öğrenci başarısı üzerinde farklılık oluşturup oluşturmadığı görülebilecektir.

Literatür incelendiğinde akademik performansla ilgili yapılan araştırmalarda, sosyoekonomik ve sosyokültürel özelliklerin genellikle akademik performansla ilgili olduğu kabul edilmektedir (Barth ve diğ., 2008; Dronkers ve Avram, 2009; Fan ve Nowell, 2011; Levine ve Painter, 2003; Nguyen, Taylor ve Bradley, 2006). Bu nedenle bu çalışmada öğrencilerin sosyoekonomik ve sosyokültürel düzeylerinin birer göstergesi olan PISA öğrenci anketinde yer alan sorulara verilen cevaplardan oluşan kültürel olanaklar, evdeki eğitim kaynakları, evde sahip olunan eşyalar, bilgi ve iletişim teknolojileri kaynakları, maddi varlık değişkenleri ortak değişkenler olarak seçilmiştir.

Bu çalışmada, deneysel olmayan araştırmalarda işlem etkisinin değerlendirilmesinde kullanılan etkili bir analiz olan EPE ile ilgili kuramsal çerçevenin sunulması ve eğitim alanında EPE'nin nasıl yapılacağına ilişkin bir örnek uygulama yapılması amaçlanmaktadır. Eğilim puanı eşleştirmesinin daha önce belirtildiği gibi olumlu yönleri olsa da ikincil araştırmalarda ve eğitim araştırmalarında oldukça az karşılaşılan bir analizdir (Marts, 2016). Bu durum çalışmanın önemini arttırmaktadır. Bu eksiklikten yola çıkarak bu çalışmada, eğitim alanında ikincil analizler yapma olanağı sunan PISA 2015 verisindeki değişkenler kullanılmıştır.

Bu çalışmada kültürel olanaklar, evdeki eğitim kaynakları, evde sahip olunan eşyalar, bilgi ve iletişim

teknolojileri kaynakları, maddi varlık değişkenleri kontrol altına alındığında öğrencilerin okuma, fen ve matematik okuryazarlığı puanlarının okul türüne göre istatistiksel olarak anlamlı bir farklılık oluşturup oluşturmadığı eğilim puanları eşleştirmesi ile belirlenmeye çalışılmıştır. Çünkü eğitim alanında böyle bir karşılaştırma için deneysel bir tasarım mümkün görülmemekte, özel ve devlet okullarında öğrenim gören öğrencilerin akademik performanslarını potansiyel olarak etkileyebilecek birçok önemli özellik eşdeğer olarak düşünülmektedir (Fan ve Nowell, 2011). Ayrıca iyi tanımlanmış gruplar oluşturulduğunda gözlemsel çalışmalar deneysel çalışmalarla benzerlik göstermekte ve işlem etkisinin ampirik olarak araştırılması söz konusu olabilmektedir (Hahs-Vaughn ve Onwuegbuzie, 2006). Bu çalışmada kullanılan model Şekil 1'deki gibidir.



Şekil 1. Model

Yöntem

Bu çalışmada öğrencilerin okuma, fen ve matematik okuryazarlığı puanlarının öğrencilerin öğrenim gördüğü okul türüne göre anlamlı bir farklılık gösterip göstermediğini belirlemek amacıyla gruplar arasında bir karşılaştırma

yapılmaktadır. Bu nedenle bu araştırma nicel araştırma türlerinden biri olan nedensel karşılaştırma türündedir.

Çalışma Grubu

Çalışmada kullanılan veriler PISA 2015 Türkiye örnekleminde elde edilen verilerdir. Çalışmada kullanılan değişkenlere göre veri seti incelenerek çalışma grubu oluşturulmuştur. Türkiye örnekleminde okul türü için altı tane bağımsız özel okul, bir tane devlete bağlı özel okul ve 179 tane devlet okulu bulunmaktadır. Bir tane okulun ise okul türünün belirtilmediği gözlenmiş ve bu okul koduna sahip 38 öğrenciye ait veriler, veri setinden çıkarılmıştır. Devlete bağlı özel okulların sayısının yalnızca bir olması nedeniyle bu okuldaki öğrenciler de (36) veri setinden çıkarılmıştır. Bu şekilde okullar devlet okulu ve özel okul olacak şekilde iki kategorili olarak biçimlendirilmiştir. Daha sonra kontrol değişkeni (ortak değişken) olarak belirlenen değişkenlerde kayıp verileri olan öğrenciler (188) veri setinden silinmiştir. Kayıp veriler silindikten sonra her bir okul türünde kaç öğrenci olduğu incelenerek PISA 2015 uygulamasına devlet okulundan katılan 5429 ve özel okuldan katılan ise 204 öğrenci olduğu belirlenmiştir. Rubin (2007) eğilim puanı analizinin büyük örneklerde daha iyi sonuç vereceğini belirtmiştir. Baser (2006) da işlem grubundaki kişilerle daha iyi eşleşme bulma olasılığını arttırmak için kontrol grubundaki kişi sayısının mümkün olduğunca geniş tutulması gerektiğine değinmiştir. Bu nedenle eğilim puanı belirlenirken bu 5633 (5429 devlet ve 204 özel) kişilik veri seti kullanılmıştır.

Veri Toplama Aracı

Veriler PISA 2015 uygulamasındaki öğrenci anketi, okul anketi ve okuryazarlık testleri üzerinden elde edilen sonuçlardan oluşmaktadır. Öğrencilerin okuma, fen ve matematik okuryazarlıklarını belirlemek için kullanılan bilişsel testlerden elde edilen verilerin yanı sıra, okul türünü belirlemek için okul anketinden ve ortak değişken seçmek için öğrenci anketinden yararlanılmıştır. Öğrenci anketinden seçilen

ortak değişkenlere ilişkin aşağıda kısa bilgiler sunulmuştur (OECD, 2017):

Kültürel Olanaklar (CULTPOSS): Öğrencinin evinde klasik eserler, şiir kitapları, sanat eserleri, müzik, sanat veya tasarım kitapları, müzik enstrümanı gibi ürünlerin olup olmadığına ilişkin maddelere (ST011Q07TA, ST011Q08TA, ST011Q09TA, ST011Q16NA ve ST012Q09NA kodlu beş madde) verilen öğrenci cevaplarından oluşturulan bir indekstir.

Evdeki Eğitim Kaynakları (HEDRES): Öğrencinin evinde çalışma masasının, çalışmak için sessiz bir ortamın, okul ödevleri için kullanabileceği bir bilgisayarın ve kitapların, eğitim amaçlı yazılım programlarının ve sözlüklerin bulunup bulunmadığının sorulduğu maddelere (ST011Q01TA, ST011Q03TA, ST011Q04TA, ST011Q05TA, ST011Q10TA, ST011Q11TA ve ST011Q12TA kodlu yedi madde) oluşturulan bir indekstir.

Evde Sahip Olunan Eşyalar (HOMEPOS): Öğrencinin evindeki koşullara ve eşyalara ilişkin maddelere (ST011, ST012 ve ST013 kodlu 25 madde) verilen öğrenci cevaplarından oluşturulan bir indekstir.

Bilgi ve İletişim Teknolojileri Kaynakları (ICTRES): Öğrencinin evinde eğitimsel yazılımlar, internet bağlantısı, bilgisayar ve e-kitap okuyucu gibi bilgi iletişim kaynaklarının olup olmadığına ilişkin maddelere (ST011Q05TA, ST011Q06TA, ST012Q05NA, ST012Q06NA, ST012Q07NA ve ST012Q08NA altı kodlu madde) verilen öğrenci cevaplarından oluşturulan bir indekstir.

Maddi Varlık (WEALTH): Ailenin refah düzeyinin gösteren bu indeks otomobil, internet bağlantısı ve öğrenciye ait bağımsız odanın varlığı gibi ailenin gelir düzeyi ve ekonomik durumu ile ilgili maddelere (ST011Q02TA, ST011Q06TA, ST011Q17TA, ST011Q18TA, ST011Q19TA, ST012Q01TA, ST012Q02TA, ST012Q03TA, ST012Q05NA, ST012Q06NA, ST012Q07NA ve ST012Q08NA kodlu 12 madde) verilen öğrenci cevaplarından oluşturulan bir indekstir.

Bu indeksler oluşturulurken kullanılan ST011 kodlu maddeler evet/hayır, ST012 kodlu maddeler hiç/bir/iki/üç ya da daha fazla, ST013 kodlu madde ise 0-10 kitap/11-25 kitap/26-100 kitap/101-200 kitap/201-500 kitap/500 kitaptan daha fazla şeklinde cevap kategorileri içermektedir. Öğrencilerin bu maddelere verdikleri cevaplar doğrultusunda elde edilen veriler, madde tepki kuramıyla ölçeklendirilmiştir ve 25 maddeden beş farklı indeks oluşturulmuştur (OECD, 2017).

Çalışmada kullanılan bilişsel testler ise okuma, fen ve matematik okuryazarlığı testleridir. Bu testlerde yer alan maddelerin bazıları seçme gerektiren bazıları ise yazma gerektiren maddelerdir. Ancak PISA'daki bilişsel testlerde yer alan her madde her öğrenci tarafından yanıtlanmamaktadır. Bu nedenle öğrencilerin gözlenen yanıtları üzerinden gözlenemeyen yanıtları tahmin edilir ve bu tahmini yaparken madde tepki kuramı ile bir ölçekleme geliştirilir. Bu şekilde her öğrenci için 10 tane olası değer kestirimi (plausible value - PV) yapılır (OECD, 2013). Bu çalışmada bilişsel testlerden elde edilen okuma, fen ve matematik okuryazarlığı puanları için bu şekilde oluşturulan 10 tane olası puan bağımlı değişken olarak kullanılmıştır. Bu bağımlı değişkenlerin kullanıldığı analizler on kere tekrarlanıp sonuçların ortalaması alınarak raporlama işlemi gerçekleştirilmiştir.

Bağımsız değişken olarak ele alınan okul türü değişkenini belirlemek için okul anketinden yararlanılmıştır. Özel okulda öğrenim gören öğrenciler 1, devlet okulunda öğrenim gören öğrenciler 0 olacak şekilde iki kategorili olarak kullanılmıştır. Ayrıca bu çalışma deneysel bir çalışma olmasa da EPE'de işlem ve kontrol grubu terimleri kullanıldığı için işlem grubu olarak özel okul, kontrol grubu olarak ise devlet okulu ele alınmıştır. İşlem etkisi olarak ele alınan durum ise öğrencinin okul türünün okuryazarlık puanları üzerinde fark yaratıp yaratmadığıdır.

Verilerin Analizi

Eğilim puanları analizleri için R, Stata, SAS, SPSS gibi farklı istatistiksel yazılımlar bulunmaktadır. Bu çalışmada

analizlerin tamamı SPSS 24 programı ile gerçekleştirilmiştir. Çünkü SPSS sosyal bilimlerde oldukça sık kullanılan ve araştırmacıların kullanımına alışkın oldukları bir yazılımdır. Analizleri gerçekleştirmek SPSS'in R eklentisini kullanılmış ve böylece R'daki Matchit paketindeki işlevler SPSS'e dâhil edilmiştir.

Eşleştirme öncesi ve sonrası gruplar arasındaki farklılıkları incelemek için ise bağımsız örneklem için t testinden yararlanılmıştır. Bu analiz ile ortalama, standart sapma, t değerleri ve Cohen'in d (etki büyüklüğü) istatistiği hesaplanmıştır. Ayrıca iki gruptaki grup üyeliğini açıklayan arka plandaki faktörleri belirlemek ve eşleştirme tekniğinde kullanılan bir eğilim puanı oluşturmak için de lojistik regresyon analizi kullanılmıştır. Lojistik regresyon analizi ile işlem ve kontrol gruplarına üyeliği öngören önemli bağımsız değişkenler incelenmiştir.

Bulgular

Bu çalışmada belirlenen ortak değişkenlerin eşleştirme değişkeni olarak ele alınması ile eğilim puanı eşleştirmesi yapılarak gruplar daha dengeli hale getirildikten sonra öğrencilerin okuma, fen ve matematik okuryazarlık puanlarının okul türüne göre anlamlı farklılık gösterip göstermediği incelenmiştir. Eğilim puanı analizi birkaç adımlı bir analizdir ve eğilim puanı eşleştirmesi için bu çalışmada izlenen adımlardan sırasıyla bahsedilmiştir.

Eğilim puanı eşleştirmesi gerçekleştirilmeden önce işlem etkisinin kestirimi

Çalışmada ilk olarak eşleştirme işlemi gerçekleştirilmeden önce öğrencilere ait farklı okuryazarlık puanlarının öğrenim görülen okul türüne göre istatistiksel olarak anlamlı farklılık gösterip göstermediği kontrol edilmiştir. Elde edilen sonuçlara göre öğrencilerin okuma puanları okul türüne göre incelendiğinde puanların özel okul lehine anlamlı bir farklılık gösterdiği belirlenmiştir ($t_{(5631)}=3.41$, $p<.01$). Özel okulda

öğrenim gören öğrencilerin okuma okuryazarlık puanlarına ait ortalama (446.84), devlet okulunda öğrenim gören öğrencilerin ortalamalarından (427.41) daha yüksek çıkmış olup standart ortalama farklılıkları $d=0.25$ bulunmuştur. Bu nedenle ortalama bir etki büyüklüğünden söz edilebilir.

Fen okuryazarlık puanları okul türüne göre incelendiğinde yine puanların özel okul lehine anlamlı bir farklılık gösterdiği belirlenmiştir ($t_{(5631)}=4.93$, $p<.01$). Özel okulda öğrenim gören öğrencilerin fen okuryazarlık puanlarına ait ortalama (450.25), devlet okulunda öğrenim gören öğrencilerin ortalamalarından (423.01) daha yüksek çıkmış olup standart ortalama farklılıkları $d=0.34$ bulunmuştur. Bu nedenle ortalama bir etki büyüklüğünden söz edilebilir

Son olarak öğrencilerin matematik okuryazarlık puanları okul türüne göre incelendiğinde de yine puanların özel okul lehine anlamlı farklılık gösterdiği sonucuna varılmıştır ($t_{(5631)}=5.73$, $p<.01$). Özel okulda öğrenim gören öğrencilerin matematik okuryazarlık puanlarına ait ortalama (449.77), devlet okulunda öğrenim gören öğrencilerin ortalamalarından (417.02) daha yüksek çıkmış olup standart ortalama farklılıkları $d=0.40$ bulunmuştur. Bu nedenle ortalama bir etki büyüklüğünden söz edilebilir.

Sonuç olarak üç puan türünün de öğrencinin öğrenim gördüğü okula göre anlamlı farklılık gösterdiği, özel okulda öğrenim gören öğrencilerin her üç alanda da okuryazarlık puanları devlet okulunda öğrenim gören öğrencilerden daha yüksek olduğu belirlenmiştir. Ancak araştırmada bir karşılaştırma grubu kullanılmasına rağmen bu gruplar arasındaki farklılıklar, bu bulgunun yorumlanmasını önleyebilecek sistematik farklılıkların bir sonucu da olabilir (Lane ve diğ., 2012). Bu durumun önüne geçmek ve bu iki grup arasında daha iyi bir karşılaştırma yapmak için eğilim puanı eşleştirmesi kullanılabilir.

İşlem değişkeni ve ortak değişkenlerin seçimi

EPE’de eğilim puanları oluşturmak için bir grup değişkeni ve bir veya daha fazla ortak değişken gerekmektedir. Gruplama değişkeni, bir kişinin ait olduğu grubu belirten değişken (ör. işlem ve kontrol) iken ortak değişkenler, grupları eşitlemeye çalışılan öğelerdir (Randolph ve diğ., 2014). Bu çalışmada da grup değişkeni, okul türüdür (özel okul=1, devlet okulu=0). Ortak değişkenlerin seçiminde ise literatürde yapılan çalışmalardan yardım alınmıştır. Eğitim alanında yapılan çalışmalarda ailenin sosyal, ekonomik ve kültürel yapısının öğrencilerin akademik performansını etkilediği görülmektedir (Buchmann ve DiPrete, 2006; Kim, 2004; Koğar, 2015; Muller, 2018). Bu nedenle bu çalışmada CULTPOSS, HEDRES, HOMEPOS, ICRES ve WEALTH değişkenleri ortak değişkenler olarak seçilmiştir.

Ortak değişkenler belirlendikten sonra gruplar arasında bir eşitsizlik söz konusu değilse EPE’nin yapılması gereksiz olur. Rosenbaum (2002) işlem ve kontrol grupları arasında sadece istatistiksel olarak önemli ölçüde farklılıklar bulunan ortak değişkenlerin kullanılması gerektiğini belirtmiştir. Bu nedenle gruplar arasındaki ön farklılıkları belirlemek için bağımsız örneklem t testi yapılmıştır. Bunun yanı sıra seçilen değişkenlerin grup değişkeni üzerindeki etki büyüklükleri hesaplanmıştır. Tablo 1, iki grubun tanımlayıcı istatistiklerini ve beş değişken için yapılan bağımsız örneklem t testi sonuçlarını içermektedir.

Tablo 1. Beş ortak değişken üzerinde eşleştirilmemiş gruplara ilişkin t testi sonuçları

	Okul Türü				t	sd	p	d
	Özel (N=204)	Devlet (N=5429)	X	SS				
Ortak Değişkenler	X	SS	X	SS				
CULTPOSS	.47	.93	-.29	.86	12.338	5631	.000	.85
HEDRES	.14	.91	-.61	1.13	9.382	5631	.000	.73
HOMEPOS	.08	1.15	-1.49	1.09	20.261	5631	.000	1.40
ICRES	-.05	1.04	-1.23	.93	17.764	5631	.000	1.20

WEALTH	-	1.13	-1.54	.97	21.057	5631	.000	1.39
	.08							

Tablo 1’de sunulan istatistikler incelendiğinde seçilen değişkenlerin okul türüne (grup değişkenine) göre anlamlı farklılık gösterdiği görülmektedir. Bu nedenle ortak değişkenlere ait puan ortalamalarının okul türündeki iki grup üzerinde dengesiz dağıldığı söylenebilir. Ayrıca okul türü için standartlaştırılmış ortalama farka (d) dayanan etki büyüklükleri de göz önüne alındığında seçilen değişkenlerin geniş etki büyüklüğüne sahip olduğu görülmektedir. Bu durumda bu değişkenlerin öğrencilerin okuma, fen ve matematik okuryazarlık puanlarını etkileyebileceği düşünülebilir. Lane ve diğ. (2012) ortak değişkenler üzerinde gruplar arasındaki farkların büyük olması durumunda ANCOVA yapılmasının uygun olmadığını söyleyerek kendi çalışmalarında da ortak değişkenlerin etkisini kontrol altına almak için ANCOVA yerine eğilim puanı eşleştirmesi yapmışlardır. Bu çalışmanın bulgularına bakıldığında da gruplar arasında farklılıkların büyük olduğu ve ANCOVA yapılmasının uygun olmadığı söylenebilir. Bu nedenle okul türüne göre okuryazarlık puanları incelenmeden önce bu ortak değişkenlerin dengesizliğini gidermek için eğilim puanı eşleştirmesi yapılmalıdır. Modelde hangi bağımsız değişkenlerin (ortak değişkenler) bağımlı değişkenle (okul türü) ilişkili olduğunu belirlemek için lojistik regresyon gerçekleştirilmiştir. Elde edilen sonuçlar Tablo 2’deki gibidir.

Tablo 2. Özel okulda olma olasılığına ilişkin beş ortak değişkenle kurulan lojistik regresyon sonuçları

Ortak Değişkenler	B	SH	Wald	sd	p	Odds Oranı	Odds Oranları için %95’lik Güven Aralığı	
							Alt Değer	Üst Değer
CULTPOSS	-.427	.297	2.070	1	.150	.652	.365	1.167
HEDRES	-.532	.217	6.020	1	.014	.588	.384	.899
HOMEPOS	2.827	.866	10.651	1	.001	16.888	3.093	92.219
ICRES	-.648	.278	5.436	1	.020	.523	.304	.902

WEALTH	-.371	.659	.317	1	.573	.690	.189	2.511
--------	-------	------	------	---	------	------	------	-------

Bu beş ortak değişkenin bir öğrencinin özel okulda olma olasılığını belirlemek için bir model kurulmuştur (Model $\chi^2_{(5, 454)}=197.83$, $p<.001$, Cox & Snell $R^2=.353$, Nagelkerke $R^2=.473$). Ki-kare değerinin anlamlı olması, yordayıcı olarak kullanılan değişkenlerin dâhil olduğu modelin iyi olduğunu belirtir. Kullanılan yordayıcı değişkenlerle değişkenliğin %35'i ile %47'sinin açıklandığı sonucuna varılmıştır. Hosmer ve Lemeshow testine göre model desteklenmiştir (Model $\chi^2_{(8, 454)}=11.24$, $p>.05$). Oluşturulan modele göre özel okuldaki öğrencilerin %69.6'sı, devletteki öğrencilerin ise %79.2'si doğru sınıflandırılmış olup toplam doğru sınıflandırılma yüzdesi 74.9 olarak belirlenmiştir. Bu bulgular, bu modelin bu veri setinin eğilim puanı eşleştirmesi için makul olduğunu göstermektedir.

Eşleştirme algoritması

Bu çalışmada eğilim puanları eşleştirmede en yakın komşu eşleştirme kullanılmıştır. Bu eşleştirme algoritması, işlem grubundaki kişileri eğilim puanları kestirimine göre sıralar ve her bir kişiyi kontrol grubundaki en yakın eğilim puanı olan kişiyle eşleştirir. Eşleştirme oranı (ratio matching) 1:1, 2:1, 5:1 gibi farklı oranlarda olabilse de bu çalışmada işlem grubundaki her kişi kontrol grubundan yalnızca bir kişiyle eşleşecek şekilde 1:1 eşleştirilme seçilmiştir. Thoemmes ve Kim (2011) de işlem grubundaki katılımcıyla en benzer tahmini eğilim puanına sahip kontrol grubundaki katılımcının bire bir eşleştirildiği 1:1 en yakın komşu eşleştirme tekniğinin en yaygın olarak kullanılan yalın ve basit bir teknik olduğunu belirtmiştir. Ancak bu teknikle özellikle gruplar arasındaki birimlerin sayısı eşit değilse veri sayısı azalmaktadır.

Modelde ortak destek alanı dışındaki birimlerin atılması (discarded) veya eşleşen algoritmaya dâhil edilmesi tercih edilebilir. Ancak ortak değişkenler üzerindeki dengenin güçlenmesini sağlayabilmek ve diğer grupta benzer birimlerin bulunmadığı değişkenler üzerinde çok farklı olan bir gruptaki birimlerle eşleştirilmesini önleyebilmek amacıyla birimler atılabilir (Thommes, 2012). Bu nedenle bu çalışmada ortak

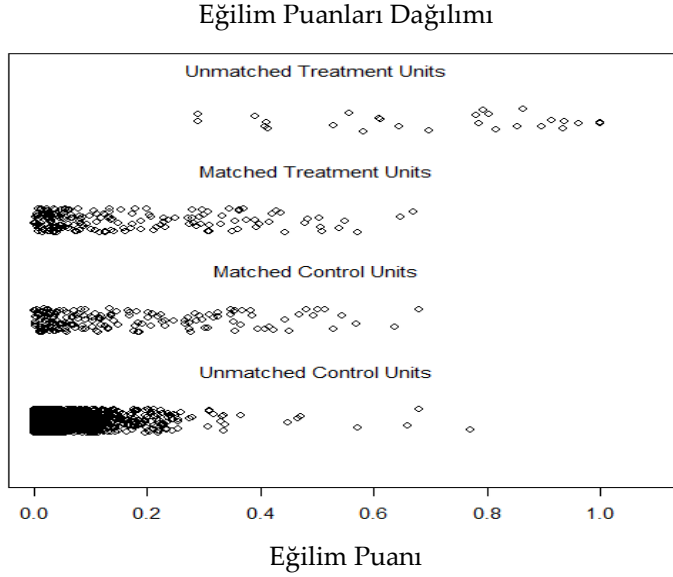
destek sağlamayan birimlerin her iki gruptan da atılması seçeneği kullanılmıştır. Ayrıca iyi bir eşleşme işleminin gerçekleştiğinden emin olmak için iki birimin kestirilen eğilim puanları bakımından birbirinden ayrı olabileceği maksimum mesafe olarak isimlendirilen “caliper” .15 olarak ayarlanmıştır. Böylece çok farklı eğilim puanları kestirimlerine sahip olan birimlerin eşleşmelerinin yani kötü eşleştirmelerin önüne geçmek amaçlanmıştır (Thommes, 2012).

Bu çalışmada 1:1 en yakın komşu eşleştirilmesi, ortak destek (common support). uygun olmayan birimlerin atılması ve caliper değerinin .15 olarak belirlenmesi ile yapılan EPE sonucunda gruptardaki kişi sayıları Tablo 3’te gösterilmiştir.

Tablo 3. Eşleştirme işlemi sonrasında gruptardaki kişi sayısı

	Kontrol (Devlet Okulu)	İşlem (Özel Okul)
Toplam Kişi Sayısı	5429	204
Eşleştirilen Kişi Sayısı	177	177
Eşleştirilmeyen Kişi Sayısı	4774	13
Atılan Kişi Sayısı	478	14

1:1 eşleştirme yönteminin kullanıldığı bu çalışmada her iki gruptan eğilim puanları benzer veya yakın olan 177 kişi olduğu belirlenmiştir. Modelde ortak destek bölgesinin dışında kalması nedeniyle veya uygun bir eşleşme bulunamayacağı için atılan öğrenciler bulunmaktadır. Kontrol grubundan 478, işlem grubundan ise 14 öğrenci dışarda tutulduğunda kontrol grubundan eşleştirme yapılmayan 4774 ve işlem grubundan 13 kişi olduğu görülmektedir. Bu tablonun daha net anlaşılmasına yardımcı olacağı düşüncesiyle her dairenin bir kişinin eğilim puanını gösterdiği jitter grafiğine Şekil 2’de yer verilmiştir.



Şekil 2. Eğilim puanlarına ilişkin jitter grafiği

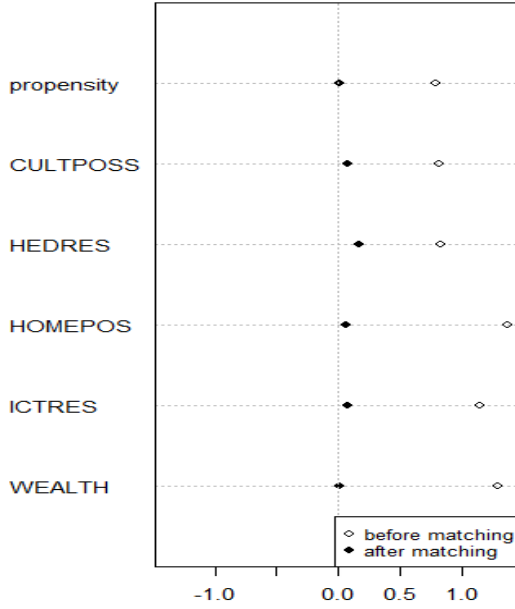
Şekil 2 işlem birimlerinden eşleşmeyen birimlerin olduğunu göstermektedir. Orta kısma bakıldığında işlem birimleri ile eşleşen kontrol birimleri arasında yakın bir eşleşme olduğu görülmektedir. Son kısım ise ilk kısım gibi ilerleyen kısımlardaki analizlerde kullanılmayacak eşleştirilmemiş kontrol ünitelerini göstermektedir.

Eğilim puanı eşleştirmesi sonrasında gruplar arasındaki dengenin değerlendirilmesi

Eğilim puanı eşleştirme işleminin dengeli olup olmadığını değerlendirmek için farklı stratejiler kullanılmıştır. Eşleştirilen 177 kişinin ortak değişkenler üzerindeki dengesinin eşleştirme prosedürüyle gerçekten elde edilip edilmediğini kontrol etmek amacıyla eşleştirme öncesi ve sonrası gruplar arasındaki ortalamalara ve standartlaştırılmış ortalama farklılıklara ilişkin bulgulara Tablo 4'te yer verilmiştir. Ayrıca Şekil 3'te eşleştirme işleminden önce ve sonra tüm ortak değişkenler için standartlaştırılmış ortalama farklılıkların (Cohen'in d) bulunduğu noktasal grafik (dot plot) sunulmuştur.

Tablo 4. Eşleştirme öncesi ve sonrası gruplara ilişkin betimsel istatistikler

Ortak Değişkenler	İşlem Grubu		Kontrol Grubu		Std. Ortalama Farklılıklar	
	Ortalama		Ortalama		Farklılıklar	
	Önce	Sonra	Önce	Sonra	Önce	Sonra
propensity	.223	.151	.029	.150	.785	.006
CULTPOSS	.469	.348	-.288	.276	.810	.076
HEDRES	.139	.086	-.614	-.062	.829	.163
HOMEPOS	.082	-.185	-1.494	-.255	1.369	.060
ICTRES	-.047	-.253	-1.235	-.330	1.146	.075
WEALTH	-.082	-.353	-1.545	-.365	1.295	.011



Şekil 3. Önce ve sonra tüm ortak değişkenler için standartlaştırılmış ortalama farklılıkların bulunduğu noktasal grafik

Tablo 4 ve Şekil 3 birlikte incelendiğinde eşleştirme sonrası gruplar arasındaki standart ortalama farklılıkların azaldığı görülmektedir. Rubin (2001), iki grup arasındaki ortalama eğilim puanında standartlaştırılmış farkın sifıra yakın olması gerektiğini önermektedir. Tablo 4'teki bulgulara göre bu

ölçütün karşılandığı görülmektedir. Dolayısıyla eğilim puanı eşleştirme işlemi sonrasında gruplar arasında bir denge olduğu söylenebilir. Eşleştirme işlemi gerçekleştirildikten sonra en yüksek mutlak ortalama fark HOMEPOS değişkeni için bulunmuştur.

Dengeleme işlemi, tek değişkenli ve çok değişkenli yaklaşımların bir kombinasyonu kullanılarak ortak değişkenler arasında bakımından incelenmiştir. Okul türü değişkenine göre ortak değişkenlerden üzerinden eşleştirme öncesi yapılan bağımsız örneklem t testi sonuçları her bir ortak değişken için istatistiksel olarak anlamlı bulunmuştur (bkz. Tablo 1). Eşleştirme sonrasında ise bu analiz tekrar edildiğinde elde edilen bulgular Tablo 5'teki gibidir.

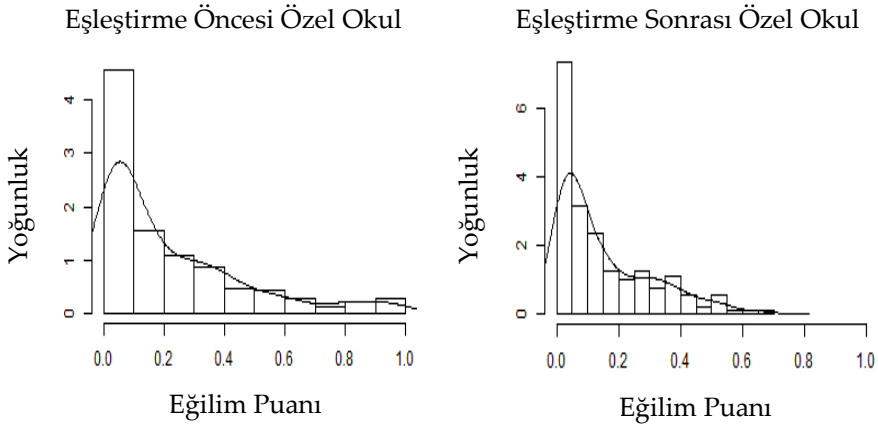
Tablo 5. Eşleştirme sonrası ortak değişkenler üzerindeki dengeler

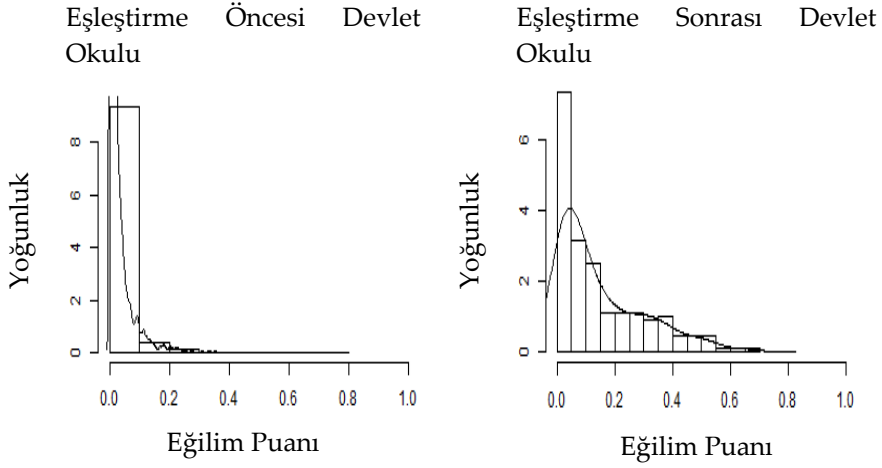
Ortak Değişkenler	Okul Türü				t	sd	p	d
	Özel (N=177)		Devlet (N=177)					
	X	SS	X	SS				
CULTPOSS	.35	.88	.28	.84	.781	352	.435	.08
HEDRES	.09	.91	-.06	.96	1.486	352	.138	.16
HOMEPOS	-.19	.89	-.25	.95	.715	352	.475	.07
ICRES	-.25	.88	-.33	.94	.803	352	.423	.09
WEALTH	-.35	.86	-.37	.86	.141	352	.888	.02

Eğilim puanı eşleştirme sonrasında standartlaştırılmış ortalama farkın .25'ten büyük olup olmadığı kontrol edilir (Cochran ve Rubin, 1973) ve bu farkın sifıra yakın olması önerilir (Rubin, 2001). Tablo 5'ten bu çalışmada belirtilen değerden büyük hiçbir değer olmadığı ve bu değerlerin sifıra yakın olduğu görülmektedir. Bu nedenle gruplar arasında ortak değişkenler üzerinde dengesizlik olmadığı sonucuna varılmıştır. Ayrıca gruplar arasındaki ortak değişkenlerin önemli farklılıkları kaldırıldığında eşleştirme başarılı kabul edilir (Pufahl ve Weiss, 2007).

SPSS eşleştirmeden sonra oluşturulan gruplar arasındaki dengeyi değerlendirmek için iki farklı çok değişkenli test sonucu vermektedir. Bunlardan ilki eşleştirmeden sonra herhangi bir değişken veya herhangi bir doğrusal değişken kombinasyonunda önemli ölçüde dengesizlik olup olmadığını eş zamanlı olarak değerlendirmek için Hansen ve Bowers (2008) tarafından geliştirilen χ^2 testidir ve bu testin anlamlı çıkmaması dengesizlik olmadığına işaret eder. Bu çalışmada toplam χ^2 denge testi anlamlı bulunmamıştır $\chi^2_{(5)} = 3.934$, $p = .559$. Bir diğer çok değişkenli dengesizlik ölçüsü Blackwell, Iacus, King ve Porro (2009) tarafından geliştirilen $L1$ ölçüsüdür. Tüm ortak değişkenlerin tamamen ortak dağılımlarını (full joint distribution) eş zamanlı olarak test eden bu istatistik için iyi veya kötü dengeyi gösteren bir kesme noktası yoktur (Thommes, 2012). 0 ile 1 arasında değer alan bu ölçünün eşleştirilmiş örnekleme, eşleştirilmemiş örnekleme daha küçük bir değer ulaşılması istenir. Bu çalışmada da $L1$ ölçümü eşleştirme işlemi gerçekleştirilmeden önce .984 olarak hesaplanmışken eşleştirme sonrası .797 olarak bulunmuştur.

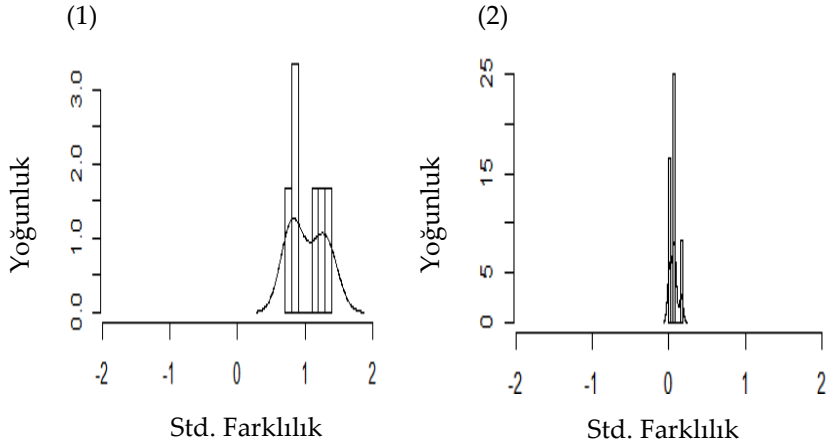
Tüm bu aşamalar sonucunda işlem ve kontrol grubu olarak ele alınan özel okuldaki ve devlet okulundaki öğrencilerin eşleştirme öncesi ve sonrası eğilim puanlarına ilişkin dağılım Şekil 4'teki gibidir.





Şekil 4. Eşleştirme işlemi öncesi ve sonrası özel okulu ve devlet okulu eğilim puanları dağılımı

Şekil 4'ten de görüldüğü üzere sol tarafta bulunan eşleştirme öncesindeki histogram grafikleri gruplar arasında oldukça farklıdır. Sağdaki eşleştirmeden sonra oluşturulan histogramlar ise iki grupta birbirine oldukça benzerdir. Şekil 5'te de eşleşme öncesi ve sonrası standartlaştırılmış farkların çekirdek yoğunluğu tahminleri (kernel density estimates) verilmektedir.



Şekil 5. Eşleştirmeden önce (1) ve sonra (2) standartlaştırılmış farklılıkların üst üste gelen çekirdek yoğunluğu tahminlerine sahip histogramlar

Şekil 5'e göre veri setinde eşleştirme sonrası yoğunluk değerlerinde artış, standartlaştırılmış farklarda ise bir azalış gözlenmiştir. Özetle hem sayısal hem de görsel veriler eşleşmenin başarılı olduğunu gösterir.

Eğilim puanı eşleştirilmesi gerçekleştirildikten sonra işlem etkisinin kestirimi

Eğilim puanları ve ortak değişkenlere göre gruplar arasında denklik sağlandıktan sonra gruplar bağımlı değişken bakımından incelenebilir hale gelir. Bu eşleştirilmiş gruplar arasında bulunan herhangi bir farklılığın gerçek işlem etkisini daha fazla yansıttığı düşünülür ve bu işlem etkisi, işlem grubundaki ortalama işlem etkisi (average treatment effect on the treated-ATT), kontrol grubundaki ortalama işlem etkisi (average treatment effect on the controlled-ATC) veya popülasyondaki ortalama işlem etkisi (average treatment effect on the population-ATE) olarak incelenebilir (Lane ve diğ., 2012). Bu çalışmada işlem etkisi, işlem grubundaki ortalama işlem etkisi (ATT) ile kontrol grubundaki ortalama işlem etkisi (ATC) arasındaki farkın kıyaslamasına dayalı olarak incelenmiştir. Bu amaçla gruplar arasında eşleştirme işlemi gerçekleştirildikten sonra bağımlı değişken olan öğrencilerin okuma, fen ve matematik okuryazarlık puanlarının okul türlerine göre incelenmesi, bir başka deyişe işlem etkisinin (treatment effect) belirlenmesi için bağımsız örneklem t testi yapılmıştır. Elde edilen bulgular Tablo 6'daki gibidir.

Tablo 6. Eşleştirme işlemi sonrasındaki t testi sonuçları

	Okul Türü				t	sd	p	d
	Özel (N=177)		Devlet (N=177)					
Değişkenler	X	SS	X	SS				
Okuma	444.03	79.55	445.86	79.69	-	352	.217	.02
Fen	447.69	82.13	449.26	78.39	-.188	352	.692	.02
Matematik	445.43	83.66	447.57	78.62	-.245	352	.747	.03

Tablo 6 incelendiğinde eşleştirme öncesi bağımsız örneklem t testinden farklı olarak okul türünün, öğrencilerin

okuma, fen ve matematik okuryazarlık puanları üzerinde istatistiksel olarak anlamlı bir etkisinin olmadığı görülmektedir. Eşleştirme öncesi her okuryazarlık puanına ilişkin özel okul ortalamaları daha yüksek iken eşleştirme sonrası devlet okuluna ilişkin okuryazarlık ortalamalarının daha yüksek olduğu belirlenmiştir. Ancak bu farklılık istatistiksel olarak anlamlı olmadığından öğrencilerin her üç okuryazarlık puanları üzerinde işlem etkisinin yani okul türünün önemli olmadığı belirlenmiştir ($p>.05$). Bu nedenle etki büyüklükleri yorumlanmamıştır.

Tartışma ve Yorum

Bu çalışmada okul türünün öğrencilerin akademik başarıları üzerindeki “doğru” etkisini tahmin etmek için eğilim puanı eşleştirmesi yapılmıştır. Çünkü standart regresyon yaklaşımlarına göre daha az istatistiksel varsayımları olan EPE’ye dayalı olarak daha iyi bir karşılaştırma grubu bulunabileceği düşünülmüştür. ANCOVA ile de ortak değişkenlerin etkisi giderilebilse de Dehejia ve Wahba’nın (1999) belirttiği gibi farklı gruplardaki öğrenciler arasında, ortak değişkenlerin dağılımları benzer olduğunda ANCOVA yapmak etkilidir. Gruplar çok farklı olduğunda eğilim puanı tekniği ile daha iyi sonuçlara ulaşılmaktadır. Bu çalışmada da grupların ortak değişkenler bakımından büyük farklılık göstermesi nedeniyle EPE yapılmıştır.

Yapılan analizler sonucunda eşleştirme işlemi öncesinde öğrencilerin okuma, fen ve matematik okuryazarlık puanlarının öğrenim görülen okul türü değişkenine göre anlamlı farklılık gösterdiği belirlenmiştir. Ancak bu durum gruplarda yer alan öğrencilerin başka ortak özellikleri arasındaki ilişki nedeniyle de bu şekilde bir sonuç vermiş olabilir. Bu durumun okul türünden ziyade öğrencinin sahip olduğu sosyoekonomik ve sosyokültürel olanaklardan da kaynaklanmış olabileceği düşüncesiyle çalışmada CULTPOSS, HEDRES, HOMEPOS, ICRES ve WEALTH değişkenleri ortak değişken olarak ele alınmıştır. Başlangıçta CULTPOSS, HEDRES, HOMEPOS,

ICRES ve WEALTH olarak belirlenen ortak değişkenler bakımından grupların dengeli olmadığı, bu dengeleme işleminin eğilim puanı eşleştirmesi ile iyi bir şekilde gerçekleştiği belirlenmiştir. Eşleştirme işlemi sonrasında okul türü değişkenine göre öğrencilerin okuma, fen ve matematik okuryazarlık puanları üzerinde anlamlı farklılık olmadığı, yani bir işlem etkisi görülmediği belirlenmiştir. Bu çalışmada eğilim puanı eşleştirmesi sonrasında oluşturulan gruplar arasında bağımlı değişken bakımından anlamlı bir farklılık olmadığı belirlenmiştir. Benzer şekilde eğilim puanı eşleştirmesi sonrasında gruplar arasında bağımlı değişken bakımından anlamlı farklılıkların olmadığı çalışmalar bulunmaktadır (Adelson, 2013; Vandenberghe ve Robin, 2003; Dronkers ve Avram, 2009). Bu bulgudan farklı olarak alanyazında eşleştirme işlemi sonrasında işlem etkisinin bulunduğu çalışmalar da bulunmaktadır (Fan ve Nowell, 2011; Leow, Marcus, Zanutto ve Boruch. 2004).

Sonuçlar

Sistematik hatayı azaltmaya yardımcı olan eğilim puanı analizi son yıllarda gözleme dayalı çalışmalarda kullanılan önemli bir yöntemdir. Deneysel olmayan gözlemsel çalışmalarda, deneysel çalışmalarda olduğu gibi bireylerin gruplara rastgele atanması üzerinde araştırmacının bir kontrolü bulunmamaktadır (D'Agostino, 1998). Bu durum, gözlemsel verilere dayalı çalışmalarda neden-sonuç ilişkisi hakkında etkili yorumlar yapabilme üzerine bir kısıtlılık oluşturabilmektedir. EPE, gözlemsel çalışmalarda farklı gruplarda yer alan örneklemelerin gözlenen ortak değişkenler bakımından daha dengeli olmasını sağlayarak neden-sonuç ilişkisi üzerine olan yorumların kuvvetlenmesine yardımcı olmaktadır.

Kandemir ve Bülbül (2018) karşılaştırılacak gruplara atanan birimlerde araştırmacının kontrolünün olmaması durumunun yapılacak tahminlerin de hatalı olmasına yol açtığını belirterek gözlemsel veriler üzerinden yaptıkları analizler sonucunda eğilim puanlarının sapsmaları azaltarak

sonuçları daha anlamlı hale getirdiğini belirlemişlerdir. Bu çalışma sonucunda da grupların ortak değişkenler üzerinde dengelenmesini sağlayan bu tekniğin oldukça kullanışlı olduğu yorumu yapılabilir. Ancak diğer çoğu istatistiksel tekniklerde olduğu gibi bazı sınırlılıkları bulunmaktadır. EPE'nin iyi sonuç vermesi uygun ortak değişkenlerin belirlenmesine bağlıdır. Ancak araştırmacıların bütün ilgili değişkenleri belirlediklerine ve bunlara ulaşabildiklerine emin olmaları konusuna ilişkin şimdilik bir çözüm getirilmemiştir (Fan ve Nowell, 2011). EPE oluştururken kullanılacak değişkenler için bir havuz oluştururken dikkatli davranılmalıdır (Yanovitzky, Zanuttob ve Hornik, 2005). EPE alt sınıflara ayırma ya da tabakalandırma gibi işlemler gerektirdiğinden örneklem büyüklüğünün geniş olması gerekmektedir. Aksi takdirde bazı alt sınıflardaki birim sayısı küçük olacak ve bu durum bu alt sınıflar için grup ortalaması ve grup farkının kararsız tahminine yol açarak bu tür analizlerden elde edilen bulguların geçerliliğini tehlikeye sokabilir. Bununla birlikte, bu yaklaşımın araştırma uygulamasında sınırlamalar ve uygulama zorlukları olduğu için deneysel olmayan tasarım veya gözlemsel araştırmalar için bütün sorunları giderici bir yol olarak düşünmemelidir (Fan ve Novell, 2008).

EPE, özellikle sağlık alanında yaygın olarak kullanılmakta olup (Austin, 2011; Beal ve Kupzyk, 2014; Boening ve diğ., 2003; Chen ve diğ., 2008; Grzybowski ve diğ., 2003; Hall, Summers ve Obenchain, 2003) eğitim alanında kullanımına ait çalışmalara daha az rastlanılmaktadır (Frye, 2014; Kelcey, 2011; Shapiro ve Trevino, 2004). Slavin (2002) eğitim araştırmalarına ilişkin literatürde eğilim puanları eşleştirmesinin yetersiz kaldığını belirtse de literatüre bakıldığında günümüzde bu konuda yapılan çalışmalarda artış gözlemlendiği söylenebilir. Ayrıca Kandemir ve Bülbül (2018) çalışmaları sonucunda özellikle tıp alanında oldukça sık olarak kullanılan eğilim puanı analizlerinin sosyal bilimlerde de güvenle kullanılabileceğini belirtmişlerdir. Bu durumlardan yola çıkılarak yapılan bu araştırmanın eğitim alanı için önemli olduğu sonucuna varılmıştır.

Öneriler

Bu çalışmada SPSS üzerinden analizler yürütüldüğü ve SPSS'te duyarlılık analizi (sensitivity analysis) şu anda mevcut olmadığı için bu analizler yapılamamıştır. Çalışmanın kullanılan programdan kaynaklanan bu sınırlılığı R, Stata gibi farklı programlar kullanılarak ortadan kaldırılabilir. Bu araştırmada PISA verileri üzerinden ikincil bir analiz gerçekleştirilerek eğilim puanı eşleştirmesi yapılmıştır. Benzer bir araştırma, araştırmacıların kendileri tarafından toplanan verilerle gerçekleştirilebilir.

Bu çalışmada eşleştirme işlemi için 1:1 en yakın komşu eşleştirmesi kullanılmıştır. Diğer eşleştirme işlemleri ile benzer çalışmalar tekrar edilebilir. Daha sonra bu eşleştirmelerden elde edilen sonuçlar karşılaştırılarak en iyi sonucu veren eşleştirme işlemiyle analizlere devam edilebilir.

Eğitim araştırmalarında büyük ölçekli deneysel çalışmalar gerçekleştirmek pek mümkün değildir ve gerçek deneysel desen oluşturmak zordur. Bu nedenle rastgele oluşturulamayan gruplar üzerinde çalışılması gereken ya da gözlemsel veriler elde edilen eğitim araştırmalarında eğilim puanı eşleştirmesi yapılması önerilmektedir.

Kaynaklar

- Adelson, J. L. (2013). Educational research with real-world data: Reducing selection bias with propensity scores. *Practical Assessment, Research & Evaluation, 18*(15), 1-11.
- Altunok, E. Ç., Bekiroğlu, N. ve Genceli, M. (2015). Gözleme dayalı çalışmalarda propensity skor: Bir uygulama. I. Kan (Ed.), *VIII. Ulusal Biyoistatistik Kongresi* (ss. 191-202) içinde, 20-22 Eylül 2005, Uludağ Üniversitesi, Bursa.
- Austin, P. C. (2011). A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behavioral Research, 46*(1), 119-151.
- Barth, R. P., Guo, S., & McCrae, J. S. (2008). Propensity score matching strategies for evaluating the success of child and

- family service programs. *Research on Social Work Practice, 18(3)*, 212-222.
- Baser, O. (2006). Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health, 9(6)*, 377-385.
- Beal, S. J. & Kupzyk, K. A. (2014). An introduction to propensity scores: What, when, and how. *The Journal of Early Adolescence, 34(1)*, 66-92.
- Blackwell, M., Iacus, S., King, G. & Porro, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal, 9(4)*, 524-546.
- Boening, A., Friedrich, C., Hedderich, J., Schoettler, J., Fraund, S., & Cremer, J. T. (2003). Early and medium-term results after on-pump and off-pump coronary artery surgery: a propensity score analysis. *The Annals of Thoracic Surgery, 76(6)*, 2000-2006.
- Bostian, B. E. (2008). *Avoiding remedial education: Academic effects on college transfer students*. (Doctoral dissertation). ProQuest Dissertations and Theses. (Accession Order No. AAT 3320959)
- Bridgewater, F. D. (2013). *The effects of school type on kindergarten reading achievement: Comparing multiple regression to propensity score matching*, Unpublished Master of Thesis. University of Wisconsin-Milwaukee.
- Büyüköztürk, Ş. (2007). *DeneySEL desenler: Öntest-sontest kontrol grubu desen ve veri analizi* (2. baskı). Ankara: Pegem A Yayıncılık
- Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22(1)*, 31-72.
- Chen, Y. S., Lin, J. W., Yu, H. Y., Ko, W. J., Jerng, J. S., Chang, W. T., Chen, W. J., Huang, S. C., Chi, N. H., Wang, C. H., Chen, L. C., Tsai, P. R., Wang, S. S., Hwang, J. J. & Lin, F. Y. (2008). Cardiopulmonary resuscitation with assisted extracorporeal life-support versus conventional cardiopulmonary resuscitation in adults with in-hospital

- cardiac arrest: An observational study and propensity analysis. *The Lancet*, 372(9638), 554-561.
- D'Agostino, R. (1998). Tutorial in biostatistics propensity score methods for bias reduction in the comparison of a non-randomized control group. *Statistic in Medicine*, 17, 2265-2281
- Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448), 1053-1062.
- Dronkers, J., & Avram, S. (2009). Choice and effectiveness of private and public schools in seven countries. A reanalysis of three PISA data sets. *Zeitschrift für Pädagogik*, 55(6), 895-909.
- Dronkers, J., & Robert, P. (2008a). Differences in scholastic achievement of public, private government-dependent, and private independent schools: A cross-national analysis. *Educational Policy*, 22(4), 541-577.
- Dronkers, J., & Robert, P. (2008b). School choice in the light of the effectiveness differences of various types of public and private schools in 19 OECD countries. *Journal of School Choice*, 2(3), 260-301.
- Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly*, 55(1), 74-79.
- Frye, B. J. (2014). *Using propensity score matching to model retention of developmental math students in community colleges in North Carolina* (Doctoral dissertation). North Carolina State University
- Graesser, A. C. (2009). Inaugural editorial for Journal of Educational Psychology. *Journal of Educational Psychology*, 101, 259-261.
- Grunwald, H. E., & Mayhew, M. J. (2008). Using propensity scores for estimating causal effects: A study in the development of moral reasoning. *Research in Higher Education*, 49(8), 758-775.

- Grzybowski, M., Clements, E. A., Parsons, L., Welch, R., Tintinalli, A. T., Ross, M. A., & Zalenski, R. J. (2003). Mortality benefit of immediate revascularization of acute ST-segment elevation myocardial infarction in patients with contraindications to thrombolytic therapy: a propensity analysis. *JAMA, 290*(14), 1891-1898.
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and analysis*. Oaks, CA: Sage Publications.
- Hahs-Vaughn, D. L., & Onwuegbuzie, A. J. (2006). Estimating and using propensity score analysis with complex samples. *The Journal of Experimental Education, 75*(1), 31-65.
- Hall, J. A., Summers, K. A., & Obenchain, R. L. (2003). Cost and utilization comparisons among propensity score-matched insulin lispro and regular insulin users. *Journal of Managed Care Pharmacy, 9*(3), 263-268.
- Hair, M. D. (2015). Propensity score matching in SPSS: How to turn an audit into a RCT. 12 Şubat 2019 tarihinde <http://www.spssusers.co.uk/Events/2015/HAIR2015.pdf> adresinden erişildi.
- Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica, 46*(4), 931-959.
- Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcome Methodology, 2*, 259-278.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis, 15*(3), 199-236.
- Jiang, F., & McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education, 37*(3), 554-576.

- Kandemir, A. Ş. ve Bülbül, Ş. (2018). Gözleme dayali çalışmalarında propensity skor ve bir uygulama. *Trakya University Journal of Social Science*, 20(2), 89-104.
- Kelcey, B. (2011). Multilevel propensity score matching within and across schools. *Journal of the American Statistical Association*, 99, 609-618.
- Koğar, H. (2015). PISA 2012 Matematik okuryazarlığını etkileyen faktörlerin aracılık modeli ile incelenmesi. *Eğitim ve Bilim*, 40(179), 45-55.
- Lane, F. C., To, Y. M., & Shelley, K. (2012). An illustrative example of propensity score matching with education research. *Career and Technical Education Research*, 37(3), 187-212.
- Lee, D. (2010). The early socioeconomic effects of teenage childbearing: A propensity score matching approach. *Demographic Research*, 23, 697-736.
- Levine, D. I., & Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *Review of Economics and Statistics*, 85(4), 884-900.
- Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review*, 29(6), 530-558.
- Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23, 2937-2960.
- Marts, J. L. (2016). *Understanding student success and institutional outcomes in service-learning coursework at a North Carolina community college: A propensity score study*. (Doctoral dissertation) North Carolina State University
- Newgard, C. D., Hedges, J. R., Arthur, M., & Mullins, R. J. (2004). Advanced statistics: The propensity score—a method for estimating treatment effect in observational research. *Academic Emergency Medicine*, 11(9), 953-961.
- Nguyen, A. N., Taylor, J., & Bradley, S. (2006). The estimated effect of Catholic schooling on educational outcomes

- using propensity score matching. *Bulletin of Economic Research*, 58(4), 285-307.
- OECD (2013). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed*. OECD publishing.
- OECD (2017). *PISA 2015 technical report*. OECD publishing.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79(387), 516-524.
- Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.
- Rosenbaum, P. R. (2002). *Observational studies*. In *Observational studies* (pp. 1-17). Springer, New York, NY. 7 Şubat 2019 tarihinde <http://propensityscoreanalysis.pbworks.com/f/BehStatObserv.study.rosenbaum05.pdf> adresinden erişilmiştir.
- Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8S), 757-763.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine*, 26(1), 20-36.
- Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*, 33(1), 62-87.
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A

- randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, 103(484), 1334-1344.
- Shapiro, J., & Trevino, J. M. (2004). *Compensatory education for disadvantaged Mexican students: An impact evaluation using propensity score matching*. The World Bank.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31, 15-21.
- Thoemmes, F. (2012). *Propensity score matching in SPSS*. arXiv preprint arXiv:1201.6385. 5 Şubat 2019 tarihinde <https://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf> adresinden erişilmiştir.
- Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, 46(1), 90-118. DOI:10.1080/00273171.2011.540475
- Titus, M. A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a master's degree. *Research in Higher Education*, 48(4), 487-521.
- U.S. Department of Education, Institute of Educational Sciences. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: Institute of Education Sciences.
- Vandenberghe, V. & Robin, S. (2003). Private, private government-dependent and public schools. An international efficiency analysis using propensity score matching. *Public Economics*. 15 Ocak 2019 tarihinde <http://econwpa.repec.org/eps/pe/papers/0308/0308002.pdf> adresinden erişilmiştir.
- Yanovitzky, T., Zanutto, E., & Hornik, R. (2005). Estimating causal effects of public health education campaigns using propensity score methodology. *Evaluation and Program Planning*, 28(2), 209-220.

The Use of Propensity Score Matching Analysis in Educational Research[†]

Esin Yılmaz Koğar*

Niğde Ömer Halisdemir University, Turkey

Received: 16.10.2018 - Revised: 07.04.2019 - Accepted: 24.04.2019

Citation: Yılmaz Koğar, E. (2019). The use of propensity score matching analysis in educational research. *Amasya Education Journal, 8(1)*, 24-61.

Summary

Problem Statement: In this study, it is aimed to present the theoretical framework related to propensity score matching-PSM which is an effective analysis used in the evaluation of the effect of the process in non-experimental research and to make a sample application on how to make PSM in the field of education. For this purpose, when CULTPOSS, HEDRES, HOMEPOS, ICRES and WEALTH variables are controlled, it has been tried to determine whether the school type has a statistically significant difference on reading, science and mathematics literacy. Is there any significant difference among the students' reading, science and mathematics literacy scores of the pre-matching and post-matching according to the type of school?

Purpose of the Study: In this study, it is aimed to present the theoretical framework related to PSM, which is an effective technique to evaluate the effect of the process in non-experimental research, and

*Corresponding Author: Phone: +90 388 2254422, E-mail: esinyilmaz@ohu.edu.tr

[†]A part of this article was presented in the 6th International Congress on Educational and Psychological Assessment in Psychiatry/Kosova, September 5-8, 2018.

ISSN: 2146-7811, ©2019

to make a sample application on how to make PSM in the field of education.

Method(s): The data were consisted of the results obtained from the PISA 2015 student survey, school survey and literacy tests. Data were analyzed with SPSS 24 program. In order to examine the differences between the groups before and after matching, independent samples t test was used. With this analysis, mean, standard deviation, t values and Cohen's d (effect size) statistic were calculated. Logistic regression analysis and important independent variables predicting membership in the treatment and control groups were examined. Nagelkerke R2 analysis, chi-square, beta coefficient, $p < .05$ values that indicate the predictors of bias were calculated by logistic regression analysis.

Findings and Discussions: As a result of the analyzes, it was determined that there was a significant difference on the reading, science and mathematics literacy scores of the students according to the school type variable in pre-matching process. However, it was thought that the groups were not stable in terms of common variables determined as CULTPOSS, HEDRES, HOMEPOS, ICRES and WEALTH at the beginning. For this reason, it is tried to perform a balance between groups with propensity score matching. As a result of the analysis, it was determined that the balancing process was performed well with the propensity score matching. After the matching process, it was determined that there was no significant difference on reading, science and mathematics literacy scores of the students according to the type of school. Therefore, it was determined that there was no treatment effect in the study.

Conclusions and Recommendations: As propensity score matching can be used in cases where the assumptions of regression-based methods are not met, it is recommended that this method of analysis be included in educational research. Researchers conduct a similar investigation without performing a secondary analysis on a PISA data. Other programs can be used for propensity score matching analyzes.

Keywords: Propensity Score, Covariate, Matching