International Journal of Human Sciences
ISSN:1303-5134

Volume: 11 Issue: 2 Year: 2014

Study of test equating on the common item non-equivalent group design

Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması

Süleyman Demir[1]
Neşe Güler[2]

Abstract

This research aims at testing the statistical equivalence of different forms of a test which are administered at the same time. For our purposes, an equating design with shared items was used for non-equivalent groups. Non-equivalent groups design with common items is used for problems that might arise in relation to the reliability and implementation of tests in which different forms are applied. The data set of the research was obtained from responses given by students participating in the PISA 2009 application within Turkey's sample. The data collected from the 761 students of 15 age group who had answered the 3rd and 10th booklets of the science studies literacy test were analyzed through Tucker Linear equating, Levine linear equating, frequency prediction and Braun-Holland linear equating methods. The weighted mean error squares averages indices that were obtained through equating procedures were 0.046 for the Tucker- linear equating, 0.072 for the Levine- linear equating, 0.049 for frequency prediction, and 0.034 for the Braun-Holland linear equating. It was observed based on the WMSE coefficient that the Braun-Holland linear equating method was the most

Özet

Bu araştırmanın amacı aynı anda uygulanan bir teste ait farklı formların istatistiksel eşitliğini sınamaktır. Bu amaç için denk olmayan gruplar için ortak maddeli eşitleme deseni kullanılmıştır. Ortak maddeli denk olmayan gruplarda ortak test deseni; farklı formların uygulandığı testlerin güvenliği ve uygulamasıyla ilgili meydana gelebilecek problemlerden dolayı kullanılmaktadır. Araştırmanın veri setini, PISA 2009 uygulamasına Türkiye örnekleminde katılmış olan öğrencilerin vermiş oldukları cevaplar oluşturmaktadırlar. Fen Bilimleri okuryazarlık testinin 3. ve 10. kitapçıklarını cevaplayan 15-yaş grubu 761 öğrenciden elde edilen veriler Tucker doğrusal eşitleme, Levine doğrusal eşitleme, frekans tahmin ve Braun-Holland doğrusal eşitleme yöntemlerine göre analiz edilmiştir. Eşitleme işlemleri sonucunda elde edilen ağırlıklandırılmış hata kareleri ortalaması indeksleri ise Tucker-Doğrusal Eşitleme için 0,046; Levine-Doğrusal Eşitleme için 0,072; Frekans Tahmin Eşit Yüzdelikli eşitleme için 0,049 ve Braun-Holland Doğrusal Eşitleme için ise 0,034 olarak bulunmuştur. Ağırlıklandırılmış hata kareleri ortalaması katsayılarına göre Braun-Holland Doğrusal

[1]Research Assistant, Sakarya Univesity, Faculty of Education, Department of Educational Measurement and Evaluation. suleymand@sakarya.edu.tr
[2]Associate Prof. Dr., Sakarya Univesity, Faculty of Education, Department of Educational Measurement and Evaluation. gnguler@gmail.com

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

appropriate for the equating of booklets 3 and 10 in the PISA 2009 Science Studies sub-test

Keywords: Non-Equivalent Groups with Anchor Test, Tucker Linear equating, Levine linear equating, frequency prediction, Braun-Holland linear equating, PISA 2009

(Extended English abstract is at the end of this document)

Eşitleme yönteminin PISA 2009 Fen Bilimleri alttestindeki 3 ve 10 numaralı kitapçıkların eşitlenmesi için en uygun yöntem olduğu görülmektedir.

Anahtar Kelimeler: Ortak maddeli denk olmayan gruplar deseni, Tucker doğrusal eşitleme, Levine doğrusal eşitleme, frekans tahmin, Braun-Holland doğrusal eşitleme, PISA 2009

Giriş

Gelişmiş ve gelişmekte olan dünya ülkelerinin çoğunda olduğu gibi Türkiye'de de son yıllarda hem ulusal, ekonomik, sosyal ve kültürel değerlerini hem de dünyanın evrensel değerlerini dikkate alarak, eğitim alanında reform niteliği taşıyan çalışmalar yapılmaktadır. Milli Eğitim bakanlığı eğitim-öğretim programlarında yapmış olduğu bu reformların ihtiyacı karşılama derecesini ulusal testlerin yanı sıra uluslararası testlerle de ölçmektedir. Ulusal boyutta; Seviye Belirleme-SBS, Yükseköğretime Geçiş-YGS, Lisan Yerleştirme-LYS sınavları yapılırken, uluslararası boyutta yapılan çalışmalar arasında ise Uluslararası Fen ve Matematik Çalışması-TIMSS, Okuma Becerilerinde Gelişim-PIRLS ve uygulaması OECD tarafından yapılan Uluslararası Öğrenci Değerlendirme Programı-PISA sınavları bulunmaktadır.

15 yaş öğrencilere uygulanan PISA, ilköğretim eğitiminin sonuna doğru, bilgi ve becerileri doğrudan ölçerek gençlerin yetişkin hayatına hazırbulunuşluk düzeylerini ve belli bir kapsam dâhilinde eğitim sisteminin etkinliğini inceler (OECD, 2009). Ülkelerin birbiri ile kıyaslanmasına olanak tanıyacak değerlendirmelerin dünya çapında kabul görmesi gerekir. Bu da ancak büyük organizasyonlar tarafından, titizlikle yapılacak çalışmalar sonunda mümkün olabilir. PISA'da ölçülen beceriler yaşam boyu öğrenme becerileri içerisinde olup, bu becerilerin katılımcı ülkeler tarafından ne derecede geliştirilebildiğine dair sağlıklı veriler edinilmesine katkı sağlar (Akkuş, 2008). PISA sonuçlarıyla, öğrencilerin temel becerilerine ilişkin değerli bulgular edinmek ve yaşam boyu öğrenme becerileri açısından kendimizi diğer ülkelerle kıyaslamak mümkündür.

PISA uygulamasına katılan öğrenciler, uygulama kapsamında yer alan tüm maddeleri cevaplamazlar. PISA 2009'da kullanılan maddeler, 13 madde demeti halinde gruplandırılmıştır. Her bir madde demetinin yanıtlanma süresi 30 dakikadır. PISA 2009'da yedi "okuma becerisi", üç "matematik" ve üç de "fen" madde demeti bulunmaktadır. Bu madde demetleri belirli bir döngü düzenine göre 13 kitapçığa yerleştirilmiştir. Her bir kitapçıkta dört madde demeti yer alır ve her öğrenci seçkisiz yöntemle belirlenen 13 kitapçıktan birini cevaplar. Kitapçıklarda öğrencilere farklı

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

soruların sorulması öğrencilerin birbirleri ile karşılaştırılmasını zorlaştırmaktadır. Öğrencilerin birbirleri ile karşılaştırılabilmesi için 13 kitapçığın bulunduğu PISA uygulaması için öğrencilerin aldıkları puanların birbirine eşitlenmesi gerekmektedir.

Test eşitleme, bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlanabilir (Angoff, 1971). Birbirine paralel ve eşit güvenirlik düzeyine sahip iki farklı testten elde edilen standart puanlar birbirine eşitse, bu formlardan elde edilen iki puanın birbirine eşit olduğu iddia edilebilir (Angof, 1982).

Test geliştiricilerin/testi uygulayan kurumların istatistiksel açıdan ve kapsam açısından aynı veya birbirine benzer testler oluşturma çabalarına rağmen bir oturumda farklı test formları ve farklı maddeler kullanıldığı sürece testlerin güçlük düzeylerinde belirli düzeyde farklılıklar olacaktır (Tanguma, 2000). Bu tür büyük ölçekli test programları çeşitli yasal, psikometrik ve pratik konuları göz önünde tutmalıdır. Bu konulardan birisi bir testin, aynı özelliği ölçen farklı formlarını oluşturmaktır. İki farklı oturumda aynı testi farklı ya da aynı bireylere uygulamak pek olası değildir. Böyle bir uygulama, testi sonra alanların, önce alanlara göre kesin bir şekilde daha avantajlı olmasını sağlayacak ve test güvenirliğini tehdit edecektir ve doğal karşılanamaz(Kan, 2010).

Fakat iki testin psikometrik açıdan birbirine eşit (eşdeğer)olduğu kanıtlandığı sürece, aynı testin farklı versiyonlarının, farklı bireylere uygulanabilmesi yasal olarak savunulabilir. Bütün bunların yanında test geliştiricilerin veya kurumların tamamıyla birbirine paralel, fakat farklı maddelerden oluşan ve testi alan her bir birey için aynı veya benzer sonuçlar üreten testler yapılandırmaları oldukça zordur ve beklenemez(Kan, 2010).

Literatür incelendiğinde, genel olarak yatay ve dikey olmak üzere iki tür eşitleme yönteminden söz edilmektedir. Yatay eşitleme, bir testin iki farklı formu arasında yapılan eşitlemedir. Yatay eşitlemede, testlerin benzer güçlükte olması, grupların da benzer yetenek dağılımına sahip olması gerekmektedir. Dikey eşitleme ise, aynı bilgi ya da becerileri ölçmek için hazırlanan ve standardize edilmiş farklı zorluk derecelerindeki testlerin eşitlenmesidir. Bu tür eşitlemede, uygulanan testlerin ve testi alanların yetenek düzeyleri farklıdır. Örneğin; öğrenci başarı gelişiminin yıllara göre izlenmesi dikey eşitleme kapsamında yer almaktadır (Hambelton ve Swaminathan 1985; Kelecioğlu, 1994).

Test Eşitleme Desenleri

Test eşitleme desenleri ihtiyaç duyduğumuz bilginin hangi tür olduğu, iki puan dağılımının nasıl olduğu, grupların hangi formu alması gerektiği ve hangi yöntemin daha ekonomik olduğu göz

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

önünde bulundurularak kurulur (Livingston, 2004; Bozdağ ve Kan, 2011).Eşitliği test edilecek iki farklı forma ait bilgilerin elde edilebileceği üç yol vardır:

- Her iki formu aynı bireylere uygulamak,
- İki formu ayrı ayrı farklı iki gruba uygulamak,
- Aynı özelliği ölçen farklı bir test formunu uygulamak.

Bir teste ait iki ayrı formu ilişkilendirmek için bu üç farklı yol, beş çeşit eşitleme deseni ile gerçekleştirilebilir. Her desen kendi içinde avantajlara ve dezavantajlara ayrıca kullanıldıkları istatistiksel yöntemle ilişkili olarak sağlanması gerekli varsayımlara sahiptir (Livingston, 2004).

Tek Grup Deseni

Tek grup test eşitleme deseni, her iki testin aynı gruba uygulanmasıyla elde edilir. Uygulama yapılacak grubun hedeflenen grubun içinden olması gerekmektedir. Ancak hedeflenen grubu temsil etmesi zorunlu değildir. Örneğin, uygulamanın yapılacağı örneklem, evrene göre daha başarılı veya daha başarısız olabilir. Bir değişken açısından evrenden farklı olan bir örneklemin kullanılması önemli değildir. Önemli olan örneklemde yer alan öğrencilerin yapılan her iki uygulama için aynı yetenek düzeyinde olmasıdır (Livingston, 2004).

Dengelenmiş grup deseni

Test eşitleme çalışmalarında genel olarak tek grup düzeneğini kullanmak uygun değildir. Bu problemin üstesinden gelebilmek için testi alan öğrenciler iki gruba ayrılırlar. Grup 1'e, Birinci gruba öncelikli olarak Y formu (yeni form) daha sonrasında X formu (referans form)uygulanır. İkinci gruba ise öncelikli olarak X formu daha sonrasında ise Y formunun uygulaması yapılır. Testi alan öğrencilere uygulaması yapılan iki test uygulaması arasındaki süre, öğrencilerin bilgi düzeylerinde değişme olmayacak şekilde ayarlanmalıdır. Oluşturulan iki grup birbirine benzer olmalıdır (Livingston, 2004).

Dengelenmiş grup düzeneği uygulamalarının avantajı tek grup düzeneğinde olduğu gibi testin uygulandığı öğrenci grubunun sayısının az olması durumunda kesin sonuçlar vermesidir. Dezavantajı ise genellikle bu eşitleme düzeneği veri toplama için özel bir çalışma gerektirmektedir. Ayrıca tek grup deseninde olduğu gibi bu desende de grupların hedef evreni yansıtmalarına gerek yoktur. Gruplar hedef evrenden daha başarılı ya da daha başarısız olabilirler (Livingston, 2004; Kollen ve Brennan, 2004).

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

Ortak maddeli denk olmayan gruplar deseni

Ortak maddeli denk olmayan gruplar deseni; farklı formların uygulandığı testlerin güvenliği ve uygulamasıyla ilgili meydana gelebilecek problemlerden dolayı kullanılmaktadır. Bu uygulama düzeninde, iki farklı test ve bu testlerde ortak maddeler bulunmaktadır ve bu formlar iki ayrı gruba uygulanmaktadır. Eğer ortak maddelerden elde edilen puanlar toplam puana ilave edilir ve ortak maddeler diğer maddeler ile birlikte uygulanırsa içsel (internal), ortak maddeler toplam puana ilave edilmez ve ayrı bir form olarak uygulanırsa dışsal (external) ortak maddeli denk olmayan gruplar deseni olarak adlandırılır (Kolen ve Brennan, 2004).

Kilmen tarafından yapılan çalışmada test eşitlemede kullanılan yöntemlerin çeşitli koşullarda karşılaştırılmasını amaçlamaktadır. Bu bağlamda Madde Tepki Kuramına dayalı "ortalama-ortalama", "ortalama-standart sapma", "Haebara" ve "Stocking-Lord" eşitleme yöntemlerinden kestirilen eşitleme hatalarının, yetenek dağılımı (benzer ve farklı yetenek dağılımı) ve örneklem büyüklüğü (500-1000 kişilik) değişkenlerine dayalı olarak karşılaştırılması amaçlanmıştır. Araştırmada, 1-0 şeklinde puanlanan 3 parametreli modele uyumlu simülatif veriler üzerinde "ortak maddeli eşitlenmemiş gruplar eşitleme deseni" kullanılmıştır. Bu araştırma ile Türkiye'de, her yıl veya yılda birden fazla uygulanarak sonuçları önemli kararlar (seçme, yeterlik belirleme) için kullanılan ölçme uygulamalarında kullanılmayan test eşitleme uygulamalarına dikkati çekmek amaçlanmıştır (Kilmen, 2010).

Test Eşitleme Yöntemleri:

Klasik Test Kuramına dayalı; ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme olmak üzere üç eşitleme yöntemi bulunmaktadır. Ortalama eşitlemede, eşitlenecek olan testlerin güçlükleri arasında fark olduğu, fakat bu farkın puan ölçeği boyunca sabit kaldığı kabul edilmektedir. Bu yöntem, eşitlenecek testlerdeki puanların ortalamaya olan uzaklıklarının eşit kabul edilmesi esasına dayanmaktadır (Kolen ve Brennan, 2004).

Ortalama eşitleme, test eşitleme yöntemlerinden en az parametre kullanılanıdır. Ortalama eşitlemede X ve Y bir testin iki farklı formu olmak üzere, X ve Y formunun ortalamaları arasındaki fark sabit kabul edilir. Eşitlenen puanlar aynı başarı düzeyine işaret etmektedir (Kolen, 2007). Diğer bir deyişle bireyin X testinden almış olduğu puana x, Y testindeki eşitlenmiş puanına y dersek;

$$x - \overline{X} = y - \overline{Y} \qquad (1)$$

Bu eşitlikte y değişkeni yalnız bırakıldığında;

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

$$x - \overline{X} + \overline{Y} = y \qquad (2)$$

eşitliği elde edilmiş olur.

Doğrusal eşitleme: Ortalama eşitlemede iki ayrı form arasındaki farkın sabit olduğu kabul edilmektedir. Doğrusal eşitleme yönteminde ise test formaları arasındaki farklılıklar dikkate alınmaktadır. Örneğin bir testin güçlük düzeylerinin farklı olduğu A ve B formları için doğrusal eşitleme yöntemi kullanılabilmektedir. Doğrusal eşitleme yapabilmek için kullanılan formül z standart puan formülü ile belirlenmektedir (Kolen ve Brennan, 2004).

$$\frac{y - \overline{Y}}{S_y} = \frac{x - \overline{X}}{S_x} \qquad (3)$$

Bu eşitlikte Y formundaki değerin X formu referans alınarak yapılan eşitleme sonrasında y yeni puanı göstermek üzere;

$$y = s_y \left[\frac{x - \overline{X}}{s_x} \right] + \overline{Y} \qquad (4)$$ eşitliği elde edilmiş olur.

Bu eşitlikte görüleceği üzere farklı formlara ait standart sapma değerlerinin (s_x ve s_y) eşit olması halinde doğrusal eşitleme ve ortalama eşitleme yöntemlerinden elde edilecek sonuçlar aynı olur (Kolen ve Brennan, 2004).

Doğrusal eşitleme yöntemi, farklı test formlarını alan grupların aynı yetenek düzeyinde olması halinde uygulanır. Testin farklı formlarını alan gruplar, aynı yetenek düzeyinde değilse, farklı doğrusal eşitleme yöntemlerinin uygulanması önerilir. Örneğin, ortak maddeli denk olmayan gruplar için standart doğrusal eşitleme uygun değildir. Bu problemin üstesinden gelebilmek için farklı doğrusal eşitleme yöntemleri geliştirilmiştir (Kan 2011).

Denk olmayan gruplar için doğrusal eşitleme yöntemi

Ortak maddeli denk olmayan gruplar düzeneği iki ayrı gruptan oluşmasına rağmen eşitleme işlemi tek bir grup üzerinden yapılmaktadır. Bunun için referans ve yeni formun uygulandığı gruplardan bir grup oluşturulmaktadır. Braun ve Holland (1982) oluşturulan bu grubu sentetik grup olarak adlandırmıştır (akt., Kolen ve Brennan, 2004).

Doğrusal eşitleme yönteminin, z standart puan formülü ile belirlendiği daha önce de belirtilmişti. İki grubun birleşimiyle oluşturulan sentetik grup için eşitleme ise aşağıda verildiği gibidir.

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

$$y = s_{y(s)}\left[\frac{x-\overline{X}_s}{s_{x(s)}}\right]+\overline{Y}_s \qquad (5)$$

$s_{y(s)}$ = sentetik gruptan elde edilen yeni formun standart sapması,

$s_{x(s)}$ = sentetik gruptan elde edilen referans formunun standart sapması,

$\overline{Y}_s$ = sentetik gruptan elde edilen yeni formun aritmetik ortalaması,

$\overline{X}_s$ = sentetik gruptan elde edilen referans formunun aritmetik ortalamasıdır.

Eşitlik 5'teki sentetik gruba ait parametrelerin tanımlaması ise aşağıda verilen eşitliklerde gösterilmiştir.

$$\overline{X}_s = w_1\overline{X}_1 + w_2\overline{X}_2 \qquad (6)$$

$$\overline{Y}_s = w_1\overline{Y}_1 + w_2\overline{Y}_2 \qquad (7)$$

$$s^2_{x(s)} = w_1 s^2_{x(1)} + w_2 s^2_{x(2)} + w_1 w_2\left[\overline{X}_1 - \overline{X}_2\right]^2 \qquad (8)$$

$$s^2_{Y(s)} = w_1 s^2_{Y(1)} + w_2 s^2_{Y(2)} + w_1 w_2\left[\overline{Y}_1 - \overline{Y}_2\right]^2 \qquad (9)$$

Ortak maddeli denk olmayan gruplarda X formu (referans form) Grup 1'e, Y formu (yeni form) Grup 2'ye uygulanmamıştır. Dolayısıyla $\overline{X}_2$, $s^2_{x(2)}$, $\overline{Y}_1$ ve $s^2_{Y(1)}$ parametreleri doğrudan hesaplanamaz. Test eşitleme yöntemlerinin pratikte uygulanabilmesi için bu parametrelerin hesaplanabilmesi gerekmektedir. Bu amaç doğrultusunda bazı yöntemler (Tucker, Levine yöntemi vs.) geliştirilmiştir. Bu yöntemler bazı varsayımları kabul ederek $\overline{X}_2$, $s^2_{x(2)}$, $\overline{Y}_1$ ve $s^2_{Y(1)}$ parametrelerini hesaplama olanağı vermektedirler(Kolen ve Brennan, 2004).

6'dan -9'a kadar olan eşitliklerindeki grup 1 ve grup 2'nin ağırlıklandırma katsayısı olarak belirlenen w_1 ve w_2 toplamları 1'e eşit olan pozitif sayılardır. w_1 ve w_2'nin hesaplanması ile ilgili olarak üç özel kullanım bulunmaktadır. Bunlardan ilki Gulliksen (1950) tarafından hazırlanmış olan $w_1 = 1$ ve $w_2 = 0$ oranları; ikincisi Angoff (1971) tarafından düzenlemiş $w_1 = \frac{N_1}{N_1+N_2}$ ve $w_2 = \frac{N_2}{N_1+N_2}$ (N_1= Grup 1'deki birey sayısı, N_2= Grup 2'deki birey sayısı olmak üzere)

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

formüllerinden elde edilmiş oranlar; üçüncüsü ise ağırlıklandırmaların eşit olduğu durum olarak $w_1 = w_2 = 0,5$ oranları kullanılmaktadır.

Tucker doğrusal eşitleme yöntemi

Ledyard Tucker'a atfedilen, Tucker doğrusal eşitleme yöntemi Gulliksen tarafından 1950 yılında açıklanmıştır. Bu yöntem 6'dan-9'a kadar olan eşitliklerindeki parametreleri açıklamak için iki varsayım kabul etmiştir. Birinci olarak yeni ve referans formlardaki maddeler ile ortak maddelerden elde edilen doğru grafiğinin Grup 1 ve Grup 2 için aynı olduğu varsayımı; ikincisi ise referans ve yeni formlardan elde edilen varyansların Grup 1 ve Grup 2 için eşitliği varsayımıdır (Kolen ve Brennan, 2004).

Levine doğrusal eşitleme yöntemi

Levine doğrusal eşitleme yöntemi 1955 yılında Levine tarafından geliştirilmiştir. Bu yöntem 6'dan-9'a kadar olan eşitliklerindeki parametreleri açıklamak için üç varsayım kabul etmiştir. Birincisi T_X, T_Y ve T_O sırasıyla X, Y ve ortak madde formundan elde edilen doğru puan sayıları olmak üzere T_X ile T_O ve T_Y ile T_O arasındaki korelasyonun yüksek olması ve Grup 1 ve Grup 2 için eşitliği varsayımı; ikincisi T_X ile T_O ve T_Y ile T_O puanlarından elde edilen doğruların Grup1 ve Grup 2 için eşitliği varsayımı; üçüncü olarak ise referans form, yeni form ve ortak madde formları içim ölçmenin hata varyansının her iki grup içinde eşit olduğu varsayımıdır.

Eşit yüzdelikli eşitleme

Doğrusal eşitleme yönteminin bir takım sınırlılıkları bulunmaktadır. Örneğin, bir testin iki farklı formundan elde edilen puanların ranjı değişiyorsa, bir formdaki puanın diğer formdaki karşılığı bulunmayabilir. Örneğin 100 maddelik iki test formu doğrusal eşitleme ile eşitlendiğinde, zor olan X formundan 99 puan alan bireyin, kolay olan Y formundaki puan karşılığı 103 puan olarak hesaplanabilir (Livingston, 2004).

Eşit yüzdelikli eşitleme yönteminde form X'in dağılımı, form Y'nin dağılımına iki testin yüzdelik sıraları hesaplanarak eşitlendiği için X formundaki her puana karşılık Y formunda bir puan karşılık gelmektedir. İki form aynı puan dağılımına sahip değil ise eşit yüzdelikli eşitleme yöntemi önerilir (Zhu, 1998). İki testin puan dağılımlarının farklı olması durumunda, eşitleme kesinliği, eşit yüzdelikli eşitleme ile sağlanır (Angoff, 1971).

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

Eşit yüzdelikli eşitleme, iki basamakta gerçekleştirilir. İlk olarak X ve Y formlarına ilişkin toplamalı frekans dağılımlarına göre sıraya koyulur. İkinci basamakta ise, elde edilen bu toplamalı frekans dağılımlarına göre eşitlenmiş puanlar bulunur(Kolen ve Brennan, 2004).Eşit yüzdelikli eşitleme yöntemi 1982 yılında Braun ve Holland tarafından açıklanmıştır. Bu açıklamaya göre eşitleme formülü eşitlik 10'daki gibidir.

$$e_Y(x) = G^{-1}[F(x)] \qquad (10)$$

$e_Y(x)$: Y formuna eşitlenmiş puanları

F ve G: X ve Y formlarına ait toplamalı frekans fonksiyonu.

Doğrusal eşitleme yönteminde olduğu gibi eşit yüzdelikli eşitleme yönteminde de standart eşitleme yöntemleri ortak maddeli denk olmayan gruplardan elde edilen verilerin eşitlenmesi için kullanılamamaktadır. Ortak maddeli denk olmayan grupların puanlarının eşitlenebilmesi için farklı eşitleme yöntemleri kullanılmaktadır.

Frekans Tahmin Yöntemi

Ortak maddeli gruplardan elde edilen test sonuçlarının eşitlenebilmesi için Grup 1 ve Grup 2'ye ait bireylerden sentetik grup adı verilen yeni bir grubun oluşturulması gerekmektedir. *f* ve *g* sırasıyla referans ve yeni formdan elde edilen frekans fonksiyonları olmak üzere sentetik gruba ait frekans değerlerinin belirlenmesi için 11 ve 12 numaralı eşitlikte verilen formüller kullanılmaktadır.

$$f_S(x) = w_1 f_1(x) + w_2 f_2(x) \qquad (11)$$

$$g_S(y) = w_1 g_1(y) + w_2 g_2(y) \qquad (12)$$

Ancak X formunun Grup 2'ye Y formunun Grup 1'e uygulanmamasından dolayı "frekans tahmin yöntemi için bazı varsayımların kabul edilmesi gerekmektedir. Bunun için öncelikli olarak koşullu dağılımın (conditional distributions) tanımlaması verilecektir.

Koşullu Dağılım (ConditionalDistributions): *f* ve *h* sırasıyla X ve V formları için frekans (olasılık) fonksiyonu olarak belirlenmiş olsun. Bu durumda *f(x)* X formundaki *x* değerinin frekansı/(olasılığı), *h(v)* ise V formundaki *v* değerinin frekansı/(olasılığı) olmak üzere *f(x,v)* X formu için *x* değeri ve V formu için *v* değerinin olasılıklarının birlikte olması olasılıkları olacaktır. X formuna ait x değeri belli iken V formuna ait v değerinin frekansını belirlemek için 13 numaralı eşitlik kullanılmaktadır.

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

$$f(x|v) = \frac{f(x,v)}{h(v)} \qquad (13)$$

Frekans tahmin yönteminde Grup 1 ve Grup 2'ye ait koşullu frekans dağılımlarının eşit olduğu varsayımı kabul edilir. Bu koşullar altında X ve Y formlarının sentetik gruptaki frekans fonksiyonları sırasıyla eşitlik 14 ve 15'teki gibi olur.

$$f_S(x) = w_1 f_1(x) + w_2 \sum_v f_1(x|v) h_2(v) \qquad (14)$$

$$g_S(y) = w_1 \sum_v g_2(y|v) h_1(v) + w_2 g_2(y) \qquad (15)$$

Sentetik gruba ait eşit yüzdelikli test eşitleme formülü ise eşitlik 16'daki gibi ortaya çıkmaktadır.

$$e_{Ys}(x) = G_s^{-1}[F_s(x)] \qquad (16)$$

Braun-Holland doğrusal eşitleme yöntemi

Braun ve Holland 1982 yılında geliştirmiş oldukları test eşitleme yönteminde, frekans tahmin yöntemindeki varsayımları, doğrusal eşitleme yöntemine adapte etmişlerdir. Bu eşitleme işlemi yapılırken ortalama ve standart sapma değerlerinden yararlanılmaktadır. Ayrıca Braun ve Holland eşitleme işlemini yapabilmek için Tucker doğrusal eşitleme yönteminde bulunan varsayımları kabul etmişlerdir (Kolen ve Brennan, 2004).

Bu açıdan düşünüldüğünde; Braun-Holland eşitleme yöntemi, Tucker doğrusal eşitleme yönteminin genelleştirilmiş bir halidir. X ve Y formlarından elde edilen puanlar ile ortak madde formundan elde edilen puanlar arasındaki grafiğin doğrusal olduğu varsayımının kabul edilmediği durumlarda da kullanılabilmektedir. Diğer bir deyişle Braun-Holland yönteminin farklılığı varsayımlardan birincisinin kabul edilmediği durumlarda da kullanılabilmesidir. Ancak Braun-Holland eşitleme yöntemi Tucker doğrusal eşitleme yönteminden daha karmaşık hesaplama sistemine sahiptir. Bu yüzden pratikte daha az kullanılan bir yöntemdir (Kolen ve Brennan, 2004).

Eşitleme alanında yapılan birçok çalışmada çeşitli eşitleme hataları hesaplanmıştır. Bu hataların farklılığı, kavramsal çatılarının farklı olmasından kaynaklanır. Bunlardan biri de ağırlıklandırılmış hata kareleri ortalamasıdır. Eşitleme yöntemlerindenelde edilmiş puanların hata miktarını belirlemek için her bir ham puan ve bu ham puanakarşılık gelen eşitlenmiş puanlar ağırlıklandırılmış hata kareleri ortalaması (AHKO) ile karşılaştırılır (Skaggs ve Lissitz, 1986; Kelecioğlu, 1994; Şahhüseyinoğlu, 2005; Bozdağ ve Kan 2010).

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

$$AHKO = \frac{\sum_{i=1}^{k-1} f_i \left(X_E - X_{krit} \right)^2}{\sum_{i=1}^{k} f_i S_Y^2}$$

k : *Y* testindeki madde sayısı.

S_Y^2 : Y testindeki ham puanların varyansı.

X_{krit} : Y testindeki i. ham puan.

X_E :Eşitleme yöntemleriyle elde edilen ve X testindeki i. ham puana eşit olan puan.

f_i : Y testindeki i. ham puan frekansı.

Kolen ve Whitney (1982), çalışmalarında eşit yüzdelikli eşitleme, doğrusal eşitleme, bir- ve iki-parametreli Madde Tepki Kuramına (MTK) dayalı eşitleme yöntemlerinin yeterliliğini karşılaştırmışlardır. Araştırmada kullanılan veriler Genel Eğitim Gelişim Testleri (Tests of General Educational Development-GED)kullanılarak ortak test desenine dayalı olarak toplanmıştır. Araştırma sonucunda doğrusal eşitleme ile elde edilen puanların ortak testin olası puan dağılımının üstünde olduğu bulunmuştur. Alt puanlar için üç-parametreli yöntemin en küçük, üst puanlarda ise en yüksek eşdeğer puanlar sağlayan yöntem olduğuna ulaşılmıştır. Eşit yüzdelikli eşitlemenin ise puan aralığı boyunca en düzensiz eşitleme ilişkisi üreten yöntem olduğu sonucuna varılmıştır.

Skaggs ve Lissitz (1986), yaptıkları Monte Carlo çalışmasında, farklı psikometrik özelliklere sahip testleri, yaygın olarak kullanılan dört eşitleme yöntemini (doğrusal, eşit yüzdelikli eşitleme, Rasch ve üç-parametreli model) kullanarak eşitlemişlerdir. Eşitleme yöntemlerinin sonuçlarının karşılaştırılmasında, ağırlıklandırılmış ve ağırlıklandırılmamış hata kareleri ortalamaları hesaplanmıştır. Sonuç olarak, bu çalışmada eşit yüzdelikli eşitleme ve üç-parametreli MTK yönteminin daha kabul edilebilir sonuçlar verdiği görülmüştür. Ancak eşitlenmek istenen testler psikometrik özellikleri açısından büyük farklılıklar gösterdiğinde, eşit yüzdeliklieşitleme yönteminin kullanılması önerilmektedir.

Kelecioğlu (1994), yaptığı çalışmasında 1990, 1991 ve 1992 yıllarında yapılan ÖSS sınavlarına ait puanları eşit yüzdelikli, doğrusal, Rasch modeli ve iki parametreli lojistik model yöntemlerini kullanarak eşitlemiş ve bu yöntemler arasından uygun olanı önermiştir. Çalışmada ortak test deseni kullanılmış ve eşitleme hataları, eşitlenmiş puanlarla test puanlarının hata kareleri ortalamalarını hesaplanarak bulmuştur. Araştırma sonucunda, Türkçe testleri için en uygun

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

yöntemin doğrusal eşitleme, Sosyal Bilimler ve Matematik testleri için Rasch modeli ile eşitleme, Fen Bilimleri testleri için eşit yüzdelikli eşitleme olduğu sonucuna ulaşılmıştır..

Davier, Holland ve Thayer (2002), eşdeğer olmayan gruplar için ortak madde test deseninde zincirleme eşitleme ve son tabakalı eşitleme yöntemleriyle grup değişmezliği (population invariance) sayıltısını incelemiştir. Bu araştırma sonucunda eşitlenmek istenen gruplar aynı dağılıma sahip ve ortak madde testi eşitlenmek istenen testlerin küçük birversiyonuysa, zincirleme eşitleme ve son tabakalı eşitleme yöntemlerinin sonuçlarının yaklaşık olarak aynı olduğu gözlenmiştir.

Yöntem

Araştırma Grubu ve Veri Toplama Araçları

PISA 2009 uygulaması ülkemizde 2009 yılının Nisan ayı içerisinde yapılmıştır. PISA 2009 uygulamasında, 12 istatistiki bölge biriminden 56 il ve okul türlerine göre PISA uluslararası merkez tarafından seçkisiz yöntemle belirlenen toplam 170 okuldan 4996 öğrenci yer almıştır. PISA uygulamasında bulunan 13 kitapçıktan fen bilimleri sorularının en fazla olduğu kitapçıklar analize dahil edilmiştir. 3 numaralı kitapçıkta ve 10 numaralı kitapçıkta toplam 35'er madde bulunduğu için analize bu iki kitapçıkla devam edilmiştir. Her iki kitapçıkta fen bilimlerine ait iki farklı soru demeti bulunmaktadır. Birinci kitapçıkta 17 ikinci kitapçıkta ise 18 adet soru bulunmaktadır. Dolayısıyla bu çalışmada Fen Bilimleri okuryazarlık testinin 3. ve 10. kitapçıklarını cevaplayan 15-yaş grubu 761 öğrenciden elde edilen veriler kullanılmıştır.

Verilerin analizi

Verilerin analizi üç aşamada gerçekleştirilmiştir. Birinci aşamada, eşitleme koşullarının sağlanıp sağlanmadığı test edilmiştir. İkinci aşamada, eşitleme yöntemleri kullanılarak, eşitlenmiş puanlar elde edilmiştir. Üçüncü aşamada ise, her bir eşitleme yöntemine ait hata kareleri ortalamaları hesaplanmıştır.

I. Aşama: Eşitlemenin yapılabilmesi için bazı koşulların karşılanması gerekmektedir. Literatür incelendiğinde eşitlenecek testlerin genel olarak üç koşulu mutlaka karşılaması gerektiği belirtilmiştir. Bunlar; aynı ve tek bir yapıyı ölçme, eşit güvenirliğe ve benzer güçlüğe sahip olmadır (Angoff, 1984; Dorans ve Holland, 2000; Kolen ve Whitney, 1982: Kelecioğlu 1994).

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

Tablo 1. K3 ve K10 kitapçıklarına ait faktör analizi sonuçları

	K3		K10	
	Özdeğer	V.A.O (%)	Özdeğer	V.A.O (%)
1	5,583	16,918	4,961	15,034
2	1,443	4,372	1,531	4,64

Tablo 1'de yer alan verilere göre K3 ve K10 alt testleri sonuçları üzerinde yürütülen faktör analizi sonucunda her iki teste ve birleştirilmiş teste ilişkin özdeğerler ve varyans açıklama oranları (V.A.O) birbirine çok yakın değerlere sahiptir. Bu özdeğerler ve varyans açıklama oranları 1. faktörden sonra keskin bir düşüş göstermekte ve ilk faktör ile kendisinden sonra gelen ilk faktöre (2. faktör) ait V.A.O ile özdeğerler arasında en az 4 kat fark olduğu göze çarpmaktadır. Bu bulgu testlerin baskın tek faktöre sahip olduğuna ilişkin kanıt olarak kullanılabilir (Hambleton ve Swaminathan, 1985). Birçok durumda tek boyutluluk varsayımının karşılanabilmesi için testin baskın tek bir faktöre sahip olması yeterli görülmekte ve bu baskın faktör, testle ölçülmek istenen özellik olarak tanımlanmaktadır.

Eşitlenmek istenilen K3 ve K10 alt testlerinin ortalama güçlükleri arasında anlamlı bir fark olup olmadığı iki oran farkı testiyle incelenmiştir (Baykul, 1996). Alt testlerin ortalama güçlükleri arasındaki farkın test edilmesine ilişkin bulgular Tablo 2'de verilmiştir.

Tablo 2. K3 ve K10 alt testlerinin ortalama güçlüklerinin karşılaştırılması

	$\overline{X}$	K	$\overline{p}$	*t*	*p*
K3	15,473	33	0,47	0,724	0,472
K10	14,404	33	0,44		

Tablo 2 incelendiğinde, eşitlenecek K3 ve K10 alt testlerinin ortalama güçlükleri arasında .05 anlamlılık düzeyinde anlamlı bir fark olmadığı görülmektedir. Bu sonuca dayalı olarak eşitlenmek istenilen alt testlerin ortalama güçlüklerinin eşit olma koşulunu sağladığı söylenebilir.

Eşitlenmek istenilen K3 ve K10 alt testlerinin, eşit güvenirlikte olup olmadığını belirlemek için her bir alt testin Cronbach α güvenirlik katsayıları hesaplanmıştır. Güvenirlik katsayılarının her birikorelasyon katsayısı olarak kabul edilmiş ve güvenirlik katsayılarına Fischer'ın Z dönüşümü yapılmıştır. İki güvenirlik katsayısı arasında fark olup olmadığı Fischer'ın Z istatistiği ile test edilmiştir (Akhun, 1984).Güvenirlik katsayıları arasındaki farkın test edilmesine ilişkin bulgular

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

Tablo 3'de yer almaktadır.

Tablo3. K3 ve K10 alt testlerinin güvenirliklerinin karşılaştırılması

	KR-20	Z_r	Z
K3	0,836	1,208	1,54
K10	0,796	1,088	

Tablo 3 incelendiğinde,K3 ve K10 alt testlerinin güvenirlikleri arasında .05 anlamlılık düzeyinde anlamlı bir fark olmadığı görülmektedir. Bu sonuca göre eşitlenmek istenilen alt testlerin eşit güvenirliğe sahip olma koşulunu sağladığı söylenebilir.

II. Aşama: Bu aşamada PISA 2009 Fen Bilimleri okur-yazarlığı alt testinin K3 ve K10 alt testleri arasında eşitleme çalışması, CIPE programı (Kolen, 2003) ile yapılmıştır. Tablo 4'de K3 alt testindeki madde puanları,Tucker, Levine ve Frekans Tahmin yöntemi ile hesaplanan eşitlenmiş puanlar ve orijinal puan ile eşitlenmiş puan arasındaki farklar verilmiştir.

Tablo 4. Tucker, Levine, Braun-Holland ve Frekans Tahmin yöntemi ile hesaplanan eşitlenmiş puanlar

K10 Formu	Tucker-Doğrusal Eşitleme		Levine-Doğrusal Eşitleme		Frekans Tahmin Yöntemi		Braun-Holland	
	Fark	Eşit. Puan	Fark	Eşit. Puan	Fark	Eşit. Puan	Fark	Eşit.
0	3,57	3,57	4,38	4,38	0,16	0,16	3,03	3,03
1	3,38	4,38	4,15	5,15	0,47	1,47	2,87	3,87
2	3,19	5,19	3,93	5,93	0,78	2,78	2,72	4,72
3	3,00	6,00	3,71	6,71	1,60	4,6	2,56	5,56
4	2,81	6,81	3,48	7,48	1,45	5,45	2,40	6,40
5	2,62	7,62	3,26	8,26	1,90	6,90	2,25	7,25
6	2,43	8,43	3,04	9,04	1,83	7,83	2,09	8,09
7	2,24	9,24	2,81	9,81	1,86	8,86	1,93	8,93
8	2,05	10,05	2,59	10,59	2,28	10,28	1,77	9,77
9	1,86	10,86	2,37	11,37	2,37	11,37	1,62	10,62
10	1,67	11,67	2,14	12,14	2,48	12,48	1,46	11,46
11	1,48	12,48	1,92	12,92	1,96	12,96	1,30	12,30
12	1,29	13,29	1,70	13,70	1,43	13,43	1,15	13,15
13	1,10	14,10	1,47	14,47	0,99	13,99	0,99	13,99

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

14	0,91	14,91	1,25	15,25	0,70	14,7	0,83	14,83
15	0,72	15,72	1,03	16,03	0,44	15,44	0,67	15,67
16	0,53	16,53	0,80	16,80	0,43	16,43	0,52	16,52
17	0,34	17,34	0,58	17,58	0,35	17,35	0,36	17,36
18	0,15	18,15	0,36	18,36	0,21	18,21	0,20	18,20
19	-0,04	18,96	0,13	19,13	0,16	19,16	0,05	19,05
20	-0,23	19,77	-0,09	19,91	-0,07	19,93	-0,11	19,89
21	-0,42	20,58	-0,31	20,69	-0,41	20,59	-0,27	20,73
22	-0,61	21,39	-0,54	21,46	-0,63	21,37	-0,43	21,57
23	-0,80	22,20	-0,76	22,24	-0,95	22,05	-0,58	22,42
24	-0,99	23,01	-0,98	23,02	-1,42	22,58	-0,74	23,26
25	-1,18	23,82	-1,21	23,79	-1,51	23,49	-0,90	24,10
26	-1,37	24,63	-1,43	24,57	-1,44	24,56	-1,05	24,95
27	-1,56	25,44	-1,65	25,35	-1,57	25,43	-1,21	25,79
28	-1,75	26,25	-1,88	26,12	-1,54	26,46	-1,37	26,63
29	-1,94	27,06	-2,10	26,90	-0,94	28,06	-1,53	27,47
30	-2,13	27,87	-2,32	27,68	-1,63	28,37	-1,68	28,32
31	-2,32	28,68	-2,55	28,45	-1,16	29,84	-1,84	29,16
32	-2,51	29,49	-2,77	29,23	-0,70	31,30	-2,00	30,00
33	-2,70	30,30	-3,00	30,00	-0,23	32,77	-2,16	30,84

Tablo 4'de yer alan sonuçlara dayalı olarak, orijinal puan ile eşitlenmiş puan arasındaki farkların Tucker-Doğrusal Eşitleme yöntemi için -2,70 ile 3,57; Levine Doğrusal Eşitleme yöntemi için -3,00 ile 4,38; Frekans Tahmin yöntemi için -1,57 ile 2,48 arasında değerler aldığı görülmektedir. Ayrıca eşitleme yöntemleri analizi sonucunda elde edilen bulgular, doğrusal ve eşit yüzdelikli eşitleme sonucu elde edilen puanlar ile orijinal puanlar arasındaki farkın düşük puanlar için negatif, daha yüksek puanlar için ise pozitif olduğunu göstermektedir.

III. Aşama: Tucker-Doğrusal Eşitleme; Levine-Doğrusal Eşitleme, Braun-Holland Doğrusal Eşitleme ve Frekans Tahmin yöntemlerinden hangisinin daha uygun olduğunu belirlemek için dört yönteme ait ağırlıklandırılmış hata kareleri ortalamaları (AHKO katsayıları) hesaplanmıştır. Elde edilen bulgular Tablo 5'te verilmiştir.

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

Tablo 5. Eşitleme yöntemlerine ait Ağırlıklandırılmış Hata Kareleri katsayıları

Eşitleme Yöntemleri	AHKO katsayıları
Tucker-Doğrusal Eşitleme	0,046
Levine-Doğrusal Eşitleme	0,072
Frekans Tahmin Eşit Yüzdelikli	0,049
Braun-Holland Doğrusal Eşitleme	0,034

Tablo 5' te verilen AHKO katsayılarına göre Braun-Holland Doğrusal Eşitleme yönteminin PISA 2009 Fen Bilimleri alttestindeki 3 ve 10 numaralı kitapçıkların eşitlenmesi için en uygun yöntem olduğu görülmektedir. Bu yöntemi ise sırasıyla Tucker-Doğrusal eşitleme yöntemleri, Frekans Tahmini Eşit Yüzdelikli eşitleme yöntemi, Levine-Doğrusal eşitleme yöntemi izlemektedir.

Sonuç ve Öneriler

Ülkemizde de uygulanan Uluslararası sınavlardan biri olan ve OECD tarafından düzenlenen Uluslararası Öğrenci Değerlendirme Programı (PISA), 15 yaş öğrencilere uygulanmakta ve ilköğretim eğitiminin sonuna doğru, bilgi ve becerileri doğrudan belirlemek ve eğitim sisteminin etkinliğini incelemek amacı taşımaktadır (OECD, 2009). Uluslararası önem taşıyan bu tür sınavlarda, testin yansız maddelerden oluşarak geçerli bir ölçme aracı olup olmadığının ve farklı kitapçıklardan elde edilen puanların istatistiksel eşitliğinin incelenmesi gerekmektedir. Bu çalışmada da öncelikle bu amaca hizmet eden test eşitleme desen ve yöntemleri açıklanmaya çalışılmıştır. Daha sonra bu yöntemlerden; Braun-Holland doğrusal eşitleme, Tucker doğrusal eşitleme ve frekans tahmin eşit yüzdelikli yöntemler PISA 2009 uygulamasında yer alan 3. ve 10. kitapçıklarından elde edilen verilerin eşitliğinin araştırılmasında kullanılmıştır. Çalışmadan elde edilen sonuçlar, orijinal puanlar ile Braun-Holland doğrusal eşitleme yöntemi ile elde edilen eşitlenmiş puanlar arasındaki farklılığın ve bu yönteme ilişkin ağırlıklandırılmış hata kareleri ortalamalarının daha düşük olduğunu göstermektedir. Bu bulgulara dayalı olarak da Braun-Holland Doğrusal eşitleme yöntemin eşitlemede kullanılmasının daha uygun olduğu yorumu yapılabilir. Bu çalışmadan elde edilen sonuca dayalı olarak; PISA ve benzeri uluslararası sınavlarda yer alan farklı kitapçıkların ve ulusal düzeyde gerçekleştirilen sınavların farklı formlarının eşitlenmesine ve farklı yöntemlerin uygulanmasına ilişkin daha geniş kapsamlı çalışmaların yapılması ileriki çalışmalar için önerilebilir.

Kaynakça

Akhun, İ. (1984). İki Korelasyon Katsayısı Arasındaki Farkın Manidarlığının Test Edilmesi. Ankara.

Akkuş, N. (2008). Yaşam Boyu Öğrenme Becerilerinin Göstergesi olarak 2006 PISA Sonuçlarının Türkiye Açısından Değerlendirilmesi. Yayımlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Angoff, W. H. (1971). Scale, norms and equivalent scores. In R. L. Thorndike (Eds.) Educational Measurement (2nd. Ed.) Washington D.C; American Council of Education.

Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. In P.W. Holland ve D. B. Rubin (Ed). Test Equating. New York: Academic Press.

Angoff, W. H. (1984). Scales, norms and equivalent scores. New Jersey: Educational Testing Service.

Baykul, Y. (1996). İstatistik: Metodlar ve uygulamalar (3.baskı). Ankara: Anı Yayıncılık

Bozdağ, S. ve Kan, A. (2010). Şans Başarısının Test Eşitlemeye etkisi. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 39, 91-108.

Dorans, J. N. ve Holland, P. W. (2000). Population in variance and the equitability of tests: Basic theory and the linear case. Journal of Educational Measurement, 37, 281-306.

Davier, A. A., Holland, P.W. ve Thayer, D. T. (2002). Population in variance and chain versus post-stratification equating methods. In N. J. Dorans (Ed.), Population in variance of score linking: Theory and applications to Advanced Placement Program® examinations(ETS RR-03-27, pp. 19-36). Princeton, NJ: Educational Testing Service.

Hambleton, R. K. ve Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, Kluwer Academic Publishers Group.

Kan A. (2010). Test Eşitleme: Aynı Davranışları Ölçen, Farklı Madde Formlarına Sahip Testlerin İstatistiksel Eşitliğinin Sınanması. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi1(1) 16-21.

Kan A. (2011). Test Eşitleme: OKS Testlerinin İstatistiksel Eşitliğinin Sınanması. Eğitim ve Bilim 36 (160) 38-51.

Kelecioğlu, H. (1994). Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Kolen, M. J. ve Whitney, D. R. (1982). Comprison of four procedures for equating the tests general educational development. Journal of Educational Measurement, 19(4), 279–293.

Kolen, M. J. ve Brennan, R. L. (1995). Test equating: Methods and practices. New York: Springer.

Kolen, M. J. (2003). CIPE: Common item program for equating (CIPE) (version 2.0). University of Iowa: Center for Advanced Studies in Measurement and Assessment (CASMA).

Kolen, M. J. (2007). Data Collection Designs and Linking Procedures. In Dorans, N. J. Pommerich, M. Holland, P. W. (Eds.), Linking and Aligning Scores and Scales. USA: Springer.

Kolen, M. J. ve Whitney, D. R. (1982). Comprison of four procedures for equating the tests general educational development. Journal of Educational Measurement, 19(4), 279–293.

Kilmen, S. (2010). Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinden Kestirilen Eşitleme Hatalarının Örneklem Büyüklüğü ve Yetenek Dağılımına göre Karşılaştırılması. Yayınlanmamış doktora tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Livingstone, S. A. (2004). Equating test scores(Without IRT). Educational Testing Service.

OECD, (2009) Organization for Economic Cooperation and Development 2009. PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science. Paris: OECD.

Skagg, G. ve Lissitz R. W. (1986). An Exploration of the Robustness of Four Test Equating Models. Applied Psychological Measurement. 10, 303-317.

Şahhüseyinoğlu, D. (2005). İngilizce Yeterlik Sınavı Puanlarının Üç Farklı Eşitleme yöntemine Göre Karşılaştırılması. Yayımlanmamış Doktora Tezi, Hacettepe Üniversitesi SBE.

Tanguma, J. (2000). Equating test scores using the linear method: A primer. Paper presented at the annual meeting of the Southwest Educational Research Association. Dallas, TX.

Zhu, W. (1998). Test equating: What, why, how? Research Quarterly for Exercise andSport, 69(1), 11-23.

Extended English Abstract

Despite the efforts made by test developers/test administering institutions to construct identical or similar tests in terms of statistics and of content, differences are bound to occur in test difficulty as long as different test forms and different test items are used in a session (Tanguma, 2000). There are two different kinds of equating: horizontal and vertical equating. Vertical equating refers to the process of equating tests administered to groups of students with different abilities, such as students in different grades (years of schooling). Horizontal equating refers the equating of tests administered to groups with similar abilities; for example, two tests administered students in the same grade in two consecutive calendar years. Different tests are used to avoid practice effects. This research aims at testing the statistical equivalence of different forms of a test which are administered at the same time. For our purposes, an equating design with shared items was used for unequalized groups. Non-equivalent groups design with common items is used for problems that might arise in relation to the reliability and implementation of tests in which different forms are applied. In this design, Form X and Form Y have a set of items in common, and different groups of examinees are administered the two forms. For example, a group tested one year might be administered Form X and a group tested another year might be administered Form Y. This design has two variations. When the score on the set of common items contributes to the examinee's score on the test, the set of common items is referred to as *internal*. The internal common items are chosen to represent the content and statistical characteristics of the old form. For this reason, internal common items typically are interspersed among the other items in the test form. When the score on the set of common items does *not* contribute to the examinee's score on the test form, the set of common items is referred to as *external*. Typically, external common items are administered as a separately timed section (Kolen & Brennan, 2004).

Method

The data set of the research was obtained from responses given by students participating in the PISA 2009 application within Turkey's sample. The data collected from the 761 students of 15 age group who had answered the 3rd and 10th booklets of the science studies literacy test were

Demir, S., & Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin test eşitleme çalışması. *International Journal of Human Sciences*, *11*(2), 190-208. doi: 10.14687/ijhs.v11i2.2870

analyzed through Tucker Linear equating, Levine linear equating, frequency prediction and Braun-Holland linear equating methods. The data analysis was performed in three stages. Thus, at the first stage, whether or not the equating conditions were satisfied was tested. At the second stage, the equating methods were used and the equalized scores were obtained. At the third stage, the error squares averages for each method were calculated. Some assumptions needed to be met in order for equating to be achieved. A literature review shows that three conditions should absolutely be met for the test to be equalized: Namely, the measurement of a single and the same structure, equal reliability, and similar difficulty (Angoff, 1984; Dorans and Holland, 2000; Kolen and Whitney, 1982: Kelecioğlu, 1994).

Results

Results obtained in consequence of the equating procedure demonstrated that the differences between the original score and the equalized score were in the -2.70 – 3.57 range for the Tucker-linear equating method, in the -3.00 - 4.38 range for the Lenine- linear equating method, and in the -1.57 – 2.48 range for the Frequency prediction method. Moreover, the findings obtained through the equating methods analysis showed that the difference between the scores that were obtained through linear and equal percentage equating was negative for lower scores whereas it was positive for higher scores. The weighted error squares averages indices that were obtained through equating procedures were 0.046 for the Tucker- linear equating, 0.072 for the Levine- linear equating, 0.049 for frequency prediction, and 0.034 for the Braun-Holland linear equating. It was observed based on the AHKO coefficient that the Braun-Holland linear equating method was the most appropriate for the equating of booklets 3 and 10 in the PISA 2009 Science Studies sub-test. This method was followed by Tucker-linear equating method, Frequency prediction with equal percentage equating method, and Levine-linear equating method.

Conclusion and Recommendations

The results obtained in the study demonstrated that the difference between the original scores and the scores achieved through Braun-Holland linear equating method and the weighted error squares coefficient for the method was small. Based on these findings, it might be stated that the use of Tucker-linear equating method would be more suitable. Consequently, it might be recommended that further research with regard to equating the different booklets available in PISA and in international tests alike in different forms and by using different methods be conducted.