

## PISA 2012 Matematik Testlerinden İki Kitapçığın Gözlenen Puan Eşitleme Yöntemleri İle Eşitlenmesi

### Equation of Two PISA 2012 Mathematics Tests Through Observed Score Equating Methods

Sami PEKTAŞ

Mustafa KILINÇ

Başvuru Tarihi: 11.12.2016

Yayına Kabul Tarihi: 19.12.2016

DOI: 10.21764/efd.49376

**Özet:** Bu çalışmada kitapçıklardan ortak maddeli iki kitapçık seçilerek klasik test kuramına dayalı eşitleme yöntemlerinden ortalama, lineer, eşit yüzdellikli eşitleme ile eşitlenmiş puanların karşılaştırılması amaçlanmıştır. Testler ortak maddeli denk grup deseni kullanılarak eşitlenmiştir. Araştırma kapsamında eşitlenen testlerin aynı yapıyı ölçtükleri, güvenilirlik katsayılarının, varyanslarının, ortalama güçlüklerinin ve puan ortalamalarının denk olduğu yani aralarında anlamlı bir farklılık olmadığı saptanmıştır. Farklı eşitleme yöntemleri uygulanarak elde edilen bulgulara göre ortalama eşitleme yapılarak elde edilen eşitlenmiş Form X puanı ile eski form olan Form Y puanı karşılaştırıldığında sadece ortalamalarının eşitlendiği görülmektedir. Lineer (doğrusal) eşitleme yapıldığında ortalama ve standart sapmalarının eşitlendiği görülmektedir. Eşit yüzdellikli eşitleme yapıldığında ise ortalama, standart sapma, çarpıklık ve basıklık değerlerinin eşitlendiği görülmektedir. İki kitapçık üzerinden formların eşitlenmesi işlemleri dikkate alındığında puan dağılımlarının en doğru şekilde eşitlenmesini sağlayan yöntemin eşit yüzdellikli eşitleme olduğu sonucuna ulaşılmıştır.

**Abstract:** The aim of study is to compare the scores equated using mean, linear, and equipercentile methods based on classical test theory by choosing two booklets with common items. These tests were equated with common-item equivalent group design. The contents, structures, reliability coefficients, variances, mean difficulties, and score means of these two tests were observed as equal to each other, which means that there was no significant difference between these two tests. According to the results obtained by using different equating methods, it has been found that only these tests' means were equated when Form X and Form Y scores were compared by using mean equating method. Through linear equating, the means and standard deviations of the tests were equated. By means of equipercentile equating, standard deviations, means, and skewness and kurtosis of the tests were equated. After all the equating processes applied to the two booklets, it can be said that the best method to equate the score distributions in the most correct way was equipercentile equating method.

**Keywords:** *PISA, Mean Equating, Linear Equating, Equipercentile Equating.*

**Anahtar Sözcükler:** *PISA, Ortalama Eşitleme, Doğrusal Eşitleme, Eşit Yüzdellikli Eşitleme*

## Giriş

Dünyanın var olması ile başlayan ve toplumun gelişimine büyük katkısı olan kavramlardan birinin eğitim olduğu söylenebilir. Gelişime çabucak ayak uydurabilen, araştıran, sorgulayan, öz güven duygusu yüksek bireyler yetiştirmek eğitim ile mümkündür. Bu bağlamda öğrencilere uygulanan ulusal ve uluslararası ölçekli sınavların, eğitimin kalitesini artırmaya yönelik neler yapılabileceğine ilişkin tespitleri doğru bir şekilde ortaya çıkarması gerekmektedir.

Öğrencilerin başarı düzeylerini artırmak, eğitim politikalarının öğrenci üzerindeki etkisini görmek, eğitim sistemini daha işlevsel hale getirebilmek ve eğitim kalitesini yükseltmek için Türkiye'nin kurucu üyesi olduğu Ekonomik İşbirliği ve Kalkınma Teşkilatı (OECD)'nin, Uluslararası Öğrenci Değerlendirme Programı (PISA), 2000 yılından itibaren uygulanmaktadır (MEB, 2015). Araştırma kapsamında incelenen PISA, OECD ülkelerinde zorunlu eğitimin sonuna yaklaşan 15 yaşındaki öğrencilerin topluma aktif katılımları için bazı temel bilgi ve becerileri ne derece edindiklerini ölçmeyi amaç edinen bir çalışmadır (MEB, 2015).

PISA, bireylerin matematik, fen bilimleri, okuma becerisi gibi temel becerileri gerektiren zorunlu eğitimin sonunda, cemiyete tam olarak iştirak etmesi için bu kazanımlara ne seviyede sahip olduklarını değerlendirmektedir (OECD, 2006). Türkiye, ilk defa PISA 2003 ve son olarak PISA 2015'e katılım göstermiştir. PISA 2003 ve PISA 2012 uluslararası ölçekli sınavda matematik ağırlıklı konu alanı konumundadır. Diğer senelerden farklı olarak matematik konu ağırlıklı uygulanan PISA programlarında, sınavlara katılan öğrencilerin problem çözme becerileri de ölçülmüştür. PISA matematik soruları çoğunlukla matematiksel okuryazarlık ve problem çözme ile ilişkilidir.

Türkiye örneğinde PISA tarafından uygulanan standart 13 kitapçık olduğu görülmektedir. Bu kitapçıkların birbirine denk olan gruplar cevaplandırmıştır. Kitapçıkları cevaplayan grupların karşılaştırılması için uygulanan kitapçıkların eşitlenmesi ve bu bağlamda değerlendirme yapılması gerektiği tespit edilmiştir. Bu kitapçıkların eşitlenebilmesi için birbirleri ile ortak madde içeren kitapçıklar eşitlenmesi bu araştırmanın temel amacıdır. Test eşitleme yöntemleri incelendiğinde gözlenen puanlara dayalı eşitleme yöntemlerinden ortalama, lineer ve eşit yüzdellikli eşitleme yöntemleri aşağıdaki gibi açıklanmaya çalışılmıştır. Test eşitleme bir formun birim sistemini diğer formun birim sistemine dönüştürme yoluna denir (Angoff, 1971). Test eşitleme iki amaç ile gerçekleştirilmektedir. Farklı test formlarını alan bireyler arasında yanlılığı önlemek ve farklı formlardan alınan puanların aynı ölçek üzerinde incelemek ve incelenen puanların anlamını korumak test eşitlemenin başlıca amaçlarıdır (Barnard, 1996).

#### Ortalama Eşitleme

Ortalama eşitlemede tek farklılığın aritmetik ortalamalardan kaynaklandığı yani formların güçlüklerinden kaynaklandığı düşünülür. Ortalama eşitleme, kendi ortalamalarından eşit uzaklıktaki iki formun puanlarının eşit olması üzerine kurulmuştur. Yani grubun kendi aritmetik ortalamasından aynı mesafede uzak olan puanların eşit olduğu söylenir. Ortalama eşitleme, (1) numaralı eşitlikte olduğu gibi ifade edilebilir:

$$\begin{aligned}x - \mu(x) &= y - \mu(y) \\m_Y(x) = y &= x - \mu(x) + \mu(y)\end{aligned}\quad (1)$$

Ortalama eşitleme yeni formu eski forma dönüştürmek için sabit bir puanın eklenmesini içerir. Ortalama eşitlemede, eski formun aritmetik ortalaması ile yeni formun eşitlenmiş puanlarının aritmetik ortalaması ve yeni formun ham puanları ile eşitlenmiş formun dönüştürülmüş puanlarının standart sapması eşittir (Angoff, 1971; Felan, 2002; Kolen and Brennan, 2014).

#### Lineer (Doğrusal) Eşitleme

Lineer eşitleme, ölçek puanı boyunca değişkenlerin iki formu arasındaki zorluk farklılıklarını belirlemeye olanak sağlar. Lineer eşitleme sonucunda üretilen değerler puan düzeyine göre test formlarındaki zorluğun değişimindeki farklılığı gösterir. Lineer eşitleme ortalamalardaki farklılıkların standart sapmaya bölümünün eşitliği üzerine kurulur. Lineer dönüşüm eşitlenmiş iki formdaki standart sapma puanlarının (z-puanlarının) eşitliğini kullanır. Lineer eşitleme (2) numaralı eşitlikte olduğu gibi ifade edilebilir:

$$\frac{x - \mu(x)}{\sigma(x)} = \frac{y - \mu(y)}{\sigma(y)}$$

$$I_Y(x) = y = \sigma(y) \left[ \frac{x - \mu(x)}{\sigma(x)} \right] + \mu(y)$$

$$I_Y(x) = y = \frac{\sigma(y)}{\sigma(x)} x + \left[ \mu(y) - \frac{\sigma(y)}{\sigma(x)} \mu(x) \right] \quad 2)$$

İki formun standart sapmaları eşit ise ortalama ve lineer eşitleme dönüşümleri aynı sonucu üretir. Lineer eşitlemede eski form puanları ile yeni formun eşitlenmiş puanlarının aritmetik ortalaması ve standart sapması eşittir (Angoff, 1971; Felan, 2002; Kolen and Brennan, 2014; Woldbeck, 1998).

### Eşit Yüzdellikli Eşitleme

Eşit yüzdellik eşitlemede, formdan forma zorluktaki farklılığı tanımlamada bir eğri kullanılır. Bu durum eşit yüzdellik eşitlemeyi lineer eşitlemeden daha genel bir fonksiyon yapar. Form Y ölçeğine dönüştürülen form X'e ait puanların dağılımı, evrendeki form Y puanlarının dağılımına eşitse eşit yüzdellikli eşitleme fonksiyonu ortaya çıkmıştır. Eşit yüzdellik eşitleme fonksiyonu form Y deki puanlar gibi aynı yüzdellik aralıklara sahip form X deki puanlar tanımlanarak geliştirilir. Eşit yüzdellik eşitleme gerçekleştirildiğinde hem ortalama hem de lineer eşitleme varsayımlarının sağlandığı söylenebilir. Eşit yüzdellik eşitleme yöntemi ile eşdeğerlerin bulunmasında iki yöntem izlenir: Grafıksel Süreç ve Analitik Süreç (Kolen and Brennan, 2014).

### Grafıksel süreç

Eşitleme yöntemlerinden test eşitleme sürecinde eşit yüzdellikli eşitleme işlemi yapılması aşmasında grafıksel süreç üç adımda gerçekleşir. Aynı grafık üzerinde her bir formun yüzdellik aralığı çizilir, Form X puanının Form Y eşdeğerini bulmak için Form X puanının yüzdellik aralığı bulunur ve aynı yüzdellik aralığındaki Form Y eşdeğeri bulunur (Angoff, 1971).

### Analitik süreç

*Eşit yüzdellikli eşitlemede analitik süreç (3) numaralı eşitlikte olduğu gibi ifade edilebilir.*

$$e_Y(x) = Q^{-1}[P(x)]$$

$$e_Y(x) = \frac{\frac{P(x)}{100} - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + (y_U^* - 0.5)$$

$$P(x) = 100 \left[ F(x-1) + \frac{f(x)}{2} \right] \quad (3)$$

$f(x)$ ,  $x$  puanını elde eden bireylerin oranıdır.

$F(x)$ ,  $x$  ve aşağısındaki değerlerin kümülatif oranıdır.

$P(x)$ , yüzdellik aralık, bir tamsayı puanı için  $x$  in aşağısındaki bireylerin yüzdesi artı  $x$  deki bireylerin yüzdesinin  $\frac{1}{2}$  sidir.

$y_u^*$  belirlemek için  $x$  in form  $X$  dağılımındaki yüzdelik aralığı bulunur. Form  $Y$  dağılımındaki yığılmalı frekans dağılımından, bu yüzdelik aralığının bulunduğu yer belirlenir ve buradaki aralıktaki büyük değeri  $y_u^*$  alınır.

Eşit yüzdelikli eşitleme analitik süreçte, hem form  $X$  için  $f(x)$ ,  $F(x)$  ve  $P(x)$  değerlerini ve form  $Y$  için  $g(y)$ ,  $G(y)$  ve  $Q(y)$  değerlerini hesaplamak gerekir. Sonra belirli bir puanın yüzdelik aralığına bakılabilir. Her bir Form  $X$  puanı için Form  $Y$  de karşılık gelebilecek eşitlenmiş  $X$  puanları hesaplanır. Analitik süreçte elde edilen değer grafiksel sürece göre elde edilen değerden daha güvenilirdir (Angoff, 1971; Felan, 2002; Kolen and Brennan, 2014; Woldbeck, 1998). Klasik test kuramına dayalı olarak yapılan araştırmalar aşağıdaki şekilde incelenmiştir.

Öztürk ve Anıl (2012)'ın Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavına (ALES) ilişkin farklı tarihlerde farklı formlarının eşitlenmesine ilişkin yaptıkları araştırmada, 2008 ilkbahar ve sonbahar ALES sayısal-1 testi puanlarını doğrusal ve eşit yüzdelikli eşitleme yöntemlerini kullanarak eşitlemek ve bu yöntemlerden uygun olanını önermektir. Elde edilen bulgular, doğrusal ve eşit yüzdelikli eşitleme sonucu ham puanlar ile eşitlenmiş puanlar arasında doğrusal bir ilişki olduğunu göstermiştir. Doğrusal eşitleme sonucu testlerin güçlük düzeyinin puan ölçeği boyunca farklılaştığı; eşit yüzdelikli eşitleme sonucu sonbahar testinin alt puanlarda ilkbahar testinden daha zor, üst puanlarda ise daha kolay olduğu görülmüştür. Araştırmanın sonunda, testlerin eşitlenmesi için en uygun yöntemin eşit yüzdelikli eşitleme olduğu sonucuna ulaşılmıştır.

Şahhüseyinoğlu (2006)'nun yaptığı İngilizce yeterlik sınavında kullanılan testlerin birbirine eşdeğer olup olmadığı, eşdeğer değilse, test eşitleme yöntemleri ile farklı sınavlara ait alınan puanların birbirine dönüştürülmesi ve testlerin birbirilerinin aynı olması sağlanması hedeflenmiştir. Eşitlemede kullanılan yöntemler, Klasik Test Kuramına dayalı eşit yüzdelikli ve doğrusal eşitleme yöntemleri ve Örtük Özellikler Kuramına dayalı Rasch parametresinin karakteristik eğri yöntemleri kullanılmıştır. Hacettepe Üniversitesi İngilizce Hazırlık Bölümünün yaptığı farklı yıllara ait puanlar eşitlenmiştir. Sonuç olarak Rasch yöntemi ve eşit yüzdelikli eşitleme yöntemlerinin uygun yöntemler olduğu bulunmuştur.

Bozdağ (2007)'in yaptığı yüksek lisans tez çalışmasında, şans başarısının test eşitlemeye etkisi olup olmadığını incelemektedir. Araştırmada eşitleme düzeneklerinden, tek grup düzeneği ve klasik eşitleme yöntemlerinden, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılmıştır. Ayrıca OKS 2003 ve 2005 Türkçe alt testlerine ait puanları bu yöntemlere göre eşitlemek ve bu yöntemler arasında uygun olanını önerme amaçlanmaktadır. Şanstan arındırılmış ve arındırılmamış testlerin ikisi için de testlerin ortalama güçlükleri, güvenilirlikleri, ortalamaları ve varyansları arasında anlamlı bir fark olmadığı görülmüştür. En uygun eşitleme yöntemi ise eşit yüzdelikli eşitleme yöntemi olarak kabul edilmiştir.

Yapılan araştırmalardan incelendiğinde, PISA uygulamasında yer alan alt testlerin ortak maddeler baz alınarak Klasik Test Kuramına dayalı eşitleme yöntemlerinden ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılarak en doğru eşitlemenin hangi yöntemle olabileceğine ilişkin çalışmanın olmaması açısından önemli görülmektedir. Bu bağlamda araştırmaya ilişkin belirlenen amaç aşağıdaki şekilde ifade edilmiştir.

#### Araştırmanın Amacı

Bu araştırmada Klasik Test Kuramına dayalı eşitleme yöntemlerinden ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılarak PISA 2012 matematik alt testlerinin eşitlenmesi amaçlanmaktadır.

## Yöntem

### Çalışma Grubu

PISA 2012 Türkiye örnekleminde 7 numaralı kitapçığı (X formu) cevaplayan 370 kişi, 6 numaralı kitapçığı (Y formu) cevaplayan 372 kişi bulunmaktadır. Bu kitapçıklara verilen cevaplar incelendiğinde X formunda bulunan 35 puan alan bir uç değer veri setinden çıkarılmıştır. Son haliyle, araştırma grubunu X formunda yer alan 369 birey ve Y formunda yer alan 372 birey oluşturmuştur.

### Veri Toplama Aracı ve Eşitleme Deseni

Araştırma kapsamında test eşitleme işlemlerini gerçekleştirmek için PISA 2012 verilerinden Türkiye örneklemine bağlı veri setinden kitapçık 7 ve kitapçık 10 matematik kitapçıklarına ait veriler incelenmiştir. Frekans analizi yapılarak kitapçıklarda bulunan soruları cevaplayan bireyler çalışmaya alınmıştır. Bu iki kitapçıkta 12 ortak madde olmak üzere 7'inci kitapçıkta toplam 35 madde bulunmaktadır ve kitapçık 7 eski form (Y formu) olarak belirlenmiştir. Kitapçık10'da toplam 36 madde bulunmaktadır ve yeni form (Form X) olarak belirlenmiştir. Belirlenen iki kitapçığın eşitlenmesinde ortak maddeli denk grup deseni kullanılmıştır.

### Verilerin Analizi

PISA 2012 Türkiye örneklemine ait matematik testinde 0-1-2 şeklinde puanlanan maddeler, 0-1 puanlanacak şekilde 2'ler 1 olacak şekilde değerlendirmeye alınmıştır. Doğru cevaplar 1 ve yanlış cevaplar 0 olarak kodlanmıştır. Genel itibari ile doğru cevap puanları toplanarak X formuna ve Y formuna ilişkin betimsel istatistikler belirlenmiştir. Kitapçıkların kapsam olarak aynı olduğu PISA 2012 raporlarında görülmektedir. Tek boyutluluk varsayımlarının sınaması için 1/0 puanlanan yapı olduğundan dolayı Factor Version 10.3.01 64 Bits (Lorenzo-Seva and Ferrando, 2015) programı ile Tetrakorik korelasyon matrisine dayalı olarak açımlayıcı faktör analizi yapılmıştır.

Testlerin ortalama güçlük indeksleri ve test varyanslarının hesaplamasında TAB (Brooks and Johanson, 2003) programı kullanılmıştır. Bu araştırma için sadece açımlayıcı faktör analizi yapılması yeterlidir. Doğrulamalı faktör analizine gerek yoktur. Her iki testten elde edilen puanların eşit güvenilirlik düzeyine sahip olup olmadığını belirlemek üzere, öncelikle her iki testten elde edilen puanlara ait KR-20 güvenilirlik katsayıları hesaplanmış daha sonra bu katsayılar arasında istatistiksel olarak fark olup olmadığı Fischer'in Zr dönüşümü kullanılarak test edilmiştir (Angof, 1971; Jaeger 1981).

Ayrıca her iki teste ait varyans ve ortalama güçlük düzeyleri hesaplanmış ve her iki teste ait bu istatistikler arasında anlamlı bir fark olup olmadığı t testleri aracılığıyla belirlenmiştir. KTK'ya dayalı eşitleme yöntemleri uygulanırken Excel ve RAGE-RGEQUATE programları kullanılmıştır.

### **Varsayımların sınanması**

#### *Tek boyutluluk varsayımı*

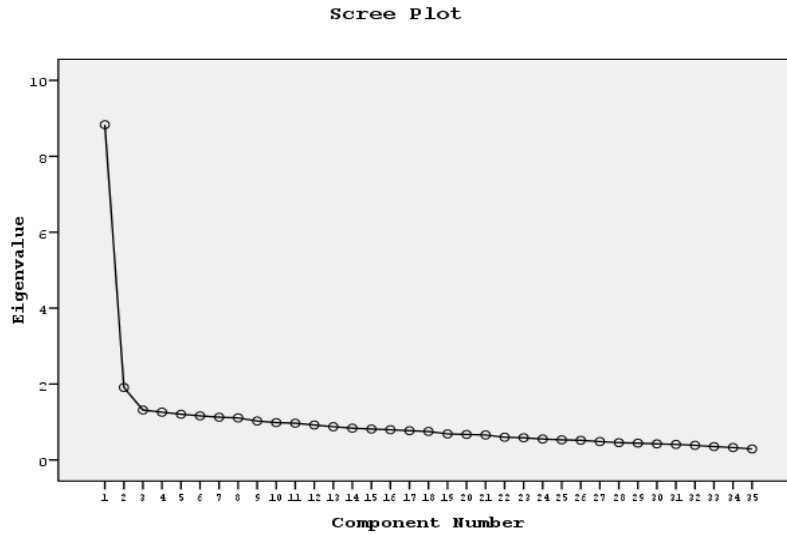
İlk olarak eşitlenmek istenen testlerin tek boyutlu olup olmadıkları ve aynı yapıyı ölçüp ölçmedikleri belirlenmeye çalışılmıştır. Tetrakorik korelasyon matrisine dayalı açıklayıcı faktör analizi sonuçları incelendiğinde:

Form Y (Kitapçık 7) için, Kaiser-Meyer-Olkin (KMO) test 0,91, Bartlett's istatistiği 3001.3 (sd=630; p=0.000010), bu değerlere bakıldığında faktör analizi yapılabileceği söylenebilir.

Tablo 1. Form Y (Kitapçık 7) İçin Başlangıç Özdeğerleri

Faktör	Başlangıç Öz-Değer		
	Toplam	Varyans %	Yığılmalı %
1	8,833	25,238	25,238
2	1,909	5,453	30,691
3	1,314	3,754	34,445
4	1,257	3,593	38,038
5	1,204	3,439	41,477
6	1,161	3,317	44,794
7	1,126	3,216	48,010
8	1,107	3,164	51,173
9	1,025	2,928	54,101

Tablo-1'e bakıldığında faktörlerin başlangıç öz değerlerine göre kritik değer olan 1 öz değerinden daha yüksek olan 9 faktör olduğu fakat birinci faktörün öz değerinin ikinci faktörün öz değerinin 5 katından fazla olduğu ikinci faktörün öz değerinin ise diğer faktörlerin öz değerlerine yakın olmasından dolayı tek faktör varsayımının sağlanabileceği söylenebilir. Faktörlerin başlangıç öz-değerlerine ait saçılma diyagramı Şekil 1.'de görülmektedir.



Şekil 1. Form Y Faktörlerinin Başlangıç Öz-Değerlerine Ait Scree-Plot (Saçılma Diyagramı)

Şekil-1 incelendiğinde faktörlerin başlangıç öz-değerleri arasındaki farkların yukarıdaki gibi olduğu görülmektedir. Saçılma diyagramına göre başlangıç öz-değerlerinde büyük bir kırılma noktasının olduğu ve diğer kırılma noktalarının çok küçük olmasından dolayı Form-Y kitapçığının tek faktörlü bir yapıya sahip olduğu görülmektedir.

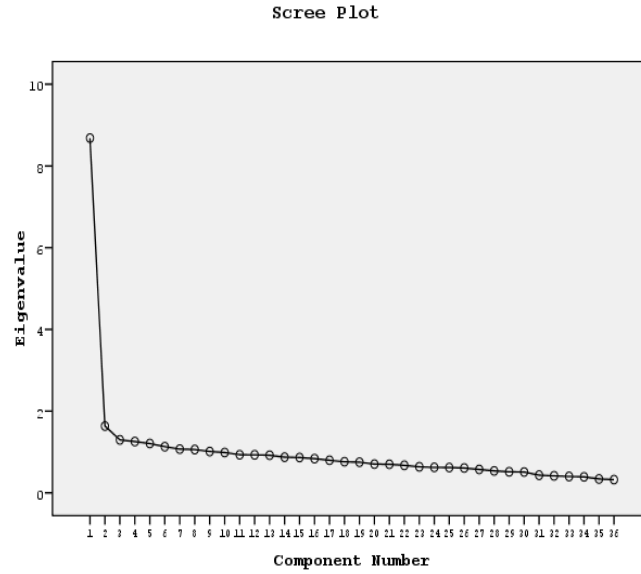
Form X (Kitapçık 10) için, Kaiser-Meyer-Olkin (KMO) test 0,91, Bartlett's istatistiği 3001.3 (sd=630; p=0.000010), bu değerlere bakıldığında faktör analizi yapılabileceği görülmektedir. Sadece açımlayıcı faktör analizi yapmak yeterlidir. Doğrulayıcıya gerek yoktur.

Tablo 2. Form X (Kitapçık 10) İçin Başlangıç Özdeğerleri

Faktör	Başlangıç Öz-Değer		
	Toplam	Varyans %	Yığılmalı %
1	8,683	24,119	24,119
2	1,635	4,541	28,660
3	1,297	3,603	32,263
4	1,255	3,486	35,749
5	1,207	3,353	39,102

6	1,130	3,140	42,242
7	1,069	2,969	45,211
8	1,061	2,946	48,157
9	1,012	2,811	50,968

Tablo-2'ye bakıldığında faktörlerin başlangıç öz değerlerine göre kritik değer olan 1 öz değerinden daha yüksek olan 9 faktör olduğu fakat birinci faktörün öz değerinin ikinci faktörün öz değerinin 5 katından fazla olduğu ikinci faktörün öz değerinin ise diğer faktörlerin öz değerlerine yakın olmasından dolayı tek faktör varsayımının sağlanabileceği söylenebilir. Faktörlerin başlangıç öz-değerlerine ait saçılma diyagramı Şekil 2.'de görülmektedir.



Şekil 2. Form X Faktörlerinin Başlangıç Öz-Değerine Ait Scree-Plot (Saçılma Diyagramı)

Şekil-2 incelendiğinde faktörlerin başlangıç öz-değerleri arasındaki farkların yukarıdaki gibi olduğu görülmektedir. Saçılma diyagramına göre başlangıç öz-değerlerinde büyük bir kırılma noktasının olduğu ve diğer kırılma noktalarının çok küçük olmasından dolayı Form-X kitapçığının tek faktörlü bir yapıya sahip olduğu görülmektedir.

#### Kitapçıkların güvenilirliklerin eşit olma durumu

Güvenirlik katsayıları iki korelasyon katsayısı olarak kabul edilmiş ve güvenilirlik katsayılarına Fischer'in Z dönüştürmesi yapılmıştır. İki güvenilirlik katsayısı arasında manidar bir fark olup olmadığı Fischer'in Z istatistiği ile test edilmiştir. Z istatistiğine ilişkin sonuçlar Tablo-3'de görülmektedir.

Tablo 3. Kitapçıkların Güvenirlikleri Arasındaki Farklılığa İlişkin Z İstatistiği Sonuçları

Kitapçıklar	KR-20	Zr	Z
Form Y (Kitapçık 7)	0,86	1,24	
Form X (Kitapçık 10)	0,84	1,26	0,571

\* $p < ,05$

Tablo-3'e bakıldığında Form Y ve Form X kitapçıklarının Kuder-Richardson 20 iç tutarlılık katsayısı ile hesaplanan güvenilirlik katsayıları arasındaki farklılığın incelendiği görülmektedir. Güvenirlik katsayıları Fisher'in Zr oran dönüşümü yapılarak aralarındaki farklılık  $Z=0,571$ ,  $p > ,05$ 'e göre anlamlı bir farklılık olmadığı görülmektedir. Kitapçıkların güvenilirlikleri arasında manidar olarak farklılık olmadığı gözlenmiştir.

#### Kitapçıkların güçlük düzeylerinin karşılaştırılması



İki kitapçık arasındaki güçlük düzeyleri arasındaki farklılıkların incelenmesinde t istatistiğinden yararlanılmıştır. T istatistiğine ilişkin sonuçlar Tablo-4’de görülmektedir.

Tablo 4. *Kitapçıkların Güvenirlikleri Arasındaki Farklılığa İlişkin T İstatistiği Sonuçları*

Kitapçıklar	Ortalama Güçlük	t	p
Form Y (Kitapçık 7)	0,38		
Form X (Kitapçık 10)	0,40	,529	0,598

\*p<,05

Tablo-4’e bakıldığında Form Y ve Form X kitapçıklarının ortalama güçlük düzeylerine ilişkin katsayılar arasındaki farklılığın incelendiği görülmektedir. Ortalama güçlük düzeylerine ait katsayılar arasında t=0,529, p>,05’e göre anlamlı bir farklılık olmadığı görülmektedir. Kitapçıkların ortalama güçlükleri arasında manidar olarak farklılık olmadığı gözlenmiştir.

#### *Kitapçıkların varyanslarının karşılaştırılması*

İki kitapçık arasındaki varyans değerleri arasındaki farklılıkların incelenmesinde t istatistiğinden yararlanılmıştır. t istatistiğine ilişkin sonuçlar Tablo-5’te görülmektedir.

Tablo 5. *Kitapçıkların Varyansları Arasındaki Farklılığa İlişkin T İstatistiği Sonuçları*

Kitapçıklar	Varyans	t	p
Form Y (Kitapçık 7)	57,00		
Form X (Kitapçık 10)	51,41	,432	0,667

\*p<,05

Tablo-5’e bakıldığında Form Y ve Form X kitapçıklarının varyans düzeylerine ilişkin katsayılar arasındaki farklılığın incelendiği görülmektedir. Varyans düzeylerine ait katsayılar arasında t=0,432, p>,05’e göre anlamlı bir farklılık olmadığı görülmektedir. Kitapçıkların varyansları arasında manidar olarak farklılık olmadığı gözlenmiştir.

## **Bulgular ve Yorum**

### Ortalama Eşitleme

Kitapçıklara ait eşitleme için dönüştürülmüş ve dönüştürülmemiş puanların momentleri gözlenen puan eşitleme yöntemlerinden ortalama eşitleme yöntemi ile karşılaştırılması Tablo 6.’da görülmektedir.

Tablo 6. *Form X ve Form Y Kitapçıklarının Ortalama Eşitleme Yöntemine Dayalı Eşitlenmiş Puanlar*

	Ortalama	Standart Sapma	Çarpıklık	Basıklık
Form X	14,26	7,55	0,53	-0,49
Form Y	12,86	7,17	0,75	-0,21
Eşitlenmiş X	12,86	7,55	0,53	-0,49

Tablo 6’ya bakıldığında kitapçıklardan Form X ve Form Y’nin ortalama, standart sapma, çarpıklık ve basıklık değerleri karşılaştırıldığında eşit olmadığı görülmektedir. Kitapçıklara ilişkin dört momente ait hesaplanan değerler karşılaştırıldığında momentlerin eşitlenmediği görülmektedir. Aynı bilgiyi ölçen iki kitapçık arasında farklılık olduğu bu kitapçıkların eşitlenmesi için gözlenen puan eşitleme yöntemlerinden ortalama eşitleme yöntemi kullanılarak Form X puanı ile Form Y puanı eşitlenmeye çalışılmıştır.



Bu yöntem kullanıldığında dönüştürülmüş puanlar incelenerek tekrardan moment puanları karşılaştırılmıştır. Eşitlenmiş X puanı ile Form Y puanları karşılaştırıldığında sadece ortalama puanların eşit olduğu diğer üç momentin eşit olmadığı görülmektedir. Ortalama eşitleme yöntemi kullanılarak kitapçıklar arasında sadece bir momentin eşitlenebildiği görülmektedir.

#### Lineer (Doğrusal) Eşitleme

Kitapçıklara ait eşitleme için dönüştürülmüş ve dönüştürülmemiş puanların momentleri gözlenen puan eşitleme yöntemlerinden lineer eşitleme yöntemi ile karşılaştırılması Tablo 7.'de görülmektedir.

Tablo 7. Form X ve Form Y Kitapçıklarının Lineer Eşitleme Yöntemine Dayalı Eşitlenmiş Puanlar

	Ortalama	Standart Sapma	Çarpıklık	Basıklık
Form X	14,26	7,55	0,53	-0,49
Form Y	12,86	7,17	0,75	-0,21
Eşitlenmiş X	12,86	7,17	0,53	-0,49

Tablo 7'ye bakıldığında kitapçıklardan Form X ve Form Y'nin ortalama, standart sapma, çarpıklık ve basıklık değerleri karşılaştırıldığında eşit olmadığı görülmektedir. Kitapçılara ilişkin dört momente ait hesaplanan değerler karşılaştırıldığında momentlerin eşitlenmediği görülmektedir.

Aynı bilgiyi ölçen iki kitapçık arasında farklılık olduğu bu kitapçıkların eşitlenmesi için gözlenen puan eşitleme yöntemlerinden lineer eşitleme yöntemi kullanılarak Form X puanı ile Form Y puanı eşitlenmeye çalışılmıştır. Bu yöntem kullanıldığında dönüştürülmüş puanlar incelenerek tekrardan moment puanları karşılaştırılmıştır.

Eşitlenmiş X puanı ile Form Y puanları karşılaştırıldığında sadece ortalama puanların ve standart sapmaların eşit olduğu diğer iki momentin eşit olmadığı görülmektedir. Lineer eşitleme yöntemi kullanılarak kitapçıklar arasında sadece iki momentin eşitlenebildiği görülmektedir.

#### Eşit Yüzdellikli Eşitleme

Kitapçıklara ait eşitleme için dönüştürülmüş ve dönüştürülmemiş puanların momentleri gözlenen puan eşitleme yöntemlerinden eşit yüzdellikli eşitleme yöntemi ile karşılaştırılması Tablo 8.'de incelenmektedir.

Tablo 8. Form X ve Form Y Kitapçıklarının Eşit Yüzdellikli Eşitleme Yöntemine Dayalı Eşitlenmiş Puanlar

	Ortalama	Standart Sapma	Çarpıklık	Basıklık
Form X	14,26	7,55	0,53	-0,49
Form Y	12,86	7,17	0,75	-0,21
Eşitlenmiş X	12,86	7,17	0,76	-0,20

Tablo 8'e bakıldığında kitapçıklardan Form X ve Form Y'nin ortalama, standart sapma, çarpıklık ve basıklık değerleri karşılaştırıldığında eşit olmadığı görülmektedir. Kitapçılara ilişkin dört momente ait hesaplanan değerler karşılaştırıldığında momentlerin eşitlenmediği görülmektedir. Aynı bilgiyi ölçen iki kitapçık arasında farklılık olduğu bu kitapçıkların eşitlenmesi için gözlenen puan eşitleme yöntemlerinden eşit yüzdellikli eşitleme yöntemi kullanılarak Form X puanı ile Form Y puanı eşitlenmeye çalışılmıştır. Bu yöntem kullanıldığında dönüştürülmüş puanlar incelenerek tekrardan moment puanları karşılaştırılmıştır.

Eşitlenmiş X puanı ile Form Y puanları karşılaştırıldığında sadece ortalama puanları, standart sapmaları eşit olduğu çarpıklık ve basıklık değerlerinin birbirine çok yakın olduğu görülmektedir. Eşit yüzdelikli eşitleme yöntemi kullanılarak kitapçıklar arasında dört momentin eşitlenebildiği görülmektedir.

## Sonuç ve Öneriler

PISA 2012 matematik alt testinde yer alan iki kitapçığın Türkiye verileri kullanılarak ortak maddeler aracılığı ile eski formdan elde edilen puanlar ile yeni formdan elde edilen puanların birbirine denk olabilmesi için gözlenen puan eşitleme yöntemlerinden KTK'ya dayalı eşitleme yöntemleri kullanılmıştır. Eşitleme yöntemlerinden elde edilen bulgulara ilişkin sonuçlar aşağıdaki gibidir:

Ortalama eşitleme yöntemi kullanılarak hesaplanan dört momente ait eşitlenmiş puanlar ile eski form olan Form Y'ye ait dört moment karşılaştırıldığında sadece ortalamaların birbirine eşit olduğu geriye kalan üç momentin eşit olmadığı sonucuna ulaşılmıştır.

Lineer (doğrusal) eşitleme yöntemi kullanılarak hesaplanan dört momente ait eşitlenmiş puanlar ile eski form olan Form Y'ye ait dört moment karşılaştırıldığında ortalama ve standart sapmalarının birbirine eşit olduğu geriye kalan iki momentin eşit olmadığı sonucuna ulaşılmıştır. Eşit yüzdelikli eşitleme yöntemi kullanılarak hesaplanan dört momente ait eşitlenmiş puanlar ile eski form olan Form Y'ye ait dört moment karşılaştırıldığında ortalama, standart sapma, çarpıklık ve basıklık değerlerinin birbirine eşit olduğu sonucuna ulaşılmıştır.

İki kitapçık üzerinden formların eşitlenmesi işlemleri dikkate alındığında puan dağılımlarının en doğru şekilde eşitlenmesini sağlayan gözlenen puan eşitleme yönteminin eşit yüzdelikli eşitleme yöntemi olduğu sonucuna ulaşılmıştır. Eşit yüzdelikli eşitlemede puan aralıklarının daha küçük olmasından dolayı iki formun puan dağılımları birbirine denk gelir (Angoff, 1971; Felan, 2002; Kolen and Brennan, 2014; Woldbeck, 1998).

Ulusal ve uluslararası ölçekli sınavların farklı dönemlerde uygulanması ve hesaplanan puanların karşılaştırılması için testlerin eşitlenmesine ihtiyaç duyulmaktadır. Bu sınavlardan PISA 2012 Matematik alt testine ait iki kitapçıktan elde edilen puanlar eşitlenmiştir. Bu süreçte Klasik test kuramına dayalı eşitleme yöntemlerinden eşit yüzdelikli eşitleme yönteminin daha doğru bir yöntem olduğu sonucuna ulaşılmıştır. Bu sonuç benzer araştırmalarla da desteklenmektedir (Bozdağ, 2007; Öztürk ve Anıl, 2012; Şahhüseyinoğlu, 2006).

Bu çalışmanın sonuçları doğrultusunda ülkemizde yapılan birçok seçme amaçlı sınavın farklı versiyonlarının benzer sonuçlar üretip üretmediğinin kontrol edilmesi gerektiği söylenebilir. PISA gibi uluslararası yapılan sınavlarda ve alt testlerinde birden fazla kullanılan kitapçıkların eşitlenmesi sürecinde farklı test eşitleme düzenekleri (dengelenmiş grup, eşdeğer grup, iç anchor, dış anchor vb.) çalışılarak KTK ve MTK dayalı test eşitleme yöntemleri ile eşitleme işlemleri gerçekleştirilebilir.

## Kaynakça

- 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.
- Barnard, J. (1996). In search for equity in educational measurement: Traditional versus modern equating. Methods, ASEESA. *In National Conference, Pretoria South Africa.*

- Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi* (Yüksek Lisans Tezi). <https://tez.yok.gov.tr/> adresinden erişilmiştir.
- Brooks, G. P., & Johanson, G. A. (2003). TAP: Test analysis program. *Applied Psychological Measurement*, 27(4), 303-304.
- Felan, George D. (2002). Test equating: Mean, linear, equipercentile, and item response theory. *ERIC*. <http://eric.ed.gov/?id=ED462436> adresinden erişilmiştir.
- Harman, H. H. (1962). *Modern Factor Analysis*, 2nd Edition. University of Chicago Press, Chicago.
- Jaeger, M. R. (1981). Some exploratory indices for selection of a test equating method. *Journal of Educational Measurement*. 18(1), 23-38.
- Kelley, T. L. (1935). *Essential traits of mental life*. Harvard University Press. Cambridge.
- Kolen, M. J. & Brennan, R. L. (2014). *Test equating, scaling, and linking*. New York: Springer.
- Lorenzo-Seva, U & Ferrando, P. J. (2015). *Factor version 10.3.01*. Rovira i Virgili University. Spain: Tarragona
- MEB (2015). *PISA 2012 araştırması ulusal nihai raporu*. 18.02.2016 tarihinde [http://pisa.meb.gov.tr/?page\\_id=22](http://pisa.meb.gov.tr/?page_id=22) adresinden erişilmiştir.
- OECD (2006). *Assessing scientific, reading ad mathematical literacy: A framework for PISA 2006*, Paris: OECD Publications.
- Öztürk, N., & Anıl, D. (2012). Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma. *Eğitim ve Bilim*, 37(165).
- Şahhüseyinoğlu, D. (2006). İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(31).
- Woldbeck, Tanya (1998). Basic concepts in modern methods of test equating. *ERIC*. <http://eric.ed.gov/?id=ED417215> adresinden erişilmiştir.

### Extended Abstract

It can be said that education has been one of the most important concepts that contribute to the development of the human society since the beginning of the human history in the world. Education makes it possible to train individuals who can quickly adapt to developments and improvements, question the situations in which they are and the events that they go through, search on whatever they want to learn and have a high level of self-confidence. In this context, international and national exams should reveal the notifications that can be done to increase the quality of education correctly.

Programme of International Student Assessment (PISA), which was organized by Organization of Economic Cooperation and Development (OECD) in 2012 in order to increase students' academic achievement, determine the effect of educational programmes on students, make educational systems more functional and increase the quality of education, has 13 booklets for Mathematics Subtests. The aim of this study is to compare the scores equated through average, linear, equipercentile method based on classical test theory by choosing two booklets with common items.

Test equating means adjusting scores on one test form to scores on another test form so that scores on the forms can be used interchangeably. Test equating is done for two different purposes. These purposes are to avoid the bias among the individuals that take different test forms and examine the scores obtained from different test forms on the same scale and remain the significance of these scores. It is thought that the only difference in mean equating among the observed score equating methods is caused by arithmetic mean, that is, this difference is caused by the difficulties of the test forms.

Mean equating is based on the equation of the mean differences of the two test forms on the same score scale. In other words, scores that are an equal distance from their means are set equal. Linear equating makes it possible to determine the difficulty difference between the variables of the two test forms along the scale scores. The scores obtained through linear equating show the difference in the difficulty change in the test forms according to the score levels. Linear equating is based on the equation of the differences in the mean to the standart deviation units. Linear formation uses the equation of the standart deviation scores in the two test form that are equated. In the equipercentile equating, a curve is used to define the difference in difficulty from one test form to another test form.

This situation makes equipercentile equating a more general function than linear equating. The equating function is the equipercentile equating fuction if the score distribution in the Form X transformed into Form Y scale is equal to the score distribution in the Form Y in the population. Equipercentile equating function is developed by defining the scores in the Form X like the scores in the Form Y that has the same percentile distance. When equipercentile equating is provided, both mean and linear equating assumptions are said to be provided. Two methods are followed to find the equivalence through equipercentile equating method: The Graphical Process and the Analytical Process.

It is aimed to equate the Turkish students' scores obtained from two test forms considered as the old and new forms with the same number of common items chosen among the 13 test booklets in mathematical subtests in PISA 2012 through mean, linear, equipercentile method based on classical test theory.

In the scope of the research, the data taken form the Test Booklet 7 and the Test Booklet 10 in the mathematics subtests from the Turkey sampling in PISA 2012 were examined to fulfill the test equating functions. The individuals who answered the items in these booklets were chosen through frequency analysis. There were 12 common items in these two items. The Booklet 7 was chosen as the old test form Y with its 35 items and the Booklet 10 was chosen as the new test form X with its 36 items. In this context, the common item balanced group design was used in the study. The new Form X was answered by 370 individuals and the old Form Y was answered by 372 individuals.

The items scored as 0-1-2 was changed into the ones scored as 0-1 and the "2" scores were regarded as 1. The correct answers were coded as 1 and the wrong answers were coded as 0. Generally, the descriptive statistics about the Form X and the Form Y were calculated by summing up the correct answer scores. The fact that the booklets have the same content can be seen in the PISA 2012 reports. Tetrachoric Factor Analysis through Factor Version 10.3.01 64 Bits program was conducted to test the one dimensionality assumption because 1/0 scoring structure was observed. TAB program was used to calculate the test mean difficulty indexs and test variances. Confirmatory factor analysis through LISREL was conducted to confirm the one factor structure.

To determine whether the scores obtained from the two test forms have the same reliability level, KR-20 reliability coefficient about the scores was calculated and then whether there is a statistical difference between these two coefficients was tested through Fischer's Zr transformation. Furthermore, variances and mean difficulty levels about two test forms were calculated and whether there is a significant difference between these statistics was tested through t-tests. Excel and RAGE-RGEQUATE programs were used while equating methods based on classic test theory were being applied.

The fact that the booklets have the same content is seen in PISA 2012 reports. Considering the tetrachoric explanatory factor analysis results, according to the tetrachoric explanatory factor analysis results for Form X (Booklet 10), it was seen that factor analysis can be conducted and one factor is compatible. When RMSR

scores were examined, one factor structure was confirmed as they were lower than the critical scores. According to the tetrachoric explanatory factor analysis results for Form Y (Booklet 7), it was seen that factor analysis can be conducted and one factor is compatible. When RMSR scores were examined, one factor structure was confirmed as they were lower than the critical scores.

KR-20 reliability coefficient for each subtest was calculated to test whether the tests to be equated have the same reliability level. It was observed that there is no significant difference between the reliability level of the tests. It was seen that there is no significant difference between the test mean difficulty level and test mean scores when compared through t-test and there is no significant difference between the variances of the test forms when compared through F-tests. According to the results from the confirmatory factor analysis conducted to confirm one factor structure about the Form X and Form Y, it was seen that the fit index scores were at an acceptable level.

The equating methods based on the classical test theory among the observed score equating methods were utilized to equate the scores of the old test form and the new test form chosen from the PISA 2012 Mathematics Subtest Booklets applied in Turkey by focusing on the common items. The results about the findings that were obtained through equating methods are as follows:

When the equated scores about the 4 moments calculated through mean equating method were compared to the scores about the 4 moments in the Form Y chosen as the old form, it was observed that only the mean scores were equal to each other but the other 3 moments scores were not equal to each other.

When the equated scores about the 4 moments calculated through linear equating method were compared to the scores about the 4 moments in the Form Y chosen as the old form, it was observed that only the mean scores and the standart deviation scores were equal to each other but the other 2 moments scores were not equal to each other.

When the equated scores about the 4 moments calculated through equipercentile equating method were compared to the scores about the 4 moments in the Form Y chosen as the old form, it was observed that the mean scores, standart deviation scores, skewness and kurtosis scores were equal to each other.

Considering the equating functions conducted for two test forms, it was understood that the equipercentile equating method was the observed score equating method that proves the most correct score distribution equating. In equipercentile equating, since the score ranges are less, the score distributions of the two test forms were equal to each other.

In light of the results of this study, it can be suggested that whether the different versions of many exams in Turkey produce similar results should be checked. To equate the international tests and their subtests booklets like PISA, different test equating models can be studied and the equating can be realized through test equating methods based on classical test theory and item response theory. Further studies can be conducted by comparing the equating results obtained through a structured random group design with the results obtained through the equating model structured in this study.