

Comparing Linear Equating and Equipercetile Equating Methods Using Random Groups Design

Hülya Kelecioğlu¹ and Neşe Öztürk Gübeş²

^{1,2} Hacettepe University, Department of Educational Sciences, Turkey

ARTICLE INFO

Article History:

Received 02.04.2012
 Received in revised form
 24.09.2012
 Accepted 28.09.2012
 Available online
 10.04.2013

ABSTRACT

In this study, raw scores which taken from 9th grade 2009 ÖBBS Form C of Social Sciences were equated to 2009 ÖBBS Form A of Social Sciences with linear equating and three different (unsmoothed, presmoothed and postsmoothed) equipercetile equating methods. The random equating error of different equating methods was compared by Root Mean Squared Difference (RMSD) index and mean of bootstrap standard errors (MBSE). The results indicated that there was a linear relationship between equivalent scores of Form A and raw score of Form C, and Form C was easier than Form A along the score scale. It was seen that results based different equipercetile equating method were consistent with each other, there was a nonlinear relationship between them and test difficulty varied along the score scale. Finally, although the least MBSE and RMSD coefficients were got from linear equating because of nonlinear relationship between raw scores and equivalent scores, concluded that the most appropriate equating method was postsmoothed equipercetile equating which had relatively the lowest random equating error among equipercetile equating methods.

© 2013 IOJES. All rights reserved

Keywords:

Linear equating, equipercetile equating, presmoothing, postsmoothing, random error.

Extended Summary

Purpose

To provide test security and to eliminate negative effect of using the same test form in different test administrations, multiple test forms are used in large scale testing. Although different forms for given test are built to be very similar two forms of a test can't be expected to be precisely equivalent in level of difficulty. Therefore, test equating is necessary for getting comparable test scores from different test forms.

ÖBBS is a national large scale examination for determining students' achievement in Turkey. ÖBBS is administered by Ministry of National Education and the aim of this examination is assessing primary and secondary school students' achievement routinely. For providing test security, four test booklets (A, B, C, D) are used in the administrations. Although these test booklets are built to be very similar content and difficulty levels, they may have different difficulty levels. The aim of this study is equating raw scores of 9th grade 2009 ÖBBS Social Sciences Form C to Form A with linear equating and three different equipercetile equating methods and to determine which equating method is the most appropriate for equating these scores.

² Corresponding author's address: Hacettepe University, Department of Educational Sciences, Ankara/Turkey.

Telephone: +90 312 297 85 50

Fax: +90 312 299 20 27

e-mail: neseozturk@hacettepe.edu.tr

Method

The population of this study is 9th grade students in secondary education. Social Sciences subtests in ÖBBS had Form A, B, C and D and A-B and C-D test forms were constructed parallel with respect to measuring the same trait (EARGED, 2010). This research conducted with using random groups design so that samples of this study consisted of 16670 and 15743 9th grade students which taken Form A and Form C.

Social Sciences tests have 15 items. While administering test forms, a spiraling procedure was carried out. Form A, B, C and D alternated when the test booklets packaged. When test booklets handed out, the first examine received Form A, the second examine Form B, third examine Form C and the fourth examine Form D and so on. This spiraling procedure leded randomly two (A-C, B-D) equivalent groups.

This research was conducted with equating Form C raw scores to Form A raw scores using linear equating and unsmoothed equipercentile equating, presmoothed (log-linear method with $C=5$) equipercentile equating and postsmoothed ($S=0.01$) equipercentile equating. Four equating methods were compared with mean of bootstrap standard error of equating (MBSE) and Root Mean Squared Difference (RMSD) coefficients. *The Equating Error* (Hanson, 2004) computer program was used to conduct the equating and getting mean of bootstrap standard errors of equating and 500 bootstrap replications were used.

Results

The results of linear equating showed that equivalent scores of Form A got values between -0.478 and 14.790 points. And all raw scores of Form C were higher than Form A equivalent scores. According to this result, it could be said that there was a linear relationship between Form C raw scores and Form A equivalent scores and Form C was easier than Form A along the score scale.

The results based different equipercentile equating methods were consistent with each other, there was a nonlinear relationship between them and test difficulty varied along the score scale. While Form A equivalent scores were higher than Form C raw scores in 0-1 and 13-15 score range but in 2-12 score range Form A equivalent scores are lower than Form C raw scores. According to this result, difficulty of forms varied along score scale and in 0-1 and 13-15 score range Form A was easier than Form C but in 2-12 score range more difficult. There was a nonlinear relationship between raw scores and equivalent scores.

Based bootstrap standard error of linear equating, in middle of raw score range bootstrap standard errors got the minimum values but maximum values in extreme points. With respect to bootstrap standard error of different equipercentile equating there were fluctuations.

Lastly, for evaluating random equating error RMSD and MBSE values were examined, results indicated that while linear equating had the least RMSD (0.3410) and MBSE (0.0434) values, unsmoothed equipercentile had the largest RMSD (0.4329) and MBSE (0.0522) values.

Discussion

Based results of linear equating, there were equivalent scores out of raw score range. Kolen and Brennan (1995), indicated that in number-correct scored equated scores from linear equating could be out of range of possible observed scores and there were two way to handle with this problem. Firstly, the highest equated score might be allowed to exceed the highest raw score or secondly the conversion could be truncated at the highest and lowest scores.

Equipercentile equating have equated scores within the range of possible scores (Kolen and Brennan, 1995). In this study, equated scores from unsmoothed equipercentile, presmoothed equipercentile equating and postsmoothing equipercentile equating had values in the range of 0.149-15.133, 0.035-15.135 and 0.022-15.244.

Results of equating methods compared based RMSD and MBSE values while linear equating had the smallest random equating error, unsmoothed equipercentile equating had the largest random error. The results of similar studies (Butler and Hanson, 1997; Zhu, 1998) were confirmed this result. Another result of this study was that smoothing methods could help reducing random equating error. Hanson, Zeng and

Colton (1994) found that presmoothing and postsmoothing improving estimated equipercentile function and reducing equating error in random group design.

Based results for equating Form A and C of Social Sciences test linear equating had the smallest random equating error. Kolen and Brennan (1995) emphasized that although linear equating has the smallest random error it doesn't mean that it is better than equipercentile equating. If relationship is nonlinear then equipercentile might provide a more accurate estimate of the population equivalent, even when it has a much larger random error than linear equating. So that provided results from unsmoothed, presmoothed and postsmoothed equipercentile equating had nonlinear relationship with raw scores. Hence, it could be said that equipercentile equating methods yielded more accurate result in estimating population equivalent scores.

Conclusion

Finally, concluded that for equating A and C test forms postsmoothed equipercentile equating which had the relatively smallest random error among other equipercentile equating methods was the most appropriate method.

Random Grup Deseni ile Yapılan Doğrusal Eşitleme ve Eşit Yüzdelikli Eşitleme Yöntemlerinin Karşılaştırılması

Hülya Kelecioğlu¹ ve Neşe Öztürk Gübeş²

^{1,2}Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara/Türkiye.

MAKALE BİLGİ

Makale Tarihi:
Alındı 02.04.2012
Düzeltilmiş hali alındı
24.09.2012
Kabul edildi 28.09.2012
Çevrimiçi yayınlandı
10.04.2013

ÖZET

Bu çalışmada, 2009 yılında yapılan Öğrenci Başarılarının Belirlenmesi Sınavı (ÖBBS)'nin sosyal bilimler alt testi C kitapçığından elde edilen ham puanlar doğrusal eşitleme ve üç farklı eşit yüzdelikli eşitleme (düzgünleştirilmemiş, ön-düzgünleştirilmiş ve son-düzgünleştirilmiş eşit yüzdelikli eşitleme) yöntemleri ile A kitapçığından elde edilen puanlara eşitlenmiştir. Farklı eşitleme yöntemlerinin sahip olduğu random hatalar RMSD katsayısı ve bootstrap standart hatalarının ortalaması (BSHO) hesaplanarak karşılaştırılmıştır. Araştırmanın sonucunda doğrusal eşitleme ile elde edilen A formu eşdeğer puanları ile C formu ham puanları arasında doğrusal bir ilişki olduğu, puan ölçeği boyunca C formunun A formundan daha kolay olduğu görülmüştür. Üç farklı eşit yüzdelikli eşitleme yönteminin birbiriyle tutarlı sonuçlar ürettiği, A formu eşdeğer puanları ile C formu ham puanları arasında eğrisel bir ilişki olduğu ve testin güçlük düzeyinin puan ölçeği boyunca farklılık gösterdiği tespit edilmiştir. Araştırmanın sonucunda, en küçük BSHO ve RMSD katsayısına doğrusal eşitleme yöntemi ile elde edilen eşdeğer puanların sahip olduğu görülmüştür. Ancak, ham puanlar ile eşitlenmiş puanlar arasında eğrisel bir ilişki olmasına dayalı olarak puanların eşitlenmesi için göreceli olarak daha az random hataya sahip olan son-düzgünleştirilmiş eşit yüzdelikli eşitlemenin en uygun eşitleme yöntemi olduğu kararına varılmıştır.

© 2013 IOJES. Tüm hakları saklıdır

Anahtar Kelimeler:

Doğrusal eşitleme, eşit yüzdelikli eşitleme, ön-düzgünleştirme, son-düzgünleştirme, random hata.

Giriş

Test güvenliğini sağlamak ve bir testin aynı formunun birden fazla uygulanmasının olumsuz etkisini ortadan kaldırmak için büyük ölçekli sınavların birbirini izleyen uygulamalarında genellikle paralel test formları kullanılmaktadır. Aynı kararı vermek için kullanılacak sınavların güvenlik nedeniyle farklı formlarının kullanılması formların denkliği sorununu ortaya çıkarmaktadır. Bir testin farklı formlarını alan öğrencilerin puanları, formların denkliği sağlamadığı sürece karşılaştırılmaz. Çünkü bu durumda gözlenen puan farklılıkları, öğrencilerin yetenek düzeylerinden kaynaklanabileceği gibi formların eşit güçlükte olmamasından da kaynaklanabilir. Denk olmayan formlarla yapılacak karşılaştırmalarda bu iki durum ayırt edilemez. Bu nedenle, aynı amaçla kullanılacak sınavların farklı formlarından elde edilen puanları karşılaştırabilmek için test puanlarının eşitlenmesine ihtiyaç duyulur.

Test eşitleme, bir testin güçlük düzeyi farklı formlarından elde edilen puanlarının birbiri yerine kullanılabilmesini sağlamak amacıyla yapılan bir dizi sayısal düzenlemedir (Braun ve Holland, 1982). Bu düzenlemelerde amaç, bir testin kolay formunu alan bireyin avantajını ya da zor formunu alan bireyin dezavantajını ortadan kaldırarak puanların karşılaştırılabilmesini sağlamaktır. Testlerin eşitlenmesi sürecinde farklı yöntemler ve eşitleme desenleri kullanılabilir.

Dorans ve Holland (2000), eşitleme yöntemlerini gerçek puanlara ve gözlenen puanlara dayalı eşitleme yöntemleri olarak iki sınıfa ayırmıştır. Ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme gözlenen puanlara dayalı eşitleme yöntemlerini oluşturmaktadır (Kolen ve Brennan, 2004; Livingston, 2004). Ortalama eşitlemede, eşitlenecek olan testlerin güçlük düzeyleri arasında fark olduğu fakat bu farkın puan

² Sorumlu yazarn adresi: Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara/Türkiye
Telefon: +90 312 297 85 50
Faks: +90 312 299 20 27
e-posta: neseozturk@hacettepe.edu.tr

ölçeği boyunca sabit olduğu kabul edilmektedir. Örneğin, üst gruptaki cevaplayıcılara X formu Y formundan 2 puan daha kolay gelmişse alt gruptaki cevaplayıcılara da 2 puan daha kolay gelecektir. Ortalamaları μ_X ve μ_Y olan X ve Y formları için ortalama eşitleme (1) numaralı eşitlikte görüldüğü gibi ifade edilir (Kolen ve Brennan, 1995).

$$x - \mu_X = y - \mu_Y \quad (1)$$

Doğrusal eşitleme, iki testin ortalama ve standart sapma dışındaki tüm özelliklerinin eşit olması varsayımına dayanmaktadır (Crocker ve Algina, 1986; Moses ve Holland, 2007; Woldbeck, 1998). Crocker ve Algina (1986), eş değer puanları X ve Y formlarında aynı z puanına denk gelen puan çiftleri olarak tanımlamışlardır. μ_X ve μ_Y X ve Y formlarının ortalamaları; S_X ve S_Y X ve Y formlarının standart sapması olmak üzere doğrusal eşitleme aşağıda verilen (2) numaralı eşitlikle yapılır.

$$\frac{x - \mu_X}{S_X} = \frac{y - \mu_Y}{S_Y} \quad (2)$$

Eşit yüzdelli eşitleme bir puan dağılımında aynı yüzdelik sıraya denk gelen puanları belirlemeye dayanmaktadır. Angoff (1984), eşit güvenilirlik derecesiyle aynı özelliği ölçen X ve Y formlarından elde edilen puanların eşit yüzdelik sırasına sahip olanların eşdeğer olarak kabul edilebileceğini belirtmiştir. Eşit yüzdelikli eşitleme, iki aşamalı bir süreçten oluşmaktadır. İlk aşamada, eşitlenecek iki formun puan dağılımının yığılmalı frekans dağılımı tablolaştırılır ya da yığılmalı frekans dağılımlarının grafiği çizilir. İkinci aşamada elde edilen yığılmalı frekanslarda aynı yüzdelik sıraya denk gelen puanlar eşitlenir. Örneğin; 75. yüzdelik sırada X formunda 13 puan, Y formunda 14 puan yer alıyorsa, eşit yüzdelikli eşitlemede bu iki puanın eşit olduğu kabul edilir (Kolen, 1984).

Eşit yüzdelikli eşitlemede temel problem, gerçek test durumlarından elde edilen puan dağılımlarının düzensiz olması ve bir puanı alan bireylerin yüzdesinin kademeli olarak puan arttıkça değişmemesi, dalgalanmalar olmasıdır. Örnekleme ait puan dağılımının grafiği çizildiğinde genellikle dağılımda bir takım düzensizlikler olduğu görülür ve bu düzensiz dağılımlar önemli düzeyde örnekleme hatası içerebilir. Puan dağılımlarındaki düzensizlikler eşit yüzdelikli eşitlemede de düzensizliğe yol açmakta, elde edilen sonuçların diğer testi alan gruplara genellenememesine neden olmaktadır. Bu problemle başa çıkmanın bir yolu gözlenen puan dağılımını aynı konuma, dağılıma ve şekle sahip fakat düzensiz olmayan bir dağılım ile değiştirmektir. Bu tekniğe ise genel olarak düzgünleştirme (smoothing) denilmektedir (Kolen, 1991; Livingston, 2004).

Düzgünleştirme yöntemleri, düzgünleştirilmemiş eşit yüzdelikli eşitlemeden daha az random hata içeren düzgünleştirme fonksiyonları üretmek için tasarlanır. Düzgünleştirme yönteminin kullanılması sonucu x_i puanı düzgünleştirilmemiş eşit yüzdelikli eşitlemeye nazaran daha az random hata içerir. Düzgünleştirme, eşitlemeden önce yapıldığında ön-düzgünleştirme (presmoothing), eşitlemeden sonra yapıldığında son-düzgünleştirme (postsmoothing) adını alır (Hanson, Zeng ve Colton, 1994; Kolen ve Brennan, 2004).

Uygulamada en çok kullanılan ön-düzgünleştirme yöntemlerinden biri log-linear düzgünleştirme yöntemidir (Liu, 2011). Log-linear düzgünleştirme yöntemi, örneklem yoğunluğu logunun polynomial fonksiyona uyumunu dikkate alır. Bir polynomial model aşağıdaki dağılım takip edilerek bir modele uyar:

$$\log[N_x f(x)] = w_0 + w_1 x + w_2 x^2 + \dots + w_c x^c \quad (3)$$

(3) numaralı eşitlikte yoğunluk logu bir alt sıralı C'nin polynomial derecesi ile açıklanmıştır. Örneğin, C=2 için $\log[N_x f(x)] = w_0 + w_1 x + w_2 x^2$ olacaktır ve model 2 dereceli bir polynomialdir. Modeldeki w parametresi maksimum olabilirlikle kestirilebilir. Sonuç olarak, elde edilen uygun dağılım moment korunumu özelliğine sahiptir ve uyumlu dağılımın ilk C momentleri örneklem dağılımı ile aynıdır. Örneğin, eğer C=2 ise uyumlu dağılımın ortalama ve standart sapması örneklem dağılımının ortalama ve standart sapması ile aynıdır (Kolen ve Brennan, 2004).

Polynomial log-linear yönteminde ne kadar düzgünleştirme yapılacağını kontrol eden polinomun derecesinin (C) seçimi oldukça önemlidir. C parametresi, genellikle 1'den 10'a kadar bir değer alır ve düzgünleştirilmiş modellerin gözlenen modele uyumlu olup olmadığı incelenerek seçilir (Liu, 2011).

Bir diğer düzgünleştirme yöntemi son-düzgünleştirmedir. Ön-düzgünleştirme yöntemi ile puan dağılımları düzgünleştirilirken son-düzgünleştirme yöntemleriyle eşit yüzdelli eşdeğer puanlar düzgünleştirilmektedir (Kolen ve Brennan, 1995). Kübik spline yöntemi literatürde en çok kullanılan son-düzgünleştirme yöntemidir. Kesikli puanlar için x_i , spline fonksiyonu (4) numaralı eşitlikte görüldüğü gibi tanımlanırken; eğriliği en aza indirmek için (5) numaralı eşitlik ile gösterilen sınırlama getirilir (Kolen ve Brennan, 1995):

$$d(y) = v_{0i} + v_{1i}(x - x_i) + v_{2i}(x - x_i)^2 + v_{3i}(x - x_i)^3 \quad (4)$$

$$\frac{\sum_{i=low}^{high} \left[\frac{d_Y(x_i) - e_Y(x_i)}{se[e_Y(x_i)]} \right]^2}{x_{high} - x_{low} + 1} \leq S \quad (5)$$

$se[e_Y(x_i)]$ eşit yüzdelli eşitlemenin kestirilen standart hatasıdır. Eşitlemenin standart hatası düzgünleştirilmiş ve düzgünleştirilmemiş ilişkiler arasındaki farkı standartlaştırmak için kullanılır. Standart hata küçük ise düzgünleştirilmiş ve düzgünleştirilmemiş ilişkiler yakın olur, değilse ilişkiler birbirinden uzaklaşır. S ($S \geq 0$) parametresi araştırmacı tarafından ayarlanır ve düzgünleştirmenin derecesini kontrol eder (Kolen ve Brennan, 1995).

Bir eşitleme çalışmasını yürütürken göz önünde bulundurulması gereken önemli ölçütlerden biri istatistiksel hatadır. Eşitleme hatasının olabildiğince küçük olması seçilen eşitleme yönteminin uygunluğunun bir göstergesidir. Random ve sistematik olmak üzere iki tür eşitleme hatası vardır. Random eşitleme hatası; eşitleme ilişkisini belirlemede yer alan ortalama, standart sapma ve yüzdellik sıra gibi parametrelerin kestirildiği örnekleme bağlı olarak oluşan hatadır. Örneklem büyüklüğü arttıkça random hata azalır. Sistematik hata ise eşitleme ilişkisinin kestirildiği yöntemin yanlılık içermesi sonucu oluşur. Örneğin, random gruplar deseninde grupların karşılaştırılmasını sağlayan sarmal sürecin tam olarak uygulanmaması sistematik hataya neden olabilir. Eşitleme sürecine karşın random hatanın miktarını eşitlemenin standart hatasını kullanarak belirlemek mümkün iken sistematik hatanın miktarını belirlemek çok daha zordur (Kolen, 1988; Kolen ve Brennan, 1995).

Random hata genellikle eşitlemenin standart hatası bir başka deyişle her bir puanda eşitleme hatasının standart sapması olarak nitelendirilmektedir. Kolen ve Brennan (1995) eşitlemenin standart hatasını şu şekilde kavramsallaştırmışlardır: Herhangi bir büyüklükteki hipotetik bir örneklem, evrenden çekilir. Örneklem dayalı olarak, eşitleme süreci uygulanarak eş değer puanlar elde edilir. Bu aşamaların birçok defa tekrarlanması sonucu elde edilen eş değer puanların standart sapması eşitlemenin standart hatasını oluşturur.

Yeniden örnekleme (resampling) için literatürde sıkça başvurulan yöntemlerden biri bootstrap yöntemidir. Tek bir örneklemden bootstrap ile standart hataları kestirilirken şu basamaklar takip edilir (Kolen ve Brennan, 1995):

1. N büyüklüğünde bir örneklem ile başlanır.
2. Yerine koyularak bu örneklemden N büyüklüğünde random bir örneklem çekilir. Bu örnekleme bootstrap örnekleme denir.
3. Eş değer puanlar bootstrap örnekleme için hesaplanır.
4. 2 ve 3. Basamaklardaki işlemler R kez tekrarlanır.
5. R tane bootstrap örnekleme üzerinden elde edilen eşdeğer puanların standart sapması hesaplanır. Bu standart sapma eşitlemenin kestirilen bootstrap standart hatasıdır.

Eğitim testlerinin eşitlenmesi Amerika'da, Kanada'da, Avustralya'da ve çoğu Avrupa ülkesinde düzenli aralıklarla yapılan önemli bir uygulamadır. Türkiye'de Ölçme, Seçme ve Yerleştirme Merkezi ve Milli Eğitim Bakanlığı benzer içerik ve formatta fakat bir testin tamamen farklı sorulardan oluşan formlarını üreten merkezlerdir. Ülkemizde öğrenci başarısının belirlenmesinde uluslararası düzeyde yapılan PISA, TIMSS ve PIRLS sınavlarının yanında ulusal düzeyde MEB tarafından ilköğretim ve ortaöğretim düzeylerinde okul öğrenmelerinin rutin olarak izlenmesine yönelik Öğrenci Başarılarının Belirlenmesi Sınavı (ÖBBS) yapılmaktadır.

İlköğretim düzeyinde, ilki 2002 yılında olmak üzere üçer yıl arayla 2005 ve 2008 yıllarında yapılan ÖBBS (Öğrenci Başarılarının Belirlenmesi Sınavı) ile ilköğretim düzeyindeki 4, 5, 6, 7 ve 8. sınıf öğrencilerinin Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve ilköğretim düzeyindeki başarı düzeyleri izlemek amaçlanmıştır. Bu izleme sonuçlarına göre öğretmen yetiştirmeden ders materyallerini düzenlemeye, öğretim yöntemleri ile çalışma alışkanlıklarına, öğretmenlerin kendilerini geliştirmelerine yönelik dönütler elde edilmektedir (EARGED, 2010).

İlköğretimdekine benzer şekilde ortaöğretim öğrencilerinin okuldaki öğrenmelerinin izlenmesi ve okul öğrenmelerini etkilediği düşünülen bazı değişkenler ile ilişkilerinin belirlenmesi amacıyla Ortaöğretim Geliştirme Projesi geliştirilmiştir. Proje, ortaöğretim okullarının genel ortaöğretim ile mesleki ve teknik ortaöğretim kategorisi altında çeşitliliğin çok olması nedeniyle, derslerde birlikteliğin çok olduğu 9. ve 10. Sınıflar üzerinde yürütülmektedir. Proje, öğrencilerin bu sınıflardaki Türk edebiyatı, dil ve anlatım, matematik ve geometri, fen bilimleri ile ilgili dersleri (fizik, kimya, biyoloji), sosyal bilimler ile ilgili dersleri (tarih, coğrafya) ve İngilizce derslerindeki öğrenmeleri kapsamaktadır. Ortaöğretim Geliştirme Projesi kapsamında ilk ÖBBS, 2009 yılında Türkiye genelinde uygulanmıştır (EARGED, 2010).

ÖBBS'nin uygulamasında test güvenliğini sağlamak amacıyla kapsam ve güçlük düzeyi birbirine olabildiğince benzer A, B, C ve D olmak üzere dört sınav kitapçığı kullanılmaktadır. EARGED (2010), farklı kitapçıklarından elde edilen puanları aynı test ile elde edilen ölçümler olarak kabul edip testlerin eşitlenmesine gerek duymamıştır. Ancak, ÖBBS'de her kitapçıkta yer alan testler farklı sorulardan oluştuğu için test geliştirme sürecinde mevcut kontrolün derecesine bağlı olarak formların güçlük düzeyi birbirinden farklılık gösterebilmektedir. Farklı sorulardan elde edilen puanları doğrudan karşılaştırmak ve bu puanlara dayanarak öğrencilerin başarı düzeyine ilişkin yorumlar yapmak hatalı olabilir. Bu nedenle farklı test formlarından alınan puanların eşitlenmesine ihtiyaç vardır.

Bu araştırmanın amacı, ortaöğretim 9. sınıf ÖBBS sosyal bilimler alt testi C kitapçığından elde edilen puanları doğrusal eşitleme, düzgünleştirilmemiş, ön-düzgünleştirilmiş ve son-düzgünleştirilmiş eşit yüzdelikli eşitleme ile A kitapçığına eşitleyerek bu testlerin eşitlenmesi için en uygun olan yöntemi belirlemektir.

Yöntem

Evren ve Örneklem

Araştırmanın evrenini Türkiye'de resmi ortaöğretim okullarının 9. sınıfına devam eden öğrenciler oluşturmaktadır. ÖBBS kapsamında kullanılan 9. sınıf testleri Türk edebiyatı-dil ve anlatım, matematik, fen bilimleri, sosyal bilimler ve İngilizce alt testlerinden oluşmaktadır. Testlerin her biri için A, B, C ve D olmak üzere dört ayrı kitapçık bulunmaktadır. Bu kitapçıklardan A ile C ve B ile D ölçülen özellikler bakımından birbirine paralel şekilde hazırlanmıştır (EARGED, 2010). Bu çalışmada random gruplara dayalı eşitleme deseni kullanılmıştır. Bu nedenle araştırmanın örneklemini birbirine paralel şekilde hazırlandığı belirtilen A ve C kitapçığını alan dokuzuncu sınıf öğrencileri oluşturmuştur. A kitapçığını alan 16670 ve C kitapçığını alan 15743 öğrenci bulunmaktadır.

Araştırmanın Verileri ve Eşitleme Deseni

Araştırmanın verileri ÖBBS kapsamında ortaöğretim 9. sınıf öğrencilerinin tarih ve coğrafya derslerindeki kazanımlarını ölçmek amacı ile hazırlanan sosyal bilimler testinden alınan puanlardır. Sosyal bilimler testi soruları A, B, C ve D olmak üzere dört ayrı kitapçıkta yer almaktadır ve her kitapçıkta 15 soru vardır. Bu kitapçıklardan A ile C ve B ile D ölçülen özellikler bakımından birbirine paralel olarak hazırlanmıştır. Araştırma kapsamında A ve C kitapçıkları seçilerek, C kitapçığından elde edilen puanlar doğrusal ve eşit yüzdelikli eşitleme yöntemleri ile A kitapçığına eşitlenmiştir.

Kitapçıklardaki sorularda ortak madde bulunmamakta, her kitapçık diğerlerinden farklı 15 soru içermektedir. ÖBBS uygulamasında öğrencilere ikişerli olarak birbirine paralel olarak hazırlanan (A ile C ve B ile D) dört kitapçık öğrencilere sarmal bir şekilde uygulanmıştır. Kitapçıklar ilk öğrenciye A, ikinciye B,

üçüncüye C ve dördüncüye D verilecek şekilde sırayla dağıtılmıştır. Sarmal bir süreç takip edildiği için uygulamada random olarak eş değer iki grup (A ile C ve B ile D) oluşturulmuştur. ÖBBS uygulaması random grupla eşitleme koşullarını sağladığından C formu A formuna random grup deseni kullanılarak eşitlenmiştir.

Eşitleme Yöntemleri ve Değerlendirme Ölçütleri

Araştırmada, 2009 ÖBBS sosyal bilimler alt testinin C formunu A formuna eşitlemede doğrusal eşitleme ve eşit yüzdelli eşitleme yöntemleri kullanılmıştır. Eşit yüzdelli eşitleme, düzgünleştirme yapmadan, ön-düzgünleştirme (5 parametrelili log-linear) ve son düzgünleştirme (s=0,01) yapılarak uygulanmıştır. Dört eşitleme yöntemine dair eşitlemenin hatası, bootstrap standart hata ortalamaları (BSHO) ve RMSD (Root Mean Squared Difference) katsayısı ile hesaplanmıştır.

$$BSHO = \sqrt{\sum_i f(x_i) [se^2 \hat{e}_Y(x_i)]} \quad (6)$$

$se^2 \hat{e}_Y(x_i)$: Eşdeğer puanının kestirilen standart hatasının karesi

$$RMSD = \sqrt{\frac{\sum f_i (X'_i - X_i)^2}{\sum f_i}} \quad (7)$$

f_i : C formunda i. puanın frekansı

X'_i : Eş değer puan; X_i : i. ham puan

Formların eşitlenmesi ve bootstrap eşitleme hatalarının hesaplanmasında *Equating Error* (Hanson, 2004) bilgisayar programı kullanılmıştır. Yürütülen analizlerde tekrar sayısı 500 olarak alınmıştır.

Bulgular

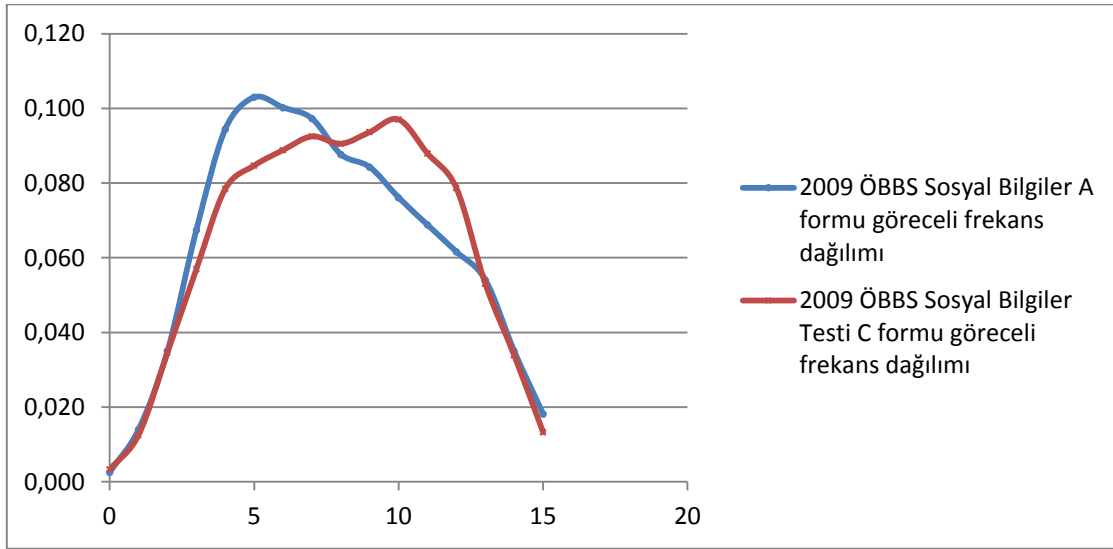
Veri analizinin ilk aşamasında, 2009 ÖBBS 9. sınıf sosyal bilimler A ve C formlarına ilişkin betimsel istatistikler hesaplanmış ve elde edilen bulgular Tablo 1'de verilmiştir.

Tablo 1. Betimsel istatistikler

İstatistikler	A Formu	C Formu
N	16670	15743
Ortalama	7.65	7.98
Standart sapma	3.48	3.42
Çarpıklık Katsayısı	0.19	-0.03
Çarpıklık Katsayısının Std. Hatası	0.02	0.02
Basıklık Katsayısı	-0.86	-0.88
Basıklık Katsayısının Standart Hatası	0.04	0.04
Ortalama Güçlük	0.51	0.53
KR-20	0.76	0.75

Tablo 1'deki bilgilere göre A formuna ait puan dağılımının çarpıklık katsayısının pozitif olmasına dayalı olarak dağılımın normale göre sağa çarpık; C formuna ait puan dağılımına ait çarpıklık katsayısının negatif olmasına dayalı olarak dağılımın normale göre sola çarpık olduğu söylenebilir. Her iki forma ait puan dağılımlarına ait basıklık katsayılarının negatif olmasına dayalı olarak ise dağılımların normale göre basık olduğu söylenebilir.

Bu durumu Şekil 1'deki A ve C formlarına ait ham puanların göreceli frekans dağılımlarına göre çizilen grafikte de görmek mümkündür.



Şekil 1. 2009 ÖBBS sosyal bilgiler A ve C testlerinin ham puan dağılımlarına ilişkin göreceli frekans dağılımı

Eşitleme Koşullarının Test Edilmesi

Eşitleme çalışmasını yürütebilmek için testlerin bir takım koşulları sağlaması gerekir. Literatür incelendiğinde Klasik Test Kuramına dayalı olarak eşitlenecek testlerin genel olarak üç koşulu mutlaka karşılaması gerektiği belirtilmiştir. Bu koşullar; tek boyutluluk, eşit güvenilirliğe ve benzer güçlüğüne sahip olmaktır (Angoff, 1984; Dorans ve Holland, 2000; Kolen ve Whitney, 1982). Aşağıda sırayla 2008 ÖBBS sosyal bilgiler A ve C formlarının bu üç koşulu sağlayıp sağlamadığı incelenmiştir.

2009 ÖBBS sosyal bilgiler A ve C formlarının tek boyutlu olup olmadığı faktör analizi yapılarak incelenmiştir. Testlerden elde edilen madde puanlarının iki kategorili olması sebebiyle faktör analizi tetrakorik korelasyon matrisi üzerinden yürütülmüştür. Test formlarının ilk üç faktörüne ait elde edilen bulgular Tablo 2’de verilmiştir.

Tablo 2. 2009 ÖBBS sosyal bilgiler A ve C formlarına ait faktör analizi sonuçları

Faktör	Özdeğer	A FORMU		Özdeğer	C FORMU	
		Varyans Yüzdesi	Yığılmalı Yüzde		Varyans Yüzdesi	Yığılmalı Yüzde
1	5.06	33.75	33.75	4.89	32.61	32.61
2	0.98	6.55	40.30	1.06	7.04	39.65
3	0.96	6.46	46.76	0.94	6.28	45.92

Tablo 2’deki bilgilere göre A formunda varyansın %33.75’i, C formunda ise varyansın %32.61’i birinci faktör tarafından açıklanmaktadır. Büyüköztürk (2007), tek faktörlü ölçeklerde açıklanan varyans oranının %30 ve daha fazla olmasının yeterli olacağını belirtmiştir. Ayrıca, A formunda özdeğeri 1’den büyük faktör sayısının 1 iken C formunda özdeğeri 1’den büyük faktör sayısının ikidir. Ancak, iki formda da birinci faktörle ikinci faktör arasındaki farkın yaklaşık dört kat olması her iki formun genel bir faktöre sahip olabileceğinin bir diğer göstergesidir. Tüm bu bulgular ışığında, 2009 ÖBBS sosyal bilgiler A ve C formlarının tek boyutluluk varsayımını sağladığı söylenebilir.

2009 ÖBBS sosyal bilgiler A ve C formlarının güvenilirliklerinin eşit olup olmadığı test edilmiştir. Bunun için A ve C formlarının güvenilirlikleri korelasyon katsayısı olarak kabul edilmiş ve güvenilirlik katsayılarına Fischer’in Z_r dönüşümü yapılmıştır. İki güvenilirlik katsayısı arasında fark olup olmadığı ise

Fischer'in Z istatistiği ile test edilmiştir (Akhun, 1984). A ve C formunun güvenilirlik katsayıları arasında 0.05 düzeyinde anlamlı bir fark bulunamamıştır ($Z=0.06$; $p=0.95$). Bu bulgu, A ve C formlarının eşit güvenilirlikte olma koşulunu sağladığını göstermektedir.

2009 ÖBBS sosyal bilimler A ve C formlarının ortalama güçlükleri arasındaki fark iki oran farkı testi (Baykul, 1996) ile incelenmiş ve 0.05 düzeyinde formların güçlükleri arasında anlamlı bir fark bulunamamıştır ($z=0.12$; $p=0.90$). Bu bulguya göre A ve C formlarının ortalama güçlükleri arasında anlamlı bir fark olmadığı ve her iki formun ortalama güçlüklerinin eşit olduğu görülmüştür.

Eşitlemeye İlişkin Bulgular

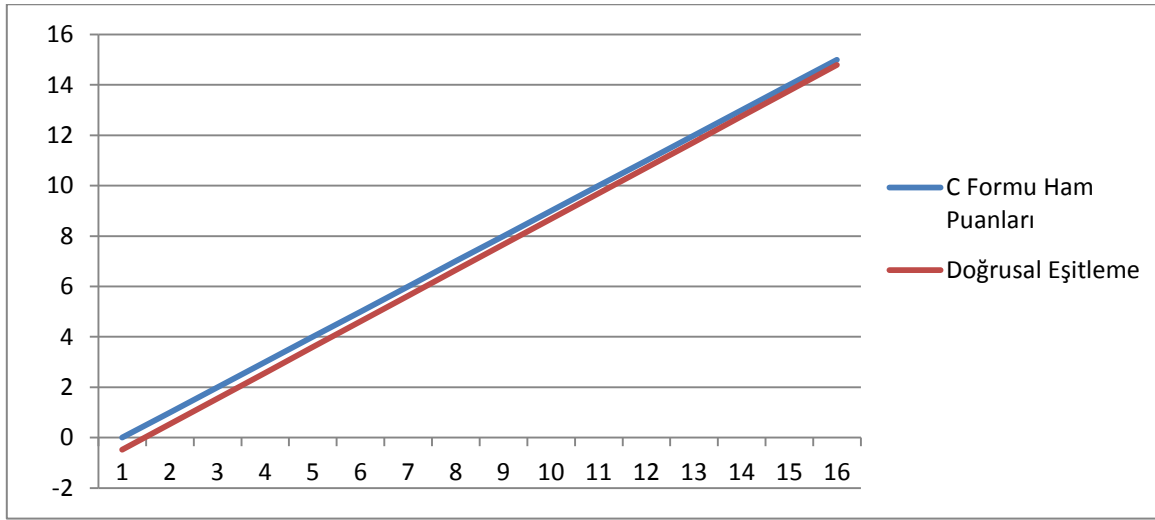
2009 ÖBBS sosyal bilimler C formu doğrusal eşitleme, düzgünleştirilmemiş eşit yüzdelikli eşitleme, $C=5$ log-linear ön-düzgünleştirme ve $s=0.01$ son-düzgünleştirme yapılarak eşitlenmiştir. Tablo 3'te C formu ham puanları, her bir yöntem ile 500 kez tekrar yapılarak elde edilen A formu eşdeğer puanları ve bootstrap standart hataları (BSH) verilmiştir.

Tablo 3. C formu ham puanlarına karşılık gelen A formunun eşdeğer puanları ve BSH değerleri

C Formu Ham Puan	DE ^a	BSH ^b (DE)	EYE-D ^c	BSH (EYE-D)	EYE-SD ^d (s=0,01)	BSH (EYE-SD)	EYE-ÖD ^e (C=5)	BSH (EYE-ÖD)
0	-0.478	0.060	0.149	0.136	0.022	0.020	0.035	0.086
1	0.540	0.056	1.003	0.065	1.059	0.055	1.024	0.065
2	1.558	0.051	1.976	0.049	1.967	0.044	1.981	0.044
3	2.575	0.048	2.907	0.039	2.866	0.039	2.904	0.037
4	3.593	0.044	3.794	0.038	3.750	0.039	3.793	0.038
5	4.611	0.042	4.643	0.044	4.624	0.041	4.650	0.041
6	5.629	0.040	5.485	0.050	5.507	0.045	5.495	0.046
7	6.647	0.039	6.388	0.052	6.422	0.050	6.370	0.048
8	7.665	0.038	7.326	0.054	7.391	0.055	7.316	0.053
9	8.683	0.039	8.356	0.059	8.432	0.057	8.358	0.056
10	9.701	0.041	9.486	0.061	9.547	0.056	9.507	0.059
11	10.719	0.043	10.719	0.063	10.718	0.054	10.732	0.059
12	11.736	0.047	11.980	0.057	11.899	0.053	11.954	0.054
13	12.754	0.050	13.123	0.049	13.049	0.051	13.114	0.049
14	13.772	0.054	14.156	0.047	14.158	0.049	14.180	0.045
15	14.790	0.059	15.133	0.034	15.244	0.054	15.130	0.029

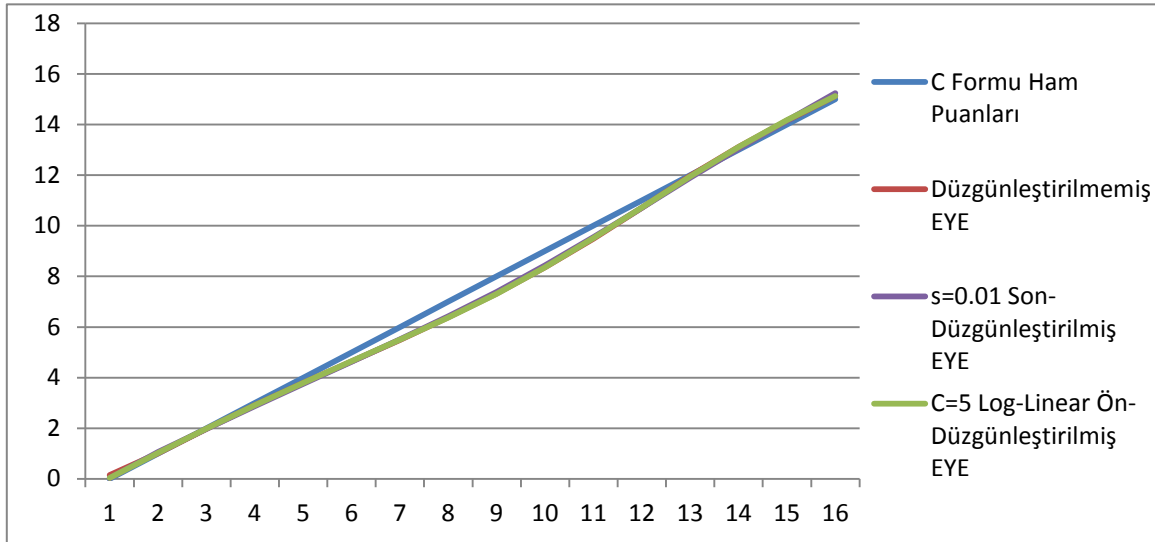
^aDE: Doğrusal eşitleme; ^bBSH: Bootstrap standart hatası; ^cEYE-D: Eşit yüzdelikli eşitleme-düzgünleştirilmemiş; ^dEYE-ÖD: Eşit yüzdelikli eşitleme-ön düzgünleştirme; ^eEYE-SD: Eşit yüzdelikli eşitleme-son düzgünleştirme.

Tablo 3'teki bilgilere göre C formuna ait ham puanlar 0-15 puan aralığında değerler alırken doğrusal eşitleme yöntemi ile elde edilen puanların -0.478 ile 14.790 arasında değerler aldığı görülmektedir. Doğrusal eşitleme sonucu, 2009 ÖBBS sosyal bilimler C formu ham puanlarının tümünün A'nın eşdeğer puanlarından büyük olduğu görülmektedir. Bu bulguya dayalı olarak, C formunun ham puanları ile A formunun eş değer puanları arasında doğrusal bir ilişki olduğu ve puan ölçeği boyunca C formunun A formundan daha kolay olduğu söylenebilir. Şekil 2'deki grafikte de, C formu ham puanları ile doğrusal eşitleme sonucu elde edilen A formu eşdeğer puanları arasında doğrusal bir ilişki olduğu görülmektedir.



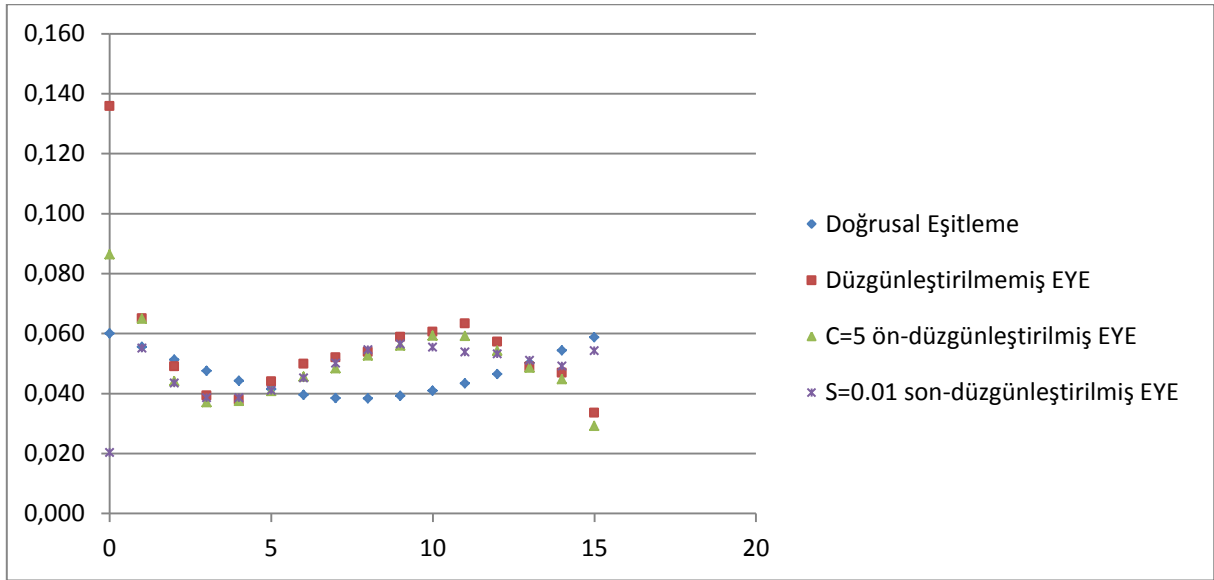
Şekil 2. C formu ham puanları ile doğrusal eşitleme sonucu elde edilen eşdeğer puanlar

Tablo 3'te yer alan bilgilere göre düzgünleştirilmemiş, $C=5$ log linear ön-düzgünleştirme ve $s=0,01$ son-düzgünleştirme yapılmış eşit yüzdelikli eşitleme yöntemleri ile elde edilen eş değer puanların ham puan ranjının dışında değerler aldığı görülmektedir. Üç eşit yüzdelikli eşitleme yöntemi ile elde edilen A formu eşdeğer puanları 0-1 ve 13-15 puan aralığında C formu ham puanlarından büyük değerler alırken 2-12 puan aralığında C formu ham puanlarından daha küçük değerler almıştır. Bu bulguya dayalı olarak formların güçlük düzeyinin ölçek boyunca farklılaştığını; 0-1 ve 13-15 puan aralığında A formunun C formundan daha kolay, 2-12 puan aralığında ise A formunun C formundan daha zor olduğu söylenebilir. C formu ham puanları ile A formu eşdeğer puanları arasındaki eğrisel ilişkiyi Şekil 3 'te verilen grafikte de görmek mümkündür.



Şekil 3. C formu ham puanları ve üç eşit yüzdelikli eşitleme yöntemi ile elde edilen eşdeğer puanlar

Tablo 3'te yer alan bir diğer bilgi doğrusal eşitleme ve üç eşit yüzdelikli eşitleme yöntemine ilişkin elde edilen bootstrap standart hata değerleridir. Doğrusal eşitlemeye ilişkin bootstrap standart hatalarının puan dağılımının orta noktalarında en aza inerken uç noktalara gidildikçe arttığı görülmektedir. Bu durumu Şekil 4'te ham puanlara karşılık gelen bootstrap standart hataları grafiğinde de görmek mümkündür.



Şekil 4. Dört yönteme ilişkin kestirilen bootstrap standart hatalarının dağılımı

Düzgünleştirilmemiş, C=5 log-linear ön düzgünleştirilmiş ve s=0.01 son-düzgünleştirme sonucu oluşan bootstrap standart hataları incelendiğinde ise üç eşit yüzdelikli eşitleme yöntemine ilişkin bootstrap standart hatalarının puan ölçeği boyunca yer yer artıp azalarak dalgalanmalar oluşturduğu görülmektedir. Bu durumu Şekil 2’de verilen grafikte de görmek mümkündür. Eşitleme yöntemlerine karışan random hatayı değerlendirebilmek için RMSD ve bootstrap standart hatalarının ortalaması (BSHO) hesaplanmış ve Tablo-4’te verilmiştir.

Tablo 4. Random eşitleme hatası ve bootstrap standart hatalarının ortalaması

	RMSD	BSHO
Doğrusal Eşitleme	0.3410	0.0434
Düzgünleştirme yapılmadan eşit yüzdelikli eşitleme	0.4329	0.0522
C=5 log linear ön-düzgünleştirme	0.4320	0.0494
S=0.01 son-düzgünleştirme	0.4046	0.0493

Tablo 4’teki bilgilere göre en küçük RMSD (0.3410) ve BSHO (0.0434) katsayılarına doğrusal eşitleme yöntemine göre eşitlenmiş puanlar sahip iken en büyük RMSD ve BSHO katsayılarına ise düzgünleştirilmemiş eşit yüzdelikli eşitleme yöntemiyle elde edilen eşitlenmiş puanlar sahiptir. Bu bulguya dayalı olarak en az random hata içeren eşitleme yöntem doğrusal eşitleme iken en fazla random hata içeren yöntemin düzgünleştirilmemiş eşit yüzdelikli eşitleme olduğu söylenebilir.

Tartışma

Bu araştırma ile doğrusal eşitleme, düzgünleştirilmemiş eşit yüzdelikli eşitleme, C=5 ön-düzgünleştirilmiş eşit yüzdelikli eşitleme ve s=0.01 son-düzgünleştirilmiş eşit yüzdelikli eşitleme ile 2009 ÖBBS sosyal bilimler C formu, A formuna eşitlenerek elde edilen bootstrap standart hata ortalamaları ve RMSD katsayılarına göre dört yöntem random eşitleme hatalarına göre karşılaştırılmıştır.

Doğrusal eşitleme sonucunda elde edilen A formunun eş değer puanlarının puan ranjının dışında değerler aldığı görülmüştür. Doğrusal eşitleme ile elde edilen eşdeğer puanların ham puan aralığının dışına çıkması doğrusal eşitlemeye özgü bir özelliktir. Jaeger (1981), doğrusal eşitleme yönteminin doğası gereği eşitlenmiş puanlar ile ham puanların puan aralıklarının tam olarak örtüşmediğini belirtirken bu durumu

Livingston (2004), doğrusal eşitlemenin karakteristik bir özelliği olarak nitelendirmiştir. Kolen ve Brennan (1995), doğru sayısı temel alınarak puanlanan testlerin, doğrusal eşitleme ile elde edilen eşitlenmiş puanlarının testin puan aralığının dışında değerler almasının beklendik bir durum olduğunu belirtmiş ve bu sorunla başa çıkmak için iki yol önermiştir. Birinci yol; ham puan aralığını aşmış eşitlenmiş puanlara izin verilebilir bir başka deyişle en yüksek ve en düşük ham puanı aşmış olan eşitlenmiş puan kabul edilebilir ya da ikinci yol olarak en yüksek ve en düşük eşitlenmiş puanlar veri setinden çıkartılabilir. Kolen ve Brennan (1995), eşit yüzdelikli eşitleme yürütüldüğünde her zaman eşitlenmiş puanların;

$$-0.5 \leq e_y(x) \leq K_Y + 0.5 \quad (8)$$

K_Y : Y testindeki madde sayısı ve $e_y(x)$ eşdeğer puan

8 numaralı eşitlikle belirtilen aralıkta değerler aldığını, böylece eşit yüzdelikli eşitlemenin olası ham puan ranjının içinde eşdeğer puanlar verdiğini belirtmiştir. Bu çalışmada, (8) numaralı eşitliğin doğrulanması için eşit yüzdelikli eşitlemeden elde edilen puanların $-0.5 \leq e_y(x) \leq 15.5$ aralığında değerler alması gerekir. Bulgular incelediğinde, düzgünleştirilmemiş eşit yüzdelikli eşitleme ile elde edilen eşdeğer puanlar 0.149-15.133; ön-düzgünleştirme ile elde edilen eşdeğer puanlar 0.035-15.135 ve son-düzgünleştirme yöntemi ile elde edilen eşdeğer puanlar 0.022-15.244 aralığında değerler alarak (8) numaralı eşitliğin doğrulandığı gözlenmiştir.

Doğrusal eşitleme ile elde edilen bootstrap standart hatalara ilişkin dağılımlar incelendiğinde, puan dağılımının orta noktalarında bootstrap standart hataların en küçük değerleri alırken puan dağılımının uç noktalara gidildikçe bootstrap standart hatalarının arttığı görülmüştür. Skaggs (2005), çalışmasında benzer bulguya ulaşmış ve bu durumun oluşma sebebini doğrusal eşitlemede eşitlemenin ham puan dağılımındaki sadece ortalama ve standart sapmaya dayalı olarak yapıldığı dolayısıyla iki testin ortalamasına yakın puanlarda doğrusal eşitlemenin en tutarlı sonuçları ürettiği şeklinde açıklamıştır.

Doğrusal eşitleme ve üç farklı eşit yüzdelikli eşitleme yöntemlerine dayalı olarak elde edilen RMSD ve BSHO katsayıları karşılaştırıldığında en az random hata içeren yöntem doğrusal eşitleme iken en fazla random hata içeren yöntemin düzgünleştirilmemiş eşit yüzdelikli eşitleme olduğu görülmüştür. Literatürde yapılan çalışmalarda da benzer bulgular elde edilmiştir. Butler ve Hanson (1997), yaptıkları çalışmalarında büyük örneklemelerde (N=1000), göreceli olarak düzgünleştirilmemiş eşit yüzdelikli eşitlemenin doğrusal ve düzgünleştirilmiş eşit yüzdelikli eşitlemeye göre daha fazla hata içerdiğini bulmuşlardır. Zhu (1998), tek grup desenine dayalı olarak yürüttüğü çalışmasında doğrusal eşitleme, düzgünleştirilmemiş ve son-düzgünleştirilmiş eşit yüzdelikli eşitleme yöntemlerine ilişkin elde ettiği RMSD katsayıları karşılaştırmıştır. Çalışmasının sonucunda, en az random hata içeren yöntem doğrusal eşitleme iken en fazla random hata içeren yöntemin düzgünleştirilmemiş eşit yüzdelikli eşitleme olduğu bulgusuna ulaşmıştır.

Bu araştırma bulgularına göre elde edilen bir diğer sonuç, düzgünleştirme yöntemlerinin random hatanın azalmasına yardımcı olduğudur. Random gruplar desenine dayalı olarak yapılan benzer bir çalışmada (Hanson, Zeng ve Colton, 1994), ön-düzgünleştirme ve son-düzgünleştirme yöntemlerinin random grup deseninde eşit yüzdelikli eşitleme fonksiyonunun kestirimini geliştirdiği ve eşitleme hatasının azalmasına yardımcı olduğu sonucuna ulaşılmıştır. Literatürde farklı veri toplama desenleriyle yapılan çok sayıda araştırmada da (Livingston, 1993; Kolen, 1984) ön-düzgünleştirme ve son-düzgünleştirme yöntemlerinin random hatayı azalttığı görülmüştür.

Elde edilen bulgulara dayalı olarak, 2009 ÖBBS sosyal bilimler A ve C formlarının eşitlenmesinde en az random eşitleme hatasına sahip olan yöntemin doğrusal eşitleme olduğu gözlenmiştir. Kolen ve Brennan (1995), doğrusal eşitlemenin daha az random hata içermesinin puanların eşitlenmesi için eşit yüzdelikli eşitlemeden daha iyi bir yöntem olduğu anlamına gelmediğini, eğer ilişkiler doğrusal değil ise doğrusal eşitlemeden daha büyük standart hataya sahip olmasına rağmen evrendeki eşdeğer puanların eşit yüzdelikli eşitleme ile daha doğru kestirildiğini belirtmişlerdir. Çünkü, böyle bir durumda doğrusal eşitlemeye başvurulması sistematik hatanın oluşmasına neden olur. Livingston (2004), eşitleme çalışmasının yürütüldüğü iki formun puan dağılımları aynı şekle sahip ise bir başka deyişle her iki formdaki puanların yüzdelik sıraları birbirine eşit ise ancak o zaman doğrusal eşitleme ve eşit yüzdelikli eşitlemenin aynı sonuçlar vereceğini belirtmiştir. Bu araştırmadan elde edilen bulgular, formların puan dağılımlarının birbirinden farklı olduğunu göstermiştir. Dolayısıyla, C formunun ham puanları düzgünleştirilmemiş, C=5

log-linear düzgünleştirilme ve $s=0.01$ son-düzgünleştirme yapılarak elde edilen A formu eşdeğer puanları arasındaki ilişkinin doğrusal olmadığı görülmüştür. Elde edilen bu bulgulara dayalı olarak evrendeki eşdeğer puanları kestirmede eşit yüzdelli eşitlemenin daha doğru sonuçlar vereceği söylenebilir. Bu nedenle, 2009 ÖBBS sosyal bilimler A ve C formlarının eşitlenmesi için göreceli olarak diğer eşit yüzdelli eşitleme yöntemlerine göre daha az random hata içeren son-düzgünleştirme yönteminin uygun olduğu sonucuna ulaşılmıştır.

Bu araştırmada 2009 ÖBBS sosyal bilimler C formu A formuna random gruplar deseni kullanılarak doğrusal eşitleme ve üç eşit yüzdelli eşitleme yöntemi kullanılarak eşitlenerek random hatalar karşılaştırılmıştır. Benzer bir araştırma farklı eşitleme desenleri (örneğin; tek grup deseni, eşdeğer olmayan gruplar ortak test deseni) kullanılarak tekrarlanıp elde edilen random hatalar karşılaştırılabilir. Bu araştırma, Klasik Test Kuramı'na dayalı iki yöntem olan doğrusal eşitleme ve eşit yüzdelli eşitleme ile sınırlıdır. Madde Tepki Kuramı'na dayalı eşitleme yöntemleri ile bir başka eşitleme çalışması yürütülüp elde edilen sonuçlar ile Klasik Test Kuramı'na dayalı eşitleme yöntemlerinden elde edilen sonuçlar karşılaştırılabilir.

Kaynakça

- Akhun, İ. (1984). İki korelasyon katsayısı arasındaki manidarlığın test edilmesi. *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 17, 1-7.
- 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.
- Braun, H. I, & Holland, P. W. (1982). Observed- score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland&D.B. Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic Press.
- Butler, O. D., & Hanson, B. A. (1997). Examination of presmoothing and postsmoothing methods in equating a direct writing assessment. Reports-Evaluative. <http://www.eric.ed.gov/PDFS/ED412239.pdf>
- Büyüköztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı* (8. Baskı). Ankara: Pegem A Yayıncılık.
- Crocker, L.,& Algina, J. (1986). *Introduction to classical&modern test theory*. New York: Harcourt Brace Jovanovich College Publishers.
- Dorans, J. N., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Measurement*, 37, 281-306.
- Eğitim, Araştırma ve Geliştirme Daire Başkanlığı (EARGED). (2010). *Ortaöğretim ÖBBS raporu 2009*. Ankara, Milli Eğitim Bakanlığı.
- Hanson, B. (2004). Equating Error: A Program for Computing Equating Error Using the Bootstrap (Version 2.0) [*Computer Software*]. Iowa, Lindquist Center S.
- Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comprison of presmoothing and postsmoothing methods in equipercentile equating*. ACT Report Series, 94-4, American College Testing: Iowa City.
- Jaeger, R. M. (1981). Some exlatory indices for selection of a test equating method. *Journal of Educational Measurement*, 18(1), 23-38.
- Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9(1), 24-44.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29-36.
- Kolen, M. J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement*, 28(3), 257-282.
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating methods and practices*. New York: Springer.

- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd. ed.). New York: Springer.
- Kolen, M.J., & Whitney, D. R. (1982). Comparison of four procedures for equating the tests general educational development. *Journal of Educational Measurement*, 19(4), 279-293.
- Liu, C. (2011). A comparison of statistics for selecting smoothing parameters for loglinear presmoothing and cubic spline post smoothing under a random groups design (Unpublished doctoral dissertation). Available from Iowa Research Online. (UMI No. 1013).
- Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement*, 30(1), 23-39.
- Livingston, S. A. (2004). *Equating test scores (Without IRT)*. Educational Testing Service.
- Moses, T., & Holland, P. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (ETS Research Rep. No. RR-07-15). Princeton, NJ: ETS.
- Petersen, N. S. (2007). Equating: best practices and challenges to best practices. N.J. Dorans, M.Pommerich & P.W. Holland. (Ed.) *Linking and aligning scores and scales* (pp. 89-106). USA: Springer Science&Business Media.
- Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement*, 42 (4), 309-330.
- Woldbeck, T. (1998, April). *Basic concepts in modern methods of test equating*. Paper presented at the annual meeting of the Southwest Psychological Association, New Orleans.
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23.