

Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma*

A Study on Equating Academic Staff and Graduate Education Entrance Examination Scores

Neşe ÖZTÜRK** Duygu ANIL***

Hacettepe Üniversitesi

Öz

Bir sınavın farklı tarihlerde farklı formlarının uygulanması, formların en azından güçlük düzeyi bakımından birbirinden farklılık gösterebileceği sorununu ortaya çıkarmaktadır. Bir sınavın farklı formlarından elde edilen puanları karşılaştırabilmek için puanların eşitlenmesi gerekir. Bu araştırmanın amacı, 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.

Anahtar Sözcükler: Test eşitleme, doğrusal eşitleme, eşit yüzdelikli eşitleme.

Abstract

Applying different forms of a single test on different times may cause at least different difficulty level problems. To compare the scores which are provided from different forms need to be equated. The scope of this research study was to equate 2008 spring and fall ALES quantitative-1 scores with using linear and equipercentile equating methods and to suggest most appropriate method. The findings showed that there was a linear relation between raw scores and equated scores. A result of linear equating difficulty of tests were varying along the score scale; a result of equipercentile equating fall test was more difficult than spring test at the low scores and easier than spring at the top scores. At the end of the study, it is seen that the most appropriate method for equating test scores was equipercentile equating.

Keywords: Test equating, linear equating, equipercentile equating

Summary

Purpose

The scope of this study was to equate 2008 spring and fall ALES quantitative-1 test scores with linear and equipercentile equating methods and suggest the most appropriate equating method.

* Bu çalışma, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme ABD için hazırlanan yüksek lisans tezinin bir kısmından özetlenmiştir.

** Arş. Gör. Neşe ÖZTÜRK, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme A.B.D, e-posta: neseozturk@hacettepe.edu.tr

*** Doç. Dr. Duygu ANIL, Hacettepe Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme ABD, e-posta: aduygu@hacettepe.edu.tr

AKADEMIK PERSONEL VE LISANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

Results

Population of this study was 40777 people who taken both 2008 spring and fall ALES examinations. Sample of this study was determined by computer randomly from population 5% and size of sample was 2040 people.

The data used in this study were 2008 spring and fall ALES quantitative-1 tests and item scores. Data analysis was completed in four steps. Firstly, a descriptive statistical analysis was completed for each test; secondly whether or not each of test provided assumptions of equating methods was examined; in third step equated scores were provided by linear and equipercentile equating method. Lastly, for deciding which equating method was more appropriate, equating error was calculated by WMSE index.

According to findings; both tests were unidimensional and measured the same construct with equal reliability and difficulty. As a result of linear equating, found that there was linear relation between raw scores and equated scores which referred mathematically as $Y = 1,161X - 3,542$. The findings showed that equated scores had values out of raw score scale and the difference between equated scores and raw scores was increasing in the extreme points of score distribution. As a result of equipercentile equating although there were fluctuations, seen that there was a linear relation between raw scores and equated scores which referred mathematically as $Y = 0,8711X + 2,9261$. The findings showed that equated scores didn't have any values out of raw score scale. To decide which equating method was the most appropriate for equating quantitative-1 test scores the equating error was calculated by WMSE index. According to findings, the equating error for linear equating was 0.026 and for equipercentile equating was 0.025.

Discussion

The results of linear equating showed that there was a linear relation between equated scores and raw scores. And equated scores had values out of raw score scale. Jaeger (1981) indicated that due to linear equating nature equated scores and raw scores didn't overlap and Livingston (2004) described this situation as characteristic of linear equating. Also Kolen and Brennan (1995) referred that equated scores from linear equating could sometimes be out of the range of possible observed scores and they suggested two way to handle with this problem. One way was to allow to top and bottom to "float". The second way was truncating the conversion at the highest and lowest score. The findings showed that difference between converted scores were highest in extreme points of score scale. Jaeger (1981) indicated that difference between equating functions were usually highest in the score scale extreme points. Also the study which was conducted by Kan (2010) supported this result. Difference between equated scores from linear equating equated scores and raw scores were equal only while standard deviations and means were equal (Angoff, 1984; Lord, 1980). In this study, although there was no significant difference between means, difference between variances could be reason of difference between converted scores.

As a result of equipercentile equating seen that despite of fluctuations there was a linear relation between converted scores. Kolen and Brennan indicated that if the score distributions of equated forms are identical in shape in other words they only differ in mean and standart deviation in this case linear equating and equipercentile equating relationships are identical. In this study, distributions of 2008 spring and fall ALES quantitative -1 tests are different beyond mean and standart deviation. The fluctuations of equating relationship between raw score and equated scores could be explained with non- identical score distributions.

The results of equipercentile equating as Kolen and Brennan (1995) point out equated scores didn't have values out of the range of raw scores. From the fluctuations between raw scores and equated scores, it could be said that differences between variances and standard deviations could affect equated scores. Also due to not implementing any smoothing method to raw score distribution, random error could mix while estimating equivalent scores.

Kolen and Brennan (1995) stated that using equipercentile equating for example X form could be more difficult than Y form at high and low scores, but less difficult at the middle scores. According to this study, regarding equated scores were bigger than raw scores in the bottom and smaller in the top of score scale, we could say that at the low scores 2008 fall ALES quantitative-1 test was more difficult than spring test; at the top scores fall test easier than spring test.

Conclusion

At the end of the study, it was found that there was a linear relation between equated scores and raw scores. The result of linear equating found that difficulty of tests were varying along the score scale and the result of equipercentile equating found that fall test was more difficult than spring test at the low scores and easier than spring test at the top scores. Finally, it could be said that the most appropriate method for equating quantitative test scores was equipercentile equating.

Giriş

Ülkemizde, Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından hazırlanan ve uygulanan, üç yıl geçerliği olan merkezi bir sınavdır. Bir yılda, ilkbahar ve sonbahar olmak üzere iki kez yapılan bu sınavda test güvenliği ve gizliliğini sağlamak amacıyla her uygulamada ölçülen davranışlar değişmemekle birlikte farklı sorulardan oluşan farklı formlar kullanılmaktadır.

Bir sınavın farklı tarihlerde farklı formlarının uygulanması, formların en azından güçlük düzeyi bakımından birbirinden farklılık gösterebileceği sorununu ortaya çıkarmaktadır. Örneğin, bir sınavın farklı iki formunu alarak sınav olan A ve B öğrencilerinden A öğrencisinin B öğrencisinden daha yüksek puan almasının nedenlerinden biri A öğrencisinin daha başarılı olması olabileceği gibi, diğer bir sebebi de A öğrencisinin aldığı formun B öğrencisinin aldığı formdan daha kolay olması olabilir. Test geliştirme uzmanları, her ne kadar içerik ve istatistiksel özellikler bakımından birbirine olabildiğince benzer formlar hazırlamaya çalışsa da formların tam olarak birbirine eşdeğer olması beklenemez ve formlar en azından güçlük düzeyi olarak birbirinden farklılık gösterebilir. Bu yüzden, bir testin farklı formlarından elde edilen puanları karşılaştırabilmek için puanların eşitlenmesi gerekmektedir.

Angoff (1984), test formlarının eşitlenmesini bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlamış ve dönüştürmeden sonra formların eşdeğer olacağını belirtmiştir. Kolen ve Brennan (1995) eşitlemeyi, benzer içerik ve güçlük düzeyinde geliştirilen test formları arasındaki farklılıkları düzenleyerek, bu formlardan elde edilen puanlarının birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olarak ifade etmiştir. Braun ve Holland (1982) yine benzer bir şekilde eşitlemeyi, güçlük düzeyi farklılık gösteren formlardan elde edilen puanların birbiri yerine kullanılabilmesini sağlamak için yapılan sayısal düzenlemeler olarak tanımlamışlardır. En genel tanımıyla eşitleme, test formları arasındaki güçlük farklılıklarını düzenleyerek elde edilen puanların birbiri yerine kullanılması sağlayan sürece verilen addır.

Eşitleme Koşulları

Eşitleme çalışmasını yürütebilmek için testlerin birtakım koşulları sağlaması gerekmektedir. Dorans ve Holland (2000), literatürde en çok öne çıkan beş eşitleme koşulunu şöyle sıralamışlardır: aynı yapıyı ölçme, güvenilirliklerin eşit olması, simetriklik, eşitlik ve gruptan bağımsızlık.

En çok öne çıkan ve en temel eşitleme koşulu, eşitlenecek iki testin mutlaka aynı yapıyı, özelliği veya yeteneği ölçmesi gerektiğidir. Ayrıca eşitlenecek iki testin sadece aynı özelliği ölçmesi yeterli değildir, aynı zamanda testlerin tek bir özelliği ölçmesi gerekir. Bu da eşitlenecek testlerin tek boyutlu bir yapıya sahip olması gerektiğini göstermektedir (Wolbeck, 1998). Testlerin güvenilirliklerinin eşit olması, aynı gerçek puanı ölçmeleri anlamına gelmektedir. Dorans ve Holland (2000), testlerin güvenilirliklerinin eşit olmasının yanı sıra güvenilirliklerinin

AKADEMİK PERSONEL VE LISANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

büyükliğünün de dikkate alınması gerektiğini vurgulamış ve güvenilirliği yüksek olan testlerin eşitlenmesinin daha iyi sonuçlar vereceğini belirtmişlerdir.

Simetriklik, hangi forma X, hangi forma Y denildiğine bakılmaksızın eşitlemenin aynı olmasıdır (Lord, 1980). Yani dönüşümün tersine çevrilebilir olmasıdır. Örneğin, X formundaki 26 ham puan, Y formundaki 27 ham puana dönüşüyorsa; Y formundaki 27 ham puan, X formundaki 26 ham puana dönüşmelidir. Bir test, Lord (1980) tarafından öne sürülen eşitlik özelliğine sahip ise, herhangi bir kişinin bir testin X veya Y formunu almış olmasının fark oluşturması gerekir. Lord'un eşitlik koşulu çok tartışılan ve Dorans ve Holland (2000)'in deyimleriyle üzerinde kara bulutlar olan bir konudur. Eğer testler tam olarak paralel hazırlanabilirse eşdeğer puanlar vereceği için eşitlemeye gerek kalmayacağından Lord'un eşitlik koşulunu ölçüt olarak testleri eşitlemek imkânsızdır ya da bu koşulun sağlandığı bir durumda gereksizdir. Bu yüzden, uygulamada Lord'un eşitlik koşulunu sağlamak zordur ve çoğu zaman kullanımı teoride kalmaktadır (Crocker ve Algina, 1986; Harris ve Crouse 1992; Kolen ve Brennan, 1995, 2004)

Bir diğer eşitleme koşulu ise eşitlemenin gruptan bağımsızlığıdır. Bu özellik sağlanırsa eğer eşitleme ilişkisi gruptan bağımsız olacaktır; örneğin kızlar ve erkekler için aynı eşitleme ilişkisi bulunacaktır (Hambleton ve Swaminathan, 1985; Kolen ve Brennan, 2004)

Eşitleme Yöntemleri

Klasik Test Teorisi'ne dayalı; ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelli eşitleme olmak üzere üç eşitleme yöntemi bulunmaktadır. Ortalama eşitlemede, eşitlenecek olan testlerin güçlükleri arasında fark olduğu, fakat bu farkın puan ölçeği boyunca sabit olduğu kabul edilmektedir. Bu yöntem, eşitlenecek testlerdeki puanların ortalamaya olan uzaklıklarının eşit kabul edilmesi esasına dayanmaktadır (Kolen ve Brennan, 1995). Ortalamaları μ_x ve μ_y olan eşitlenecek X ve Y formları için ortalama eşitleme (1) numaralı eşitlikte olduğu gibi ifade edilebilir:

$$y = x - \mu_x + \mu_y \quad (1)$$

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). Doğrusal eşitleme (2) numaralı eşitlikte görüldüğü gibi ifade edilebilir:

$$Y = \frac{S_y}{S_x} X + \mu_y - \frac{S_y}{S_x} \mu_x \quad (2)$$

(2) numaralı eşitlikte yer alan $\frac{S_y}{S_x}$ bileşenine "a" ve $\mu_y - \frac{S_y}{S_x} \mu_x$ bileşenine "b" denirse; eğimi "a" ve

kesme noktası "b" olan $Y = ax + b$ doğru denklemi elde edilir.

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 sırasında olanların eşdeğer olarak kabul edilebileceğini belirtmiştir.

Eşitleme Hatası

Eşitleme desenlerini ve yöntemlerini seçerken göz önünde bulundurulması gereken ölçütlerden biri eşitleme hatasıdır. Eşitleme hatasının olabildiğince küçük olması, seçilen eşitleme yönteminin uygunluğunun bir göstergesidir. Random ve sistematik olmak üzere iki çeşit eşitleme hatası vardır (Kolen, 1988; Kolen ve Brennan 1995, 2004).

Random eşitleme hatası, eşitleme ilişkisini belirlemede yer alan ortalama, standart sapma ve yüzdeler gibi parametrelerin kestirildiği örnekleme bağlı olarak ortaya çıkan hatadır. Eşitleme ilişkisi sadece random hata içerdiğinde, veriler örneklemden toplandığı için kestirilen eşitleme ilişkisi evrenden farklılık gösterecektir. Eğer evrende çalışılırsa random hata oluşmayacaktır. Örneklem büyüklüğü arttıkça random eşitleme hatası küçülmektedir. Bir diğer eşitleme hatası sistematik hatadır. Sistematik hatanın oluşma nedenlerinden biri eşitleme ilişkisinin kestirildiği yöntemin yanlılık içermesidir. Örneğin, eşitleme sürecinde kullanılan düzgünleştirme teknikleri yanlılık içerebilmektedir. En kullanışlı düzgünleştirme tekniği, random hatayı azaltırken sistematik hatayı arttırabilir. Sistematik hatanın bir diğer oluşma nedeni, eşitlemenin varsayımlarının ihlal edilmesidir (Kolen ve Brennan, 1995, 2004).

Her yıl, bir testin farklı formlarını geliştirip kullanan sınav programlarında, test eşitleme çok ciddiye alınması gereken istatistiksel bir süreçtir (Holland ve Thayer, 1985). ALES bir yılda ilkbahar ve sonbahar olmak üzere iki kez uygulanan ve geçerliği üç yıl olan merkezi bir sınavdır. ALES sayısal ve sözel alt testlerinden oluşmaktadır. Sınavın sayısal bölümünde 40'ar sorudan oluşan sayısal-1 ve sayısal-2 alt testleri; sözel bölümünde ise 80 sorudan oluşan sözel alt testi bulunmaktadır. Yükseköğretim kurumlarında öğretim görevlisi, okutman, araştırma görevlisi, uzman, çevirici ve eğitim-öğretim planlamacısı kadrolarına açıktan veya öğretim elemanı dışındaki kadrolara naklen atamalarda, ülkemizde lisansüstü eğitime girişte, yurtdışına lisansüstü eğitim için gönderilecek adayların seçiminde ilgili kurumlar ALES puanını dikkate almaktadırlar (ÖSYM, 2008). Kesme puanın değişmediği ve üç yıl geçerliği olan bu sınavda, farklı dönemlerde farklı maddelerden oluşan test formlarını alan bireylerin puanlarını karşılaştırabilmek ve puanları birbiri yerine kullanılabilmesi için test eşitleme oldukça önemlidir. Literatür incelendiğinde, ülkemizde test eşitlemeye yönelik çalışmaların çok sınırlı sayıda olduğu görülmüştür. Bu araştırmanın test eşitleme çalışmalarına bir bakış açısı katması açısından da önemli olduğu düşünülmektedir.

Bu çalışma ile 2008 ilkbahar ve sonbahar dönemlerinde yapılan ALES'in sayısal-1 testi puanlarının farklı eşitleme yöntemleri kullanılarak eşitlenmesi ve en uygun eşitleme yönteminin önerilmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıdaki araştırma sorularına cevap aranmıştır;

1. 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerine ait puanların doğrusal eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?
2. 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerine ait puanların eşit yüzdellikli eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?
3. 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerine ait puanların eşitlenmesi için doğrusal eşitleme ve eşit yüzdellikli eşitleme yöntemlerinden hangisi en uygundur?

Yöntem

Araştırmanın Türü

Bu çalışma kapsamında, 2008 ilkbahar ve sonbahar ALES sayısal-1 testi puanları doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılarak eşitlenmiştir. Bu çalışmada, var olan yöntem ve tekniklerin gerçek veri üzerinden sınanması yapıldığından araştırma temel araştırma niteliğindedir (Karasar, 2009).

Evren ve Örneklem

Bu araştırmanın evrenini, 2008 ilkbahar ve sonbahar dönemlerinde yapılan ALES'e her iki dönem katılan 40777 kişi oluşturmaktadır. Evreni temsil edecek örnekleme oluşturan

AKADEMİK PERSONEL VE LİSANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

185

kişiler bilgisayar aracılığıyla evrenin %5 i tesadüfi olarak seçilerek belirlenmiştir. Araştırmanın örneklemini 2040 kişiden oluşmaktadır.

Veri Toplama Araçları

Araştırmanın verilerini, 2008 ilkbahar ve sonbahar dönemlerinde yapılan ALES'e her iki dönem de katılan bireylerin sayısal-1 test ve madde puanları oluşturmaktadır. Sayısal-1 testi 40 sorudan oluşmaktadır ve sayısal-1 testiyle sınava girenlerin sayısal akıl yürütme becerilerini ölçmek amaçlanmaktadır (ÖSYM, 2008).

ÖSYM'den alınan bilgilere göre, 2008 ilkbahar döneminde yapılan ALES'e giren kişi sayısı 174832 iken 2008 sonbahar döneminde yapılan ALES'e giren kişi sayısı 142178'dir. 2008 yılında her iki dönem yapılan ALES'e giren (ortak aday) kişi sayısı 40777'dir ve bu kişilere ait puanlar ÖSYM tarafından TC kimlik numaralarına göre eşleştirilerek belirlenmiştir. Kişilerin testlere ait doğru cevap sayıları ve her bir alt teste ait cevapları gerekli izinler alınarak ÖSYM kayıtlarından elde edilmiştir.

Verilerin Analizi

Verilerin analizi dört aşamada gerçekleştirilmiştir. Birinci aşamada 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerinin betimsel istatistikleri hesaplanmış; ikinci aşamada testlerin eşitleme koşullarını sağlayıp sağlamadığı test edilmiş; üçüncü aşamada doğrusal ve eşit yüzdelikli eşitleme yöntemleri kullanılarak eşitlenmiş puanlar elde edilmiş ve son aşamada ise her bir eşitleme yönteminde yapılan hata miktarı hesaplanmıştır.

Bulgular

Betimsel İstatistiklerin Hesaplanması

Verilerin analizinin ilk aşamasında 2008 ilkbahar ve sonbahar ALES alt testlerinin betimsel istatistikleri hesaplanmış ve Tablo 1'de verilmiştir.

Tablo 1.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerine Ait Betimsel İstatistikler

İstatistikler	2008 İlkbahar ALES Sayısal-1	2008 Sonbahar ALES Sayısal-1
K	40	40
N	2040	2040
Ortalama	21.90	21.88
Ortanca	23.00	24.00
Tepe Değer	33	35
Std. Sapma	10.22	11.86
Varyans	104.36	140.64
Çarpıklık K.	-0.33	-0.25
Basıklık K.	-0.99	-1.27
KR-20	0.93	0.95

K: madde sayısı; N: kişi sayısı

Tablo 2.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerine Ait faktör Analizi Sonuçları

2008 İlkbahar ALES Sayısal - 1				2008 Sonbahar ALES Sayısal - 1			
Faktör	Özdeğer	Varyans Yüzdesi	Yığılmalı Yüzde	Faktör	Özdeğer	Varyans Yüzdesi	Yığılmalı Yüzde
1	20.20	50.49	50.49	1	21.02	52.56	52.56
2	3.31	8.27	58.77	2	3.50	8.76	61.32

Tablo 2'deki bilgilere göre, ilkbahar ALES sayısal-1 testinde varyansın %50.49'u ve sonbahar sayısal-1 testinde varyansın %52.56'sı birinci faktör tarafından açıklanmaktadır. Ayrıca özdeğerler ve varyans açıklama oranları arasında birinci faktörden sonra keskin bir düşüşün olması ve aralarındaki farkın yaklaşık 6 katı bulması dikkati çekmektedir. Bu sonuçlara dayalı olarak sayısal-1 testlerinin tek boyutlu olduğu, bir başka deyişle tek bir yapıyı ölçtüğü söylenebilir. Sayısal-1 testleri arasındaki ilişkiyi belirlemek için PMÇKK hesaplanmış ve iki test arasında yüksek düzeyde, pozitif ve anlamlı ($r = .89, p < .05$) bir ilişki bulunmuştur. Tüm bu sonuçlara dayalı olarak, sayısal-1 testlerinin tek bir ve aynı yapıyı, "sayısal akıl yürütme yeteneğini" ölçtüğü söylenebilir.

Testlerin Güvenirlikleri Arasındaki Farkın Test Edilmesi

Testlerin eşit güvenirlikte olup olmadığını belirlemek için, güvenirlik katsayıları iki korelasyon katsayısı olarak kabul edilmiş ve güvenirlik katsayılarına Fischer'in Z_r dönüşümü yapılmıştır. İki güvenirlik katsayısı arasında fark olup olmadığı Fischer'in Z istatistiği ile test edilmiştir (Akhun, 1984). Güvenirlik katsayıları arasındaki farkın test edilmesine ilişkin bulgular Tablo 3'te görüldüğü gibidir.

Tablo 3.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerinin Güvenirliklerinin Karşılaştırılmasına İlişkin Bulgular

Testler	KR-20	Z_r	Z
İlkbahar Sayısal-1	0.93	1.658	-0.75
Sonbahar Sayısal-1	0.95	1.832	

Tablo 3 incelendiğinde, testlerin güvenirlikleri arasında .05 anlamlılık düzeyinde manidar bir fark olmadığı görülmektedir. Bu sonuca dayalı olarak testlerin eşit güvenirliğe sahip olma koşulunu sağladığı söylenebilir.

Ortalama Güçlükleri Arasındaki Farkın Test Edilmesi

Testlerin ortalama güçlükleri arasında anlamlı bir fark olup olmadığı iki oran farkı testiyle incelenmiştir (Baykul, 1996). Elde edilen bulgular Tablo 4'te verilmiştir. Tablo 4 incelendiğinde, testlerin ortalama güçlükleri arasında .05 anlamlılık düzeyinde manidar bir fark olmadığı görülmektedir.

AKADEMİK PERSONEL VE LISANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

Tablo 4.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerinin Ortalama Güçlüklerinin) Karşılaştırılmasına İlişkin Bulgular

Testler	\bar{X}	K	\bar{p}	t	p
İlkbahar Sayısal-1	21.90	40	0.5475	0.005	0.50
Sonbahar Sayısal-1	21.88	40	0.5470		

Ortalamaları ve Varyansları Arasındaki Farkın Test Edilmesi

Testlerin ortalamaları arasında anlamlı bir fark olup olmadığı bağımlı gruplar *t*-testi ile varyanslarının eşit olup olmadığı ise *Levene's testi* ile incelenmiş ve elde edilen bulgular Tablo 5'te verilmiştir.

Tablo 5.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerinin Ortalama ve Varyanslarının Karşılaştırılmasına İlişkin Bulgular

Testler	t -Testi			Levene's Testi		
	\bar{X}	t	p	S_x^2	F	p
İlkbahar Sayısal-1	21.90	-0.15	0.88	104.36	102.631	0.00*
Sonbahar Sayısal-1	21.88			140.64		

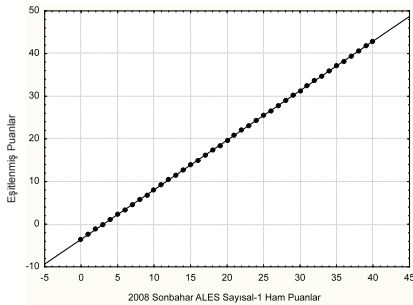
* $p < 0.05$

Tablo 5 incelendiğinde, testlerin ortalamaları arasında .05 anlamlılık düzeyinde manidar bir fark olmadığı görülmektedir. *Levene's testi* sonuçlarına göre ise testlerin varyansları arasında .05 anlamlılık düzeyinde manidar bir fark olduğu görülmektedir.

Tek ve aynı yapıyı ölçen, eşit güvenilirliğe, ortalamaya ve ortalama güçlüğüne sahip olan 2008 ilkbahar ve sonbahar ALES sayısal-1 testleri doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılarak eşitlenmiştir. Elde edilen bulgular aşağıda sırayla verilmiştir.

2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerine Ait Puanların Doğrusal Eşitleme Yöntemi ile Eşitlenmesi Sonucu Elde Edilen Bulgular

Ham puanlar ile eşitlenmiş puanların birlikte grafiği elde edilmiş ve buna ait grafik Şekil 1'de verilmiştir.



Şekil 1. 2008 İlkbahar ve Sonbahar ALES Sayısal-1 Alt Testi Ham Puanlarının Doğrusal Eşitleme Yöntemiyle Eşitlenmesi

NEŞE ÖZTÜRK, DUYGU ANIL

Şekil 1'deki grafiğe bakıldığında, ham puanlar ile eşitleme sonucu elde edilen puanlar arasındaki ilişkinin doğrusal olduğu görülmektedir. Bu ilişki matematiksel olarak $Y = 1,161X - 3,542$ doğru denklemi ile ifade edilebilir. Tablo 6'da ham puanlar ile bu puanlara karşılık gelen eşitleme sonucu elde edilen puanlar yer almaktadır.

Tablo 6.

2008 Sonbahar ALES Sayısal-1 Testi Ham Puanları ve Bu Puanlara Karşılık Gelen Doğrusal Eşitleme Yöntemiyle Elde Edilen Eşitlenmiş Puanlar

Ham Puanlar	Eşitlenmiş Puanlar	Fark
0	-3,54	3,54
1	-2,38	3,38
2	-1,22	3,22
3	-0,06	3,06
4	1,10	2,90
5	2,26	2,74
6	3,42	2,58
7	4,59	2,42
8	5,75	2,25
9	6,91	2,09
10	8,07	1,93
11	9,23	1,77
12	10,39	1,61
13	11,55	1,45
14	12,71	1,29
15	13,87	1,13
16	15,03	0,97
17	16,20	0,81
18	17,36	0,64
19	18,52	0,48
20	19,68	0,32
21	20,84	0,16
22	22,00	0,00
23	23,16	-0,16
24	24,32	-0,32
25	25,48	-0,48
26	26,64	-0,64
27	27,81	-0,81
28	28,97	-0,97
29	30,13	-1,13
30	31,29	-1,29
31	32,45	-1,45
32	33,61	-1,61
33	34,77	-1,77
34	35,93	-1,93
35	37,09	-2,09
36	38,25	-2,25
37	39,42	-2,42
38	40,58	-2,58
39	41,74	-2,74
40	42,90	-2,90

AKADEMİK PERSONEL VE LİSANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

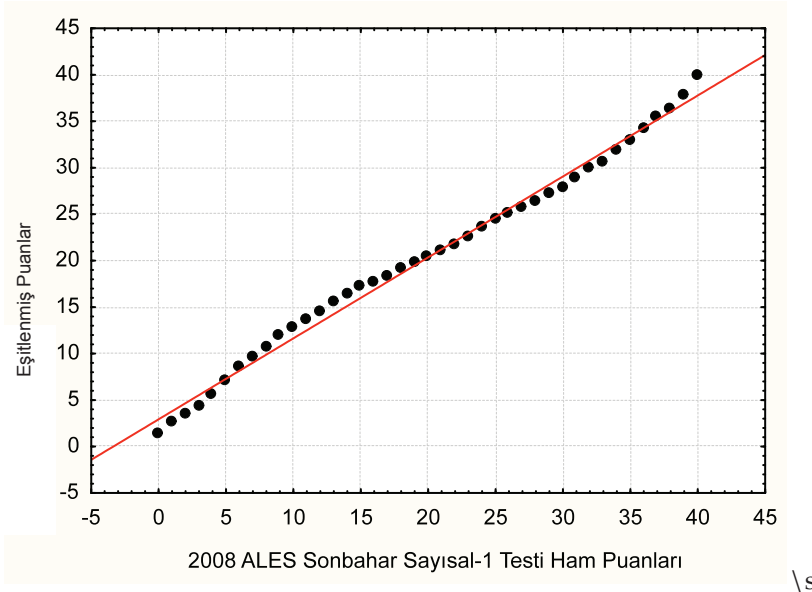
2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testlerine Ait Puanların Eşit Yüzdellikli Eşitleme Yöntemi ile Eşitlenmesi Sonucu Elde Edilen Bulgular

Tablo 6 incelendiğinde, ham puanlar 0 ile 40 aralığında değerler alırken eşitlenmiş puanların -3.54 ile 42.90 arasında değerler aldığı dikkati çekmektedir. 0 ile 21 arasındaki ham puanların eşitlenmiş puanlarının ham puanlardan küçük, 22 ham puanının eşitlenmiş puanının kendisine eşit ve 23 ile 40 arasındaki ham puanların eşitlenmiş puanlarının ham puanlardan büyük olduğu görülmektedir.

2008 sonbahar ALES sayısal-1 alt testi ham puanları 2008 ilkbahar ALES sayısal-1 testi ham puanlarına tek grup desenine dayalı eşit yüzdellikli eşitleme yöntemi kullanılarak eşitlenmiştir. Testleri eşitlemeden önce ham puanların yığılmalı yüzdeleri hesaplanarak puanların yüzdellik sıraları belirlenmiştir. Her iki alt teste ait ham puanlar, nadiren aynı yüzdellik sıraya denk geldiği için Livingston (2004) tarafından önerilen (3) numaralı eşitlikte verilen öteleme formülü uygulanmıştır.

$$\text{referansformaltpuan} + \frac{\text{yeniformyüzdellik} - \text{referansformaltyüzdellik}}{\text{referansformüstyüzdellik} - \text{referansformaltyüzdellik}} - (\text{referansformüstpuan} - \text{referansformaltpuan}) \quad (3)$$

Ham puanlar ile bu puanlara karşılık gelen eşit puanların birlikte grafiği elde edilmiş ve buna ait grafik Şekil 2'de verilmiştir.



Şekil 2. 2008 İlkbahar ve Sonbahar ALES Sayısal-1 Testi Puanlarının Eşit Yüzdellikli Eşitleme Yöntemiyle Eşitlenmesi

Şekil 2'deki grafik incelendiğinde, ham puanlar ile eşitlenmiş puanlar arasında yer yer dalgalanmalar olmasına rağmen doğrusal bir ilişki olduğu söylenebilir ve bu ilişki $Y = 0,8711X + 2,9261$ doğru denklemiyle ifade edilebilir. 2008 sonbahar ALES sayısal-1 testi ham puanları ve bu puanların eşit yüzdellikli eşitleme sonucu elde edilmiş eşitlenmiş puanları Tablo 7'de verilmiştir.

Tablo 7 incelendiğinde, eşitlenmiş puanların 1.33 ile 40 arasında değerler aldığı ve 0-40 ham puan aralığının dışına çıkmadığı görülmektedir. 0 ile 21 arasındaki ham puanların eşitlenmiş puanlarının ham puanlardan daha büyük; 22 ile 39 arasındaki ham puanların eşitlenmiş puanlarının ise ham puanlardan daha küçük olduğu görülmektedir. Sadece 40 ham puanının eşitlenmiş puanı kendisine eşittir. Ham puanlar ile eşitlenmiş puanlar arasındaki farkın 0 ile 9 ve 32 ile 40 puanlar arasında yer yer artıp azaldığı; 10 ile 21 puanlar arasında puanlar arttıkça

NEŞE ÖZTÜRK, DUYGU ANIL

azaldığı ve 22 ile 31 puanları arasında ise puanlar arttıkça farkın da arttığı görülmektedir. Ham puanlar ile eşitlenmiş puanlar arasında oluşan bu dalgalanmaları Şekil 2'deki grafikte de görmek mümkündür.

Tablo 7.

2008 Sonbahar ALES Sayısal-1 Testi Ham Puanları ve Bu Puanlara Karşılık Gelen Eşit Yüzdelikli Eşitleme Yöntemi ile Elde Edilen Eşitlenmiş Puanlar

Ham Puan	Eşitlenmiş Puan	Fark
0	1,33	-1,33
1	2,63	-1,63
2	3,50	-1,50
3	4,31	-1,31
4	5,52	-1,52
5	7,03	-2,03
6	8,49	-2,49
7	9,60	-2,60
8	10,60	-2,60
9	11,86	-2,86
10	12,73	-2,73
11	13,62	-2,62
12	14,54	-2,54
13	15,48	-2,48
14	16,41	-2,41
15	17,30	-2,30
16	17,70	-1,70
17	18,26	-1,26
18	19,19	-1,19
19	19,87	-,87
20	20,47	-,47
21	21,14	-,14
22	21,77	,23
23	22,59	,41
24	23,53	,47
25	24,45	,55
26	25,09	,91
27	25,80	1,20
28	26,45	1,55
29	27,11	1,89
30	27,94	2,06
31	28,86	2,14
32	29,87	2,13
33	30,66	2,34
34	31,86	2,14
35	32,96	2,04
36	34,16	1,84
37	35,44	1,56
38	36,38	1,63
39	37,79	1,21
40	40,00	,00

AKADEMİK PERSONEL VE LİSANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

Doğrusal Eşitleme ve Eşit Yüzdellikli Eşitleme Yöntemlerinin Uygunluğuna İlişkin Elde Edilen Bulgular

Bu yöntemlerden hangisinin daha uygun olduğunu belirlemek için iki yönetime ait ağırlıklandırılmış hata kareleri ortalamalarını veren WMSE katsayıları hesaplanmıştır. Elde edilen bulgular Tablo 8'de verilmiştir.

Tablo 8.

Doğrusal Eşitleme ve Eşit Yüzdellikli Eşitleme Yöntemlerine Ait WMSE Katsayıları

Eşitleme Yöntemleri	WMSE
Doğrusal Eşitleme	0.026
Eşit Yüzdellikli Eşitleme	0.025

Tablo 8'de verilen bilgilere göre, sayısal-1 testlerinde doğrusal eşitleme yöntemi için hesaplanan WMSE değeri 0.026 iken eşit yüzdellikli eşitleme için 0.025'tir. Bu bulguya dayalı olarak sayısal-1 testlerinin eşitlenmesi için en uygun yöntemin WMSE değeri daha küçük olan eşit yüzdellikli eşitleme olduğu söylenebilir.

Sonuç ve Tartışma

2008 sonbahar ALES sayısal-1 testinin 2008 ilkbahar ALES sayısal-1 testine doğrusal eşitleme yöntemiyle eşitlenmesi sonucu ham puanlar ile eşitlenmiş puanlar arasında doğrusal bir ilişki olduğu sonucuna ulaşılmıştır. Eşitlenmiş puanların ham puan aralığı dışında değerler aldığı görülmüştür. Jaeger (1981), doğrusal eşitleme yönteminin doğası gereği eşitlenmiş puanlar ile ham puanların puan aralıklarının tam olarak örtüşmediğini belirtmiştir ve Livingston (2004) bu durumu doğrusal eşitlemenin karakteristik bir özelliği olarak nitelendirmiştir. Kolen ve Brennan (1995) ise doğru sayısı temel alınarak puanlanan testlerin, doğrusal eşitlemeyle elde edilen eşitlenmiş puanlarının testin puan aralığının dışına çıkmasının beklenen bir sonuç olduğunu belirtmiş ve bu sorunla başa çıkmak için iki yol önermiştir. Birinci yol, ham puan sınırı aşan eşitlenmiş üst ve alt puanlara izin verilebilir. Örneğin, en yüksek ham puanı aşan eşitlenmiş puan kabul edilebilir. İkinci yol, en yüksek ve en düşük eşitlenmiş puanlar eşitleme sürecinden çıkartılabilir.

Doğrusal eşitleme sonucu elde edilen eşitlenmiş puanlarla ham puanlar arasındaki farkın, puan dağılımının uç noktalarına gidildikçe arttığı tespit edilmiştir. Jaeger (1981), eşitleme fonksiyonları arasındaki farkın doğrusal eşitlemede genellikle puan dağılımının uç noktalarında daha büyük olduğunu belirtmiştir. Bu sonucu aynı zamanda Kan (2010)'ın çalışması da desteklemektedir. Doğrusal eşitleme yöntemi kullanılarak elde edilen eşitlenmiş puanlar ve ham puanlar arasında ancak testlerin ortalama ve standart sapmalarının eşit olması durumunda fark oluşmamaktadır (Angoff, 1984; Lord, 1980). Bu çalışmada eşitlenen 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerinin ortalamaları arasında anlamlı bir fark bulunmamasına rağmen varyanslarının eşit olmaması ham puanlarla eşitlenmiş puanlar arasında farkın oluşmasının bir nedeni olabilir.

2008 sonbahar ALES sayısal-1 testinin 2008 ilkbahar ALES sayısal-1 testine eşit yüzdellikli eşitlemeyle eşitlenmesi sonucu ham puanlar ve eşitlenmiş puanlar arasında doğrusal bir ilişki olmasına rağmen yer yer dalgalanmalar olduğu görülmüştür. Kolen ve Brennan (1995), eşitlenen formlara ait puan dağılımları aynı şekle sahip ise bir başka deyişle sadece ortalama ve standart sapmaları farklılık gösteriyor ise böyle bir durumda doğrusal eşitleme ve eşit yüzdellikli eşitleme fonksiyonlarının aynı olacağını belirtmişlerdir. Bu çalışma kapsamında eşitlenen 2008 ilkbahar ve sonbahar ALES sayısal-1 testlerine ait puan dağılımlarına ait ortalama ve standart sapma dışındaki çarpıklık ve basıklık katsayılarının aynı olmaması, ham puanlar ile eşitlenmiş puanlar arasında tam olarak doğrusal bir ilişki oluşmamasının bir nedeni olarak gösterilebilir.

Eşit yüzdelikli eşitleme sonucu elde edilen eşitlenmiş puanların ham puan aralığının dışında değerler almadığı görülmüştür. Kolen ve Brennan (1995), doğrusal eşitlemede görülen eşitlenmiş puanların ham puan aralığının dışında değerler alması sorununun eşit yüzdelikli eşitlemede oluşmadığını belirtmiştir. Elde edilen bulgular da bu sonucu destekler niteliktedir. Eşitlenmiş puanlarla ham puanlar arasındaki ilişkiyi gösteren grafikte oluşan dalgalanmalardan, eşitleme yapılmadan önce puan dağılımlarında herhangi bir düzgünleştirme yönteminin uygulanmamış olmasının eşdeğer puanların kestirilmesine random hata karışmış olabileceği sonucuna ulaşılabilir. Kolen ve Brennan (1995), eşit yüzdelikli eşitlemenin formlar arasındaki güçlük düzeyini açıkladığını, örneğin üst puanlarda X formunun Y formundan daha üst ve alt puanlarda daha zor olabileceğini belirtmişlerdir. Bu araştırmada ise eşitlenmiş puanların, puan dağılımının alt noktalarında ham puanlardan daha büyük, üst noktalarında ise daha küçük olduğu görülmüştür. Bu bulguya dayalı olarak 2008 sonbahar ALES sayısal-1 testinin alt puanlarda ilkbahar sayısal-1 testinden daha zor iken, üst puanlarda daha kolay olduğu söylenebilir.

Araştırmanın sonucunda, 2008 sonbahar ALES sayısal-1 testlerinin 2008 ilkbahar ALES sayısal-1 testlerine doğrusal ve eşit yüzdelikli eşitleme yöntemleri kullanılarak eşitlenmesi sonucu, eşit yüzdelikli eşitlemeye ait hesaplanan WMSE değeri daha küçük olduğu için bu testlerin eşitlenmesinde eşit yüzdelikli eşitlemenin daha uygun olduğu görülmüştür. Bu sonucu, Kelecioğlu (1994) ve Şahhüseyinoğlu (2005)'nin yaptıkları çalışmalar da desteklemektedir.

Öneriler

Bu çalışmanın sonuçlarına dayalı olarak, farklı dönemlerde ve yıllarda elde edilen ALES puanlarının birbiri yerine kullanılabilmesi için test puanlarının eşitlenmesi önerilebilir. Aynı zamanda, ÖSYM tarafından merkezi olarak yapılan ve geçerliği birkaç yıl olan ÜDS, KPDS ve KPSS gibi sınavlardan elde edilen puanlar eşitlenebilir. Bu araştırma kapsamında, testler Klasik Test Teorisi'ne dayalı yöntemler kullanılarak eşitlenmiştir, benzer bir araştırma Madde Tepki Kuramı'na dayalı eşitleme yöntemleri kullanılarak yapılabilir. Düzgünleştirme yöntemlerinin test eşitlemeye etkisini görmek amacıyla eşitleme yapılmadan önce ve sonra farklı düzgünleştirme yöntemleri uygulanarak elde edilen eşitlenmiş puanlar karşılaştırılabilir.

Kaynakça

- Akhun, İ. (1984). İki Korelasyon Katsayısı Arasındaki Farkın Manidarlığının 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.
- 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 Educational Measurement*, 37, 281-306.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Baston: Kluwer-Nijhoff Publisihing
- Harris, D. J. (1991). A comprison of Angoff's design I and Design II for vertical equating using traditional and IRT methodology. *Journal of Educational Measurement*, 28(3), 221-235
- Harris, D. J. (2003). Equating the multistate bar examination. *The Bar Examiner*, 72(3), 12-18.

AKADEMİK PERSONEL VE LISANSÜSTÜ EĞİTİMİ GİRİŞ SINAVI PUANLARININ EŞİTLENMESİ ÜZERİNE BİR ÇALIŞMA

193

- Harris, D. J. & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education*, 6 (3), 195-240.
- Holland, P.W & Thayer, D. T. (1985). Section pre-equating in the presence of practice effects. *Journal of Educational Statistics*, 10(2). 109-120.
- Jaeger, R. M. (1981). Some expository indices for selection of a test equating method. *Journal of Educational Measurement*, 18(1), 23-38.
- Kan, A. (2010). Test Eşitleme: Aynı Davranışları Ölçen, Farklı Madde Formlarına Sahip Testlerin İstatistiksel Eşitliğinin Sınanması. Eğitimde ve Psikolojide Ölçme ve Değerlendirme II. Ulusal Kongresi'nde bildiri olarak sunulmuştur, Mersin.
- Karasar, N. (2009). *Bilimsel Araştırma Yöntemi* (19. Baskı). Ankara: Nobel Yayınları.
- Kelecioğlu, H. (1994). "Öğrenci Seçme Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma." Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29-36.
- Kolen, M. J. (2004). Population invariance in equating and linking: Concept and history. *Journal of Educational Measurement*, 41(1), 3-14.
- 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.
- Moses, T. & Holland, P. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (ETS Research Rep. No. RR-07-15). Princeton, NJ: ETS.
- Livingston, S. A. (2004). *Equating test scores (Without IRT)*. Educational Testing Service.
- Lord, F. M. (1980). *Applications of Item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- ÖSYM (2008). *Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) İlkbahar Dönemi Kılavuzu*. Ankara: ÖSYM.
- Şahhüseyinoğlu, D. (2005). "İngilizce Yeterlik Sınavı Puanlarının Üç Farklı Eşitleme Yöntemine Göre Karşılaştırılması." Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- 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.