

Test Eşitleme: OKS Testlerinin İstatistiksel Eşitliğinin Sınanması

Test Equating: Checking Statistical Equivalence of OKS Test Edition

Adnan KAN*

Gazi Üniversitesi

Öz

Bu araştırma ile farklı yıllarda uygulanan OKS test formlarının istatistiksel eşitliğini sınamak amaçlanmıştır. Çalışma çeşitli ilköğretim kurumlarında öğrenim gören 1030 sekizinci sınıf öğrencisi üzerinde yürütülmüştür. Bu amaç doğrultusunda klasik eşitleme metotlarından lineer eşitleme ve tek grup düzeneği kullanılmıştır. Aynı zamanda eşitleme işleminin bir parçası olarak tek grup düzeneği için eşitleme hatası (SHE) kestirilmiştir. İki testin eşitliğini değerlendirmek için eşitlemenin standart hatasına dayalı güven aralıkları kullanılmıştır. Sonuç olarak, eşitlik fonksiyonu ve özdeşlik fonksiyonu arasında $-0,14$ ile $0,833$ arasında değişen farklılıklar olduğu saptanmıştır. Bu farklılıkların bazı puan ranjlarında 2 SHE'den fazla olması sebebiyle 2003 ve 2005 yıllarında uygulanan OKS test versiyonlarının istatistiksel olarak birbirine eşit olmadığı ve birbirinin yerine kullanılamayacağı saptanmıştır.

Anahtar Sözcükler: Test eşitleme, doğrusal eşitleme, tek grup deseni, test yansızlığı, geçerlik.

Abstract

In this study, it was aimed to check the statistical equivalence of OKS test edition applied in different years. The data of the study was collected from 1030 8th grade students from various secondary schools. Single group design and linear equating procedure from classical equating methodologies were used in the study. At the same time, as a part of the equating process, SEEs (Standart error of equating) for single group were estimated. Confidence bands based on the SEE were used to assess equivalence of OKS test edition. As a result of the study, differences were found between equating function and identity function, differences ranging from $-0,14$ to $0,833$. Because these differences are more than two SEEs for some score range, the two OKS test edition applied 2003 and 2005 can not be considered equivalent and interchangeable.

Keywords: Test equating, linear equating, single group design, test fairness, validity

Summary

Purpose

This study was conducted to check the statistical equivalence of OKS test edition applied in different years. At the same time, the use of test equating procedure was proposed as a framework for examining whether examining score are not unfairly affected by the particular test formats relative difficulty as compared to other forms administered different test dates.

Method

Participants were 1030 students in grade 8 who were enrolled in various secondary school. Two version of OKS Turkish subtest was used as a data collection instruments. Single group design and linear equating procedure from classical equating methodologies were used in the study. At the same time, as a part of equating process, standart error of equating (SEE) were estimated for single group design. Confidence bands based on SEE were used to assess equivalence of OKS test edition.

* Doç. Dr. Adnan KAN, Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, adnankan@gazi.edu.tr

Results

In order to determine whether the two forms of OKS administered different test dates could be treated as interchangeable for the purpose to identify whether the nearly identical OKS test editions administered different test dates create advantage or disadvantage on examinees performance. In other words, whether the examinees scores are fairly affected by particular OKS test forms administered different test dates. First of all it was checked whether equating conditions are fulfilled. It was checked whether the two different version of OKS to be equated measure same construct or not, by combining the individual-item matrix of two test editions and conducting principal component analysis. It was found that both test measure the same construct. Another equation condition is that two of OKS test editions are unidimensional. In order to identify that condition factor analysis were separately conducted on test scores. As a result of this analysis it was found that both test are unidimensional. The correlation coefficient was calculated between 2003 and 2005 OKS test score and high and significant correlation was found between two test score. At the same time, descriptive statistics and reliability of test scores was calculated. According to descriptive statistics two of test score distribution are nearly identical. The mean score, mean difficulties, variances and reliabilities of two OKS test editions were compared by using t, z and F test procedure. As a result of these analysis no statistically differences have been found at 0,05 significance level. Linear equating function was used to equate raw scores on 2005 OKS form to 2003 OKS form. As a part of equating process, SEEs (standart error of equating) were estimated for single group design. As a result of this analysis, some important differences has been found between the equating function and the identity function. Differences are ranging from -0,14 to 0,833 raw score points. These differences follow a monotonically increasing trend, with largest differences occurring at the top of the range and the smallest differences occurring at the bottom of the range. Some of these differences are more than two SEEs, two version of OKS test edition administered different test dates can not be considered interchangeably.

Conclusion and Recommendation

Large testing programs are held to a high degree of scrutiny in their testing schedule. Testing programs (such as OKS, ÖSS, ÜDS) must consider various legal, psychometric and practical issues. One such a issue is the necessity of creating alternate forms of test. Though the trait tested every year remain the same, different test forms applied on individuals to ensure test security. With this research it was aimed to check statistical equivalence of OKS test edition applied different. years It is not legally offensive to administer different forms of the same test to different examinees unless the tests can be shown to be equivalent. Because differences between equating function and identity function are more than two SEEs for some score range, the two OKS test edition administered 2003 and 2005 can not be considered equivalent and interchangeable. Test assemblers would have been alerted to the fact that alternate forms affects test properties and this condition may need to be taken in to account when assembling a test. In Turkey, in order to select students and staff for various institutions, tests such as Student Selection Examination (ÖSS), High Schools Student Selection and Placement Examination (OKS), Public Personnel Selection Examination (KPSS) etc., are applied every year. It is very important and legal issue that examinees scores are not unfairly affected by the particular test formats relative difficulty as compared to other forms administered different test dates. In order to carry on the study to a further stage, the study maybe repeated by using different equating designs on different exams administered in Turkey or may be repeated for tests with more items and in bigger groups.

Giriş

Ülkemizde her yıl SBS, OKS, ÖSS gibi yılda bir kez uygulanan veya KPDS, ÜDS gibi her yıl birkaç kez uygulanan sınavlar vardır. Bu sınavlara her yıl yüz binlerce öğrenci girmekte ve hayatlarını bu sınavlara göre şekillendirmektedir. Şimdi A ve B gibi iki öğrenci olduğunu ve bu iki öğrencinin de farklı dönemlerde ÜDS'ye girdiğini, A öğrencisinin bu testin güz dönemindeki versiyonunu aldığını, B öğrencisinin ise bahar dönemi versiyonunu aldığını varsayalım. Bu aynı özelliği ölçen fakat farklı zamanlarda uygulanan ve farklı sorulardan oluşan sınavlardan A öğrencisi Ax, B öğrencisi Bx gibi bir puan almış olsun. Bu bir testin farklı zamanlarda uygulanmış versiyonlarından elde edilen puanlar birbiriyle karşılaştırılabilir mi ? Testin bir başka versiyonunu aldığı için B öğrencisinin A öğrencisine göre veya A öğrencisinin B öğrencisine göre daha avantajlı olmadığını güvenle söyleyebilir miyiz ? Bir başka ifade ile farklı zamanlarda uygulanan bu testlerin, testi alan bireylere adil davrandığını iddia edebilir miyiz ? Bu gibi sebeplerden dolayı yukarıda belirtilen birçok test programı her bir oturumda belli özellikleri ölçen bir testin farklı sorulardan oluşan ama mümkün olduğunca aynı güçlük düzeyine ve kapsama sahip birçok versiyonunu kullanır. Test geliştiricilerin veya testi uygulayan kurumların istatistiksel açıdan ve kapsam açısından aynı veya birbirine benzer testler oluşturma çabalarına rağmen her bir oturumda farklı test versiyonları ve farklı soru setleri kullanıldığı sürece, özellikle testlerin güçlük düzeylerinde bir miktar farklılıklar olacaktır (Tanguma, 2000). Geniş ölçekli testlerde testin uygulaması sırasında (test uygulanırken) yüksek düzeyde güvenlik ve gizlilik önlemlerinin alınması gerekir. Bu tür geniş test programları çeşitli kanuni, psikometrik ve pratik konuları göz önünde tutmalıdır. Bu konulardan birisi bir testin, aynı özelliği ölçen farklı formlarını oluşturmaktır. İki farklı oturumda aynı testi farklı ya da aynı bireylere uygulamak pek duyulmuş bir şey değildir. Böyle bir uygulama, testi sonra alanların, önce alanlara göre kesin bir şekilde daha avantajlı olmasını sağlayacak ve test güvenliğini tehdit edecektir ve doğal karşılanamaz. Fakat iki testin psikometrik açıdan birbirine eşit (eşdeğer) olduğu kanıtlandığı sürece, aynı testin farklı versiyonlarının, farklı bireylere uygulanabilmesi kanuni olarak savunulabilir. Bütün bunların yanında test geliştiricilerin veya kurumların tamamıyla birbirine paralel, fakat farklı sorulardan oluşan ve testi alan her bir birey için aynı veya benzer sonuçlar üreten testler yapılandırılmaları oldukça zordur ve beklenemez.

Psikometrik olarak ideal olan, bir testin farklı formlarının tamamıyla paralel olması, testi alan tüm grupların random olarak seçilmiş olması, eşit yetenek düzeyine sahip olması ve etik ve kanuni konuların test programını sınırlamamasıdır. Bütün test programlarında karşılaşılabilecek bu ve benzeri konular ve sorunlar, test eşitleme olarak bilinen süreçlerden yararlanmayı gerekli kılar (Woldbeck, 1988). Test eşitleme metotları, farklı test formlarından elde edilen puanları birbiriyle karşılaştırılabilir hale getirmek için kullanılır (Davier, Holland & Thayer, 2004 ; Holland, Sinharay, Davier, & Han, 2008; Harris, 2003).

Test Eşitleme

Angoff (1971), test eşitlemeyi bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlar. Angof (1982)'a göre eğer X ve Y gibi birbirine paralel ve eşit güvenilirlik düzeyine sahip iki farklı testten elde edilen standart puanlar birbirine eşitse, bu formlardan elde edilen iki puanın birbirine eşit olduğu *iddia edilebilir. Bir başka ifade ile bir testin iki farklı formundan elde edilen iki puan ;*

$$\frac{X - \mu_X}{S_X} = \frac{Y - \mu_Y}{S_Y} \quad (1)$$

koşulunu sağlıyorsa birbirine eşittir. Y eşitlikten çekilerek, gerekli matematiksel düzenlemeler yapılırsa ;

$$Y = \frac{S_Y}{S_X} X + \mu_Y - \frac{S_Y}{S_X} \mu_X \quad (2)$$

denklemini elde edilir.

Formülde yer alan $\frac{S_Y}{S_X}$ bileşenine "a" ve $\mu_Y - \frac{S_Y}{S_X} \mu_X$ bileşenine "b" denirse,

Eğimi "a" y eksenini kesme noktası (sabit) "b" olan $Y = ax + b$ şeklinde bir doğru denklemi elde edilir. Bu aynı zamanda doğrusal eşitlemenin de temel denklemdir. Doğrusal eşitleme, paralel olarak yapılandırılan iki test formuna ait ham puan dağılımlarının özdeş, fakat eşitlenecek test puanlarının sadece ortalama ve standart sapmalarının farklı olduğunu varsayar (Lord & Novick 1968).

Test eşitlemenin, farklı test formlarını alan bireyler arasında yanlılığı önlemek, farklı formlardan alınan puanları aynı ölçek üzerinde rapor etmek ve rapor edilen puanların anlamını korumak gibi iki önemli amacı vardır (Barnard, 1996). Farklı formlardan elde edilen puanların eşitlenmesi ile bireylerin gelişimini ölçmek, eğilimlerini belirlemek ve performanslarını karşılaştırmak mümkün olabilir. Üniversite ve kolej giriş sınavları gibi seçme gerektiren ve testin farklı formlarını içeren sınavlara ilişkin ölçme sonuçlarının eşitlenmesi ve eşitleme sonuçları, sınavın geçerliği açısından önemlidir. Çünkü başvuruların hangi formu aldıkları (kolay mı, zor mu) ve herhangi bir gruba avantaj sağlayıp sağlamadığı göz önünde tutulması gereken önemli bir konudur (Angoff, 1971). Testler birbirine eşitlenirken öncelikle eşitleme için verilerin hangi yöntemle toplanacağı (eşitleme düzeneği) ve eşitleme için hangi istatistiksel yöntemin kullanılacağına karar verilmesi gerekir (Thorndike, 1982).

Eşitleme Düzenekleri

Eşitleme düzenekleri hangi tür bilgilere ihtiyaç duyulduğu, iki puan dağılımının nasıl olduğu, grupların hangi formu alması gerektiği ve hangi yöntemin daha ekonomik olduğu göz önünde bulundurularak kurulur. Düzenek kurmanın 3 yolu vardır (Livingston, 2004; Bozdağ ve Kan, 2010). Bu yollar:

1. Bütün formlar aynı test grubuna uygulanır.
2. İki form, ölçülen özellik bakımından aynı olduğu düşünülen iki gruba uygulanır.
3. Gruplara, ölçülmek istenen özelliği ölçen, fakat bu formlardan farklı olan bir test (ankor) uygulanır. Bu üç yol, iki formun eşitlenmesi için beş farklı düzenek kurulmasını gerektirir. Bu düzeneklerin hangisinin daha iyi sonuç verdiğini söylemek güçtür, her birinin bir diğerine göre avantajları ya da dezavantajları vardır.

Ankor Kullanmadan Yapılan Eşitleme Düzenekleri

Ankor kullanmadan yapılan eşitleme; 'tek grup düzeneği', 'eşdeğer grup düzeneği' ve 'dengelenmiş grup düzeneği' olmak üzere üçe ayrılır. Bu düzenekler kullanıldığında eşitleme ilişkisinin evrene genellenebileceği varsayılır (Livingston, 2004).

Tek Grup Düzeneği

Tek grup düzeneğinde, birbirine eşitlenecek olan iki veya daha fazla test aynı gruba uygulanır. Uygulaması oldukça basit ve pratik bir yöntemdir. Aynı grubun her iki formu da alması, bu düzeneği güçlü kılan bir özelliktir ve bu sebeple bu düzenek aracılığıyla elde edilen eşitlemenin kesinliği diğer düzeneklere göre daha yüksek, eşitleme hatası ise daha düşüktür (Livingston, 2004; Felan 2002; Zhu, 1998; Dorans, 1990; Crocker ve Algina 1986). Ayrıca Livingstone (2004) bu düzeneğin kullanılması durumunda, testi alan grubun hedef evreni temsil etmesine gerek olmadığını ve testi alan gruptan elde edilen eşitleme ilişkisinin hedef evrene genellenebileceğini belirtmektedir. Bu da bu düzeneği güçlü kılan önemli bir özellik olarak göze çarpmaktadır.

Dengelenmiş Grup Düzenegi

Eşitlenmek istenen test formları, gerçekten iki farklı formda, bir başka ifadeyle, aynı testin farklı versiyonları değilse, sıra etkisi problemi tek grup düzeneginin uygulanmasını elverişsiz kılar (Livingston, 2004). Bu problemin üstesinden gelebilmek için dengelenmiş grup düzenegine başvurulur. Grup ikiye ayrılır, her iki grup eşitlenmek istenen iki formu da alır. Birinci grup ilk önce I. formu, daha sonra II. formu alırken; ikinci grup ilk olarak II. formu daha sonra I. formu alır. Livingston (2004), her iki formun alınması için aradan geçen sürenin doğru ayarlanması gerektiğine işaret eder. Yeterli zaman aralığı, aradan geçen süre içinde grupların bilgi ve yetenek düzeylerinin değişmemesini sağlayan süre anlamındadır. Bu düzenek için testleri alan grupların, testin ölçtüğü özellik bakımından birbirine benzer olması gerekir. Dengelenmiş grup düzeneginde, tek grup düzeneginde olduğu gibi formları alan grubun hedef evreni temsil etmesine gerek olmadığı ve örneklemden elde edilen eşitleme ilişkisinin evrene genellenebileceği belirtilmektedir (Livingston, 2004). Dengelenmiş grup düzeneginde, tek grup düzeneginde olduğu gibi küçük örneklem ile eşitleme kesinliği sağlanabilir. Bu düzenegin dezavantajı ise, daha fazla zaman ve özel eşitleme çalışmaları gerektirdiği için uygulamasının ve iki grubun gerçekten aynı yetenek grubundan olmasının zor olmasıdır.

Eşdeğer Grup Düzenegi

Birçok eşitleme çalışmasında aynı grubun her iki test formunu alması oldukça zordur. Bu durumda grubun testin ölçtüğü özellik bakımından, aynı yetenek ve bilgi düzeyinde olacak şekilde ikiye ayrılması gerekir. Her iki gruba farklı test formları verilir. Eşdeğer grup düzeneginde önemli olan, grupların olabildiğince ölçülen özellik açısından benzer nitelikte olmasıdır (Tanguma, 2000). Eşitleme kesinliğinin sağlanabilmesi için grubun yeterince büyük ve heterojen olması gerekir. Livingston (2004), grubu bilgi ve yetenek düzeyi yönünden ikiye ayırmanın en iyi yolunun paketleme (spiralizing the books) yöntemi olduğunu belirtir. Bu yöntemle, ölçülen özellik açısından gruplar arası farklılıklar kontrol altına alınmaya çalışılır. Eşdeğer grup düzeneginin sayılı, gözlenen eşitleme ilişkisinin hedef evrene genellenebileceğidir (Livingston, 2004). Bu düzenegin avantajlarından biri de pratiklik ve yorgunluk etkisini ortadan kaldırmasıdır. Uygulaması oldukça basittir ve geniş uygulama alanına sahiptir.

Ankor Kullanılarak Yapılan Eşitleme Düzenekleri

Pratikte test eşitleme uygulamasında kullanılacak iki grubun seçkisiz olarak seçilmesi oldukça zordur (Tanguma, 2000). Farklı zamanlarda, farklı formları alan bireyler eşdeğer olmayabilir. İki grubun seçkisiz olarak seçilememesi durumunda, gruplar arasındaki farklılıkları kontrol edebilmek için her iki gruba uygulanan testlerden farklı sorulara sahip fakat aynı özelliği ölçen ortak test uygulanır. Bu ortak teste "ankor (anchor)" test denir. Ankor test, gruplara her iki formdan farklı bir form olarak verilebileceği gibi testin bir parçası olarak da verilebilir. Bu durumda, her iki grup farklı iki formla birlikte ankor test alır. Eşitleme için en iyi ankor test, testlerin ölçtüğü niteliği ölçen ankor testtir. Ankor madde kullanılarak yapılan testlerde eşitlemenin amacına ulaşabilmesi için, ankor testin testlerle olan korelasyonunun yüksek ve iki formun uygulandığı gruplara ait puan dağılımlarının birbirine benzer olması gerekir. Testin ve soruların korunması amacıyla, aynı testin farklı formlarının kullanılması gereken durumlarda ankor düzenek önerilir (Tsai, 1997). Ankor düzenek, 'iç ankor' ve 'dış ankor' olmak üzere ikiye ayrılır.

İç Ankor Düzenek

Büyük ve kapsamlı test uygulamalarında, birçok kişi farklı zamanlarda farklı formları alır. Bu durumda grupların, hem X hem de Y formunu almasına imkân yoktur. Farklı iki grubun, farklı iki formu alması durumunda bu formlardan alınan puanlar arasındaki bağlantı ankor maddeler ile sağlanır. Ankor maddeler, testin ölçtüğü özellik bakımından grupların bilgi ve yetenek düzeyleri arasındaki farklılıkların kontrol edilmesini sağlar. İç ankor düzenekte, iki

form arasında bağlantı kuracak maddeler, formların bir parçası olarak gruplara verilir. İç ankor maddelerden alınan puanlar, toplam test puanına dahil edilir. Ankor maddelerin uzunluğunun testin %20'sinden az olmaması gerekmektedir (Angoff, 1971).

Dış Ankor Düzenek

Eşitlenmek istenen X ve Y formları dışında, ankor maddeler içeren bir U formunun da her iki gruba uygulandığı düzenektir. Ankor test, paralel formların ölçtüğü niteliği ölçmeli ve her iki gruba aynı formatta ve aynı koşullarda uygulanmalıdır. Bu durumun gerçekte sağlanması oldukça zordur. Ankor test puanları, toplam test puanına dahil edilmez, yalnızca eşitleme için kullanılır. Bu düzenek, diğer düzeneklerin uygulanmasında sorun yaşandığında uygulanır. İyi dış ankor test bulmak oldukça zordur. Ankor test, eşitlenmek istenen formların ölçtüğü özelliği ölçmüyorsa eşitleme sağlıklı sonuç vermeyecektir.

İstatistiksel Eşitleme Yöntemleri

1. Klasik Eşitleme Yöntemleri

Doğrusal eşitleme

Eşit yüzdellikli eşitleme

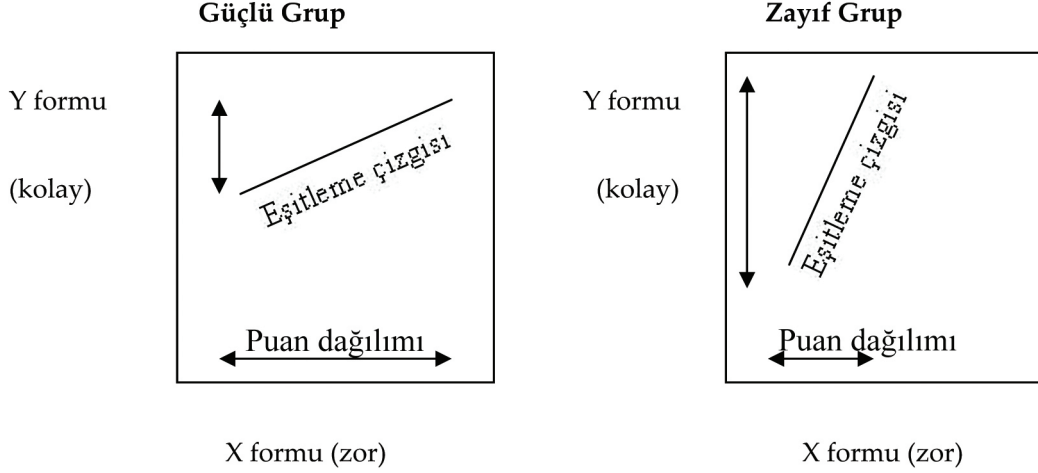
2. Madde Tepki Kuramına (IRT) dayalı Eşitleme Yöntemleri olmak üzere iki kategoride incelenebilir.

Bütün eşitleme yöntemleri, formların güçlük düzeyleri arasındaki farklılıkların giderilmesi için ham puanlar arasında istatistiksel düzeltmeler yapar (Haris, 2003). Bu çalışmada, doğrusal ve eşit yüzdellikli eşitleme yöntemleri üzerinde durulmuştur. Bu yöntemler aşağıda tanımlanmıştır.

Doğrusal Eşitleme

Doğrusal eşitleme yöntemi, farklı test formlarını alan grupların aynı yetenek düzeyinde olması halinde uygulanır. Testin farklı formlarını alan gruplar, aynı yetenek düzeyinde değilse, farklı doğrusal eşitleme yöntemlerinin uygulanması önerilir. Örneğin Ankor Düzenek için standart doğrusal eşitleme uygun değildir. Bu problemin üstesinden gelebilmek için farklı doğrusal eşitleme yöntemleri geliştirilmiştir.

İki testin puan dağılımları aynı ve güvenilirlikleri eşit ise doğrusal eşitleme yönteminin kullanılması uygundur (Thorndike, 1982; Crocker, Algina; 1986; Woldbeck, 1998). Doğrusal eşitleme uygulaması kolay ve pratik bir yöntemdir. Doğrusal eşitleme, uygulaması basit bir yöntem olsa da bazı problemlerle karşılaşılabilir. Livingston (2004) doğrusal eşitlemede yaşanan sorunlara dikkat çeker. X formunda çok yüksek veya çok düşük puanlar varsa, Y formundaki olası puan ranjının dışında kalan eşitlenmiş puanlar elde edilebilir. Örneğin 100'er maddenin bulunduğu iki form doğrusal eşitleme ile eşitlenmek istenildiğinde, zor olan X formundan 99 alan bireyin, kolay olan Y formuna dönüştürme işleminin sonucunda 103 puan alacağı hesaplanabilir. Fakat 100 sorudan 103 puan almanın açıklaması oldukça zordur (Livingston, 2004). Doğrusal eşitlemede yaşanan bir diğer problem, doğrusal eşitleme sonuçlarının gruba bağımlı olmasıdır. Bu durum, diyagram üzerinde Şekil 1'deki gibi açıklanabilir (Livingston, 2004; Bozdağ ve Kan, 2010).



Şekil 1: Doğrusal Eşitlemede Karşılaşılabilecek Problem

Şekil 1'de X formunun zor, Y formunun kolay olduğu ve bu formların ölçülen özellik bakımından zayıf ve güçlü olan iki gruba uygulandığı durum örneklenmektedir (Livingston, 2004; Bozdağ ve Kan, 2010).

Eşit Yüzdelikli Eşitleme

Doğrusal eşitleme için, eşitlenmek istenen iki formdan elde edilen puanların aynı dağılıma sahip olması gerektiği üzerinde durulur. İki form aynı puan dağılımına sahip değil ise eşit yüzdelikli eşitleme yöntemi önerilir (Zhu, 1998). X ve Y formlarından (X ve Y aynı güvenilirliğe sahip olmak üzere) elde edilen puanlar, aynı düzeltilmiş yüzdeler sıraya sahipse eşittir (Angoff, 1971). Bu eşitleme yönteminde Form X'in dağılımı, Form Y'nin dağılımına iki testin yüzdeler sıraları hesaplanarak eşitlenir. İki testin puan dağılımlarının farklı olması durumunda, eşitleme kesinliği, eşit yüzdelikli eşitleme ile sağlanır (Angoff, 1971; Thorndike, 1982). Eşit yüzdelikli eşitleme, doğrusal eşitlemeye kıyasla daha az sayıya sahiptir (Woldbeck, 1998).

Eşit yüzdelikli eşitleme, iki basamakta gerçekleştirilir. İlk olarak her iki dağılıma ait puanların yüzdeler sıraları bulunur. Bu yüzdeler sıralarıyla ham puanlara ait grafik çizilir. İkinci basamakta ise, bu grafikten yararlanılarak elle veya analitik yöntemler ile eşitlenmiş puanlar bulunur.

Eşit yüzdelikli eşitleme için puan dağılımının gergin ve sıkışık (puan aralıklarının daha küçük) olması gereklidir; böylece bir formun puan dağılımı, diğer formun puan dağılımına denk gelir (Angoff, 1971; Woldbeck, 1998). Eşit yüzdelikli eşitleme yönteminin ilk basamağında, X ve Y formundan alınan puanların yüzdeler sıraları bulunur. Bu durum Şekil 2'deki diyagram üzerinde gösterilebilir (Bozdağ ve Kan, 2010).



Şekil 2: Eşit Yüzdelikli Eşitleme Örneği

Şekil 2'de görüldüğü üzere birinci diyagram yüzdelik sıraları gösterirken, 2. diyagram düzeltilmiş yüzdelik sıraları gösterir.

Eşit yüzdelikli eşitleme yönteminin uygulanması karmaşık bir yapıya sahiptir. X formundaki her puan bazen Y formundaki yüzdelik sırasına karşılık gelmeyebilir. Bu nedenle çeşitli öteleme ve düzeltme çalışmaları yapılmaktadır. Düzeltme çalışmaları eşitlemeden önce yapılıyorsa "ön düzeltme (presmoothing)", eşitleme işleminden sonra yapılıyorsa "son düzeltme (postsmoothing)" adı verilir.

Eşit yüzdelikli eşitleme yönteminin en önemli problemi, örneklem yeterince büyük olmadığı zaman puan dağılımının çarpık olmasıdır. Eğer dağılımlar aşırı derecede farklı ise eşit yüzdelikli eşitlemenin kullanılması anlamsız hale gelir. Bu durum özellikle örneklem küçük olduğunda sorun yaratır (Woldbeck, 1998). Her iki dağılıma ait puanlar sıkıştırılmış ve gergin olmalıdır. Böylece ham puan dağılımı, diğer puan dağılımına denk gelir. Grafik yöntemi ile çarpıklık probleminin üstesinden gelinmeye çalışılır. Testi alanların sayısı arttıkça düzeltme işleminin gerekliliği azalır.

Yukarıda anlatılan her iki eşitleme yönteminin birbirine göre avantajları ve dezavantajları vardır. Eşitleme yöntemlerinin seçilmesinde dikkat edilmesi gereken noktalar şunlardır (Crocker ve Algina, 1986):

1. Koşulların sağlanabilirliği,
2. Kullanışlılık,
3. Kesinlik.

Eğer dağılımlar aynı şekle sahip ise, doğrusal ve eşit yüzdelikli eşitleme aynı sonucu verir. Eşit yüzdelikli eşitleme için daha büyük örneklem gerekir ve uygulaması daha karışıktır (Felan, 2002).

Eşit yüzdelikli veya doğrusal eşitlemeyi seçerken hangi veri toplama yönteminin kullanıldığı da önemli bir noktadır. Eşit yüzdelikli eşitleme, doğrusal eşitlemeden daha az varsayıma sahiptir (Crocker, Algina, 1986). Varsayımların az olması avantaj olmasına rağmen, eşit yüzdelikli eşitlemenin iki dezavantajı vardır.

1. Daha karmaşık ve daha az pratiktir.
2. Eşitleme hatası daha büyüktür.

Bu çalışma kapsamında, verilerin toplanması için tek grup deseni, istatistiksel yöntem olarak ise doğrusal eşitleme modeli tercih edilmiştir. Test güvenliğini sağlamak amacıyla bir testin birçok versiyonunun kullanılması gerektiği durumlarda, testlerin kapsam ve güçlük açısından paralel olması beklenir ve gereklidir (Kolen, 1988). OKS test versiyonlarının aynı beceriyi ölçtükleri,

benzer içeriklere sahip oldukları ve dahası farklı yıllarda sınavı alan bireylere eşit davranması gerektiği için eşitleme fonksiyonunun doğrusal olduğunu öngörmek yanlış olmaz (Dorans & Lawrence, 1990). Bunun dışında doğrusal eşitleme modelinin tercih edilmesini gerektirecek başkaca sebepler de vardır. Doğrusal eşitleme modeli daha pratik ve daha küçük eşitleme hatasına sahiptir (Lord, 1982; Crocker ve Algina, 1986). Birbirine paralel olduğu varsayılan ve aynı beceriyi ölçen OKS test formlarının istatistiksel eşitliğini saptamak için bir versiyondan elde edilen puanları diğer bir versiyondan elde edilen puanlara eşitlemek ve özdeşlik fonksiyonunun eşitlik fonksiyonunun etrafında yer alan güven aralığı içinde yer alıp almadığını kontrol etmek suretiyle karara varılabilir (Dorance & Lawrence, 1990).

Tek grup deseni için eşitlemenin standart hatası ($SE_{Y^*}^2$) ise N : toplam birey sayısı, σ_Y^2 : Y puanlarına ait varyans, z_x : Y^* 'ye dönüştürülen X puanlarına ilişkin Z puanları, ρ_X : iki testten elde edilen puanlar arasındaki korelasyon olmak üzere,

$$SE_{Y^*}^2 = \frac{\sigma_Y^2(1 - \rho_X)}{N} [z_x^2(1 + \rho_X) + 2] \quad (3)$$

şeklinde tanımlanabilir.

Operasyonel olarak eşitlik 3'te tanımlanan eşitlemenin standart hatası eşitleme fonksiyonunun etrafında makul bir güven aralığı tanımlamak için de kullanılabilir. Birbirine paralel olduğu varsayılan ve aynı beceriyi ölçen fakat farklı sorulardan oluşan OKS test formlarından elde edilen puanlar için makul bir güven aralığının sınırları her bir ham puan etrafında ± 2 SHE olarak tanımlanabilir (Dorans & Lawrence, 1990). Güven aralığı belirlendikten sonra farklı test versiyonlarından elde edilen puanların istatistiksel eşitliği, özdeşlik fonksiyonunun tanımlanan bu güven aralığı içerisinde yer alıp almadığını belirlemek suretiyle sınırlanabilir.

Yöntem

Çalışma Grubu

Araştırma 22 ilköğretim okulunda öğrenim gören 1030 sekizinci sınıf öğrencisi üzerinde yürütülmüştür.

Veri Toplama Araçları

Araştırmada veri toplama aracı olarak Milli Eğitim Bakanlığı Eğitim Teknolojileri Genel Müdürlüğü Ölçme ve Değerlendirme Dairesi Başkanlığı tarafından 2003 ve 2005 yıllarında düzenlenen ve Türkçeyi kullanma gücünü ölçen 25 adet çoktan seçmeli sorudan oluşan OKS Türkçe alt testleri kullanılmıştır.

Uygulama

Araştırmada kullanılan 2003 ve 2005 OKS Türkçe alt testlerine ait maddeler birleştirilerek 50 maddelik bir test oluşturulmuştur. Oluşturulan 50 maddelik bu test, 2005–2006 eğitim-öğretim yılı içerisinde yer alan çeşitli ilköğretim okullarında öğrenim gören 1030 8. sınıf öğrencisine uygulanmıştır. Birleştirilen test maddeleri, öğrencilere tek oturumda verilmiş, 50 dakika süre tanınmıştır. 2003 OKS Türkçe alt test puanları, 2005 OKS Türkçe alt test puanlarına, tek grup düzeneği kullanılarak eşitlenmiştir.

Verilerin Analizi

Verilerin analizi iki aşamada gerçekleşmiştir. Birinci aşamada eşitleme koşullarının sağlanıp sağlanmadığı kontrol edilmiştir. İkinci aşamada ise tek grup desen ve doğrusal eşitleme yöntemi kullanılarak 2003 Türkçe alt test puanları, 2005 Türkçe alt testi puanlarına eşitlenmiştir. Birinci aşamada eşitleme koşullarının sağlanıp sağlanmadığına ilişkin olarak elde edilen veriler üzerinde

bir dizi analiz yürütülmüştür. Bu kapsamda ilk olarak , 2003 Türkçe alt test puanları, 2005 Türkçe alt testi puanlarının tek boyutlu bir yapıya sahip olup olmadıklarına karar vermek üzere her iki testten elde edilen veriler üzerinde temel bileşenler analizi uygulanmıştır (Gulliksen, 1967 ; Angoff, 1971 ; Woldbeck, 1998 ; Tanguma 2000 ; Felan, 2002). Temel bileşenler analizi testlerden elde edilen madde puanlarının iki kategorili (dichotomous) olması sebebiyle tetrakorik korelasyon matrisi üzerinden yürütülmüştür. Tüm bu analizler Statistica 7.0 paket programı aracılığıyla yapılmıştır. Temel bileşenler analizi sonucunda her iki testin de tek bir faktörü ölçtüğüne ve bu faktörün de Türkçeyi kullanma gücü olduğuna karar verilmiştir. İki testten elde edilen puanlar arasındaki ilişkinin düzeyini belirlemek üzere Pearson momentler çarpımı korelasyon katsayısı hesaplanmıştır (Dorans, 2000). 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; Lord 1955; Jaeger 1981). Ayrıca her iki teste ait aritmetik ortalama, varyans ve ortalama güçlük düzeyleri hesaplanmış ve her iki teste ait bu istatistikler arasında anlamlı bir fark olup olmadığı t ve z testleri aracılığıyla belirlenmiştir.

İkinci aşamada, 2003 OKS Türkçe alt test puanlarını, 2005 OKS Türkçe alt test puanlarına eşitlemek üzere, klasik eşitleme metodlarından doğrusal eşitleme ve tek grup düzeneği kullanılmıştır. Aynı zamanda eşitleme işleminin bir parçası olarak tek grup düzeneği için eşitleme hatası (SHE) kestirilmiştir. İki testin eşitliğini değerlendirmek için eşitlemenin standart hatasına dayalı güven aralıkları kullanılmıştır.

Bulgular ve Yorum

Çalışmada öncelikle, her iki testin aynı yapıyı ölçüp ölçmediğini ve tek faktörlü yapıyı (Türkçeyi kullanma gücü) ölçüp ölçmediğini belirlemek üzere her iki testten elde edilen veriler üzerinde temel bileşenler analizi yürütülmüştür. Temel bileşenler analizi, testlerden elde edilen madde puanlarının iki kategorili (dichotomous) olması sebebiyle tetrakorik korelasyon matrisi üzerinden yürütülmüştür ve bu analize ilişkin sonuçlar Tablo 1'de verilmiştir.

Tablo 1.

2003 ve 2005 OKS Testlerine Ait Temel Bileşenler Analizi Sonuçları

Bileşen	2003 OKS		2005 OKS		BİRLEŞİK (2003 VE 2005 OKS)	
	Özdeğer	V.A.O. (%)	Özdeğer	V.A.O. (%)	Özdeğer	V.A.O. (%)
1	8,46	33,82	8,45	33,80	15,53	31,07
2	1,30	5,25	1,67	6,70	2,05	4,12

Tablo 1'deki verilere göre gerek 2003 ve 2005 gerekse bu iki testi birleştirmek suretiyle elde edilmiş OKS test sonuçları üzerinde yürütülen temel bileşenler analizi sonucunda her iki teste ve birleştirilmiş teste ilişkin özdeğerler ve varyans açıklama oranları (V.A.O.) birbirine çok yakın ve benzer değerlere sahiptir. Bu bulgu, testlerin aynı yapıyı ve özelliği ölçtüğüne ilişkin kanıt olarak kullanılabilir. Aynı şekilde bu özdeğerler ve varyans açıklama oranları 1. faktörden sonra keskin bir düşüş göstermekte ve ilk faktör ile kendisinden sonra gelen ilk faktöre (2. faktör) ait V.A.O ile özdeğerler arasında en az 5-6 kat fark olduğu göze çarpmaktadır. Bu bulgu, testlerin baskın tek faktöre sahip olduğuna ilişkin kanıt olarak kullanılabilir. Hambleton ve Swaminathan (1985) birçok durumda tek boyutluluk varsayımının karşılanabilmesi için testin baskın tek bir faktöre sahip olmasını yeterli görmekte ve bu baskın faktörü testle ölçülmek istenen özellik (trait) olarak tanımlamaktadır.

İkinci aşamada, bireylerin her iki test versiyonundan elde ettikleri puanlara ilişkin betimsel istatistikler hesaplanmış ve Tablo 2'de verilmiştir. Ayrıca her iki testin varyansları, aritmetik

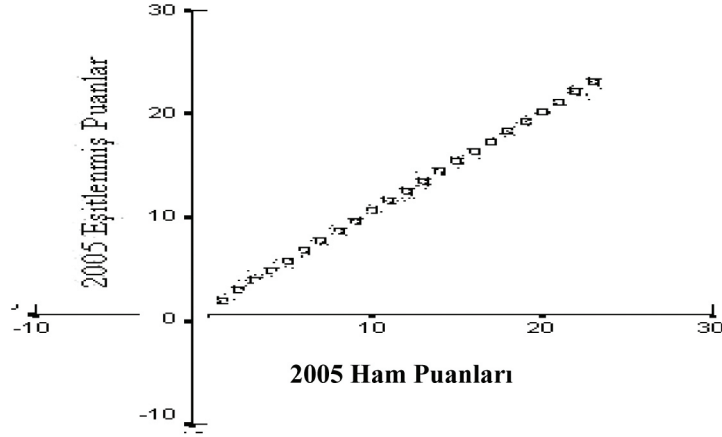
ortalamaları, ortalama güçlük düzeyleri, güvenilirlikleri arasında istatistiksel olarak fark olup olmadığı test edilmiştir.

Tablo 2.

2003 ve 2005 OKS Testlerine Ait Betimsel İstatistikler

Testler	K	N	\bar{X}	S^2_x	S_x	$KR-20$	Ortalama Güçlük	Çarpıklık Katsayıları	Basıklık Katsayıları
Türkçe 2003	25	1030	13.47	32,77	5,72	0,853	0,539	0,051	-0,999
Türkçe 2005	25	1030	13.09	30,27	5,50	0,846	0,524	0.088	-0,947

Tablo 2’de verilen betimsel istatistiklere dayanarak, her iki testten elde edilen puan dağılımlarının birbirine çok benzediği söylenebilir. Ayrıca iki teste ait aritmetik ortalamalar arasında ($t=1,553$; $p>0.05$) ; varyanslar arasında ($F=1,08$; $p>0.05$) ; ortalama güçlük düzeyleri arasında ($t=0,68$; $p>0.05$) ve güvenilirlikleri arasında ($Z=0,569$; $p>0.05$) istatistiksel olarak anlamlı bir fark bulunamamıştır. Tüm bunların dışında 2003 ve 2005 OKS Türkçe alt testlerinden elde edilen puanlar arasında yüksek ve anlamlı bir ilişki ($r=0,77$; $p<0,05$) olduğu bulgulanmıştır. Tüm bu analiz sonuçları, 2003 ve 2005 OKS Türkçe alt testlerinden elde edilen puanlar üzerinde eşitleme işleminin yapılabileceğine ilişkin kanıt olarak kullanılabilir. Tek grup düzeneğine dayalı 2005 OKS Türkçe alt testine ait ham puanlar ve bu puanlara karşılık gelen eşitlenmiş puanlara ilişkin grafik Şekil 1’de sunulmuştur.



Şekil 3 : 2003 ve 2005 OKS Test Formlarından Elde Edilen Puanlara İlişkin Doğrusal Eşitleme Grafiği

Şekil 3’te verilen grafik incelendiğinde, 2005 OKS Türkçe alt testine ait ham puanlar ve bu puanlara karşılık gelen eşitlenmiş puanlar arasında doğrusal bir ilişki olduğu görülebilir ve bu ilişki matematiksel olarak $Y=1.04X-0.147$ denklemi aracılığıyla ifade edilebilir.

2003 ve 2005 OKS Türkçe alt testlerinin istatistiksel açıdan birbirine eşdeğer olup olmadığı, birbirinin yerine kullanılıp kullanılmayacağı bir diğer ifade ile aynı özelliği ölçen ve birbirine paralel olduğu varsayılan bu test formlarının aynı sonuçları üretip üretmeyeceğini (yansızlığı) belirlemek üzere 2003 OKS Türkçe alt testinden elde edilen puanlar, 2005 OKS Türkçe alt test puanlarına eşitlenmek üzere tek grup desenine dayalı doğrusal (linear) eşitleme fonksiyonundan yararlanılmış ve eşitleme prosedürünün önemli bir parçası olan eşitlemenin standart hatası (SHE) kestirilmiştir. Tüm bu sonuçlar Tablo 3’te verilmiştir.

Tablo 3.

2003 ve 2005 OKS Test Formlarından Elde Edilen Puanlara Ait Doğrusal Eşitleme Sonuçları

Form 2 (2005 OKS) Ham Puan	Form 1 Eşit Puan	Fark	Eşitlemenin Standart Hatası (SHE)	Oran (Fark / SHE)
0	0,141	-0,141	0,245	-0,58
1	1,102	-0,102	0,235	-0,43
2	2,063	-0,063	0,224	-0,28
3	3,024	-0,024	0,215	-0,11
4	3,985	0,015	0,205	0,07
5	4,946	0,054	0,196	0,28
6	5,907	0,093	0,188	0,49
7	6,868	0,132	0,181	0,73
8	7,829	0,171	0,174	0,98
9	8,790	0,210	0,168	1,25
10	9,751	0,249	0,164	1,52
11	10,712	0,288	0,160	1,80
12	11,673	0,327	0,158	2,07
13	12,635	0,365	0,156	2,34
14	13,596	0,404	0,156	2,59
15	14,557	0,443	0,158	2,80
16	15,518	0,482	0,160	3,01
17	16,479	0,521	0,164	3,18
18	17,440	0,560	0,169	3,31
19	18,401	0,599	0,174	3,44
20	19,362	0,638	0,181	3,52
21	20,323	0,677	0,189	3,58
22	21,284	0,716	0,197	3,63
23	22,245	0,755	0,206	3,67
24	23,206	0,794	0,215	3,69
25	24,167	0,833	0,225	3,70

Tablo 3'te görülebileceği gibi eşitleme fonksiyonu ile özdeşlik fonksiyonu arasında bir miktar farklılıklar vardır. Bir diğer ifade ile Form 2'den elde edilen ham puanlarla eşitlenmiş puanlar arasında çeşitli puan düzeylerinde farklılıklar vardır. Bu farklılıklar -0,14 ile 0,833 arasında değişmekte ve düşük puanlardan yüksek puanlara doğru monotonik bir artış göstermektedir. Özellikle bu farklılıkların bazıları 2 SHE'den fazla olduğu için tanımlanan güven aralığının dışına düşmektedir. Ayrıca oran sütunundan benzer şekilde bu farklılığın büyüklüğü SHE cinsinden de görülebilir. Puan ranjının ortalarında (12, 13, 14 puan) SHE minimum değerler almasına rağmen farkın büyüklüğünü gösteren oran, düşük puanlardan yüksek puanlara doğru monotonik bir artış göstermektedir. Bunun sebebi, düşük puanlardan yüksek puanlara doğru eşitleme fonksiyonu ile özdeşlik fonksiyonu arasındaki farkın büyümesidir.

Sonuç ve Öneriler

Ülkemizde ve dünyada aynı becerileri ölçen ve her yıl yapılan birçok seçme ve yerleştirme amaçlı test (örneğin ; ÖSS, OKS, SBS, ÜDS, ALES, KPSS, SAT, GRE, TOEFL vb.) vardır. Bu testlerin sonuçları bireyler hakkında çok önemli kararlar almak için kullanılır. Bu sınavların, farklı versiyonlarını farklı zamanlarda alan bireylere eşit davranması, yanlı davranmaması

test geçerliği açısından son derece önemli bir durumdur. Bu derece önemli sınavlarda testin yansızlığına ilişkin kanıtların toplanması ve bu sınavların farklı versiyonlarından elde edilen puanların istatistiksel eşitliğinin sınanması gereklidir. Bu çalışma ile öncelikle bu ve bu türden sorunlara yönelik ve geçerliğe kanıt sağlamak amacıyla test eşitleme ve buna bağlı prosedürlerin nasıl kullanılabileceğini açıklamak hedeflenmiştir. Bu kapsamda, 2003 ve 2005 OKS Türkçe alt testlerinin istatistiksel olarak birbirine eşit olup olmadığı sınanmış ve bu yolla bu testlerin farklı versiyonlarının bu sınavları farklı zamanlarda alan bireylere adil davranıp davranmadığına bu versiyonların birbirinin yerine kullanılıp kullanılamayacağına (test yansızlığı) ilişkin kanıt sağlamak amaçlanmıştır. Sonuç olarak, elde edilen bulgular doğrultusunda, özdeşlik fonksiyonunun tanımlanan ∓ 2 SHE olarak tanımlanan güven aralığı içerisinde yer almaması sebebi ile 2003 ve 2005 OKS Türkçe alt testlerinin birbirine eşdeğer olduğunu veya birbirinin yerine kullanılabileceğini söyleyebilmek güçtür. Bir diğer ifade ile Form 2'den elde edilen ham puanlarla eşitlenmiş puanlar arasında bazı puan düzeylerinde farklılıklar ∓ 2 SHE olarak tanımlanan güven aralığı içerisinde yer almamaktadır. Tablo 2 dikkatle incelenecek olursa, Form 2'den elde edilen ham puanlarla eşitlenmiş puanlar arasında çeşitli puan düzeylerinde farklılıklar, 2 SHE'den daha küçük olduğu için 10 ve altında puan alanlar için testin hangi versiyonunu alırsa alsın benzer sonuçlar elde ederken, yani 2003 yada 2005 OKS testlerini almaları bu öğrencilerin performanslarında farklılık yaratmazken, 10 puan ve üstünde alanlar için durum tam tersine dönmekte ve dramatik bir hal almaktadır. Bir diğer ifade ile 2003 ya da 2005 OKS testlerini almaları bu öğrencilerin performanslarında farklılık yaratmaktadır. Bu durumda birbirine paralel olduğu varsayılan 2003 ve 2005 OKS Türkçe alt testlerinin birbirine paralel olduğunu ve yansız olduğunu bu çalışmanın sonuçları doğrultusunda söylemek mümkün gözükmemektedir.

Bu çalışmanın sonuçları doğrultusunda ülkemizde yapılan birçok seçme amaçlı sınavın farklı versiyonlarının benzer sonuçlar üretilip üretilmediğinin kontrol edilmesi gerektiği söylenebilir. Çünkü bir seçme testinin onu farklı yıllarda alan bireylere eşit ve adil davranması gereklidir. Aksi takdirde test, testi önce veya sonra alan yanlı davranacaktır. Bu yasal olmayan bir durumdur. Bundan sonra bu konuda çalışacak araştırmacılara farklı eşitleme düzenekleri (dengelenmiş grup, eşdeğer grup, iç anchor, dış anchor vb.) ile daha büyük gruplarda ve farklı sınavlar üzerinde (KPSS, ÜDS, KPDS vb.) çalışarak aynı amaca yönelik farklı test versiyonlarının benzer sonuçlar üretilip üretilmediğini kontrol etmeleri önerilebilir. Kolen ve Brennan (1995), test eşitleme süreçlerinin test geliştirmede çok önemli olduğunu ve tüm test geliştiren kişi ya da kurumların bu süreçlere aşina olmaları ve bunu uygulayabilmeleri gerektiğini belirtmektedir. Ayrıca Kolen ve Brennan (1995) ve Felan (2002) test eşitleme süreçlerinin, test üzerinde etkili olan faktörleri ve test yansızlığını ortaya çıkarmak ve önlemek için iyi bir yol olduğunu belirtmektedir. Zira ülkemizde çok önemli birçok sınav uygulanırken bu bir gerekliliktir. Birçok insanın hayatını etkileyen bu sınavların farklı yıllarda alanlara eşit davranmaması önemli bir sorundur ve daha büyük ölçekli çalışmalarla denetlenmesi gerekir. Bu çalışmanın bu yolda bir adım olacağı düşünülmektedir.

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.
- Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. In P.W. Holland & D. B. Rubin (Eds.). *Test Equating*. New York: Academic Press.
- Barnard, J. J. (1996). "In search for equity in educational measurement: Traditional versus modern equating methods." *Paper presented at ASEESA's national conference at the HRSC Conference Centre, Pretoria, South Africa*.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. San Diego, C.A: Harcourt Brace Jovanovich College Publisher.
- Davies, A. A., Holland, P. W. & Thayer, D.T. (2004). The chain and post stratification methods

- for observed-score equating: Their relationship to population invariance. *Journal of Educational Measurement*, 41(1), 15-32.
- Dorans, N. J. (1990). Equating methods and sampling design. *Applied Measurement In Education*, 3(1), 3-17.
- Dorans, N. J. (2000). Research notes: Distinctions among classes of linkages. The College Board, Office of Research and Development.
- Dorans, N. J., & Lawrence, I., M. (1990). Checking the statistical equivalence of nearly identical test editions. *Applied Measurement In Education*, 3(3), 245-254.
- Felan, G. D. (2002). Test equating: Mean, linear, equipercentile and item response theory. *Paper presented at the annual meeting of the Southwest Educational Research Association*. Austin, TX.
- Gulliksen, H. (1967). *Theory of mental tests*. New York: John Wiley & Sons Inc.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, Kluwer Academic Publishers Group.
- Harris, D. J. (2003). Equating the multistate bar examination. *The Bar Examiner*, 72(3), 12-16.
- Holland, P. W., Sinharay, S., Davier, A. A., & Han, N. (2008). An approach to evaluating the missing data assumptions of the chain and post-stratification equating methods for the NEAT design. *Journal of Educational Measurement*, 45(1), 17-43.
- Jaeger, M. R. (1981). Some exploratory indices for selection of a test equating method. *Journal of Educational Measurement*, 18(1), 23-38.
- Bozdağ, S., ve Kan, A. (2010). Şans Başarısının Test Eşitlemeye etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39, 91-108.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practices*, 7, 29-36
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer.
- Livingstone, S. A. (2004). *Equating test scores* (Without IRT). Educational Testing Service.
- Lord, F. M. (1955). Equating test scores- a maximum likelihood solution. *Psychometrika*, 20, 193-200.
- Lord, F. M. (1982). The standart error of equipercentile equating. *Journal of Educational Statistics*, 1, 165-192.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, M.A: Addison-Wesley.
- Tanguma, J. (2000). "Equating test scores using the linear method: A primer." *Paper presented at the annual meeting of the Southwest Educational Research Association*. Dallas, TX.
- Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin Company.
- Wooldbeck, T. (1998). "Basic concept in modern methods of test equating." *Paper presented at the annual meeting of the Southwest Psychological Association New Orleans, L. A.*
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23.