



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

ŞANS BAŞARISININ TEST EŞİTLEMeye ETKİSİNİN FARKLI EŞİTLEMeye
TEKNİKLERİYLE ARAŞTIRILMASI

Emine AYTEKİN KAZANÇ

Yüksek Lisans Tezi

Ankara, 2019

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

ŞANS BAŞARISININ TEST EŞİTLEMeye ETKİSİNİN FARKLI EŞİTLEMeye
TEKNİKLERİYLE ARAŞTIRILMASI

INVESTIGATION OF THE EFFECT OF GUESSING ON TEST EQUATING WITH
DIFFERENT EQUATING METHODS

Emine AYTEKİN KAZANÇ

Yüksek Lisans

Ankara, 2019

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Emine AYTEKİN'in hazırladıđı "Şans Başarısının Test Eşitlemeye Etkisinin Farklı Eşitleme Teknikleriyle Araştırılması" başlıklı bu çalışma j¼rimiz tarafından Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

J¼ri Başkanı

Prof. Dr. Bayram ÇETİN

İmza

J¼ri Üyesi (Danışman)

Prof. Dr. Selahattin GELBAL

İmza

J¼ri Üyesi

Prof. Dr. Nuri DOĐAN

İmza

Enstit¼ Y¼netim Kurulunun
.../.../... Tarihli ve
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 10. / 07. / 2013 tarihinde uygun gör¼lm¼ş ve Enstit¼ Y¼netim Kurulunca / / tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

Öz

Bu araştırmanın amacı, şans başarısının test eşitlemeye etkisinin olup olmadığını incelemektir. Şanstan arındırılmış veya arındırılmamış ham puanların eşitlenmesinde kullanılan yöntemlerden hangisinin daha az eşitleme hatası içerdiğini bulmak amaçlanmıştır. Araştırmada, tek grup düzeneği ve klasik eşitleme yöntemlerinden, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılmıştır. Araştırmada, 2015-2016 eğitim öğretim yılında yapılan Türkiye'den belirli okullarının katıldığı (22 okul) TEOG Değerlendirme sınavı İngilizce I ve II testlerine cevap veren 8. Sınıf ortaokul öğrencileri (N=1681) bu çalışmanın örnekleimidir. Verilerin analizi üç aşamada gerçekleştirilmiştir. Birinci aşamada, İngilizce I ve II genel değerlendirme alt testlerin eşitleme koşullarına uygunluğu için betimsel istatistikleri hesaplanmıştır. İkinci aşamada, şanstan arındırılmayan ve arındırılan puanlara doğrusal ve eşit yüzdelikli eşitleme yöntemi uygulandıktan sonra eşitlenmiş puanlar bulunmuştur. Son olarak şanstan arındırılmamış ve arındırılmış puanların eşitlenmesinde kullanılan yöntemlere ait Ağırlıklandırılmış Hata Kareleri Ortalaması (WMSE), Kareler Farkının Ortalamasının Kare Kökü (RMSD) ve İşaretli Farkların Ortalaması (MSD) değerlerine bakılmıştır. Araştırmada elde edilen bulgulara göre, şanstan arındırılmamış İngilizce test puanlarının eşitlenmesinde eşit yüzdelikli eşitlemenin; şanstan arındırılmış İngilizce alt testlerinin eşitlenmesinde ise doğrusal eşitlemenin uygun yöntem olduğu bulunmuştur.

Anahtar kelimeler: test eşitleme, doğrusal eşitleme, eşit yüzdelikli eşitleme, tek grup düzeneği, şans başarısı

Abstract

The aim of the study is to investigate whether guessing is a factor or not to equate English tests by using linear and equipercentile equating with single group design. In addition, it was aimed to equate the scores of English Assessment tests by using these traditional methods and determine which of those is proper. The data were gathered from 1681 8th grade students in various public schools in Turkey. For the data analysis process, at first descriptive statistical analysis (mean, difficulty, variance, skewness, kurtosis and reliability) was utilized to check the assumptions of equation of the tests scores. For the second phase, the equated points for scores with guessing and without guessing were given by using the methods of linear and equipercentile-equating. At the final phase, the weighted mean square error (WMSE), the root mean squared difference (RMSD) and the mean signed difference (MSD) were calculated to identify the more suitable method for equating scores. Based on the results, equipercentile-equating method yielded less weighted mean square error, root mean squared difference and mean signed difference than linear equating method when English tests were equated. At last, after elimination of the guessing factor from scores, linear equating method showed less weighted mean square error root mean squared difference and mean signed difference than the equipercentile-equating.

Key Words: test equating, linear equating, equipercentile equating, single group design, guessing, guessing

Teşekkür

Tezimin bitmesi için bilgi ve tecrübelerini paylaşan, her türlü sorularımı içtenlikle cevaplayan, karamsarlığa düştüğümde beni cesaretlendiren, zamanını bana ayıran, yönlendiren danışmanım ve sevgili hocam Prof. Dr. Selahattin Gelbal'a,

Tezimde kullandığım değişkenleri göz önünde bulundurarak önerilerde bulunan, bilgisini esirgemeyen değerli hocam Prof. Dr. Nuri Doğan'a ve geri bildirimleri ile çalışmama katkı sağlayan hocam Prof. Dr. Bayram Çetin'e,

Bu uzun tez yazım sürecinde verilerin analiz sürecinde bilgi ve tecrübesinden yararlandığım, emeğini, zamanını ve hoşgörüsünü esirgemeyen Dr. Öğretim Üyesi Çiğdem Akın Arıkan'a; verilerini benimle paylaşan Araş. Gör. Sinan Yavuz'a ve en son olarak maddi veya manevi her türlü desteklerini esirgemeyen arkadaşlarım Araş. Gör. Meltem Yurtçu, Başak Erdem Kara, Bulut Yıldıztekin ve Mine Zorlu'ya,

Yüksek lisans eğitimim boyunca beni cesaretlendiren, her dersi büyük bir coşkuyla işleyen, yeni ufuklar açan hocalarım Prof. Dr. Duygu Anıl, Dr. Öğretim Üyesi Derya Çobanoğlu Aktan ve Prof. Dr. Hülya Kelecioğlu'na,

Yapacağıma inancını yitirmeyen ve bu süreçte desteğini esirgemeyen eşim Süha Kazanç'a,

Son olarak burada adını yazamadığım hayatım boyunca beni destekleyen ve yapacağıma beni inandıran değerli öğretmenlerim ve tüm dostlarıma en samimi duygularıyla teşekkürlerimi sunuyorum.

İçindekiler

| | |
|--|------|
| Öz..... | ii |
| Abstract..... | iii |
| Teşekkür..... | iv |
| Tablolar Dizini..... | vii |
| Şekiller Dizini..... | viii |
| Simgeler ve Kısaltmalar Dizini..... | ix |
| Bölüm 1 Giriş..... | 1 |
| Problem Durumu..... | 2 |
| Çalışmanın Amacı ve Önemi..... | 3 |
| Problem..... | 4 |
| Varsayımlar..... | 5 |
| Sınırlılıklar..... | 5 |
| Tanımlar..... | 5 |
| Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar..... | 6 |
| Testin Eşitlenmesi için Gerekli Koşullar..... | 7 |
| Eşitleme Desenleri..... | 9 |
| Eşitleme Hatası..... | 14 |
| Eşitleme Yöntemleri..... | 14 |
| Şans Başarısı..... | 17 |
| İlgili Araştırmalar..... | 18 |
| Bölüm 3 Yöntem..... | 23 |
| Araştırmanın Evreni ve Örneklemi..... | 23 |
| Veri Toplama Süreci..... | 23 |
| Verilerin Analizi..... | 23 |
| Şans Başarısından Arındırılmamış İngilizce Testlerine İlişkin Varsayımlar..... | 24 |
| Şans Başarısından Arındırılmış İngilizce Testlerine İlişkin Varsayımlar..... | 27 |

| | |
|--|----|
| Eşitleme Hata Kriterleri | 29 |
| Bölüm 4 Bulgular | 31 |
| Alt problemler | 31 |
| Bölüm 5 Sonuç, Tartışma ve Öneriler | 38 |
| Kaynaklar | 41 |
| EK-A: Etik Komisyonu Onay Bildirimi | 48 |
| EK-B: Etik Beyanı | 49 |
| EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu | 50 |
| EK-Ç: Thesis/Dissertation Originality Report..... | 51 |
| EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı..... | 52 |

Tablolar Dizini

| | |
|--|----|
| Tablo 1 <i>Tek Grup Deseni</i> | 11 |
| Tablo 2 <i>Dengelenmiş Grup Deseni</i> | 12 |
| Tablo 3 <i>Eş Değer Grup Deseni</i> | 12 |
| Tablo 4 <i>İç Ortak Desen</i> | 13 |
| Tablo 5 <i>Dış Ortak Desen</i> | 14 |
| Tablo 6 <i>Alt Testlere İlişkin Faktör Analizi Sonuçları</i> | 24 |
| Tablo 7 <i>Ham Puanlara İlişkin Betimsel İstatistik Tablosu</i> | 24 |
| Tablo 8 <i>İngilizce Testlerinin Güvenirliklerinin Karşılaştırılması</i> | 25 |
| Tablo 9 <i>Ortalama Güçlükleri Arasındaki Fark Testi</i> | 25 |
| Tablo 10 <i>Ortalamalar Arasındaki Fark Testi</i> | 26 |
| Tablo 11 <i>Şans Başarısından Arındırılmış Puanların Betimsel İstatistikleri</i> | 27 |
| Tablo 12 <i>Güvenirlik Testi</i> | 27 |
| Tablo 13 <i>Ortalama Güçlükleri Arasındaki Fark Testi</i> | 28 |
| Tablo 14 <i>Ortalama ve Varyansların Kıyaslanması</i> | 28 |
| Tablo 15 <i>Eşitleme Sonucunda Elde Edilmiş Puanlar</i> | 31 |
| Tablo 16 <i>Şanstan Arındırılmış Puanların Eşitlenmiş Puanları</i> | 33 |
| Tablo 17 <i>Doğrusal ve Eşit Yüzdelikli Yönteme Ait Hata Değerleri</i> | 36 |

Şekiller Dizini

| | |
|--|----|
| Şekil 1. İki Testin eşit yüzdelikli eşitleme ile eşitlenmesi (Zhu, 1998, s. 14). | 16 |
| Şekil 2. Şanstan arındırılmamış puan dağılımı. | 32 |
| Şekil 3. Şanstan arındırılmış puanların dağılımı. | 34 |
| Şekil 4. Şanstan arındırılmamış ham puanlara göre eşitleme yöntemlerinden elde edilen fark grafiği. | 35 |
| Şekil 5. Şanstan arındırılmış ham puanlara göre eşitleme yöntemlerinden elde edilen fark grafiği. | 35 |

Simgeler ve Kısaltmalar Dizini

KTT: Klasik Test Teorisi

MSD: İşaretli Farkların Ortalaması

MTK: Madde Tepki Kuramı

RMSD: Kareler Farkının Ortalamasının Kare Kökü

WMSE: Ağırlıklandırılmış Hata Kareleri Ortalaması

Bölüm 1

Giriş

Eğitim ihtiyaçlarını ve kalitesini tespit etmek, eğitim kararlarını almak, belli bir işe eleman seçmek, öğrenci seçmek ve yerleştirmek, rehberlik hizmeti sunmak, mevcut durumu saptamak ve eğitim programının etkililiğini araştırmak gibi farklı amaçlar için bireylere testler uygulanmaktadır. Testler, bu amaçlar çerçevesinde, birey içi ve bireylerarası farkları duyarlı bir şekilde ortaya çıkarmayı amaçlar. Bireyleri seçerken ve yerleştirirken mutlaka bireylerin almış oldukları aynı sınav türlerinden puanları sıraya koyarak karşılaştırmak gerekebilir. Türkiye’de her yıl liselere, üniversitelere ya da lisansüstü programlara girmek için belirli merkezler tarafından bir ya da birden fazla sınavlar yapılmaktadır. Farklı yıllarda yapılmış olan sınavlardan alınan puanlar, aynı seçme sürecinde kullanılabilir. Bu amaçla bireyler hakkında tutarlı ve geçerli bir karar vermek için her yıl veya yılda birden fazla yapılan testlerin birbiriyle karşılaştırılmasına ihtiyaç duyulmaktadır.

Angoff (1984) test puanlarının eşitlenmesini bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlamıştır ve dönüştürmeden sonra formların eşit olacağını söylemiştir. Örneğin sıcaklık için iki birim Fahrenheit ve Celcius olarak belirlenmiştir. Bu birimlerin ortak özellikleri sıcaklık ölçmektir ve birbirine dönüştürülebilir. Eşitleme bir testten alınan puanların diğer testler için karşılaştırmak adına kullanmak için yapılan istatistiksel bir süreçtir (Kolen ve Brennan, 2004). Eş değer formunu oluşturmak için bir testin, kapsam ve yapı yönünden testlerin birbirine eş ve istatistiksel olarak birbirine benzer olmaları sağlanır. Fakat zorluk olarak eş değerde hazırlanmış testler birbirinden farklılaşır. Bu farklılığı ortadan kaldırmak adına test eşitleme yöntemi kullanılır. Böylece testler birbirinin yerine kullanılabilir hale getirilir. Felan (2002), test eşitlemenin ancak bir test için birçok formun var olması durumunda kullanılacağını, değişik formları alan kişilerin birbiriyle karşılaştırılabileceğini belirtmiştir. Linn (1993) tarafından yapılan sınıflamaya göre, test eşitleme (test equating), ilişkilendirilecek testlerin aynı yapıyı ölçmesi ve aynı yolla ölçmesi yönüyle diğer ilişkilendirme yöntemlerinden ayrılmaktadır. Buna göre test eşitleme, aynı yapıyı ölçen farklı test formlarından elde edilen puanların birbirine dönüşümünü sağlayan istatistiksel bir süreçtir. Eşitlenmiş puanlar, testin ne zaman ve kime uygulandığına bakılmaksızın aynı anlama gelir.

Böylece bir test formundan elde edilen puanlar, diğer test formundan elde edilen puanlarla karşılaştırılabilmektedir (Kolen & Brennan, 2004).

Farklı testlerden alınan puanların ortak bir ölçekte toplanabilmesi sürecine ilişkilendirme (linking) adı verilir (Huggins & Penfield, 2012). İlişkilendirme sayesinde puanlar birbirinin yerine kullanılabilir. Bu sebepten dolayı *test eşitleme* (*test equating*) kelimesi de kullanılabilir. Eşitleme yaparken eş değer formlara sahip olunan testlerin uygulanacağı örneklemin cinsiyet, dil, ırk, bulunulan coğrafya gibi özelliklerin değişmezliğine alan yazında örneklem değişmezliği adı verilir. Test eşitleme çalışması olan bu çalışmada da benzer formlardan elde edilen puanların şans başarısı değişkenine göre eşitlenmesi amaçlanmıştır.

Problem Durumu

Ülkemizde çeşitli kurumlara öğrenci veya personel seçmek amacıyla her yıl birden fazla sınav uygulanmaktadır. Ayrıca bireylerin görüşlerini ve başarılarını değerlendirmek amaçlı birden fazla test yapılmaktadır. Test gizliliğini koruyabilmek, ölçülen değişkenin yapısı gereği ve geçerlik ve güvenirlik çalışmaları yürütebilme adına testlerin birden fazla formları geliştirilmektedir (Turgut, 1979). Geliştirilen formların eşitlenmesi üç amaç için yapılmaktadır. Birincisi, farklı iki testi alan bireyler (yeteneklerinin aynı olduğu varsayılarak) arasında yanlılığı önlemek iken ikinci amaç ise aynı ölçek üzerinde her iki testten alınan puanları değerlendirebilmektir. Üçüncü olarak, farklı yetenek düzeyindeki grupların gelişimini incelemek amaçlanır. (Zhu, 1998). İlk iki amaç için yatay eşitleme; üçüncü olarak gösterilen amaç için ise dikey eşitleme yapılmaktadır. Dikey eşitlemenin içerik değiştiğinden dolayı dikey hizalama (vertical scaling) olduğu eşitleme olmadığı farklı görüşler tarafından desteklenmektedir (Kolen, 1988).

Testi üreten ve uygulayanlar tarafından testteki soru sayısının, ölçülmek istenen psikometrik ve bilişsel yapının değişmezliği sağlansa da farklı zamanlarda farklı test maddeleriyle ölçüldüğünde iki test ve puanlar arasında farklılıklar olacaktır (Tanguma, 2000). Aynı testin farklı versiyonlarından elde ettikleri puanları birbiri ile kıyaslama imkânı bulan öğrenci ve eğitimciler, öğrencilerin gelişimsel farklılığı yorumlamaları daha anlamlı olacaktır. Ayrıca eşitlenen iki test farklı bireylere de uygulanmasına olanak sağlayarak yanlılığı önleyecektir.

Çoktan seçmeli testlerin hep uygulamada kolaylık, sonuçların hızlı bir şekilde rapor edilmesine olanak ve objektif bir şekilde bireyleri seçme imkânı sağladığı için ulusal ve uluslararası birçok sınavda tercih edilmektedir. Fakat bu testler, cevaplayıcılara herhangi birini düşünmeden diğer bir deyişle şansla seçme imkânı sağlaması en büyük dezavantajlarından biridir. Şans başarısı, cevaplayıcının çoktan seçmeli testte doğru seçeneği o madde ile ölçülmek istenen bilgi ve yetenek ile değil de rastgele veya tahminle işaretlemesidir (Turgut, 1971). Şans sayesinde elde edilen puanın hatadan arındırmak için Turgut (1971)'un önerdiği düzeltme formülü uygulanmıştır. Son yıllarda yapılan çalışmalar çeşitli değişkenler ve kriterlere göre eşitleme yöntemlerinin etkililiği kıyaslanmıştır.

Çalışmanın Amacı ve Önemi

Birden fazla test eşitleme yöntemi kullanılarak bazı değişkenlerin test eşitlemeye etkisi araştırılmıştır. Bu değişkenler, örneklem büyüklüğü (Fitzpatrick & Yen, 2001; Skaggs, 2005), şans başarısı (Aşiret & Ömür-Sünbül, 2014; Bozdağ, Kan, 2010), maddelerin değişen madde fonksiyonuna (DMF) sahip olması, ankor test uzunluğu (Ricker & von Davier, 2007) sayılabilir. Örneğin, Kim ve Livingston (2010) ortak madde kullanarak küçük örneklerde beş eşitleme yönteminin doğruluğunu test etmişlerdir. Daha özeldense, bir değişken olarak şans başarısının test ve madde puanlarına etkilediği bazı çalışmalar ile belirlenmiştir (Araz, 2001; Bozdağ & Kan, 2010; Telli, 1993; Çelen,2002; Sahnüseyinoğlu, 1998). Bu çalışmaların yanı sıra, şans başarısının istatistiksel olarak test eşitlemeye etkisinden bahsedilebilir. Şans başarısından arındırılmış bir testin varyansı artarken ortalama güçlüğü azalmaktadır (Baykul, 2000). Bu durum, şans başarısının test eşitleme denklemindeki etkisini genel olarak açıklamaktadır.

Yukarıda bahsedilen test eşitleme yöntemlerinin etkililiği farklı test eşitleme düzenekleri ile birden çok çalışmalarıyla test edilmiştir (Han, Kolen & Pohlman, 1997; Hendrickson & Kolen, 2003; Lee & Ban, 2010). Bu sonuçlara göre eşdeğer grup ortak maddeli desen dışında klasik test teorisine dayalı test eşitleme yöntemlerinin Madde Tepki Kuramı'na bağlı yöntemler kadar iyi sonuçlar verdiği saptanmıştır. Ayrıca yaygın olarak seçkisiz grup düzeneği üzerinde çalışıldığı görülmektedir. Bu çalışmada da tek grup üzerinde eşitleme yapılacağı için klasik test teorisine dayalı yöntemler tercih edilmiştir.

Şans başarısının test eşitlemeye etkisi üzerine yapılan çalışmalardan biri olan Bozdağ ve Kan (2010)' in çalışmasında, şans başarısından arındırıldığında eşit yüzdellikli eşitlemenin daha uygun olduğu bulunmuştur. Aşiret ve Ömür-Sünbül (2014) çalışmasında ise şans başarısının (guessing) etkisinin farklı değişkenlerle (formların zorlukları arasındaki fark ve örneklem büyüklüğü) ilişkisine bakmıştır. Şans başarısının test eşitleme üzerinde tutarlı bir ilişkisinin gözlemlenemediği sonucuna ulaşmışlardır. Bu çalışmalar ile şans başarısının etkisi üzerinde daha çok çalışmalara ihtiyaç duyulduğu görülmüştür.

Bu bağlamda bu çalışmanın amacı, şans başarısının test eşitlemeye etkisi olup olmadığını incelemektir. Şanstan arındırılmış veya arındırılmamış ham puanların eşitlenmesinde kullanılan yöntemlerden hangisinin daha az eşitleme hatası içerdiğini bulmak amaçlanmıştır. Bir diğer ifade ile ham puanların, şans başarısından arındırılmadan ve şans başarısından arındırılarak eşitlenmesinde doğrusal eşitleme ve eşit yüzdellikli eşitleme yöntemlerinden hangisinin daha uygun olduğudur. Bu bağlamda, bu çalışmanın problemi aşağıda sunulmuştur.

Problem

Şans başarısı değişkeni dikkate alınarak, 8. Sınıf TEOG İngilizce I ve II test puanlarının, doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılarak eşitlenmesinden uygun olan eşitleme yöntemi hangisidir?

Alt problemler.

1. İngilizce testlerine ait şans başarısından arındırılmamış puanların doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

2. İngilizce testlerine ait şans başarısından arındırılmış puanların doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

3. İngilizce testlerinin ham puanlarının şans başarısından arındırılarak eşitlenmesinde eşit yüzdellikli eşitleme ve doğrusal eşitleme yöntemlerinden hangisi kullanıldığında hata payı en azdır?

Varsayımlar

1. İngilizce I testi ile İngilizce II testi arasında geçen zamanda bireylerin yetenek ve bilgi seviyelerinde deęişme olmadığı varsayılmıştır.
2. Her iki uygulama koşullarının birbirine eş olduğu varsayılmıştır.
3. Adayların sınav uygulamalarında, test maddelerini bilgilerini yansıtır ve anlamlı bir şekilde ele aldıkları düşünölmektedir.
4. Eşitlenecek iki testin İngilizce ile ilgili aynı kazanımları ölçtüęü kabul edilmiş olup kapsamlarının aynı olduğu varsayılmıştır.

Sınırlılıklar

1. Araştırma, deęerlendirme sınavına katılan okullar ve öğrencileri ile sınırlıdır.
2. Araştırma, Klasik Test Teorisine Dayalı doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri ile sınırlıdır.
3. Araştırma tek grup düzeneęi kullanılarak eşitlenmiştir.

Tanımlar

Eşitleme. Aynı yapıyı ölçen farklı test formlarından elde edilen puanların birbirine dönüşümünü sağlayan istatistiksel bir süreçtir (Linn, 1993).

Şans başarısı. Testi alan öğrencilerin doğru cevabı ölçölmek istenen bilgi ile deęil de rastgele veya şansla işaretlemelerinden kaynaklı testin geçerlięi ve güvenilirlięi düşmektedir (Baykul, 2000).

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

İlişkilendirme ve eşitlemenin istatistiksel yöntemleri birbirine benzer olmasına rağmen kullanım amaçları bakımından farklılık göstermektedirler. İlişkilendirme yapabilmek için testlerin birbiri ile farklı olması istenirken; eşitlemede içerik ve istatistiksel özellikler bakımından benzer olmasına dikkat edilmektedir (Kolen ve Breannan, 2014).

Genel olarak yatay (horizontal) ve dikey (vertical) olmak üzere iki tür eşitlemeden bahsedilmektedir (Crocker & Algina, 2006). Örneğin bir testin yıl içinde farklı zamanlarda uygulanması için farklı formlarının oluşturulması gerekir. Eğer bu formların puan dağılımı benzer olmazsa formlardan benzer puanların alınması güçleşir. İşte bir testin farklı formlarından benzer puanların elde edilmesi *yatay eşitleme* olarak tanımlanır. *Dikey eşitleme* ise farklı düzeylerde olan, örneğin farklı sınıflarda olan kişilerin yeteneklerinin karşılaştırılmasını amaçlayan tek bir ölçek geliştirilir (Crocker & Algina,2006).

Test eşitleme adımları şu şekilde verilmiştir:

1. Eşitlemenin amacına karar verilmesi
2. Alternatif formların oluşturulması (Formlar aynı yapı ve istatistiksel özelliklere sahip olmalıdır),
3. Veri toplama yönteminin ve deseninin belirlenmesi,
4. Desenin uygulanması (testin desene göre uygulanması ve verilerin de bu desene göre elde edilmesi),
5. Eşitleme tanımının belirlenmesi ve seçilmesi,
6. Bir veya birden fazla eşitleme ilişkilerini kestiren yöntemin seçilmesi ve
7. Eşitlemenin sonuçlarının değerlendirilmesidir (Kolen ve Brennan, 2004).

Eşitleme yapabilme adına iki teste ait ham puanlarını rapor etmek veya farklı zamanlarda uygulanan her iki testten elde edilen puan dağılımlarının aynı özelliğe göre birbirine dönüştürülmesi düşünülebilen yollardandır. Fakat iki testin arasındaki geçen zamana bağlı olarak öğrencilerin başarı seviyeleri ve testin zorluk düzeyi hakkında ek bilgiler verilmediğinde yorumlama güçleşmektedir (Kolen ve Brennan,

2014). Ek olarak, öğrenci başarı seviyelerinin de (her iki gruba ait ortalama puanlarının birbirine eşitlendiği durumlarda) göz ardı edilmesi, başarı seviyesi düşük veya yüksek öğrenciler için bir yanlılık oluşturabilmektedir. Kolen ve Brennan başarılı bir test eşitleme için farklı formların zorlukları arasındaki farkın düzenlenip, eşitlenmiş formların kimlerin ve ne zaman aldıklarından bağımsız olarak aynı anlama gelmesi gerektiğini belirtmiştir. Bu nedenle, eşitlenmiş puanları hesaplamadan önce bazı koşulların sağlanması gerekmektedir.

Testin Eşitlenmesi için Gerekli Koşullar

Dorrans ve Lord (1980) ise eşitlenmesi için gerekli koşulları ise aynı yapıyı ölçme, güvenilirliklerin eşit olması, simetriklik, gruptan bağımsızlık ve eşitlik olarak tanımlamıştır. Hambleton ve Swaminethan (1985)'a göre bu sınıflama eşitlik, tek boyutluluk, gruptan bağımsızlık ve simetrikliktir. Petersen, Kollen ve Hoover'a (1989) ve Kolen and Brennan'a (1995) göre eşitleme koşulları: aynı yapıyı ölçme, simetriklik, gruptan bağımsızlık ve eşitliktir. Bu araştırmacıların belirttiği bu dört özellik aşağıda detaylı bir şekilde açıklanmıştır.

Aynı yapıyı ölçme. Angoff'a (1970) göre en temel eşitleme koşulu eşitlenecek testlerin mutlaka aynı yapıyı, özelliği ve yeteneği ölçmesi gerektiğidir. Farklı içeriğe sahip ve farklı özelliği ölçen testler eşitlenemez. Örneğin, bir sözel yetenek testinden elde edilen puanlar sayısal yetenek testinden elde edilen puanlara eşitlenemez. Ek olarak, matematik konularından birinden elde edilen puan diğerinden elde edilen puana eşitleme yapılması kazanım açısından farklılık olacağı için uygun görülmez.

Tek boyutluluk. Eşitlenmesi gereken testler kullanıldığında testi oluşturan maddelerin testin sadece ölçülmek istenen yeteneği, kazanımı veya davranışı ölçmesi başka bir özelliği ya da yeteneği ölçmemesidir (Hambleton, 1985).

Güvenilirliklerin eşit olması. İki testin güvenilirliklerin birbirine yakın veya aynı olmalıdır. (Angof, 1971; Crocker & Algina, 2006; Kelecioğlu, 1994; Tanguma 2000). Ayrıca güvenilirlik katsayılarının yüksek olması eşitleme için daha uygun bir ortam hazırlamaktadır. Eğer güvenilirlikler benzer değilse veya çok düşükse test eşitleme güçleşir.

Simetriklik. İki formun birim sistemleri arasındaki dönüşüm, tesadüfi hatalardan kaynaklanan farklılıklar dışında tektir. X formundan Y formuna dönüştürülen puanlar, Y formundan X formuna dönüştürülen puanlara eşittir. (Angof, 1971; Tanguma, 2000). Yani dönüşümün tersine çevrilebilir olmasıdır. Örneğin X formundaki 50 puan Y formunda 60 puana dönüşüyorsa Y puanındaki 60 ham puanında X formundaki 50 ham puana karşılık gelmesidir.

Gruptan bağımsızlık. Eşitleme sonucunda elde edilen puanların, dönüştürmelerin yapıldığı gruptan bağımsız olmasıdır. (Angof, 1971; Kelecioğlu, 1994). Eşitleme temel olarak hangi testin X, hangi testin Y olduğuna bakılmaksızın sonucunun aynı olması gerektiğidir. İki test için eşitleme yapılmış ise o testi alan kızlar ve erkekler için de aynı eşitleme ilişkisi bulunacağını varsayar.

Eşitlik. Son olarak Lord (1980) iki formun eşit olması gerektiğini söylemiştir. Eşitlik koşulu bir testin X ve Y formlarını alan kişilerin farksız olmaları, yani bir testin X formunu alan kişinin puanının, her iki formda da eşit olmasıdır. Lord (1980) ifade ettiği gibi eşitlik koşulunun sağlanması için gereken özellikler aşağıdaki gibidir.

Farklı yetenek ve becerileri ölçen testler eşitlenemez.

Eşit güvenilirliğe sahip olmayan testler eşitlenemez.

Farklı güçlük düzeylerine sahip testlere ilişkin ham puanlar (dikey eşitleme) eşitlenemez.

Bir X testinin puanlarının θ yetenek düzeyindeki koşullu frekans dağılımı (X testine ilişkin θ puanın $f[x | \theta]$ fonksiyonu), Y testinin dönüştürülmüş x (y) puanına ilişkin $f[x(y) | \theta]$ koşullu frekans dağılımı ile aynı olmalıdır.

Paralel olmayan testler eşitlenemez; çünkü paralel olmayan testlerin hata puanları dağılımları farklıdır ve güvenilirlikleri de eşit değildir.

İki testin ortalamaları ve varyansları eşit olmalıdır.

Eşitlemenin yapılabilmesi için eşitlenecek iki testin formları arasındaki testlerin korelasyonları yüksek olmalıdır (Dorans ve Holland, 2000).

Eşitlikte ana fikir, Y formundaki herhangi bir gerçek puan için, Y Formunda gözlenen puanların dağılımının, X Formunda dönüştürülen puanların dağılımı ile aynı olması gerektiğidir (Kolen & Brennan, 2004). Eğer bu sağlanabilirse, her bir bireyin testin hangi formunu alıp almadığı önemsiz olmalıdır. Fakat bu şartlar altında

Lord, hata içeren formların eşitlemesinin zor olacağını belirtmiştir. Bu tanıma göre, birbirine paralel olan formların eşitlemesinin çalışmanın anlamsız veya gereksiz olacağı sonucuna gidilebilir. Morris (1982) bu tanımın biraz daha esnek hali olarak “zayıf eşitlik” tanımını ortaya atmıştır. Bu tanım göz önünde bulundurularak eşitlik için birinci (Form Y den elde edilen gözlenen puan dağılımlarının ve çevrilmiş Form X puanlarının ortalaması) ve ikinci sıra Form Y den elde edilen gözlenen puan dağılımlarının ve çevrilmiş Form X puanlarının varyansları) eşitlik kavramları ortaya atılmıştır.

Görüldüğü üzere, alanyazın incelendiğinde eşitleme için gerekli koşullar farklılık göstermektedir. Eşitleme modelini MTK kullanarak yapan kişiler eşitlemenin olması için tek boyutluluk özelliğinin şart olduğunu söylemişlerdir. Bununla ilgili Dorans (1990) bir testi diğerine eşitlemek için her iki testin de aynı yapıyı ölçmesinin ve her iki testin de tek bir yeteneği ölçen (tek boyutluluk) maddeleri içermesinin zorunlu olduğunu ifade eder. Fakat MTK modeli ile eşitlemeyenler ise tek boyutlulukla ilgili bir kıstas belirtmemişlerdir. Angoff (1971) eşitleme maddelerin aynı yapıyı ölçmesi ve gruplar arası değişmezlik kriterlerinin karşılanmasından bahsetmektedir.

Eşitleme Desenleri

Eşitlemede veri toplama deseni “eşitleme deseni” olarak ifade edilir. (Kolen & Brennen, 2004). Eşitleme desenleri, test formlarının sayısı, hangi sıra ile uygulanacağı ve grup sayısı olarak birbirlerinden farklılıklar içerir. Herbiri cevaplayıcıların yetenek düzeylerinin farklılığını çeşitli yollarla kontrol altına almayı hedeflemiştir (Holland, Dorans & Petersen, 2007). Eşitlenmek istenen grubun özelliği ve testlerin uygulanış şekline göre birden çok eşitleme desenleri ortaya atılmıştır (Angof, 1971; Tanguma, 2000; Kolen & Brennen, 2004; Crocker & Algina; 2006).

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 & Kan, 2010). Bu yollar:

1. Bütün formlar aynı test grubuna uygulanır.

2.İki form, ölçülen özellik bakımından benzer 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. Literatür incelendiğinde pek çok sınıflandırma yapılmıştır. İlk olarak Crocker & Algina (2006) ve Kolen & Brennan (2004) üçlü sınıflanması açıklanmıştır.

1. Tek Grup Deseni (Single Group Design)

2. Random Grup Deseni (Random Group Design)

3. Denk olmayan gruplarda ortak (Anchor) madde/test deseni (Non-equivalent groups anchor test design)

İçlerinden en zor uygulanan desen Anchor-test olmasına rağmen en bilinen olan desenin bu desen olduğunu belirtmiştir (Kolen & Brennan, 2004).

Livingston (2004)'e göre ise bu sınıflama aşağıdaki gibidir:

1. Tek grup deseni

2. Eşdeğer grup deseni

3. Dengelenmiş grup deseni

4. Anchor(ortak) Desen

Tek grup deseni. Uygulama ve hata açısından en basit eşitleme deseni denilebilir (von Davier, 2011). Tek grup düzeneğinde iki veya daha fazla test aynı gruba uygulanır. Aynı kişilerin her iki testi de alması nedeniyle testlerin zorluk düzeyleri cevaplayıcılar için benzerdir. Fakat test formlarının veriliş sırası bir hata oluşturabilir. Y formunu alırken x formunu almış olmanın verdiği yorgunluk ile y formunun cevaplayıcılar için daha zor veya x formundaki sorulara aşinalık kazanılabileceğinden y formunun cevaplayıcılar için daha kolay gelmesi olası durumlar olabilir (Kolen & Brennan, 2014). Bunu dengeleyebilmek adına, grubun yarısına eşitlenecek testlerden ilk formunu diğer yarısına ise eşitlenecek testlerden ikinci formunu vermek gereklidir. Böylelikle sıradan kaynaklanan hatalar (sıra

hatası) giderilmiş olur (Kolen & Brennan, 2004). Diğer yandan, cevaplayıcıların birden fazla teste katılması yorgunluk ve bıkkınlık oluşturacağından dolayı ikinci forma verilen cevapların puanları düşük olabilir. Puan düşüklüğü ve testlerin uygulanış sırası etkisi en aza indirgense bile bu desenin kullanılabilirliği tartışılmaktadır. (örn. Kolen ve Brennan, 2014). Çünkü seçilen örneklemin evreni temsil edip etmediği bu grup deseni için dikkate alınmazken grubun tek ve aynı yetenek düzeyinde olması gerekmektedir (Livingston, 2004). Bu desene ait gösterim aşağıdaki tablodaki gibidir (Tablo 1).

Tablo 1

Tek Grup Deseni

| Grup | Form A | Form B |
|------|--------|--------|
| G1 | X | X |

Tek grup desenine ilişkin problem 1976 yılında ordu için bireyleri seçerken karşılaşılmıştır. Bireyler arasındaki motivasyon farklılığı yeni formun uygulanması sonucunda puanların çok yüksek olmasına neden olmuştur. Böylece ordu yeni formdaki yüksek puanlara göre bireyleri seçtiğinde amaçladıkları standartların altındaki yetenekteki bireyleri seçmişlerdir.

Bu araştırma kapsamında, veriler tek grup desenine göre toplanmıştır. Uygulamada, bir sınavın farklı formlarını aynı gruba iki kez aynı anda uygulama imkânı pek mümkün olmamaktadır. İngilizce 1 ve 2 testlerinin uygulama sırasında geçen süre az da olsa sıra etkisi tam anlamıyla kontrol altına alınmamıştır.

Dengelenmiş grup deseni. Bu desen Livingston'ın sınıflandırmasında dengelenmiş grup deseni adını almaktadır. Bu desende gruplar rastgele olarak ikiye ayrılır. Bu deseni kullanmak için kullanılacak yöntemlerden biri sarmal (spiralling) yöntemidir. Bu gruplara her iki test de uygulanır fakat her gruba sırasıyla farklı test uygulanır. (Kolen & Brennan, 2004). Birinci grup A testini aldıktan sonra B testini ikinci gruba, B testini aldıktan sonra A testi uygulanır. Bu sarmal süreç A ve B testlerini alan grupların karşılaştırılmasına olanak sağlamıştır. Bu desen aşağıdaki tabloda gösterilmiştir.

Tablo 2

Dengelenmiş Grup Deseni

| Grup | Form A | Form B | Form A | Form B |
|------|--------|--------|--------|--------|
| RG1 | X | | | X |
| RG2 | | X | X | |

Bu desen gruplardan gelen farklılığı azaltarak eşitlemenin kesinliğini artırır. Bu desende gruplar aynı olmadığı için testlere çeşitli gruplardan kaynaklı yanlılıklar ortaya çıkabilir bunu dengelemek güçtür. Eşitlenmek istenen iki farklı test formlarını tek grup düzeneği ile eşitlemek mümkün değildir (Livingston, 2004). Bu nedenle dengelenmiş grup düzeneği kullanılır. Bu düzenek için testleri alan grupların, birbirine benzer yeteneklere (testte ölçülmek istenen) sahip olması gerekir. Ayrıca formların uygulanışındaki geçen zamana da dikkat edilmeli ve yeterli süre verilmesi gerekmektedir.

Eş değer grup deseni (Seçkisiz grup). Birçok eşitleme çalışmasında aynı grubun her iki test formunu almaları zaman ve uygulama açısından zordur. Bu durumda gruplar özellik ve yetenek bakımında benzer olacak şekilde ikiye ayrılır. Ölçülen özellik bakımından grubun benzer yeteneğe sahip olması önemli bir aşamadır. Her iki gruba farklı test formları verilir. Dengelenmiş gruptan temel farkı gruplar rastgele olarak değil, belli özellik veya yeteneğe göre gruplar eş olarak seçilmektedir. Ayrıca gruptaki kişi sayısının fazla olması eşitlemenin daha doğru sonuç vermesi için gereklidir. Tablo 3'te bu desene ilişkin gösterim verilmiştir.

Tablo 3

Eş Değer Grup Deseni

| Grup | Form A | Form B |
|------|--------|--------|
| EG1 | X | |
| EG2 | | X |

Livingston (2004) bu iki grubu yetenek düzeylerinin birbirine paralel olacak şekilde belirleyebilme adına paketleme (spiraling the books) yapılabileceğini belirtmiştir. Bu şekilde gruplar arası farklılığı kontrol altına alınmaya çalışılır.

Ortak (Ankor) desen. Ortak test/madde deseni, en çok kullanılan desendir (Kim, Walker, Mchale, 2010; Kolen & Brennan, 2004; Norman-Dvorak, 2009). Bu desen, denk olmayan gruplara uygulanır ve ortak maddeler üzerinde grupların performansını karşılaştırarak iki grup arasındaki eşitleme ilişkisini ortaya çıkarmada kullanılır.

Uygulama esnasında özellikle başarı testlerinde, test eşitleme uygulamasında kullanılacak iki grubun seçkisiz olarak seçilip iki gruba ayırmak oldukça zordur (Tanguma, 2000). Farklı zamanlarda, farklı formları alan kişiler benzer veya eşdeğer olmayabilir. İki grubun seçkisiz olarak seçilememesi durumunda, gruplar arasındaki farklılıkları kontrol edebilmek için aynı özelliği ölçen farklı maddeler her iki teste de eklenir. Bu ortak teste “ankor (anchor)” test denir. Bu testteki maddelerin, genel testteki ölçülmek istenen yeteneğe ve zorluk düzeyine paralel olması beklenmektedir. Ankor test gruplara her iki formdan farklı bir form olarak verilebileceği gibi testin bir parçası olarak da verilebilir. Eşitleme için ankor testin güvenilirlikleri testle paralel ve grupların puan dağılımlarının birbirine benzer olması gerekmektedir. (Kolen & Brannen, 2004). Her bir gruba uygulanan test ankor maddeleri içeriyorsa ve o ankor madde puanları test puanına ekleniyorsa “iç ortak test” adı verilir ve desen Tablo 4’te sunulmuştur. Eğer ortak maddeler farklı formlar uygulandıktan sonra ayrı bir test olarak uygulanıyorsa, ankor maddeler “dış ortak test” adını alır (Crocker & Algina, 2006). Bu desen sayesinde ankor maddeler özenli seçildiği sürece tek grup desenindeki hataları ortadan giderebilir. Dış ankor kullanılarak yapılan düzenek Tablo 5’te verilmiştir.

Tablo 4

İç Ortak Desen

| Grup | (Form A+İç Ankor) | (Form B+İç Ankor) |
|------|-------------------|-------------------|
| G1 | X | |
| G2 | | X |

Tablo 5

Dış Ortak Desen

| Grup | Form A+Dış Ankor | Form B+Dış Ankor |
|------|------------------|------------------|
| G1 | X | |
| G2 | | X |

Eşitleme Hatası

Eşitleme hatası seçkisiz ve sistematik olarak ikiye ayrılmaktadır.

Seçkisiz hata. Örneklemden elde edilen standart sapma, ortalama veya yüzdelik sıra gibi test istatistiklerinden kaynaklanabilir. Bu tip hata büyük örneklem ve doğru eşitleme düzeneği ile azaltılabilir (Felan, 2002; Kolen, 1988)

Sistematik hata. Eşitleme koşullarının bozulmasından kaynaklanır (Kolen, 1988; Zeng, 1991). Sistematik hatanın kestirilmesi daha zordur (Kolen, 1988). Bu hata, eşitleme için düzgünleştirme işleminin yanlış yapılması veya eşitleme varsayımlarının bazılarının sağlanmaması durumlarında ortaya çıkabilir (Kolen ve Brennan, 2004).

Eşitleme Yöntemleri

Klasik test teorisinde, bireyin bir teste ilişkin puanı, gerçek puanı (True Score:T) ve hata puanının (Error Score:E) toplamını oluşturup, gözlenen puan (observed score:X) olarak ifade edilmektedir.

$$X=T+E \quad (1)$$

Bu ilişkiyi veren denklem (1)'de gösterilmiştir. Bu teorinin dışında test eşitleme için kullanılan ikinci bir teori ise Madde Tepki Kuramı'na dayalı yöntemlerdir. Bu çalışmanın odağı klasik test teorisi yöntemleridir. Bu yöntemler: 1) Doğrusal Eşitleme 2) Eşit Yüzdelikli Eşitleme

Doğrusal eşitleme. Doğrusal eşitleme, test formlarının betimsel istatistik değerlerinin (ortalama ve standart sapma) haricindeki tüm özelliklerinin eşit olması

varsayımına dayanmaktadır (Crocker ve Algina, 1986, s.458). X formundan elde edilen ham puanlar ile Y formundan elde edilen puanlar arasındaki ilişki doğrusal olduğundan dolayı bu eşitlemeye doğrusal eşitleme adı verilmektedir. Aynı standart puanlara karşılık gelen puanların eşit olduğu kabul edilir (Angoff, 1971, s.564). İki formdan alınan puanların farkının sabit olmasından ziyade doğrusal eşitleme iki test formun zorluklarının farkı ile ilgilenmektedir (Kolen & Breannan, 2014). Bu yöntemle örneğin iki formun ortalamaları (μ_1 , μ_2) ve standart sapmaları (s_1 , s_2) düşünülürse;

$$\frac{X_1 - \mu_1}{s_1} = \frac{X_2 - \mu_2}{s_2} \quad (1)$$

(1) eşitliğinden X_2 yalnız bırakılacak şekilde eşitlik tekrar sağlanırsa;

$$X_2 = \frac{s_2}{s_1} X_1 + (\mu_2 - \frac{s_2}{s_1} \mu_1) \quad (2)$$

(2) eşitliği elde edilir. Bu eşitliği, doğrusal ilişkiyi veren genel denklem biçimine ($y=mx+b$) dönüştürdüğümüzde $\frac{s_2}{s_1}$ grafiğin eğimini gösterirken $\mu_2 - \frac{s_2}{s_1} \mu_1$ ise y eksenini kestiği noktayı göstermektedir.

Testin farklı formlarını alan gruplar, aynı bilgi ve yetenek seviyesinde değilse, farklı doğrusal eşitleme yöntemlerinin uygulanması önerilir (örn: Tucker veya Levine doğrusal eşitleme). Doğrusal eşitleme için paralel formlardan elde edilen puan dağılımlarının birbirine paralel ve güvenilirliklerinin eşit olması durumlarının sağlanması gerekmektedir. (Crocker, Algina; 1986; Skaggs & Lissitz; 1986; Thorndike, 1982; Woldbeck, 1998).

Bu yöntemin uygulamasının kolaylığının yanında karşılaşılan iki durum vardır. Bunlardan birincisi; x ve y formundan alınan puanlar birbirinden çok uzak değerlerdelerse, eşitleme sonucunda formlardan alınan puanların değeri belirlenen aralıkta yer almıyorsa açıklaması zor olabilir (Livingston, 2004, s.16). Bu durum, doğru eğiminin çok yüksek veya düşük olmasından veya y eksenini kestiği değerden kaynaklıdır ve dönüştürülmüş puanın sabit puan aralığının dışında kalması ile sonuçlanmaktadır. Diğer bir durum ise doğrusal eşitleme sonuçlarının gruba bağımlı olmasıdır. Bunu Livingston (2004) X formunun zor, Y formunun kolay olduğu ve bu formların ölçülen özellik bakımından zayıf ve güçlü olan iki gruba uygulanması şeklinde örneklemektedir. Bu yüzden farklı zorluktaki test formlarının zayıf olan gruba eşitlenmesi ile güçlü olan gruba eşitlenmesi arasında farklar oluşacaktır.

Eşit yüzdellikli eşitleme. İki form aynı puan dağılımına sahip değil ise eşit yüzdellikli eşitleme yöntemi önerilir (Zhu, 1998). Bu eşitleme yönteminde, iki testin yüzdellik olarak sıraları hesaplanarak eşitlenir. Doğrusal eşitlemeye kıyasla bu yöntem daha az sayılıya sahiptir (Woldbeck, 1998).

Eşit yüzdellikli eşitleme Braun ve Holland (1982) tarafından tanımlanmış olup Kolen ve Breannan (2014) tarafından denklemi adapte edilmiştir. Formül aşağıda sunulmuş olup bu çalışmada da bu denklemden faydalanarak eşitleme yapılmıştır.

X: Form X'ten alınan herhangi bir puan

Y: Form Y'den alınan herhangi bir puan

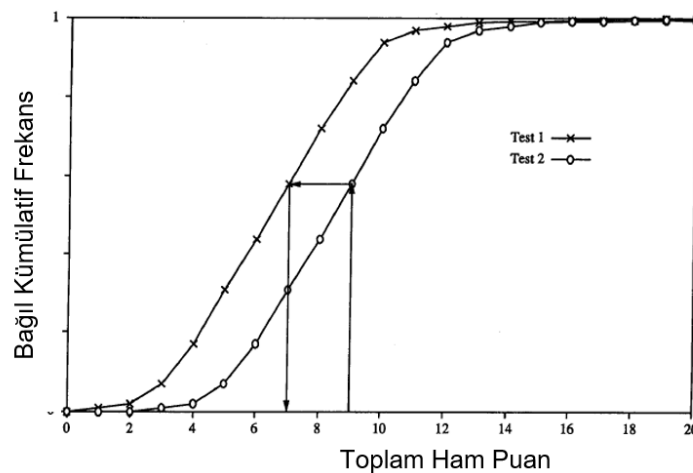
F: Popülasyondaki X'in kümülatif dağılım fonksiyonu

G: Aynı popülasyondaki y'nin kümülatif dağılım fonksiyonu

ey: Form X'ten elden edilen puanları Form Y'dekine dönüştürmek için kullanılan simetrik denklem fonksiyonu

G: ey'nin kümülatif dağılım fonksiyonu

$ey(x)=G^{-1}[F(x)]$ denkleminde ulaşılmıştır. Bu denklem göz önüne alındığında; bireylerin %20'si x testinden 30 puan ve aşağısını alırken, bireylerin %20'si Y testinden 32 puan ve aşağısını almışsa, Form X'teki 30 puan form y'deki 32 puana karşılık gelmektedir.



Şekil 1. İki Testin eşit yüzdellikli eşitleme ile eşitlenmesi (Zhu, 1998, s. 14).

Eşit yüzdellikli eşitleme sonucunda grafiği çizebilmek için önce tek ve eşdeğer gruplar için eşit yüzdellikli eşitleme, başlangıç olarak her iki dağılıma ait puanların

yüzdeler sıraları bulunur. Bu yüzdeler sıralarla ham puanlara ait grafik çizilir. İkinci basamakta ise, bu grafikten yararlanılarak elle veya analitik yöntemler ile eşitlenmiş puanlar bulunur. Böylece, "ham puan sıralaması eğrisi" oluşturulur (Zhu, 1998). Şekil 1'de 20 maddelik iki testin bağıl kümülatif frekans dağılımlarının (yüzdeler sıralaması/100) olduğu ham puan sıralama eğrisi görülmektedir. Örneğin, test 2'deki 9 puan Test 1'deki 7 puana eşittir.

Küçük örneklem kullanıldığında hata, bu eğrilerin analitik yöntemler kullanılarak düzgünleştirilmesi ile azaltılır (Heh, 2007; Kolen, 1984). Düzgünleştirme işlemi eşitlemeden önce yapılıyorsa "ön düzeltme (presmoothing)", eşitleme işleminden sonra yapılıyorsa "son düzeltme (postsmoothing)" adı verilir. Eşit yüzdelerli eşitleme yapabilmek için eşitleme ilişkisinin puan ranjını asmaması gerekir. Eğer tüm grubun puan ranjı gözlenirse bu problem ortadan kalkar (Livingston, 2004). Eşit yüzdelerli eşitlemenin bir başka şartı ise büyük örneklem gerekliliğidir. Diğer bir konu ise X formundan elde edilen puanların yüzdeler sırasının Y formundaki sıralara karşılık gelmemesi durumudur. Bu durum öteleme yöntemleri kullanılarak giderilmeye çalışılmaktadır. Livingston (2004), bu yöntemin puanların aynı yüzdelerli sıraya denk gelmemesi probleminde kesin bir çözüm olmayacağını fakat uygulamada kolaylık sağlayacağını aktarmıştır. Livingston (2004) tarafından önerilen öteleme formülü bu çalışma için uyarlanmıştır.

$$\text{İngilizce I} + \frac{\text{İngilizce II yüzdeler} - \text{İngilizce I altyüzdeler}}{\text{İngilizce I üstüydeler} - \text{İngilizce I altyüzdeler}}$$

Doğrusal eşitleme veya eşit yüzdelerli eşitlemeden hangisinin seçilmesi gerektiği Zhu (1998) konusunda testlerden elde edilen z puanlarının dağılımı birbirine paralel ise doğrusal eşitlemenin tercih edilmesini önermiştir. Aksi koşulda eşit yüzdelerli eşitlemenin uygun olacağını belirtmiştir. Ayrıca eşit yüzdelerli eşitleme için uygun örneklem büyüklüğü 1500'dür (Muraki, Hombo ve Lee, 2000).

Şans Başarısı

Testi alan öğrencilerin doğru cevabı ölçülmek istenen bilgi ile değil de rastgele veya şansla işaretlemelerinden kaynaklı testin geçerliği ve güvenilirliği düşmektedir (Baykul, 2000). Şanstan kaynaklı hataları azaltmak için Turgut (1971) test yönergelerinin değiştirilmesini, madde kökü ve seçeneklerin gözden geçirilmesini

ve puanların hatadan arındırılmasını önermektedir. Bu temelde, ham puanlar şans başarısından arındırılmıştır. Arındırma işlemi için yanlış cevapların bir kısmının doğru cevaplardan çıkarılmıştır. Böylece düzeltilmiş puan (DP) aşağıdaki formül kullanılarak hesaplanmıştır.

$$DP = D - \frac{Y}{k-1}$$

D: Bireyin toplam doğru cevap sayısı.

Y: Bireyin yanlış cevap sayısı.

k: Seçenek sayısı.

İlgili Araştırmalar

Caldwell (1984) tarafından yapılan araştırmada, ortak maddeli testler (ankor) kullanılarak "Klasik Test Kuramı"na ve "Rasch" modeline göre eşitleme yapılmıştır. Araştırmada kullanılan ortak maddeler orta ve uç güçlüklerde gruplanmıştır. Her bir grubun kullanıldığı testlerden elde edilen eşitleme sonuçları kıyaslandığında, orta güçlükteki madde setinde eşitleme hatasının en az olduğu görülmüştür. Eşitleme modelleri karşılaştırıldığında ise Rasch modelinin Klasik Test Kuramı'na dayalı eşitleme yöntemlerine göre daha hassas eşitleme yaptığı tespit edilmiştir.

Skaggs ve Lissitz (1986), yaptıkları çalışmasında, farklı psikometrik özelliklere sahip testleri, yaygın olarak kullanılan dört eşitleme yöntemini (doğrusal, eşit yüzdelikli eşitleme, Rasch ve 3 parametrelili model) kullanarak eşitlemeyi hedeflemişlerdir. Benzer şekilde, Hills, Subhiyah ve Hirsch (1988) tarafından yapılan araştırmada 1984 ve 1986 yıllarında düzenlenen sınavlarda yer alan ortak maddeler kullanılarak testlerin eşitlemesi yapılmıştır. Araştırmada, Klasik Test Kuramı'na dayalı doğrusal eşitleme ile Madde Tepki Kuramı'na dayalı 1 ve 3 parametrelili modeller karşılaştırılmıştır. Araştırma sonuçlarına göre, testlerin iç tutarlığı ve güçlüklerinin benzer olduğu durumlarda belirtilen modellerin benzer eşitleme hatalarına sahip olduklarını göstermiştir. Diğer bir sonuç ise ortak test maddelerinin sayısı kıyaslandığında 10 maddelik ortak testlerin 15, 20, 25 ve 30 maddelik ortak testlere göre daha iyi eşitleme sonucu verdiği bulunmuştur.

Zeng (1991), normallik sayılısının karşılanması ve karşılanmaması durumunda, büyük örneklem için tek grup düzeneğinde doğrusal eşitlemeye ait dört

farklı eşitleme hatasını karşılaştırmıştır. Eşitleme hatalarının kestirilmesi için, bilgisayar simülasyonları ve gerçek veriler kullanılmıştır. Araştırmada yer alan eşitleme hataları; bootstrap yöntemi, iki delta yöntemi ve normallik sayılısının karşılanmaması durumunda hesaplanan eşitleme hatasıdır. Araştırmanın sonucunda ise, dağılımın normal olduğu durumlarda iki delta yöntemine ait eşitleme hatalarının yakın sonuçlar verdiğini göstermiştir. Sonuçlara bakıldığında, delta yönteminin bootstrap yöntemine tercih edilebileceği söylenmiştir.

Kelecioğlu (1994) tarafından yapılan araştırmada 1990, 1991 ve 1992 yılları arasında yapılan Öğrenci Seçme Sınavı (ÖSS) puanlarını, Klasik Test Kuramı ve Madde Tepki Kuramı'na dayalı eşitleme yöntemlerine göre eşitleme yapmıştır. Araştırmanın örneklemini belirtilen yıllarda ÖSS sınavına katılan 4000 aday oluşturmaktadır. ÖSS testleri sözde ortak (quasi anchor) maddelerle eşitlenmiştir. Bulgulara göre, Türkçe ve Fen Bilimleri alt testleri için en uygun yöntemlerin Klasik Test Kuramı'na dayalı yöntemler iken Sosyal Bilimler ve Matematik alt testleri için Madde Tepki Kuramı'na dayalı eşitlemenin uygun olduğu belirtilmiştir.

Tsai (1997) yaptığı araştırmada seçkisiz grup düzeneğinde minimum örneklem büyüklüğünü kestirmeyi amaçlamıştır. Örneklem büyüklüğünün ne kadar olması gerektiğini bulmak için eşitleme hatası ölçüt olarak alınmıştır. Araştırmada kullanılan eşitleme yöntemleri; ortalama, doğrusal ve eşit yüzdelikli eşitlemedir.

Suanthong (1998) tarafından yapılan, simülasyon verilerine dayalı araştırmada, çeşitli faktörlerin eşitleme sonuçlarına etkileri araştırılmıştır. Bunlar, madde güçlüğü'nün standart sapması, testteki madde sayısı, ortak madde sayısı ve örneklem büyüklüğüdür. Araştırmada ortak maddeli eşitlenmemiş gruplar üzerinde Rasch modeline dayalı eşitlemeler yapılmıştır. Bu eşitlemeler sonucunda en iyi eşitlemenin, .99 güçlük sapmasına sahip, 100 maddeli test için, 20 ortak maddenin kullanıldığı 100 kişilik örneklemelerde olduğu bulunmuştur.

Şahhüseyinoğlu (2005) tarafından yapılan araştırmada, 2000, 2001 ve 2002 yıllarında Hacettepe Üniversitesi'nce düzenlenen İngilizce Yeterlik Sınavı puanları üzerinde, Madde Tepki Kuramı ve Klasik Test Kuramı'na dayalı eşitleme yöntemleri incelenmiştir. Araştırma verilerini 2000 yılında sınava giren 2233; 2001 yılına sınava giren 1989 ve 2002 yılında sınava giren 2033 kişiye ait İngilizce Yeterlik sınav sonuçları oluşturmuştur. Araştırma sonucunda Rasch modelinin ve Klasik Test

Kuramı'na dayalı yöntemlerden eşit yüzdellik yönteminin, en az hata veren yöntemler olduğu saptanmıştır.

Kilmen ve Demirtaşlı (2012) ve Kim ve Kolen'in (2004) yaptığı araştırmalarda, Madde Tepki Kuramı'na dayalı "ortalama-ortalama", "ortalama-standart sapma", "Haebara" ve "Stocking-Lord" eşitleme yöntemlerinden kestirilen eşitleme hatalarının, yetenek dağılımı (benzer ve farklı yetenek dağılımı) ve örneklem büyüklüğü (500-1000 kişilik) değişkenlerine dayalı olarak karşılaştırılması amaçlanmıştır. Buna ilişkin, bir araştırmada, 1-0 şeklinde puanlanarak üretilmiş veriler, 3 parametrelili modele uyumlu olarak "ortak maddeli eşitlenmemiş gruplar eşitleme deseni" kullanılmıştır. Araştırmanın verilerini oluşturmak için WinGen2 programı kullanılmış olup yetenek ve örneklem büyüklüğüne göre altı grup oluşturulmuştur. Simülasyon verilerine ait yetenek parametreleri BILOG-MG programı yardımıyla "beklenen a posteriori" (expected a posteriori) yöntemi kullanılarak kestirilmiştir. IRTEQ programı kullanılarak testlerin eşitlemesi için gerekli eşitleme denklemleri oluşturulup; eşitleme hatası için RMSD değeri kriter olarak alınmıştır. Araştırmanın bulgularına göre, 3 parametrelili modele uygunluk gösteren testler için; 500 ve 1000 örneklem sayısına ve de benzer ve farklı yetenek düzeyine sahip gruplarda, Stocking-Lord yöntemi ile yapılan test eşitlemede eşitleme hatasının daha küçük olduğu saptanmıştır. Araştırmada, 500 kişilik aynı yetenek düzeyinde en yüksek eşitleme hatasını veren yöntem ortalama-standart sapma (Mean-sigma) yöntemi iken diğer üç durum için ortalama-ortalama (mean-mean) yöntemi ile en yüksek hata ile eşitleme gerçekleştirilmiştir. Örneklem büyüklüğüne dayalı olarak test eşitleme hataları incelendiğinde, 1000 kişilik örneklem büyüklüğüne sahip gruplara ait eşitleme hatalarının 500 kişilik örneklem büyüklüğünden elde edilen eşitleme hatalarına nazaran daha düşük olduğu bulunmuştur. Ayrıca benzer yetenek dağılımına sahip grupların eşitlenmesi ile elde edilen RMSD puanlarının, farklı yetenek dağılımına sahip grupların eşitlenmesinden elde edilen RMSD puanlarına nazaran daha düşük olduğu belirtilmiştir.

Kan (2011) çalışmasında farklı yıllarda uygulanan OKS test formlarının eşitliğini sınamak amaçlamıştır. Çalışmanın örneklemini 1030 sekizinci sınıf öğrencisi oluşturmaktadır. 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. Puan

dağılımında hataların 2 SHE'den fazla olması sebebiyle 2003 ve 2005 yıllarında uygulanan testlerin istatistiksel olarak birbirine eşlenemediği ve kestirim yaparken birbiri yerine kullanmanın doğru olmayacağı saptanmıştır.

Demir ve Güler (2014) çalışmasında PISA 2009 iki kitapçığının (Fen Bilimleri okuryazarlık) eşitlenmesinde Braun-Holland doğrusal eşitleme, Tucker doğrusal eşitleme ve frekans tahmin eşit yüzdellikli yöntemler kullanılmıştır. Çalışmadan elde edilen sonuçlar, orijinal puanlar ile Braun-Holland doğrusal eşitleme yöntemi ile elde edilen eşitlenmiş puanlar arasındaki farklılığın ve bu yöntemle ilişkin ağırlıklandırılmış hata kareleri ortalamalarının daha düşük olduğunu göstermektedir. Bu bulgulara dayalı olarak da Braun-Holland Doğrusal eşitleme yönteminin eşitlemede kullanılmasının daha uygun olduğu yorumu yapılabilir.

Bozdağ (2007) tarafından yapılan araştırmada, şans başarısının test eşitlemeye etkisi incelenmiştir. Tek grup desende düzenlenen çalışmada, doğrusal eşitleme ve eşit yüzdellikli eşitleme yöntemleri kullanılmıştır. Araştırmanın örneklemini, 2005–2006 eğitim-öğretim yılında Mersin ili sınırları içinde yer alan farklı ilköğretim okullarında öğrenim gören 1031 8. sınıf öğrencisi oluşturmaktadır. Araştırmanın sonucunda, şanstın arındırılmamış test puanlarının eşitlenmesinde doğrusal eşitlemenin, şanstın arındırılmış test puanlarının eşitlenmesinde ise eşit yüzdellikli eşitlemenin uygun yöntemler olduğu bulunmuştur.

Şans başarısının test eşitlemeye etkisini inceleyen diğer bir çalışmada, Aşiret ve Ömür-Sünbül (2014), şans parametresi, örneklem büyüklüğü (küçük örneklemlemlerle) ve formların güçlük düzeyleri arasındaki fark değişkenlerinin eşitleme hatalarına göre kıyaslamışlardır. Simülatif rastgele oluşturulan iki test formu özdeş (identity), ortalama (mean), dairesel yay (circle arc) ve $C=2$ ve $C=3$ düzeyinde log-linear ön-düzgünleştirme yapılmış eşit yüzdellikli eşitleme yöntemleri kullanılarak eşitlenmiştir. Bu çalışmada, şans başarısının eşitleme hata kestirimindeki etkisini incelemek amacıyla, c parametre değerleri 0, 0.10, 0.20 ve 0.25 olmak üzere dört düzey olarak belirlenmiştir. Şans parametre değerleri 0 ve 0.10 olduğu durumlarda Dairesel yay yönteminin, 0.20 ve 0.25 olduğu durumlarda ise ortalama eşitleme yönteminin en düşük RMSE değerine sahip olduğu saptanmıştır. Doğrusal eşitleme yöntemi ise her parametre değer aralığında en yüksek eşitleme hatasına sahip olduğu bulunmuştur.

Yang ve Houang (1996) tarafından yapılan bir çalışmanın sonucunda şansın bir parametre olarak dikkate alınabilmesi için test formlarının puan dağılımlarının sola çarpık olması gerektiğini belirtmişlerdir. Ayrıca Li, Griffith ve Tam (1997) şans parametresi kullanıldığında Madde Tepki Kuramına dayalı ortak madde içeren ve belirli ortak madde parametreleri olan desen (FCIP) kullanılarak elde edilen sonuçları karakteristik Eğri Metodu'nun (CCM) sonuçları ile kıyaslamışlardır.

İlgili alanyazın incelendiğinde, şans başarısının test eşitlemeyi etkileyen dengesiz bir parametre olarak görülmektedir. Şans başarısı ile ilgili yapılan çalışmalarda genellikle şans başarısını bir parametre olarak alıp, Madde Tepki Kuramına göre eşitleme çalışması ile hangi metodun etkili olduğuna bakılmıştır. Şans başarısının etkisinin ortaya konması açısından Klasik test teoremine göre de hangi yöntemin etkili olduğu bakılması gerekmektedir. Buna ilişkin Bozdağ'ın (2007) çalışmasına ek olarak farklı örneklem büyüklüğü ile sistematik ve rastgele hata payları birden çok katsayı değerleri ile hesaplanmıştır.

Bölüm 3

Yöntem

Araştırmanın Evreni ve Örneklemi

İki tür evren vardır (Fraenkel & Wallen, 2010); hedef evren (target population) ve ulaşılabilir evren (accessible population). Hedef evren araştırmacının araştırması için çalışabileceği bütün evrendir ve ulaşılması hemen hemen zordur (Fraenkel & Wallen, 2010). Ulaşılabilir evren ise araştırmanın gerçekçi seçimidir yani araştırması için ulaşabildiği örneklemdir. (Fraenkel & Wallen, 2010). Araştırmada, 2015-2016 eğitim öğretim yılında yapılan Türkiye'den belirli okullarının katıldığı (22 okul) TEOG Değerlendirme sınavı İngilizce I ve II testlerine cevap veren 8. Sınıf ortaokul öğrencileri (N=1681) bu çalışmanın örneklemdir.

Veri Toplama Süreci

Araştırmada örneklem, 20 maddelik 8. Sınıf yabancı dilde (İngilizce) verilen temaya göre okuma ve yazma becerisine ilişkin çoktan seçmeli sorulardan oluşan değerlendirme testine katılan okullardaki öğrencilerin seçilmesi ile oluşturulmuştur. Tek grup deseni kullanarak, şans başarısından arındırılmayarak ve arındırılarak İngilizce I ve İngilizce II testine ilişkin puanlar eşitlenmeye çalışılmıştır.

Verilerin Analizi

Eşitleme işlemine geçilmeden önce, her iki teste ait puanlar düzeltme formülüyle şanstın arındırılmış ve iki farklı ham puan türü elde edilecektir. Güçlük düzeyleri aynı olduğu düşünülen bu testler, benzer yetenek düzeyine sahip bireylere uygulandığı için yatay eşitleme yapılmıştır. Verilerin analizi üç aşamada gerçekleştirilmiştir. Öncelikle, eşitleme koşullarının sağlanıp sağlanmadığı test edilmiş olup; eşitleme koşullarının sağlanıp sağlanmadığı için öncelikli olarak puanlarının tek boyutlu bir yapıya sahip olup olmadıklarına karar vermek üzere her iki testten elde edilen veriler üzerinde temel bileşenler analizi uygulanmıştır (Angoff, 1971; Felan, 2002; Tanguma 2000; Woldbeck, 1998). İki testten elde edilen puanlar arasındaki ilişkinin düzeyini belirlemek üzere Pearson momentler çarpımı korelasyon katsayısı, KR-20 güvenirlik katsayıları hesaplanacak ve daha sonra bu katsayılar arasında istatistiksel olarak fark olup olmadığı saptanacaktır. İkinci olarak

ise, sneuquate R paketi (Burgos, 2017) kullanılarak doğrusal ve eşit yüzdelli eşitleme yapılarak eşitlenmiş puanlar elde edilecektir. Üçüncü aşamada ise, her bir eşitleme yöntemine ait hata kareleri ortalamaları hesaplanıp hangi yöntemin daha az eşitleme hatasına sahip olacağı araştırılacaktır.

Her bir eşitleme yöntemi için eşitleme katsayıları kullanılarak EXCEL programı ile eşitleme hataları değerlendirme kriteri olarak hesaplanmıştır

Şans Başarısından Arındırılmamış İngilizce Testlerine İlişkin Varsayımlar

Tek boyutluluk. İngilizce 1-2 alt testlerine ait faktör analizi sonuçları Tablo 6'da verilmiştir.

Tablo 6

Alt Testlere İlişkin Faktör Analizi Sonuçları

| İngilizce 1 | | | | İngilizce 2 | | | |
|-------------|---------|-----------------|-------------------|-------------|---------|-----------------|-------------------|
| Faktör | Özdeğer | Varyans Yüzdesi | Açıklanan varyans | Faktör | Özdeğer | Varyans Yüzdesi | Açıklanan varyans |
| 1 | 10,463 | 0,523 | 52,3 | 1 | 10,554 | 0,527 | 52,7 |
| 2 | 1,453 | 0,072 | 7,2 | 2 | 1,248 | 0,062 | 6,2 |

Tablo 6 incelendiğinde; her iki test için de özdeğeri 1'den büyük 2 faktörün olduğu ancak hem özdeğerler hem de varyans açıklama oranları arasında birinci faktörden sonra yaklaşık 10 katlık bir düşüşün olduğu elde edilmiştir. İngilizce 1 testinin toplam varyansın %52.3'ünü ve İngilizce 2 testinin varyansın %52.7'sinin birinci faktör tarafından açıklandığı görülmektedir. Bu sonuçlar göz önüne alındığında, İngilizce testlerinin tek boyutlu olduğu söylenebilir.

Tablo 7

Ham Puanlara İlişkin Betimsel İstatistik Tablosu

| | K | N | X | Sx | S ² x | P | Kr-20 | Çarpıklık | Basıklık |
|------------|----|------|--------|-------|------------------|-------|-------|-----------|----------|
| İngilizce1 | 20 | 1681 | 14,670 | 4,537 | 20,581 | 0,734 | 0,882 | -0,932 | -0,27 |
| İngilizce2 | 20 | 1681 | 15,890 | 4,484 | 20,106 | 0,795 | 0,888 | -1,126 | 0,307 |

Ham puanlara ilişkin betimsel istatistik tablosuna (Tablo 7) göre İngilizce 1 testinin ortalaması 14,670 iken İngilizce 2 testinin ortalaması 15,890 olarak elde edilmiştir. Standart sapmalar ve varyanslar ise birbirine yakındır. İngilizce 2 testinin güçlüğü (p) 0.795 iken, İngilizce 1 testinin güçlüğü 0.734 olarak elde edilmiştir. İkinci test görece biraz daha kolaydır. Güvenirlikleri 0.882 ve 0.888 olarak elde edilmiştir ve iki testin güvenilirliği de yüksektir. Çarpıklık ve basıklık değerleri -2 ile +2 arasında olduğunda puanlar normal dağılmaktadır.

Güvenirlikleri arasındaki farkın test edilmesi. Eşitlenecek İngilizce testlerinin, güvenilirliklerinin eşit olup olmadığını belirlemek için her bir alt teste ait KR-20 güvenilirlik katsayıları hesaplanmıştır. Elde edilen bu güvenilirlik katsayıları korelasyon katsayısı olarak kabul edilmiş ve iki güvenilirlik katsayısı arasında fark olup olmadığı Fischer'in Z istatistiği ile test edilmiştir (Akhun, 1984). Güvenirlik katsayıları arasındaki farkın test edilmesine ilişkin bulgular Tablo 8'de görüldüğü gibidir.

Tablo 8

İngilizce Testlerinin Güvenirliklerinin Karşılaştırılması

| Testler | KR-20 | Zr | Z | P |
|------------|-------|-------|------|------|
| İngilizce1 | 0.882 | 1.385 | 0.80 | 0.42 |
| İngilizce2 | 0.888 | 1.412 | | |

Not: P<0.5

Tablo 8 incelendiğinde; İngilizce testlerin güvenilirliklerin eşit ve bu yüzden aralarında. 05 anlamlılık düzeyinde manidar bir fark olmadığı görülmüştür.

Ortalama güçlükleri arasındaki farkın test edilmesi. İngilizce 1 ve İngilizce 2 testlerinin ortalama güçlükleri arasında anlamlı bir farklılık olup olmadığını belirlemek için iki oran farkı testi yapılmıştır (Baykul, 1996). İki oran fark testine ilişkin bulgular Tablo 9'da verilmiştir.

Tablo 9

Ortalama Güçlükleri Arasındaki Fark Testi

| Testler | K | P | t | P |
|------------|----|-------|-------|-------|
| İngilizce1 | 20 | 0.734 | -0.45 | 0.328 |
| İngilizce2 | 20 | 0.794 | | |

Tablo 9 incelendiğinde, testlerin ortalama güçlükleri arasında .05 anlamlılık düzeyinde manidar bir fark olmadığı elde edilmiştir. Bu sonuç, testlerin ortalama güçlüklerin eşit olma koşulunu sağladığını göstermektedir.

Ortalamaları ve varyansları arasındaki farkın test edilmesi. Bu çalışmada tek grup deseni kullanıldığı için, testlerin ortalamaları arasında anlamlı bir fark olup olmadığı bağımlı gruplar *t testiyle* ve *testlerin varyanslarının eşit olup olmadığı* ise *Levene's testi* ile incelenmiştir.

Tablo 10

Ortalamalar Arasındaki Fark Testi

| Testler | Bağımlı Gruplar t Testi | | | | Levene Test | |
|------------|-------------------------|--------|------|----------|-------------|------|
| | x | t | p | Eta Kare | F | p |
| İngilizce1 | 14,67 | -7,839 | ,000 | ,018 | 2,212 | ,137 |
| İngilizce2 | 15,88 | | | | | |

İki alt test sonuçlarının ortalamaları arasında istatistiksel olarak anlamlı fark elde edilmiştir ($t=-7,839$, $p=.000$, $p<.01$). Birden çok etki büyüklüğü istatistikleri yer almasına rağmen en yaygın kullanılanı Cohen d olarak bilinen istatistiktir. Cohen (1988)'e göre 0.018 olarak bulunan etki büyüklüğü küçüktür.

İki test arasında büyük örneklem olduğu için elde edilen 1 puanlık anlamlı farkın etkisi küçüktür. Ayrıca varyansların homojenliği testi sonuçlarına göre test varyansları homojendir ($F=2.212$, $p=.137$, $p>.05$)

Tüm varsayımlar kontrol edildikten sonra şans başarısından arındırılmış ham İngilizce puanları arasındaki korelasyon 0.819 olarak elde edilmiştir. Bu korelasyon yüksek bir korelasyondur. Düzeltme formülü uygulanarak şans başarısından arındırılmamış toplam puanlar elde edildiğinden bu iki değişken arasındaki ilişki düzeyini değiştirmemiştir. Dolayısıyla korelasyon şans başarısından arındırılmamış iki İngilizce testi için de aynıdır.

Şans Başarisından Arındirilmış İngilizce Testlerine İlişkin Varsayımlar

Şans başarısında toplam puanlar deęiştii ve ham veri matrisi deęişmedięi için faktör analizi sonuçlarında farklılık yoktur. Dolayısıyla dięer varsayımlar kontrol edilmiştir.

Tablo 11

Şans Başarisından Arındirilmış Puanların Betimsel İstatistikleri

| | K | N | X | Sx | S2x | p | Kr-20 | Çarpıklık | Basıklık |
|------------|----|------|--------|-------|--------|-------|-------|-----------|----------|
| İngilizce1 | 20 | 1681 | 12,894 | 6,049 | 36,589 | 0,645 | 0,957 | -0,932 | -0,27 |
| İngilizce2 | 20 | 1681 | 14,520 | 5,979 | 35,746 | 0,726 | 0,960 | -1,126 | 0,306 |

Şans başarısından arındirilmış betimsel istatistik tablosu incelendiğinde iki testin ortalaması ve ortalama güçlüğü düşmüştür. Güvenirlikleri incelendiğine iki testin güvenilirliği de artmıştır. Çarpıklık ve basıklık deęerlerine göre puanların dağılımı normaldir.

Güvenirlikleri arasındaki farkın test edilmesi. Şans başarısından arındirilmış testlerin toplam puanlarının ve varyanslarının deęişmesinden dolayı güvenilirlikleri de deęişmiştir. Güvenirlik katsayıları arasında fark olup olmadığı tablo 12'de gösterilmiştir.

Tablo 12

Güvenirlik Testi

| Testler | KR-20 | Zr | N | Z | P |
|------------|-------|-------|------|------|------|
| İngilizce1 | 0,957 | 1,909 | 1681 | 1,07 | 0,28 |
| İngilizce2 | 0,960 | 1,946 | 1681 | | |

Şans başarısından arındirilmış test puanlarına göre ilk İngilizce testinin güvenilirliği 0.957 ikincisinin ise 0.960 olarak elde edilmiştir. Güvenirlikleri arasında fark olup olmadığı fischer testi ile analiz edilmiştir ve güvenilirlikleri arasında istatistiksel olarak anlamlı fark yoktur ($Z=1.07$, $p=.28$, $p>.05$). Yani şans başarısından arındirilmemiş testlerin güvenilirlikleri benzerdir.

Ortalama güçlükleri arasındaki farkın test edilmesi. İngilizce 1 ve İngilizce 2 testlerinin ortalama güçlükleri arasında anlamlı bir farklılık olup olmadığını belirlemek için iki oran farkı testi yapılmıştır. İki oran fark testine ilişkin bulgular Tablo 13’de verilmiştir.

Tablo 13

Ortalama Güçlükleri Arasındaki Fark Testi

| Testler | K | P | T | P |
|------------|----|-------|--------|-------|
| İngilizce1 | 20 | 0,645 | -0,376 | 0,711 |
| İngilizce2 | 20 | 0,726 | | |

Tablo 13 incelendiğinde, testlerin ortalama güçlükleri arasında .05 anlamlılık düzeyinde manidar bir fark olmadığı elde edilmiştir. Bu sonuç, şans başarısından arındırılmış testlerin ortalama güçlüklerin eşit olma koşulunu sağladığını göstermektedir.

Ortalamları ve varyansları arasındaki farkın test edilmesi. Bu çalışmada testlerin ortalamaları arasında anlamlı bir fark olup olmadığı bağımlı gruplar *t testiyle* ve testlerin varyanslarının eşit olup olmadığı ise *Levene’s testi* ile incelenmiştir.

Tablo 14

Ortalama ve Varyansların Kıyaslanması

| Testler | Bağımlı Gruplar t Testi | | | | Levene Test | |
|------------|-------------------------|---------|-------|----------|-------------|------|
| | x | T | p | Eta Kare | F | p |
| İngilizce1 | 12,894 | --7,838 | 0.000 | .018 | 2,200 | ,138 |
| İngilizce2 | 14,520 | | | | | |

İki alt test sonuçlarının ortalamaları arasında istatistiksel olarak anlamlı fark edilmiştir ($t=-7,839$, $p=.000$, $p<.01$). Birden çok etki büyüklüğü istatistikleri yer almasına rağmen en yaygın kullanılanı Cohen d olarak bilinen istatistiktir. Cohen (1988)’e göre hesaplanan 0.018 etki değeri küçüktür. İki test arasında büyük örneklem olduğu için elde edilen 1 puanlık anlamlı farkın etkisi küçüktür. Ayrıca

varyansların homojenliği testi sonuçlarına göre test varyansları homojendir (F=2.200, p=.138, p>.05).

Şans başarısından arındırılmış ve arındırılmamış İngilizce alt testlerinden elde edilen puanların eşitleme için uygun olduğu belirtilmiş ve elde edilen puanlar arasındaki ilişkinin düzeyini belirlenmiştir.

Eşitleme Hata Kriterleri

Eşitlenmiş puanların hata oranlarını belirlemek için bakılan katsayılardan biri her bir ham puan ile bu ham puana karşılık gelen eşitlenmiş puanları kıyaslayan ağırlıklandırılmış hata kareleri ortalamalarıdır (WMSE). WMSE ile eşitleme çalışmalarında kullanılan yöntemin uygunluğu değerlendirilebilmektedir (Skaggs ve Lissitz (1986), Diğer bir ölçüt ise Dorans ve Holland (2009) tarafından sunulan Kareler Farkının Ortalamasının Kare Kökü'dür (RMSD). Bu ölçüm, tek grup /eşdeğer grup desenleri içindeki tüm popülasyon için hesaplanan fonksiyon ile farklı alt popülasyonlarda hesaplanan eşitleme (veya ilişkilendirme) fonksiyonlarını karşılaştırır (Davies & Han, 2004). Diğer bir anlatımla bu kriter bireylerin sahip olduğu yetenek düzeyleri ile kestirilen yetenek düzeyleri arasındaki farkın kare ortalamalarının karekökünü temel almaktadır (Harris & Crouse, 1993). İşaretili farkların ortalaması (MSD) ise sistematik hata için kullanılan diğer bir indekstir.

WMSE, RMSD VE MSD katsayılarına ilişkin ilişkili formüller aşağıda sunulmuştur.

$$WMSE = \frac{\sum_{i=1}^{k-1} f_i (X_E - X_{crit})^2}{\sum_{i=1}^k f_i S^2_y}$$

k: Y testindeki madde sayısı

S²_y: Y testindeki ham puanların varyansı

X_{crit}: Y testindeki i. ham puan

X_E: Eşitleme sonucunda X testindeki i. ham puana karşılık gelen puan

f_i: Y testindeki i. ham puan frekansı

$$RMSD = \sqrt{\frac{\sum_{i=1}^{k-1} f_i (X_E - X_{crit})^2}{\sum_{i=1}^k f_i}}$$

$$MSD = \frac{1}{N} \sum_{j=1}^N X_E - X_{crit}$$

Yürütölen bir eşitleme çalıřmasının başarılı olabilmesi için rastgele ve sistematik eşitleme hatasını en aza indirmek gerekmektedir. Eşitleme hatası çok büyük olduğunda formları eşitlememek daha uygun bir yol olabilir (Kolen ve Brennan, 2004).

Bölüm 4

Bulgular

Bu bölümde araştırmanın alt problemlerindeki sorularına cevap verilmiştir. Bulgular alt problemlerdeki sıraya göre aşağıda sunulmuştur. Ayrıca bulgulara dayalı yorumlara yer verilmiştir.

Alt problemler

Çalışmada cevabı aranan üç alt probleme ait bulgular ve yorumları başlıklar halinde sunulmuştur.

Alt problem 1. İngilizce testlerine ait şans başarısından arındırılmamış puanların doğrusal ve eşit yüzdelikli eşitleme yöntemleri kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

Araştırmanın verilen alt problemine cevap verebilmek için öncelikle İngilizce testine ait ham puanlar hesaplanmıştır. Şans başarısından arındırılmamış test puanlarının eşitlenmesine yönelik olarak doğrusal ve eşit yüzdelikli eşitleme yöntemleri ile puanlar eşitlenmiştir. Tablo 15’de eşitleme işlemlerine ait sonuçlar verilmiştir. Ham puanlar ile her iki yöntem ile yapılan eşitleme sonuçları ile fark puanları yer almaktadır.

Tablo 15

Eşitleme Sonucunda Elde Edilmiş Puanlar

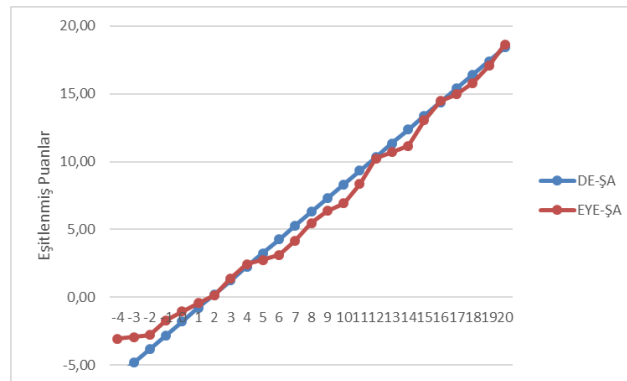
| Puan | Doğrusal Eşitleme | Fark | Eşit Yüzdelikli Eşitleme | Fark |
|------|-------------------|-------|--------------------------|-------|
| 0 | 1,39 | -1,39 | -0,50 | 0,50 |
| 1 | 2,38 | -1,38 | -0,61 | 1,50 |
| 2 | 3,37 | -1,37 | 0,83 | 1,17 |
| 3 | 4,35 | -1,35 | 3,38 | -0,38 |
| 4 | 5,34 | -1,34 | 4,53 | -0,53 |
| 5 | 6,33 | -1,33 | 5,88 | -0,88 |
| 6 | 7,32 | -1,32 | 6,91 | -0,91 |
| 7 | 8,31 | -1,31 | 8,36 | -1,36 |
| 8 | 9,30 | -1,30 | 9,77 | -1,77 |
| 9 | 10,29 | -1,29 | 10,87 | -1,87 |

| | | | | |
|----|-------|-------|-------|-------|
| 10 | 11,27 | -1,27 | 11,77 | -1,77 |
| 11 | 12,26 | -1,26 | 12,68 | -1,68 |
| 12 | 13,25 | -1,25 | 13,59 | -1,59 |
| 13 | 14,24 | -1,24 | 14,46 | -1,46 |
| 14 | 15,23 | -1,23 | 15,47 | -1,47 |
| 15 | 16,22 | -1,22 | 16,26 | -1,26 |
| 16 | 17,20 | -1,20 | 17,19 | -1,19 |
| 17 | 18,19 | -1,19 | 18,22 | -1,22 |
| 18 | 19,18 | -1,18 | 19,33 | -1,33 |
| 19 | 20,17 | -1,17 | 20,02 | -1,02 |
| 20 | 21,16 | -1,16 | 20,41 | -0,41 |

Tablo 15 incelendiğinde, ham puanların 0 ile 20 aralığında olduğu, doğrusal eşitlemeye göre elde edilen eşitlenmiş puanların 1.39 ile 21.16 arasında ve eşit yüzdelikli eşitlemeye göre elde edilen eşitlenmiş puanların ise -0.50 ile 20.41 arasında değerler aldığı görülmektedir.

Tabloya göre, doğrusal eşitlemeye göre puan ölçeği boyunca eşitlenmiş puanların ham puanlardan büyük olduğu; eşit yüzdelikli eşitlemeye göre ise 0-2 ham puanlara ait eşitlenmiş puanların ham puanlardan küçük, 3 ile 20 arasındaki eşitlenmiş puanların ise ham puanlardan büyük olduğu görülmektedir.

Ham puanlar ile eşitleme yöntemlerine göre şanstan arındırılmamış puanlara karşılık gelen eşitlenmiş puanlara ait grafik Şekil 2. 'de verilmiştir.



Şekil 2. Şanstan arındırılmamış puan dağılımı.

Şekil 4.1. incelendiğinde, İngilizce testlerinin şanstan arındırılmamış ham puanları ile eşitlenmiş puanlar arasında doğrusal bir ilişki varken, , eşit yüzdellikli eşitleme yöntemine göre elde edilen eşitlenmiş puanlar arasında doğrusallıktan sapmalar olduğu görülmektedir.

Alt problem 2. İngilizce testlerine ait şans başarısından arındırılmış puanların doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

Araştırmanın verilen alt problemine cevap verebilmek için öncelikle İngilizce testlerine ait ham puanlar hesaplanmıştır. Şans başarısından arındırılmış test puanlarının eşitlenmesine yönelik olarak doğrusal ve eşit yüzdellikli eşitleme yöntemleri ile puanlar eşitlenmiştir. Tablo 16'da eşitleme işlemlerine ait sonuçlar verilmiştir. Ham puanlar ile her iki yöntem ile yapılan eşitleme sonuçları ile fark puanları yer almaktadır.

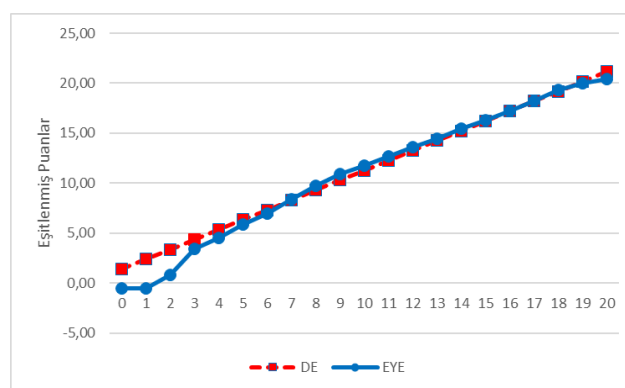
Tablo 16

Şanstan Arındırılmış Puanların Eşitlenmiş Puanları

| Puan | Doğrusal Eşitleme | Fark | Eşit Yüzdellikli Eşitleme | Fark |
|------|-------------------|------|---------------------------|-------|
| -4 | -5,84 | 1,84 | -3,05 | -0,95 |
| -3 | -4,83 | 1,83 | -2,94 | -0,06 |
| -2 | -3,82 | 1,82 | -2,78 | 0,78 |
| -1 | -2,81 | 1,81 | -1,69 | 0,69 |
| 0 | -1,80 | 1,80 | -1,06 | 1,06 |
| 1 | -0,78 | 1,78 | -0,44 | 1,44 |
| 2 | 0,23 | 1,77 | 0,14 | 1,86 |
| 3 | 1,24 | 1,76 | 1,40 | 1,60 |
| 4 | 2,25 | 1,75 | 2,45 | 1,55 |
| 5 | 3,26 | 1,74 | 2,75 | 2,25 |
| 6 | 4,27 | 1,73 | 3,11 | 2,89 |
| 7 | 5,29 | 1,71 | 4,18 | 2,83 |
| 8 | 6,30 | 1,70 | 5,49 | 2,51 |
| 9 | 7,31 | 1,69 | 6,36 | 2,64 |

| | | | | |
|----|-------|------|-------|------|
| 10 | 8,32 | 1,68 | 6,92 | 3,08 |
| 11 | 9,33 | 1,67 | 8,34 | 2,66 |
| 12 | 10,34 | 1,66 | 10,24 | 1,76 |
| 13 | 11,36 | 1,64 | 10,71 | 2,29 |
| 14 | 12,37 | 1,63 | 11,16 | 2,84 |
| 15 | 13,38 | 1,62 | 13,04 | 1,96 |
| 16 | 14,39 | 1,61 | 14,49 | 1,51 |
| 17 | 15,40 | 1,60 | 14,95 | 2,05 |
| 18 | 16,41 | 1,59 | 15,79 | 2,21 |
| 19 | 17,43 | 1,57 | 17,07 | 1,93 |
| 20 | 18,44 | 1,56 | 18,64 | 1,36 |

Tablo 16 incelendiğinde, ham puanların -4 ile 20 aralığında olduğu, doğrusal eşitlemeye göre elde edilen eşitlenmiş puanların -5.84 ile 18.44 arasında ve eşit yüzdelikli eşitlemeye göre elde edilen eşitlenmiş puanların ise -3.05 ile 18.64 arasında değerler aldığı görülmektedir. Tabloya göre, doğrusal eşitleme ile elde edilen eşitlenmiş puanların puan ölçeği boyunca ham puanlardan küçük; eşit yüzdelikli eşitleme ile elde edilen eşitlenmiş puanların -4 ile -3 puanlarında ham puanlardan büyük, -2 ile 20 arasında ise küçük olduğu görülmektedir.

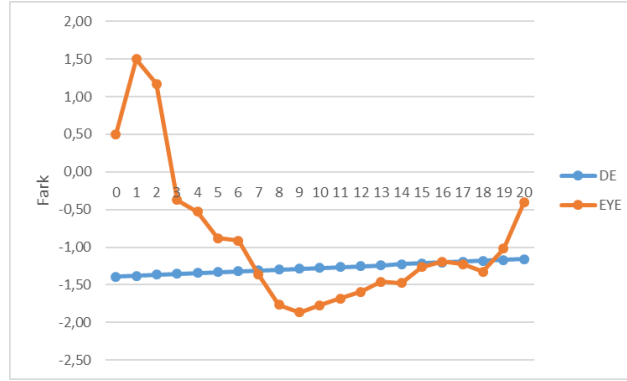


Şekil 3. Şanstan arındırılmış puanların dağılımı.

Şekil 3.'teki grafik incelendiğinde, ham puanlar ile eşitlenmiş puanlar arasında yer yer dalgalanmalar olmasına rağmen doğrusal bir ilişki olduğu söylenebilir ve bu ilişki $Y = 0,99X + 1,39$ doğru denklemiyle ifade edilebilir. Şanstan

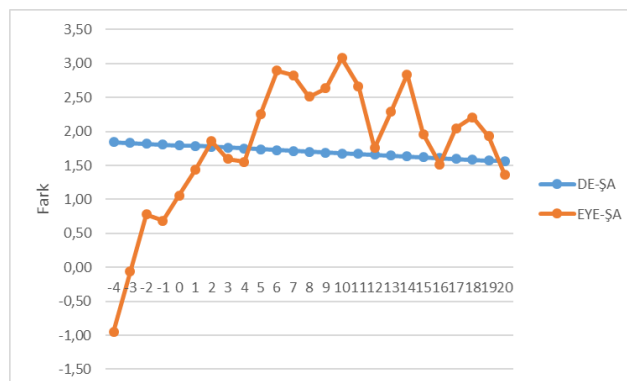
arındırılmadan eşit yüzdelikli eşitlenmiş puan dağılımlarına göre, eşit yüzdelikli eşitlemeye göre bu puan dağılımlarında doğrusallıktan daha az sapmalar olduğu görülmektedir.

Şanstan arındırılmış ve arındırılmamış ham puanlar ile eşitlenmiş puanlar arasındaki farka ait grafik Şekil 4.'te yer almaktadır.



Şekil 4. Şanstan arındırılmamış ham puanlara göre eşitleme yöntemlerinden elde edilen fark grafiği.

Şekil 4.'te yer alan grafik incelendiğinde, doğrusal eşitleme için ham puanlar ve eşitlenmiş puanlar arasındaki farkın doğrusal olduğu, eşit yüzdelikli eşitlemeye göre şanstan arındırılmamış puanlardan elde edilen eşitlenmiş puanların ise 0-2 puan aralığında pozitif, 3-20 puan aralığında ise negatif olduğu görülmektedir.



Şekil 5. Şanstan arındırılmış ham puanlara göre eşitleme yöntemlerinden elde edilen fark grafiği.

Şekil 5.'te yer alan grafik incelendiğinde, doğrusal eşitleme için ham puanlar ve eşitlenmiş puanlar arasındaki farkın doğrusal olduğu, eşit yüzdelikli eşitlemeye göre şanstan arındırılmış puanlardan elde edilen eşitlenmiş puanların ise -4 ile -3 puanlarında pozitif, -2 ile 20 puan aralığında ise negatif olduğu görülmektedir.

Alt problem 3. İngilizce testlerinin ham puanlarının şans başarısından arındırılarak eşitlenmesinde eşit yüzdelikli eşitleme ve doğrusal eşitleme yöntemlerinden hangisi kullanıldığında hata payı en azdır?

Bu çalışmada İngilizce testlerinin şanstan arındırılmış ve şanstan arındırılmamış ham puanları doğrusal ve eşit yüzdelikli eşitleme yöntemleriyle eşitlenmiştir. Şanstan arındırılmış ve şanstan arındırılmamış ham puanlar ile yapılan eşitleme yöntemlerinden hangisinin daha uygun olduğunu belirlemek ve alt problem 3'e cevaplamak için ağırlıklandırılmış hata kareleri ortalamalarını veren WMSE, kareler farkının ortalamasının karekökü olan RMSD ve işaretli farkların ortalamasını veren MSD katsayıları hesaplanmıştır. Elde edilen bulgular Tablo 4.3'te verilmiştir.

Tablo 17

Doğrusal ve Eşit Yüzdelikli Yönteme Ait Hata Değerleri

| | DE | EYE | DE-ŞA | EYE-ŞA |
|------|-------|-------|-------|--------|
| WMSE | 0.072 | 0.067 | 0.074 | 0.105 |
| RMSD | 1.456 | 1.338 | 1.646 | 1.963 |
| MSD | 0.016 | 0.011 | 0.025 | 0.027 |

Tablo 17'deki bilgilere göre, şans başarısından arındırılmış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan RMSD değeri 1.646 iken, şans başarısından arındırılmamış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan RMSD değeri 1.456'dır. Daha önce yapılan eşitleme varsayımlarına ilişkin sonuçlar şans başarısından arındırılan puanlara ait ortalama ve varyanslar arasındaki farkın arttığını fakat manidar olmadığını göstermiştir. Başka bir açıdan bakılacak olursa ortalama ve varyanslar arasındaki farkın artmasının doğrusal eşitleme ile şans başarısından arındırılmış puanların eşitleme hatasını arttırdığı söylenebilir. Buna ek olarak, eşit yüzdelikli yöntemin hata paylarına bakıldığında, Şans başarısından arındırılmış ham puanların eşit yüzdelikli

yöntemiyle eşitlenmesiyle hesaplanan RMSD değeri 1.963 iken, şans başarısından arındırılmamış ham puanların eşit yüzdelikli eşitleme yöntemiyle eşitlenmesiyle hesaplanan RMSD değeri 1.338'dir.

Şans başarısından arındırılmış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan WMSE değeri 0.074 iken, şans başarısından arındırılmamış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan WMSE değeri 0.072'dir. Şans başarısından arındırılmış ham puanların eşit yüzdelikli eşitleme yöntemiyle eşitlenmesiyle hesaplanan WMSE değeri 0.105 iken, şans başarısından arındırılmamış ham puanların eşit yüzdelikli eşitleme yöntemiyle eşitlenmesiyle hesaplanan WMSE değeri 0.067'dir. Bu değerler, RMSD şanstın arındırılmış ve arındırılmamış puanlara ilişkin eşitleme hatalarına benzerlik göstermektedir. Buna göre, ek olarak şanstın arındırıldığında doğrusal eşitleme, arındırılmadan eşit yüzdelikli eşitlemenin uygun olduğu bulunmuştur. Eşit yüzdelikli eşitleme yaparken puan aralığının sık ve gergin olması ile ham puanların eşitlenmiş puanla örtüşmesi sağlanır (Angoff, 1971). Şanstın arındırılmamış eşitlemede, elde edilen puan aralığı 1 iken benzer şekilde şanstın arındırılmış puanlarda da eşitlemede puan aralığı 1'dir. Fakat puanların ranjı gerçek puan ranjına göre daha fazla değişiklik göstermiştir. Bu durum eşit yüzdelikli eşitleme yöntemine ait hata oranının artmasına neden olduğu söylenebilir.

Şans başarısından arındırılmış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan sistematik hatayı veren MSD değeri 0.025 iken, şans başarısından arındırılmamış ham puanların doğrusal eşitleme yöntemiyle eşitlenmesiyle hesaplanan MSD değeri 0.016'dır. Şans başarısından arındırılmış ham puanların eşit yüzdelikli eşitleme yöntemiyle eşitlenmesiyle hesaplanan MSD değeri 0.027 iken, şans başarısından arındırılmamış ham puanların eşit yüzdelikli eşitleme yöntemiyle eşitlenmesiyle hesaplanan MSD değeri 0.011'dir. Bu bulguya dayalı olarak, rastgele hata ile sistematik hata bu çalışma için her bir yöntem değerlendirilmesinde benzer sonuçları vermiştir.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu çalışmada 2015 TEOG Genel Değerlendirme İngilizce alt testlerine ait puanların şans başarısından arındırılmadan ve arındırılarak eşitlenmesinde doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanılarak incelemiştir. Araştırmaya ait sonuç ve öneriler aşağıda sunulmuştur.

Araştırmanın bir sonucu olarak şans başarısından arındırılmamış ve arındırılmış puanların doğrusal eşitlemeye göre elde edilen eşitlenmiş puanların ham puan aralığı ile örtüşmediği bulunmuştur. Livingston (2004) bu durumun doğrusal eşitlemenin doğası olduğunu belirtmiştir.

Diğer bir bulgu ise, şans başarısından arındırılmamış ve arındırılmış puanlarda eşit yüzdellikli eşitlemeye göre elde edilen eşitlenmiş puanlar ile ham puanlar arasında doğrusal bir ilişki olmasına karşın grafikte bazı değer aralıklarında dalgalanmaların olduğu saptanmıştır. Bu bulgu Anıl ve Öztürk (2012)'ün çalışma bulgularına benzerlik göstermektedir. İngilizce I ve II testlerinden elde edilen puanların ortalama, standart sapma, çarpıklık ve basıklık değerlerinin aynı olmaması bu dalgalanmaların nedeni sayılabilir. Ek olarak, şanstın arındırılmış ve arındırılmamış puanların ham puan aralığının dışında değerler aldığı görülmüştür. Bu durum, eşit yüzdellikli eşitleme yapmadan önce düzgünleştirme işlemi yapılmamasından dolayı puanların kestirilmesinde hata karışmış olabileceğini desteklemektedir. Kolen ve Brennan'a (1995) göre eşit yüzdellikli eşitleme ile üst ve alt puanlarda hangi testin daha zor olduğu çıkarımı yapılabilir. Bu çalışma için İngilizce I testi alt puanlarda İngilizce II testinden daha zor olduğu söylenebilir.

Şans başarısından arındırılmış ve arındırılmamış puanların doğrusal ve eşit yüzdellikli eşitleme yöntemleri kullanıldığında hataları hesaplanmış ve hataların küçük olduğu görülmüştür. 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 alması bu düzeneği güçlü kılan bir özelliktir ve bu sebeple de bu düzenek aracılığıyla elde edilen eşitleme kesinliği diğer düzeneklere göre daha yüksek, eşitleme hatası ise daha düşüktür (Crocker ve Algina 1986; Dorans, 1990; Felan 2002; Livingston, 2004; Zhu, 1998). Bu bulgulara

dayanarak, arařtırmada tek grup d zeneđinin kullanılması, eřitleme hatasına iliřkin deđerin olduka k  k ıkmasına neden olduđu s ylenebilir.

Son olarak, hesaplanan hata deđerlerine g re, řans bařarisından arındırılmamıř İngilizce I-II alt test puanlarının dođrusal ve eřit y zdelikli y ntem kullanılarak eřitlenmesinde, eřit y zdelikliye ait eřitleme hatasının daha k  k olduđu bulunmuřtur. Bu bulgular,  zt rk (2010), řahh seyinođlu (2005) ve Tsai (1997) tarafından yapılan alıřmalarının bulgularına paraleldir. Fakat Bozdađ ve Zhu (1998) alıřmalarında arındırılmamıř puanların dođrusal eřitleme y ntemi ile eřitlendiđinde eřit y zdelikli eřitlemeye g re daha az hata verdiđini bulmuřtur. Bahsedilen alıřmanın  rneklemini 1000'e yakinken, bu alıřmada  rneklem 1500'den fazladır. Kolen ve Brennan (2004)  rneklem sayısının 1500'den daha b y k olduđu durumlarda eřit y zdelikli eřitlemenin daha dođru sonular verdiđini belirtmiřtir. Diđer bir aıdan,  rneklem b y kl đ  ile řans bařarısı arasında bir iliřki olduđu varsayımından yola ıkılırsa k  k  rneklem grubunda řans bařarisının test eřitlemeye etkisi alıřılabilir. Buna ek olarak, Kolen ve Brennan (2014), eřit puanların ham puan aralıđından -5 ila +5'e kadar farklılařabileceđini, bu da eřit y zdelikli eřitleme iin istenen bir  zellik olduđunu iddia etmektedir. Puan aralıđının -0.50 ve 20.41 olduđundan dolayı eřit y zdelikli eřitlemenin daha uygun sonu verdiđi d ř n lebilir. Fakat řans bařarisından arındırılmıř İngilizce I-II alt test puanlarının eřit y zdelikli y ntemle eřitlendiđinde hatanın daha b y k olduđu g r lm řt r. Tong ve Kolen'in (2005) ham puan dađılımlarının birbirine benzer olması durumunda eřit y zdelikli eřitleme y nteminin Madde Tepki Kuramı gerek puan ve g zlenmiř puan y ntemleri kadar bařarılı sonular verdiđini ortaya koymuřtur. Fakat bu alıřmada řans bařarisından arındırıldıktan sonra puanların birbirlerinden farklı olması eřit y zdelikli eřitleme y nteminin daha fazla hata vermesine neden olmuř olabilir. Bozdađ'ın yaptıđı alıřmada bu bulgudan farklı olarak řans bařarisından arındırılmıř test puanlarının eřitlenmesinde eřit y zdelikli eřitleme y nteminin daha az hata verdiđi bulunmuřtur. Buna ek olarak, Bozdađ'ın alıřmasında test formlarının g l k d zeyleri arasındaki fark bu alıřmada elden iki form arasındaki g l k d zeyinden daha azdır. Bu durum dođrusal eřitleme yerine eřit y zdelikli eřitlemenin daha iyi sonu vermesine sebep olmuř olabilir (Peterson, Marco & Stewart, 1982; Zhang vd., 2013). Ařiret ve  m r-S nb l (2014) alıřmasında řans bařarisının test eřitlemeye tutarlı bir etkisi g zlemlenemediđi

belirtmiştir. Bu çalışmanın bulgularına göre şans başarısından arındırılmadan eşitleme yapıldığında ve şans parametre değeri 0.10 olduğu durumlarda Dairesel yay yöntemi, diğer durumlarda ise ortalama eşitleme yönteminin kullanılabileceğini ve doğrusal eşitlemenin her bir parametre aralığında yüksek hata verdiğini belirtmişlerdir. Küçük örneklem ile elde edilen bu bulgu, bu çalışmanın bulgusu ile çelişmektedir. Fakat bu çalışmada da şans başarısının test eşitleme yöntemlerine etkisi net olarak vardır denilememektedir. Bu noktada şans başarısı ile olası diğer faktörlerin test eşitleme üzerinde etkisinin daha ayrıntılı incelenebilmesi için farklı test eşitleme teknikleri ve şans başarısı parametresi kullanarak yeni çalışmaların yapılması önerilmektedir.

Kelecioğlu (1994) tarafından yapılan araştırmanın sonucuna göre, ÖSS (1990-1993) Fen Bilimlerine ait testlerin eşitlenmesinde en uygun yöntemin eşit yüzdelikli eşitleme yönteminin olduğu saptanmıştır. Bu çalışmada ÖSS yerine TEOG Genel Değerlendirme sınavları dikkate alındığı için, ortaokul yerine lise giriş sınavlarının İngilizce alt testleri için bu çalışmanın tekrar edilip Fen Bilimleri için başarılı olan bu yöntemin İngilizce alt testleri için de tekrar sağlanıp sağlanmadığı bakılabilir.

İleriki bir çalışma olarak, Klasik Test Teoremi ile Madde Tepki Kuramı (gözlenen puan eşitleme, gerçek puan eşitleme v.b.) kullanılarak ulusal sınavlarının veya denemelerinin şanstın arındırılarak veya düzeltilerek eşitleme çalışmaları arasında bir karşılaştırma yapılabilir. Tek grup düzeneği dışında dengelenmiş grup ankor kullanılan düzenekte eşit yüzdelikli eşitleme dışındaki eşitleme yöntemlerin etkililiğine bakılabilir. Ek olarak, örneklem sayısı, şans başarısı ve öğrencilerin yetenek düzeylerine göre eşitleme modellerinin hata payları hesaplanabilir. Bu araştırmanın bulguları sadece TEOG Genel Değerlendirme 8. sınıf İngilizce testleri ile sınırlıdır. Genelleme yapabilmek için eşitleme yöntemlerinin farklı testler ve farklı değerlendirme ölçütleri kullanılarak karşılaştırılmasına ihtiyaç vardır.

Kaynaklar

- Akhun, İ. (1984). İki korelasyon katsayısı arasındaki farkın manidarlığının test edilmesi. *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 17, 1–7.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp.508–600). Washington, DC: American Council on Education.
- Angoff, W.H. (1984). *Scales, norms and equivalent scores*. New Jersey: Educational Testing Service.
- Anıl, D., & Öztürk, N. (2012). Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma. *Eğitim ve Bilim*, 37, 181-193.
- Araz, G. (2001). *Aynı davranışı ölçmeye yönelik kısa cevaplı, üç ve beş seçenekli çoktan seçmeli testlerin madde ve test özelliklerini şans başarısı ile birlikte incelenmesi* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Baykul, Y. (1996). *İstatistik: Metodlar ve uygulamalar* (3. Baskı). Ankara: Anı Yayıncılık
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara. ÖSYM Yayınları.
- Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi* (Yayımlanmamış yüksek lisans tezi). Mersin Üniversitesi Sosyal Bilimler Enstitüsü, Mersin.
- Bozdağ, S., & Kan, A. (2010). Şans başarısının test eşitlemeye etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39, 91-108.
- Braun, H. I., & Holland, P. W. (1982). Observed- score test equating: A mathematical analysis of some ETS equating procedures. In Holland, P.W., & Rubin D.B. (Ed.), *Test equating* (pp. 9-49). New York: Academic Press.
- Burgos, J. (2017). *SNSequate: Standard and nonstandard statistical models and methods for test equating* [Computer software manual]. <https://CRAN>.
- Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Yayınları.

- Caldwell, L. J. (1984). *A comparison of equating error in linear and rasch model test equating methods* (Doctoral dissertation). The Florida State University, Florida.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cook, L.L., & Petersen, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11, 225-244.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, OH: Cengage
- Çelen, Ü. (2002). *Şans başarısı için düzeltme formülü kullanılacağına ilişkin yönergenin testin psikometrik özelliklerine etkisinin araştırılması* (Yayımlanmamış yüksek lisans tezi), Ankara Üniversitesi SBE, Ankara.
- Demir, S., & Güler, N. (2014). Study of test equating on the common items non-equivalent group design. *International Journal of Human Sciences*, 11(2), 190-208.
- Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education*, 3(1), 3-17. doi:10.1207/s15324818ame0301_2.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Measurement*, 37, 281-306.
- Felan, G. (2002, February 14-16). *Test equating: Mean, linear, equipercentile, and item response theory*. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX.
- Fitzpatrick, A. & Yen, W. (2001) The effects of test length and sample size on the reliability and equating of tests composed of constructed-response items. *Applied Measurement in Education*, 14, 31-57.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston: Kluwer.

- Han, T., Kolen, M. J., & Pohlmann, J. (1997). A comparison among IRT true- and observed score equating and traditional equipercentile equating. *Applied Measurement in Education, 10*, 105-121.
- Harris, D. J. & Crouse, J. D. (1993). *A study of criteria used in equating. Applied Measurement in Education, 6*(3),195-240.
- Heh, V. K. (2007). *Equating accuracy using small samples in the random groups design* (Doctoral dissertation). Ohio University, US.
- Hills, J. R., Subhiyah, R. G., & Hirsch, T. M. (1988). Equating minimum-competency tests: Comparisons of methods. *Journal of Educational Measurement, 25*(3), 221-231.
- Holland, P.W., Dorans, N. J. & Petersen, N. S. (2007). Equating test scores. In Rao, C.R., Sinharay, S. (Ed.), *Handbook of statistics: Psychometrics* (Vol. 26, pp. 169-197). Amsterdam: Elsevier B. V.
- Huggins, A. C., & Penfield, R. D. (2012). An NCME instructional module on population invariance in linking and equating. *Educational Measurement: Issues and Practice, 31*(1), 27–40.
- Kan, A. (2011). Test eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim, 36*(160), 38-51.
- Kelecioğlu, H. (1994). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Kilmen, S., & Demirtasli, N. (2012). Comparison of test equating methods based on Item Response Theory according to the sample size and ability distribution. *Procedia - Social and Behavioral Sciences, 46*, 130–134. <https://doi.org/10.1016/J.SBSPRO.2012.05.081>.
- Kim, S., & Kolen, M. J. (2004). *STUIRT [Computer Software]*. The Center for Advanced Studies in Measurement and Assessment (CASMA), The University of Iowa, Iowa City: IA. <http://www.uiowa.edu/~casma> adresinden erişilmiştir.

- Kim, S., & Livingston, S. A. (2010). Comparisons among small sample equating methods in a common-item design. *Journal of Educational Measurement*, 47, 286–298.
- Kim, S., Walker, M. E., & McHale, F. (2010). Investigating the Effectiveness of Equating Designs for Constructed-Response Tests in Large-Scale Assessments. *Journal of Educational Measurement*, 47, 186-201.
- 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. (2004). Population invariance in equating and linking: Concept and history. *Journal of Educational Measurement*, 41(1), 3-14.
- Kolen, M. J. ve Brennan, R. L. (1995). *Test equating methods and practices*. New York: Springer.
- Kolen, M., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd Ed.). New York: Springer
- Kolen, M.J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd. ed.). New York: Springer.
- Lee, W., & Ban, J. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23, 23–48.
- Li, Y. H., Griffith, W. D., & Tam, H. P. (1997, June). *Equating multiple tests via an IRT linking design: Utilizing a single set of anchor items with fixed common item parameters during the calibration process*. Paper presented at the Annual Meeting of the Psychometric Society, Knoxville, TN.
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6, 83–102.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

- Muraki, E., Hombo, C., & Lee, Y. (2000, April). *Equating and linking of performance assessments*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Norman-Dvorak, R. L. (2009). *A comparison of kernel equating to the test characteristic curve method* (Unpublished doctorate thesis). University of Nebraska, Lincoln, Nebraska.
- Öztürk, N. (2010). *Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma* (Yayımlanmış yüksek lisans tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming and equating. In Linn, R. L. (Ed.), *Educational Measurement* (pp.221-262). New York: Macmillan.
- Petersen, N. S., Marco, G. L., & Stewart, E. E. (1982). A test of the adequacy of linear score equating models. In Holland P.W., & Rubin D. B. (Ed.), *Test equating* (pp. 71-136). New York: Academic Press.
- Ricker, K. L. & von Davier, A. A. (2007). *The impact of anchor test length on equating results in a nonequivalent groups design*. (Technical Report RR-07-44). Princeton, NJ: Educational Testing Service.
- S. Aşiret, & S.O. Sünbül. (2016). Investigating test equating methods in small samples through various factors. *Kuram ve Uygulamada Eğitim Bilim.*,16(2), 647–668.
- Skagg, G. & Lissitz R. W. (1986). An exploration of the robustness of four test equating models. *Applied Psychological Measurement*, 10, 303-317.
- Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement*, 42, 309-330.
- Şahhüseyinoğlu, D. (2005). *İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü. Ankara.
- Tanguma, J. (2000, January 27-29). *Equating test scores using the linear method: A primer*. Paper presented at the Annual Meeting of the Southwest Educational Research Association. Dallas, TX.

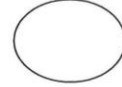
- Telli, A. (1993). *Şans başarısının madde türlerindeki madde ve test istatistiklerine etkisi* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Thorndike, R. L. (1982). Item and score conversion by pooled judgment. In P. W. Holland P. W., & Rubin D. B. (Ed.), *Test equating* (pp. 309-326). New York: Academic.
- Tong, Y., & Kolen, M. J. (2005). Assessing equating results on different equating criteria. *Applied Psychological Measurement*, 29(6), 418– 432.
- Tsai, T. H. (1997). *Estimating minimum sample sizes in random groups equating*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago.
- Turgut, F. (1979). *Ölçeleme, Norm Hazırlama, Normları Eşitleme ve Ayarlama*. Test Geliştirme Teknikleri Basılmamış Ders Notları, Ankara.
- von Davier, A. A. (Ed.). (2011). *Statistical models for test equating, scaling, and linking*. New York: Springer.
- von Davier, A. A., & Han, N. (2004). *Population invariance and linear equating for the non-equivalent groups design*. (ETS RR-04-47). Princeton, NJ: Educational Testing Service.
- Woldbeck, T. (1998). *Basic Concepts in Modern Methods of Test Equating*: Southwest Psychological Association.
- Yang, W. L., & Houang, R. T. (1996, April 8-12). *The effect of anchor length and equating method on the accuracy of test equating: comparisons of linear and IRT-based equating using an anchor-item design*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, (NY).
- Zeng, L. (1991). *Standard errors of linear equating for the single- group design* (Report No: 91–4). ACT Research Report Series.
- Zhang, X. McDermott, P.A., Fantuzzo, J. W., & Gadsden, V. L. (2013). Longitudinal stability of IRT and equivalent-groups linear and equipercentile equating. *Psychological Reports: Measures and Statistics*, 113, 1303-1325.

Zhu, W. (1998). Test equating: What, why and how?. *Research Quarterly for Exercises and Sport*, 69(1), 11–23.

EK-A: Etik Komisyonu Onay Bildirimi



Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması Etik Komisyon İzin Muafiyeti Formu*



05/ 09 / 2019

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

Tez Başlığı / Konusu: Şans Başarısının Test Eşitlemeye Etkisinin Farklı Eşitleme Teknikleriyle Araştırılması

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

(Emine Aytekin Kazanç, İmzası)

Öğrenci Bilgileri

| | |
|-----------------------|--|
| Adı Soyadı | Emine Aytekin Kazanç |
| Öğrenci No | N11225332 |
| Ana Bilim Dalı | Eğitim Bilimleri |
| Programı | Eğitimde Ölçme ve Değerlendirme Bilim Dalı |
| Statüsü | <input checked="" type="checkbox"/> Yüksek Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. |

Danışman Görüşü ve Onayı

Hayır veri kullanılması ve tezinde
kurum, kişi ismi kullanılmaması
nedeniyle ihtiyas duyulmamıştır.

(İmza)
(Prof. Dr. Selahattin Gelbal)

*Bu form sadece 16 Mart 2018 tarihinden önce önerisi sunulan tez çalışmalarında kullanılabilir. 16 Mart 2018 tarihinden sonra sunulan tüm tez önerileri için Etik Komisyonu izni gerekmektedir.

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA
Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: <http://ebe.hacettepe.edu.tr/> e-Posta: ebe@hacettepe.edu.tr

EK-B: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

..05.1.09.2013

Edyter
(İmza)
Ad SOYADI

Emine Aytekin-Kazancı

EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

05.10.2019

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Şans Başarısının Test Eşitlemeye Etkisinin Farklı Eşitleme Teknikleriyle Araştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak Turnitin adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

| Rapor Tarihi | Sayfa Sayısı | Karakter Sayısı | Savunma Tarihi | Benzerlik Oranı | Gönderim Numarası |
|--------------|--------------|-----------------|----------------|-----------------|-------------------|
| 05.10.2019 | 42 | 63,574 | 10/07/2019 | 0,08 | 1167632957 |

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Emine Aytekin-Kazanç

Öğrenci No.: N11225332

Ana Bilim Dalı: Eğitim Bilimleri Ana Bilim Dalı

Programı: Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

(Prof. Dr. Selahattin Gelbalı, İmza)

EK-Ç: Thesis/Dissertation Originality Report

05.10.2019.

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Measurement and Evaluation in Education

Thesis Title: Investigation of the Effect of Guessing on Test Equating With Different Equating Methods

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using Turnitin plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

| Time Submitted | Page Count | Character Count | Date of Thesis Defense | Similarity Index | Submission ID |
|----------------|------------|-----------------|------------------------|------------------|---------------|
| 05.10.2019 | 42 | 63,574 | 10/07/2019 | 9/08 | 1167632957 |

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

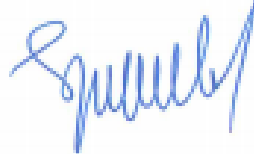
Name Lastname: Emine Aytekin Kazanç
Student No.: N11225332
Department: Educational Sciences
Program: Measurement and Evaluation in Education
Status: Masters Ph.D. Integrated Ph.D.

Signature



ADVISOR APPROVAL

APPROVED
(Prof. Dr. Selahattin Gelbal, Signature)



...../...../.....

EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- o Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- o Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

..05.. / ..09.. / 2019

 (imza)
Öğrencinin Adı SOYADI

Emine Aytekin-Kozaç

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6.1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü ana bilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezine erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metodları kullanıldığı, henüz makaleye dönüşmemiş veya patentli gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç, imkân oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü ana bilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezine erişime açılması engellenebilir.
- (3) Madde 7.1. Ulusal çıkarılan veya güvenliği değerlendirilen, emniyet, işhbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlerle ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir; gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

* Tez danışmanının önerisi ve enstitü ana bilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

