

İNGİLİZCE YETERLİK SINAVI PUANLARININ ÜÇ FARKLI EŞİTLEME YÖNTEMİNE GÖRE KARŞILAŞTIRILMASI*

A COMPARISON OF THE RESULTS OF ENGLISH PROFICIENCY EXAMS THROUGH THREE DIFFERENT EQUATION METHODS

Dr. Derya ŞAHHÜSEYİNOĞLU**

ÖZET: İngilizce yeterlik sınavları, her yıl farklı sorulardan oluşan testlerle yapılmaktadır. Bu sınavlarda kesme puanı değişmemesine karşın, sınava giren öğrenciler farklı formlardan farklı puanlar alabilmektedir. Bu çalışmada, İngilizce yeterlik sınavında kullanılan testlerin birbirine eşdeğer olup olmadığı, eşdeğer değilse, test eşitleme yöntemleriyle farklı sınavlara ait alınan puanların birbirine dönüştürülmesi ve testlerin birimlerinin aynı olmasının sağlanması hedeflenmiştir. İngilizce yeterlik sınavında kullanılan testler ortak test desenine dayalı üç farklı yöntemle eşitlenmiştir. Eşitlemede kullanılan yöntemler, Klasik Test ve Örtük Özellikler kuramlarına dayalıdır. Klasik test kuramına göre; eşit yüzdelikli ve doğrusal, örtük özellikler kuramına göre; Rasch parametresinin karakteristik eğri yöntemleri kullanılmıştır. Veriler Hacettepe Üniversitesi İngilizce Hazırlık Bölümü'nün yaptığı 2000, 2001 ve 2002 sınavlarının sonuçlarıdır. Analiz aşamasında betimsel istatistikler, örtük özellikler kuramı ve eşitleme varsayımlarından yararlanılarak farklı yıllara ait puanlar doğrusal, eşit yüzdelikli ve Rasch modeliyle eşitlenmiştir. Sonuçta, İngilizce yeterlik sınavında kullanılan testlerin eşitlenmesinde Rasch yöntemi ve eşit yüzdelikli eşitleme yönteminin uygun yöntemler olduğu bulunmuştur.

Anahtar sözcükler: Test eşitleme, paralel testler, Rasch yöntemi, İngilizce yeterlik sınavları, doğrusal eşitleme.

ABSTRACT: English proficiency exams are executed through questions that change every year. Even though cut of point does not change in these exams, students taking these test are likely to get different results different forms which puts the student in an advantageous or a disadvantageous position. In this study, the main aim is to measure If the test are equal, If not, to make different scores of different tests interchangeable so as to make the units of tests equal and to compare levels of groups of test takers of different years. To attain this aim, test used in English language proficiency exams are equalized with the help of three different test designs and two theories titled Classical test and Item Response Theories. As for the Classical test theory, equapercentage and linear, and for Item Response Theory Rasch Parametry characteristic curve method are used. Data used in this research process are gathered from the test results conducted in the years of 2000, 2001, 2002 at Hacettepe University English Language Preparatory School. As the result, it was found that while equating the tests used in English Proficiency tests, Rasch method and equapercentage equation method were applicable and proper methods.

Key words: Test equation, paralel tests, Rasch method, English proficiency tests, linear equation.

1. GİRİŞ

Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından merkezi olarak hazırlanan ve uygulanan sınav sonuçlarına göre üniversiteye gelen öğrenciler, kazanmış oldukları bölümde eğitim ve öğretime geçmeden önce, İngilizce hazırlık eğitimine alınırlar. İngilizce hazırlık programının amacı; öğrencilerin akademik çalışmalarını yürütmede ve şekillendirmede ihtiyaç duyacakları dil becerilerini (okuma, yazma, dinleme, konuşma) edinebilmelerini sağlamaktır (www.ydyo.hacettepe.edu.tr). Bu amaçla İngilizce hazırlık birimi, öğrencilerin yabancı dil bilgisi bakımından bölümlerindeki dersleri izleyebilecek düzeyde olup olmadıklarını “Yeterlik Sınavı” ile saptamaktadır. Yeterlik sınavdan belli bir puanın üzerinde alan öğrenciler

* Bu çalışma, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü'ne Ocak 2005 tarihinde sunulmuş doktora tez çalışmasından üretilmiştir.

** Dr., Bağımsız Araştırmacı, sderya@mapek.com

bölümdeki dersleri izleyebilecek yeterlikte kabul edilmektedir. Başarısız olan ve hazırlık sınıfına devam etmesi gereken öğrenciler “*Düzey Belirleme Sınavı’na*” girerek bu sınav sonucuna göre uygun bir hazırlık sınıfına yerleştirilmektedir.

İngilizce yeterlik sınavında yoklanan davranışlar değişmemekle beraber sınavın güvenliğini sağlamak amacı ile her yıl hazırlanan sınavlar farklı sorulardan oluşturulmaktadır. Her yıl farklı sorular ile tekrarlanan sınavlarda kesme puanı değişmediği halde farklı yıllara ait elde edilen puanlar aynı ölçekte yer almadığı için karşılaştırılmaları mümkün olmamaktadır. Sınavlar aynı amaç için yapılmış olsa bile sınava giren öğrenciler sınavın farklı formlarından farklı puan alabilmekte, bu durum da sınava giren öğrencilerin aldıkları puana göre avantajlı ya da dezavantajlı olmalarına neden olmaktadır. İngilizce hazırlık bölümünün farklı yıllara ait aynı amaç için geliştirilen farklı testleri, içerik ve istatistiksel özellikleri açısından benzese de testlerin güçlük derecelerinde farklılıklar vardır. Buna göre aynı testin farklı formlarını alanlar arasında bir kıyaslama olabilmesi için bu formlar üzerinde karşılaştırılabilir sonuçlar ortaya çıkaran bir yöntem olması önerilmiştir ki, bu da test sonuçlarını eşitlemektir.

Değişik test formlarından elde edilen puanların eşitlenmesi; bir formun birim sistemini diğerinin birim sistemine dönüştürmek olarak tanımlanmış ve dönüştürmeden sonra formların eşdeğer olacağı belirtilmiştir (Angoff, 1982). Böylece, dönüştürülmüş bu iki formdan ortaya çıkan sonuçlar birbirine eşit ve birbiri yerine kullanılabilir. Eşitleme tam olarak gerçekleştirildiğinde ise bir testin farklı formlarını alan iki kişinin performans ölçülerini doğrudan karşılaştırmak mümkün olmaktadır. Petersen, Marco ve Stewart (1982)’ a göre test eşitleme; test sonuçları arasında mümkün olduğu kadar birbirine yakın etkin bir eşitlik kurmak olarak tanımlanır. Braun ve Holland (1982) ise test eşitlemeyi bir testin iki formu üzerindeki sonuçlara yapılan sayısal dönüştürme olarak tanımlar.

İdeal olarak düşünüldüğünde, hazırlanan tüm test formlarının birbirine paralel olması, testi alan bireylerin tamamıyla rastgele seçilmesi ve yetenek düzeylerinin birbirine eşit olması istenen bir durumdur. Ancak, gerçekte böyle bir durumu sağlamak zordur. Böyle bir zorluk “test eşitleme” işlemlerini gerekli kılmaktadır. Test eşitlemede iki veya daha fazla testten elde edilmiş puanlar arasında istatistiksel bir işlem kurulur. Yapılan eşitleme aynı zamanda bir ölçeklemedir. İki veya daha fazla test için ortak bir ölçek kullanılmasıdır.

Genel olarak “*yatay ve dikey eşitleme*” olmak üzere iki tür eşitleme yönteminden söz edilmektedir (Crocker ve Algina 1986). *Yatay eşitleme, bir testin iki farklı formu arasında yapılan eşitlemedir. Yatay eşitlemede, testlerin benzer güçlükte olması, grupların da benzer yetenek dağılımına sahip olmasını gerektirir. Dikey eşitleme ise, aynı bilgi ya da becerileri ölçmek için hazırlanan ve standardize edilmiş farklı zorluk derecelerindeki testlerin eşitlenmesidir.* Örneğin farklı akademik başarı düzeylerinde olan kişilerin yeteneklerinin karşılaştırılmasını amaçlayan tek bir ölçek geliştirilir. Bu tür eşitlemede kullanılan testler, bir testin farklı formları olmayıp farklı düzeylerde dir. Testi alanların yetenek dağılımları da farklıdır (Hambelton ve Swaminathan 1985).

1.1 Problem

2000, 2001, 2002 yıllarına ait İngilizce yeterlik puanlarının 2000, 2001, 2002 olmak üzere ardışık iki yıla ait testlerin eşitlenmesinde eşit yüzdelliği, doğrusal, ve Rasch modellerinden hangisi bu testlerin eşitlenmesi için uygundur?

1.2 Alt problemler

1.1 2001 yılı İngilizce test puanları 2000 yılı puanlarına ortak testlere dayalı, eşit yüzdelliği eşitleme yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?

- 1.2 2001 yılı İngilizce test puanları 2000 yılı puanlarına ortak testlere dayalı, doğrusal eşitleme yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?
- 1.3 2001 yılı İngilizce test puanları 2000 yılı puanlarına ortak testlere dayalı, Rasch yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?
- 1.4 2001 yılı İngilizce test puanlarının 2000 yılı puanlarına ortak testlere dayalı eşitlenmesinde eşit yüzdelikli, doğrusal, Rasch yöntemlerinden hangisi bu testleri eşitlenmesi için en uygundur?
- 2.1 2002 yılı İngilizce test puanları 2001 yılı puanlarına ortak testlere dayalı, eşit yüzdelikli eşitleme yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?
- 2.2 2002 yılı İngilizce test puanları 2001 yılı puanlarına ortak testlere dayalı, doğrusal eşitleme yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?
- 2.3 2002 yılı İngilizce test puanları 2001 yılı puanlarına ortak testlere dayalı, Rasch yöntemiyle eşitlendiğinde elde edilen sonuçlar nasıldır?
- 2.4 2002 yılı İngilizce test puanlarının 2001 yılı puanlarına ortak testlere dayalı eşitlenmesinde eşit yüzdelikli, doğrusal, Rasch yöntemlerinden hangisi bu testleri eşitlenmesi için en uygundur?

2. YÖNTEM

2.1 Araştırma Türü

Bu çalışmada İngilizce Yeterlik Sınavı'nda kullanılan testler, klasik ve örtük özellikler kuramına dayalı eşitleme yöntemleri ile eşitlenmiş ve iki kurama dayalı elde edilen sonuçlar birbiri ile karşılaştırılmıştır. Gerçek verilere dayalı olarak yapılan bu çalışmada örneklem bilgileri bir evren genellemesi olmayıp, daha çok durum belirlemeye yönelik olduğundan temel araştırma niteliğindedir.

2.2 Araştırma Grubu

Araştırma, Hacettepe Üniversitesi İngilizce Hazırlık Bölümü'nün 2000, 2001 ve 2002 yıllarında yeterlik sınavına giren öğrencileri üzerinde yapılmıştır. Verilerin analizi için her bir testteki sıfır puan alan öğrenciler verilen testteki soruları hiç cevaplamadıkları için test sonuçlarından çıkarılmış ve veriler yeniden düzenlenmiştir. Buna göre; 2000 yılındaki 2293 öğrenciden 60'ı, 2001 yılındaki 2051 öğrenciden 62'si, 2002 yılındaki 2015 öğrenciden 25'i sınavda sıfır puan aldığı için değerlendirilmeye alınmamıştır.

2.3 Araştırma Verileri

Araştırmada kullanılan veriler Hacettepe Üniversitesi İngilizce Hazırlık Bölümü'nün yaptığı 2000, 2001 ve 2002 yıllarına ait İngilizce yeterlik sınavı sonuçlarıdır. Bu sonuçlar 2000 yılında sınava giren 2233, 2001 yılında sınava giren 1989 ve 2002 yılında sınava giren 2033 öğrenci üzerinden elde edilmiştir. Sınavda İngilizce dil becerisini ölçmeye yönelik 100 soru, aynı soruların farklı yerlerde olduğu iki kitapçık (A ve B) türü kullanılmaktadır. Araştırmada kullanılan veriler düzenlenirken, iki farklı kitapçık türüne göre elde edilen öğrenci cevapları, tek bir cevap anahtarına göre birleştirilmiştir. Sorular çoktan seçmeli ve 4 seçeneklidir. Sınav tek oturumda ve 3 saatte tamamlanmaktadır.

2.4 Verilerin Analizi

Verilerin analizi üç aşamada gerçekleştirilmiştir. Birinci aşamada betimsel istatistikler hesaplanmış, ikinci aşamada örtük özellikler kuramının varsayımlarının karşılanıp karşılanmadığına bakılmış, üçüncü aşamada ise eşitleme varsayımları kontrol edilerek farklı yıllara ait puanlar doğrusal, eşit yüzdelikli ve Rasch modeline göre eşitlenmiştir. Bunun için 2001 yılı puanları 2000 yılına, 2002 yılı puanları da 2001 yılı puanlarına eşitlenmiştir. Yapılan her bir eşitleme için hata miktarı bulunmuştur. Örtük özellikler teorisine göre parametreler hesaplanmadan önce, bu teorinin varsayımlarının karşılanıp karşılanmadığı için yapılan

hesaplamalar SPSS (10,0) paket programı, parametre hesapları için de BILOG for Windows Version 1.0 kullanılmıştır. Verilerin analizinde, öncelikle üç farklı yıla ait elde edilen testlerin betimsel istatistikleri hesaplanarak bu çalışma için uygunluğuna bakılmıştır. Tablo 2' de bunlara ait bilgiler verilmiştir.

3. BULGULAR

Tablo 1'den görüldüğü gibi 2000, 2001 ve 2002 yılına ait dağılımların basıklık katsayılarının negatif değerler alması dağılımlarının basık olduğunu göstermektedir. Çarpıklık katsayılarına bakıldığında ise 2000 ve 2001 yıllarına ait dağılımlar negatif değerler aldığından sola çarpık, 2002 yılına ait verilerine ilişkin çarpıklık katsayısı da pozitif değer aldığı için dağılım sağa çarpıktır. Basıklık ve çarpıklık katsayılarına bakıldığında dağılımların biraz basık olduğu anlaşılmaktadır. Her üç testin de güvenilirlik katsayısı birbirine eşit ve bu çalışma için yeterli düzeydedir.

Tablo 1: 2000, 2001, 2002 Ait Testlerin Betimsel İstatistikleri

	2000	2001	2002
N	2233	189	2080
K	100	100	100
Ortalama	45,30	48,38	44,07
Std. Sapma	24,19	25,46	23,21
Varyans	585,25	648,17	538,56
Çarpıklık	-0,10	-0,06	0,23
Basıklık	-0,97	-1,00	-0,96
KR-20	0,97	0,97	0,97

Çalışmanın ikinci aşamasında örtük özellikler kuramının tek boyutluluk, yerel bağımsızlık ve normallik sayıltularının sağlanıp sağlanmadığına bakılmıştır. Bu çalışmada da tek boyutluluğun test edilmesinde faktör analizi yöntemi kullanılmıştır. SPSS (10,0) paket programı ile analiz yapılmıştır. Farklı yıllara ait testlerin analiz sonuçları sırasıyla aşağıda verilmiştir. 2000 yılına ait testte toplam açıklanan varyansın % 27,024'ü nün birinci faktör, % 2,504'nin ikinci faktör, % 1,847'sinin de üçüncü faktör tarafından açıklandığı görülmüştür 2001 yılına ait testte toplam açıklanan varyansın % 21,692'nün birinci faktör, % 3,108'nin ikinci faktör, % 1,868'inin üçüncü faktör tarafından açıklandığı görülmüştür. testin 23. bileşenden sonra faktör yükü 1'in altına düşmektedir. 2002 yılına ait testte toplam açıklanan varyansın % 20,759'nün birinci faktör, % 3,119'nin ikinci faktör, % 1,585'inin üçüncü faktör tarafından açıklandığı görülmüştür. Her üç testin faktör analizi sonuçlarında, birinci faktörün özdeğeri diğer faktörlerin özdeğerlerinden yüksektir. Birinci faktörden sonra özdeğerler düşmekte diğer faktörlerin özdeğerleri arasındaki farklar azalmaktadır. Elde edilen bu sonuçlara göre testlerin tek boyutlu olduğu kabul edilmiştir. Testin tek boyutlu çıkması, örtük özellikler kuramının varsayımlarından tek boyutluluk ve yerel bağımsızlık varsayımlarının karşılandığını kabul etmek için yeterli bir sonuçtur. Bu çalışmada da tek boyutluluk varsayımı kabul edildiği için testin yerel bağımsızlık varsayımının karşılandığı kabul edilmiş, herhangi istatistiksel bir değerlendirme yapılmamıştır. Testlere ait çarpıklık ve basıklık katsayılarına bakıldığında 2000 ve 2001 yıllarına ait dağılımlar negatif değerler aldığından sola çarpık, 2002 yılına ait verilerine ilişkin çarpıklık katsayısı da pozitif değer aldığı için dağılım sağa çarpıktır. Basıklık katsayılarına bakıldığında dağılımların basık olduğu görülmüştür. Bu çalışmada testlerin eşitlenmesine geçilmeden önce, verilerin eşitleme şartlarını sağlayıp sağlanmadığına bakılmıştır. geçilmeden önce, verilerin eşitleme şartlarını sağlayıp sağlanmadığına bakılmıştır. Bir önceki aşamada örtük özellikler kuramının varsayımları kontrol edilirken eşitlenecek testlerin aynı değişkeni ölçtüğü şartı faktör analizi sonuçları ile ortaya konulmuştur.

Testlerin ortalama güçlükleri arasındaki fark z testi ile test edilmiştir. Eşitlenecek testlerin güvenilirlikleri birbirine eşit olduğu için farklılığı ortaya koyacak herhangi bir istatistiksel değerlendirme yapılmamıştır. 2000 yılı ile 2001 yılı ve 2001 yılı ile 2002 testlerinin ortalama güçlükleri arasında 0,05 düzeyinde manidar bir farklılık olmadığı görülmüştür. Eşitlenecek testlerin ortalama güçlüklerinin eşit olması koşulu sağlanmıştır. KR-20 güvenilirlik katsayıları hesaplanmış ve her üç testin güvenilirlik katsayısı 0.97 olarak bulunmuştur. Güvenirlikleri birbirine eşit olduğu için testlere karışan hata miktarları da eşit düzeyde olduğu kabul edilmiştir. Eşitleme koşullarından testlerin eşit güvenilirlikte olma koşulu sağlanmıştır. Eşitleme koşullarının sağlanıp sağlanmadığı kontrol edildikten sonra, seçilen eşitleme deseni gereği ortak testlerin oluşturulması yoluna gidilmiştir.

Test puanlarının eşitlenmesinde testlerin ve grupların düzenlenmesine göre çeşitli desenler oluşturulmaktadır. Bunlardan en sık kullanılanı, ortak test uygulamasına dayalı desendir. Ortak testlerin oluşturulmasında ise şöyle bir yol izlenmiştir. Eşitlenecek olan testlerin faktör analizi yoluyla tek boyutlu olup olmadıkları kontrol edildikten sonra, uzmanlar yardımıyla farklı yıllara ait testlerdeki her bir maddenin ölçtüğü davranışlar çıkarılmıştır. Üç testte de aynı değişkeni ölçtüğü düşünülen maddelerden faktör yükleri en fazla olanlar (faktör yükü .40 ve üzerinde olanlar) seçilerek üç ayrı ortak test oluşturulmuştur. Buna göre 100 maddelik testlerden 28'er maddelik ortak testler oluşturulmuştur. Seçilen her maddenin faktör yükü .60 ve üzerindedir. Ortak test uzunluğu için, testlerin uzunluğunun %20'sinden az olmaması ya da en az 20 maddeyi içermesi ölçütünün yerine getirilmesine dikkat edilmiştir. Oluşturulan ortak testler ile maddelerin ölçtüğü değişken ve testlerin paralelligi kontrol edilmeye çalışılmıştır. Bunun için testlerin aritmetik ortalama ve standart sapmalarının birbirine yakın olup, aralarında fark olmamasına dikkat edilmiştir.

Ortak testler oluşturulduktan sonra ortak testlere ait betimsel istatistikler hesaplanmış, Tablo 2'de bunlara ait sonuçlar verilmiştir. Ortak testlerin güvenilirlik, ve ortalama güçlük bakımından tüm testi temsil edip etmediğine bakılmıştır. Ortak testlerin güvenilirliklerinin tüm testin güvenilirliği ile eşit olup olmadığını test etmek için Sperman Brown formülünden yararlanılmıştır (Baykul 2000). Elde edilen güvenilirlik katsayıları madde cevaplarından elde edilen KR-20 güvenilirliği ile karşılaştırılmıştır. İki güvenilirlik katsayı arasındaki fark Fisher'in z testi ile test edilmiştir (Akhun, 1984). Ortak testlerin ortalama güçlük bakımından tüm testi temsil edip etmediklerine bakmak için iki oran arasındaki farkın test edilmesinde kullanılan z testi kullanılmıştır.

Tablo 2: Ortak Testlere Ait İstatistikler

	S2000	S2001	S2002
N	2233	1989	2080
K	28	28	28
Ortalama	15,41	14,89	14,39
Std. sapma	9,45	8,96	9,14
Varyans	89,41	80,23	83,62
Çarpıklık	-0,35	-0,42	-0,04
Basıklık	-1,11	-1,00	-1,17
KR-20	0,94	0,93	0,93

Tablo 2'de görüldüğü gibi testlerin aritmetik ortalama ve standart sapmaları birbirine yakındır. "Tüm testlerde" olduğu gibi dağılımlar basık ve aynı zamanda sola çarpıktır. 2001 ve 2002 yılına ait testlerin güvenilirlikleri birbirine eşit, 2000 yılına ait testin güvenilirlik katsayısı .94'dür. Ortak testlerin paralelliginin ortaya konması için aritmetik ortalama ve standart sapmaları arasındaki farklar test edilmiştir. Testlerin aritmetik ortalamaları arasında manidar bir fark bulunmamıştır. Varyanslarına bakıldığında ise üç ortak testin varyanslarının birbirine yakın değerler verdiği görülmüştür. Bu sonuçlara göre ortak testlerin

paralelliğinin sağlandığı kabul edilmiştir. Eşitlemeler ortak testler yoluyla yapılacağı için, ortak testlerin güvenilirlikleri ve güçlük düzeyleri bakımından “Tüm testleri” temsil etmesi gerekir. Spearman - Brown formülü hesaplanan ortak testlerin güvenilirliklerinin ile madde cevaplarından elde edilen KR-20 güvenilirliği ile karşılaştırılması yapılmıştır. Spearman Brown formülünden elde edilen güvenilirlik katsayıları ile madde cevaplarından elde edilen KR-20 güvenilirlik katsayıları arasındaki fark ilişkisiz iki korelasyon katsayısının test edilmesinde kullanılan Fisher’in z testi ile test edilmiştir. İki güvenilirlik katsayısı arasında manidar bir fark bulunamamıştır. Ortak testlerin güçlük düzeyleri bakımından ait oldukları “Tüm testleri” temsil edip etmediklerini görmek için güçlük düzeylerine ait karşılatırmalar yapılmıştır. sonuçlara göre, ortak testler ile ait oldukları “Tüm testlerin” ortalama güçlük düzeyleri arasında manidar bir fark bulunmamıştır. Buna göre ortak testlerin tüm testleri ortalama güçlük düzeyi yönünden de temsil ettikleri kabul edilmiştir. bu değerlendirmelerden sonra testler, doğrusal, eşit yüzdelikli ve örtük özellikler kuramına dayalı olan rasch modeli ile eşitlenmiştir. Her bir eşitleme yöntemi için şu aşamalar sağlanmıştır:

3.1 Doğrusal Eşitleme:

Doğrusal eşitlemede, Woodroff’un geliştirdiği denk olmayan gruplar için ortak maddelere dayalı doğrusal eşitleme yöntemi kullanılmıştır. 2001 yılı 2000 yılı puanlarına, 2002 yılı puanları da 2001 puanlarına eşitlenmiştir. Bunun için

$$A_1 = A_2 + V_1$$

$$B_1 = B_2 + V_2$$

eşitlikleri kullanılmıştır.

A_1 , eşitlenecek testin 100 maddelik bölümünü ifade etmiş ve 100 maddelik testin aritmetik ortalamaları, varyansları, güvenilirlikleri hesaplanmıştır. V_1 , 28 soruluk ortak testi ifade etmiş ve 28 maddelik testin aritmetik ortalamaları, varyansları, güvenilirlikleri hesaplanmıştır. A_2 ise 28 soru çıkarıldıktan 100 madde içinde kalan 72 maddelik bölümü temsil etmiş ve 72 maddelik testin aritmetik ortalamaları, varyansları, güvenilirlikleri hesaplanmıştır. 2001 yılı 2000 yılına eşitlenirken A_1 için 2000 testine ait puanlar alınmış, B_1 için 2001 yılına ait test puanları dikkate alınmıştır ve Woodroff ‘un geliştirdiği denk olmayan gruplar için ortak maddelere dayalı doğrusal eşitleme denkleminde bulunan değerler yerine konularak bir regresyon denklemi elde edilmiştir. Regresyon denkleminde 2000 yılından 2001 yılına ait puanlar, 2001 yılından da 2000 yılına ait puanlar kestirilmiştir.

3.2 Eşit Yüzdelikli Eşitleme

Eşit yüzdelikli eşitlemede eşitlenecek her bir test için ortak testlerin yüzdelik puanları elde edilmiştir. Eşitlenecek iki test için ortak testlerle aynı yüzdeliğe sahip olan puanlar bulunmuştur. Puanlar elde edildikten sonra iki puan grubu arasında doğrusal dönüşüm denklemi elde edilmiştir. Doğrusal dönüşüm denklemler yardımıyla 2000 yılı 2001 yılı puanlarına, 2001 yılı puanların da 2002 puanlarına eşitlenmiştir. Doğrusal dönüşüm denklemlerinde puanlar yerine koyularak kestirimler yapılmıştır. Öncelikle 28 maddelik ortak testlerin yüzdelik puanları bulunmuştur. Her üçü yıla ait ortak testlerin yüzdelik puanları bulunduktan sonra eşitlenecek testler ile aralarındaki dönüşüm denklemleri elde edilmiştir. Bu denklemlerden 2000 yılından 2001 yılına ait puanlar, 2001 yılından da 2000 yılına ait puanlar kestirilmiştir.

3.3 Örtük Özellikler Teorisi ile Eşitleme:

Örtük özellikler kuramına dayalı rasch modeli ile eşitleme yapılmıştır. Eşitlenecek testlerin her biri için madde parametreleri ve yetenek ölçüleri kestirildikten sonra ortak testlerin maddelerin parametrelerinden, Rasch modelinin b parametresi kullanılmıştır. Bu eşitlik aracılığıyla X testinin madde parametreleri Y testinin madde parametrelerine dönüştürülmüş daha sonra kestirilen ve dönüşüm sonucu elde edilen madde parametreleri kullanılarak eşitlenecek testler için gerçek puanlar kestirilmiştir ve bu gerçek

puanlar birbirine eşit kabul edilmiştir. Bunun için öncelikle her bir test için madde ve yetenek parametreleri BILOG paket programı yardımıyla hesaplanmıştır. Madde ve yetenek parametrelerinin Rasch modeli için uyumuna bakılmıştır. Rasch modelinin b parametresi dönüşümü için karakteristik eğri yöntemindeki denklemler kullanılarak dönüşümler sağlanmıştır. Dönüşüm sonucu elde edilen b parametresi BILOG paket programında yeni bir veri tanımlaması yardımıyla yerine konularak θ değerleri yeniden kestirilmiştir. Dönüşümden önce elde edilen yetenek parametreleri ile dönüşüm sonucu elde edilen yetenek parametreleri farklı yıllara ait karşılaştırılmıştır.

3.4 2001 yılı puanlarının 2000 yılı puanlarına üç farklı yöntemle eşitlenmesi:

2001 yılı puanlarının 2000 yılı puanlarına eşit yüzdelli ortak testlere dayalı eşitleme yöntemi ile eşitlendiğinde puanlar arasında büyük farkların olmadığı görülmüştür. Bu yöntemde, aynı evren için eşitlenecek testlerin ham puanlarından elde edilen dönüşümün aynı yüzdelik sırasına gelirse bu puanların eşit olduğu kabul edilir (Hambelton 1985). Puanlar arasındaki farkların küçük olması eşitlenen puanların aynı yüzdelik sırasına geldiğini göstermektedir. Doğrusal eşitlikte ortalama ve standart sapmada farklılık gösteren iki testin eşitlenmesi varsayımı vardır. Bu iki testin ham puanları dağılımı denk olmak zorundadır (Angoff 1971). Bu yöntemde $X = Y$ koşulu, ancak iki dağılımın aritmetik ortalamaları ve standart sapmaları eşit olduğunda sağlanır. Bu iki şart sağlanmazsa, ham puanlar arasında doğrusal bir dönüşümden söz edilemez. Bu çalışmada da eşitlenecek olan testlerin aritmetik ortalamaları ve varyansları arasında istatistiksel olarak manidar bir farklılık olmasına rağmen, betimsel olarak değerlerin birbirine yakın olması nedeniyle çalışmaya devam edilmiştir. Buna göre doğrusal eşitleme yönteminde puanlar arasında büyük farklar çıkmasının nedeni puan dağılımlarından kaynaklandığı düşünülebilir.

Rasch yöntemine dayalı eşitlemede her iki yıla ait yetenek ölçüleri arasındaki ilişki doğrusaldır. θ değerleri -3 ile +3 arasında değerler almış ve farklı yıllara ait elde edilen yetenek ölçüleri arasında büyük farklar olmadığı görülmüştür. İki yetenek ölçüsü arasındaki ilişki düzeyine pearson momentler çarpımı ile bakılmış, katsayının .95 olduğu bulunmuştur. Buna göre bireylerin, farklı yıllara ait farklı testi almış olsalar da aynı yetenek düzeyinde yer alabilecekleri ileri sürülebilir θ değerleri arasındaki farkların büyük olmadığı görülmüştür. θ değerleri -2,61 ile 2,55 sınırları arasında değerler almıştır. Rasch yöntemine dayalı eşitlemede, örtük özellikler kuramının normallik şartını sağlamadığı göz önüne alındığında, testlerin bu varsayımın yerine getirilememesinden etkilenmediği söylenebilir. Örtük özellikler kuramının normallik sayılışı karşılanmasına da θ değerleri arasındaki farkın az olmasının nedeninin θ değerlerinin ranj aralığının darlığından kaynaklandığı düşünülebilir.

Bu çalışmada klasik test teorisine ve örtük özellikler kuramına dayalı eşitleme yöntemleri kullanılmış ve bu yöntemlerden uygun olan belirlenmeye çalışılmıştır. Bu alt problemde de üç yöntemde ait hata kareleri ortalaması bulunmuş ve hata kareleri ortalaması en küçük olan yöntemin bu çalışmada en uygun eşitleme yöntemi olduğuna karar verilmiştir. Eşit yüzdelli eşitleme için hata kareleri ortalaması 0,003, doğrusal model için 0,265, Rasch modeli için 0,001'dir. Bulunan değerlere bakıldığında doğrusal model için yapılan eşitlemenin hata kareleri ortalaması diğerlerinden daha yüksektir. Eşit yüzdelli ve Rasch yöntemine dayalı eşitlemede bulunan hata kareleri ortalaması birbirine yakın ve düşük olsa da bu testler için en uygun yöntemin Rasch yöntemine dayalı eşitleme olduğu savunulabilir.

3.5 2002 yılı puanlarının 2001 yılı puanlarına üç farklı yöntemle eşitlenmesi:

2002 yılı puanlarının 2001 yılı puanlarına eşit yüzdelli ortak testler dayalı eşitleme yöntemi eşitlendiğinde 72 puana kadar tüm puanlar birbirine eşit ve 72 puan üstünde de 1 puan görülmüştür. Puanlar arasındaki farkların küçük olması eşitlenen puanların aynı yüzdelik sırasına geldiğini göstermektedir. 2002 yılı ile 2001 yılı arasındaki ortak testlere dayalı eşityüzdelli eşitlemede puan farkları diğer iki yılın eşit yüzdelli eşitleme sonucundan daha azdır. Elde edilen sonuçlardan, Hacettepe Üniversitesine gelen öğrenci

kaynağında bir farklılık olmamasının, hazırlanan sınavların içeriklerinin birbirine benzer olmasının ve puan dağılımlarının benzerliği göz önünde bulundurulduğunda puanlar arasındaki farkın azlığı bu puanların birbirine eşitlenebileceğini göstermektedir.

2002 puanlarının 2001 yılı puanlarına eşitlendiğinde ham puanlar ve eşitlenmiş puanlar arasındaki grafik, diğer eşitleme yöntemlerinde olduğu gibi doğrusaldır. puan dağılımlarına bakıldığında 21. puana kadar eşitlenmiş puanlar arasında 1 puan fark vardır. 21 ve 31. puanlar arası birbirine eşit 31. puandan puanlar arttıkça eşitlenmiş puanlarda bir düşme olduğu görülmüştür. Doğrusal eşitlikte ortalama ve standart sapmada farklılık gösteren iki testin eşitlenmesi varsayımı vardır. Bu iki testin ham puanları dağılımı eşit olmak zorundadır (Angof, 1971). Bu yöntemde, $X = Y$ koşulu, ancak iki dağılımın aritmetik ortalamaları ve standart sapmaları denk olduğunda sağlanır. Bu iki şart sağlanmazsa, ham puanlar arasında doğrusal bir dönüşümden söz edilemez. Bu çalışmada da eşitlenecek olan testlerin aritmetik ortalamaları ve varyansları arasında istatistiksel olarak manidar bir farklılık olmasına rağmen betimsel olarak değerlerin birbirine yakın olması nedeniyle çalışmaya devam edilmiştir. Buna göre doğrusal eşitleme yönteminde puanlar arasında büyük farklar çıkmasının nedenini puan dağılımlarından kaynaklandığı düşünülebilir.

Rasch yöntemine dayalı eşitlemede her iki yıla ait yetenek ölçüleri arasındaki ilişki doğrusaldır. θ değerleri -3 ile +3 arasında değerler almış ve farklı yıllara ait elde edilen yetenek ölçüleri arasında büyük farklar olmadığı görülmüştür. İki yetenek ölçüsü arasındaki ilişki düzeyine pearson momentler çarpımı ile bakılmış, katsayının .90 olduğu bulunmuştur. Buna göre bireylerin, farklı yıllara ait farklı testi almış olsalar da aynı yetenek düzeyinde yer alabilecekleri söylenebilir. θ değerleri arasındaki farkların büyük olmadığı görülmüştür. θ değerleri -2,75 ile 2.64 sınırları arasında değerler almıştır. Rasch yöntemine dayalı eşitlemede, örtük özellikler kuramının normallik şartını sağlanmadığı göz önüne alındığında, testlerin bu varsayımın yerine getirilememesinden etkilenmediği söylenebilir. Örtük özellikler kuramının normallik sayıltısı karşılanmasına da θ değerleri arasındaki farkın az olmasının nedeninin Rasch modelinde madde uyum testi sonucunda 100 madden 82'sinin olmasının bu modele uygun bir veri yapısından kaynaklandığı düşünülebilir.

Bu çalışmada Klasik Test Teorisine ve Örtük Özellikler Kuramına dayalı eşitleme yöntemleri kullanılmış ve bu yöntemlerden uygun olan belirlenmeye çalışılmıştır eşit yüzdellikli eşitleme için hata kareleri ortalaması 0,0002, doğrusal model için 0,012, rasch modeli için 0,007'dir. Bulunan değerlere bakıldığında doğrusal model için yapılan eşitlemenin hata kareleri ortalaması diğerlerinden daha yüksektir. Eşit yüzdellikli ve rasch eşitlemesine dayalı bulunan hata kareleri ortalaması birbirine yakın ve düşük olsa da bu testler için en uygun yöntemin eşityüzdellikli eşitleme yöntemine dayalı eşitleme olduğu düşünülebilir.

4. SONUÇ

- İngilizce 2001 yılı testi, 2000 yılı testine, eşit yüzdellikli eşitleme, doğrusal ve rasch modeline dayalı eşitlendiğinde, rasch modelinin eşitleme hatasının diğer yöntemlerden daha düşük olduğu görülmüştür.
- İngilizce 2002 yılı testi, 2001 yılı testine, eşit yüzdellikli eşitleme, doğrusal ve rasch modeline dayalı eşitlendiğinde, eşit yüzdellikli ve rasch modelinin eşitleme hatasının doğrusal eşitleme yönteminden daha düşük olduğu görülmüştür.
- 2001 yılı İngilizce testinin 2000 yılına, 2002 yılı İngilizce testinin 2001 yılına eşit yüzdellikli eşitleme yöntemiyle eşitlenmesinde, eşitlenmiş puanlar ile ham puanlar arasında doğrusal bir ilişki olduğu görülmüştür. 2002 yılının 2001 yılına eşit yüzdellikli eşitlenmesindeki hatanın 2001 yılının 2000 yılına eşitlenmesindeki hatasından daha düşük olduğu görülmüştür.

- 2001 yılı İngilizce testinin 2000 yılına, 2002 yılı İngilizce testinin 2001 yılına doğrusal eşitleme yöntemiyle eşitlenmesinde, eşitlenmiş puanlar ile ham puanlar arasında doğrusal bir ilişki olmasına rağmen puanlar arasındaki fark diğer eşitleme yöntemlerine göre daha fazladır. 2002 yılının 2001 yılına doğrusal doğrusal yöntemle eşitlenmesindeki hatanın, 2001 yılının 2000 eşitlenmesindeki hatasından daha düşük olduğu görülmüştür.
- 2001 yılı İngilizce testinin 2000 yılına, 2002 yılı İngilizce testinin 2001 yılına Rasch modeli dayalı eşitlenmesinde, gerçek ve kestirilen yetenek ölçüleri arasında da doğrusal bir ilişki ve hatanın az olduğu görülmüştür.
- Doğrusal eşitleme yöntemindeki hatanın fazla olması bu çalışmada kullanılan testlerin aritmetik ortalama ve varyansların arasındaki farktan kaynaklandığı sonucuna varılmıştır.
- Eşitleme yöntemlerinin her bir test için yıllara göre değişmemesi, bu yöntemlerin testler içinde bir kararlık gösterdiğini ortaya koymuştur.
- Bu çalışmada ortak testlere dayalı eşitleme yapılması eşitleme yöntemlerinde hesaplanan hata miktarlarını olumlu yönde etkilemiştir.
- Eşitleme sonucunda eşitleme yöntemlerinin yıllara göre değişimi dikkate alındığında 2001 ve 2002 yıllarına ait testlerin birbirine yakın sonuçlar verdiği görülmüş ve bu yıllarda sınava giren öğrencilerin aldıkları puanlara göre doğru kurlara yerleştirmiş olabileceği sonucuna varılmıştır.
- Eşit yüzdelikli eşitleme yönteminin de Rasch yöntemine dayalı eşitleme kadar uygun bir yöntem olduğu ileri sürülebilir.
- İngilizce Hazırlık atlama sınavında kesme puanı 65 olarak kullanılırsa, 2001 yılı testinin 2000 yılı testine eşit yüzdelikli eşitleme ile eşitlendiğinde, 2000 yılındaki 65 puan, 2001 yılında 67 puana karşılık gelmektedir. Buna göre 2001 yılında sınava giren öğrencilerden 51'i 2000 yılına göre kesme puanın altında kalmaktadır. 2002 yılı testi 2001 yılına eşit yüzdelikli eşitleme ile eşitlendiğinde 65 puan, 65 puan karşılık gelmektedir. Avantajlı ya da dezavantajlı durumda bir öğrenci grubu bulunmamaktadır.
- İngilizce Hazırlık atlama sınavında kesme puanı 65 olarak kullanılırsa, 2001 yılı testinin 2000 yılı testine doğrusal eşitleme ile eşitlendiğinde, 2000 yılındaki 65 puan, 2001 yılında 53 puana karşılık gelmektedir. Buna göre 2001 yılında sınava giren öğrencilerden 283'i 2000 yılına göre kesme puanın üzerindedir. 2002 yılı testi 2001 yılına doğrusal eşitleme ile eşitlendiğinde 65 puan 62 puan karşılık gelmektedir. 2002 yılında sınava giren öğrencilerden 76'sı 2001 yılına göre kesme puanın üzerindedir.

5. ÖNERİLER

- Bu çalışmada kullanılan ortak testler, testlerin bir parçası olarak hazırlanmıştır. Ortak testler bu testlere paralel olduğu düşünülen başka testlerden oluşturulabilir. Testlerin madde sayısı, tüm testin %20 sini temsil edecek şekilde 28 sorudan oluşturulmuştur. Ortak test uzunluğunun eşitleme yöntemleri üzerindeki etkileri çalışılabilir.
- Bu çalışmada üç ayrı grup ve üç ayrı test arasında ortak testlere dayalı eşitleme deseni üzerinde çalışılmıştır. İngilizce hazırlık biriminin kullandığı testlerle farklı eşitleme desenleri üzerinde çalışılıp, başka eşitleme yöntemleri denenebilir.
- Bu çalışmada eşitleme denklemleri şans başarısından arıtılmamış ham puanlardan elde edilmiştir. Şans başarısının eşitlemeye etkisini görmek amacıyla benzer çalışmalar düzeltilmiş puanlar üzerinde yapılabilir.

- Bu çalışmada 0 ile 1 puanlanan testler üzerinde eşitlemeler üzerinde çalışılmıştır. Açık uçlu sorulardan oluşan testler üzerinde eşitlemeler üzerine çalışılabilir. Bu çalışmada Rasch modeline dayalı eşitlemeler yapılmıştır. İki ve üç parametrelili logistik modeller için de eşitlemeler kullanılarak karşılaştırmalar yapılabilir.

KAYNAKÇA

- Akhun, İ. (1984) . İki Korelasyon Katsayısı Arasındaki Farkın Manidarlığının Test Edilmesi. *A.Ü. Eğt. Bil. Fak. Dergisi*, 17-12.
- Angoff, W. H. (1971). Scales, Norms and Equivalent Scores. R. L Thorndike (Ed.) *Educational Measurement*. Washington D.C: American Council on Education.
- Angoff, W. H. (1982). *Summary and Derivations of Equating Methods Used at ETS, in Holland Test Equating*. New York: Academic Press, 55- 59.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*, Ankara: ÖSYM Yayınları.
- Braun, H. I. ve Holland, P.W. (1982). *Observed score test equating. A mathematical analysis of some ETS equating procedure*. New York: Academic Press.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Theory*. CBS College Publishing.
- Hambelton, G. (1985). *Item Response Theory: Principles and Application*, Kluwer, Nijhoff Kluwer Academic Publisher Group.
- Petersen, N. S., Macro, G. L ve Stewart, E. E. (1982). *A test adequacy of linear scores equating methods*. New York: Academic Press.
- Woodroff, D. (1986). Derivations of Observed Score Linear Equating Methods Based On Test Score Models for the Common Item Nonequivalent Population Design. *Journal of Educational Statistics*, 11, 245-257.
- Woodroff, D. (1989). A comparison of Three Linear Equating Methods for the Common Item Nonequivalent: Population Design. *Applied Psychological Measurement* 13, 257 – 261.

EXTENDED ABSTRACT (Uzun İngilizce Özet)

English proficiency exams are executed through questions that change every year. Even though cut of point does not change in these exams, students taking these test are likely to get different results different forms which puts the student in advantageous or a disadvantageous position. In this study, the main aim is to measure If the test are equal, If not, to make different scores of different tests interchangeable so as to make the units of tests equal and to compare levels of groups of test takers of different years. To attain this aim, test used in English language proficiency exams are equalized with the help of three different test designs and two theories titled Classical test and Item Response Theories. As for the Classical test theory, equapercentage and linear, and for Item Response Theory Rasch Parametry characteristic curve method are used. Data used in this research process are gathered from the test results conducted in the years of 2000, 2001, 2002 at Hacettepe University English Language Preparatory School. As the result, it was found that while equating the tests used in English Proficiency tests, Rasch method and equapercentage equation method were applicable and proper methods.

English proficiency exams are executed through questions that change every year. Even though cutoff point does not change in these exams, students taking these tests are likely to get different results from different forms which puts the student in an advantageous or a disadvantageous position. In this study, the main aim is to measure if the tests are equal, if not, to make different scores of different tests interchangeable so as to make the units of tests equal and to compare levels of groups of test takers of different years.

To attain this aim, tests used in English language proficiency exams are equalized with the help of three different test designs and two theories titled Classical test and Item Response Theories. As for the Classical test theory, equapercentage and linear, and for Item Response Theory Rasch parametry characteristic curve method are used.

Data used in this research process are gathered from the test results conducted in the years of 2000, 2001, 2002 at Hacettepe University English Language Preparatory School. These results are gathered from test takers numerically 2233 students in 2000, 1989 students in 2001, and 2033 students in 2002. In the analysis of the data, those who scored 0 (zero) are not taken into consideration because they could not answer any question.

Data analysis was completed in three steps. In the first step a descriptive statistical analysis was completed and in the second one hypotheses of Item Response Theory was investigated and in the third step equalization was completed via the calculation of the scores of different years in terms of equapercentage, rasch and linear concepts. As for the analysis of all three tests through IRT qualities, it is found that the tests are one dimensional and not having shown normal distribution. While the equation conditions were checked, it was also tested if the tests were measuring the same variable, were in the same level of difficulty, sharing same reliability, and were parallel. As these analyses suggested, 0.05 difference between the mean of the level of difficulty and reliability. After having seen the fact that the tests fulfilled the necessary level of equation, sample and shared tests were prepared after which they were re-tested to make sure they carried the qualities necessary for equation and parallelisms were drawn amongst these tests. After these tests were formed they were measured through three different test measurement procedures mentioned above and it was found that the test with the smallest number of the error square was accepted to be the most applicable for this research.

As the result, it was found that while equating the tests used in English proficiency tests, Rasch method and equapercentage equation method were applicable and proper methods.