

Thurstone ve Bradley-Terry İkili Karşılaştırma Yargılarına Dayalı Ölçekleme Yöntemlerinin Karşılaştırılması*

Osman TAT** – Duygu ANIL

Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri ABD, Ankara Türkiye

* Bu çalışma, V. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

Öz

Günümüzde sekiz tane önemli küresel ve özellikle gelişmekte olan ülkelerde uzmanlık alanlarını, bölgesel farklılıkları dikkate alan 60'tan fazla yerel yükseköğretim kurumları sıralama listesi vardır. Yükseköğretim kurumları listelerinin hazırlanmasında listeleri hazırlayan kuruluşların tercih ettikleri kriterler ve yöntemler birbirinden oldukça farklı olabilmektedir. Thurstone'nun karşılaştırmalı yargı kanununun farklı varsayımlar altında kullanılan beş hali olduğunu bilinmektedir. Thurstone'nun ikili karşılaştırmaya dayalı ölçekleme yöntemlerinden en bilinenleri V. ve III. hal denklemleridir. İkili karşılaştırma yargılarına dayalı ölçekleme tekniklerinden bir diğeri ise psikolojik deneyler, spor müsabakaları ve genetik gibi geniş bir kullanım alanı olan Bradley-Terry modelidir. Bu çalışmada iki farklı ölçekleme yöntemi ile üniversite sıralama kuruluşlarının kullandıkları kriterler önem düzeyine göre sıralanmış ve elde edilen kriter sıralamasının tutarlılığı incelenmiştir. Çalışma, nicel araştırma tekniklerinin kullanıldığı temel araştırma modelinde desenlenmiştir. Araştırmanın çalışma grubunu Türkiye'deki kamu ve özel/vakıf yükseköğretim kurumlarında akademisyen olarak çalışmakta olan 247 kişi oluşturmaktadır. III. Hal Denklemi ve Bradley-Terry modeli ile yapılan ölçeklemede akademisyenlerin SCI, SSCI ve AHCII İndekslerinde Taranan Makale Sayısı'nı en önemli kriter olarak gördüğü buna karşın Üniversitenin Nobel Sahibi Üye veya Mezun Sayısı'nı ise en önemsiz kriter olarak gördükleri belirlenmiştir. Her iki ölçekleme yönteminde de üniversite sıralama kriterlerinin sıralarının aynı olduğu tespit edilmiştir.

Anahtar Kelimeler: Ölçekleme, Thurstone, Bradley-Terry, İkili Karşılaştırma, Thurstone III. Hal Denklemi.

Makale Bilgileri:

Gönderim / Received:

03.10.2016

Kabul / Accepted:

13.11.2016

**Sorumlu Yazar /

Corresponding Author:

Hacettepe Üniversitesi,

Eğitim Fakültesi,

Ankara, Türkiye

osmntt@gmail.com

A Comparative Analysis of Thurstone and Bradley-Terry Scaling Techniques

Abstract

Higher education institution rankings are monitored seriously by universities, sponsors of private and state institutions, potential students, parents and also by public. Nowadays there are eight

important global ranking lists and more than 60 local lists considering regional differences. The methods and criteria preferred by the institutions publishing these league tables are specific to the institutions. There are five cases of the Thurstone's law of comparative judgement that can be benefited from under different conditions. The best known cases of Thurstone's law of comparative judgement are case III. and case V. Bradley and Terry's model which has a broad use from sports to genetics is another paired comparative scaling method. In this study university ranking criteria were scaled using two different methods and consistence of emerged order of criteria was investigated. In this study it is aimed to compare the results of Thurstone's III. Case and Bradley-Terry model. The research is a descriptive research in which quantitative research techniques used. The research group consists of 247 academicians working at state and private higher education institutions in Turkey. According to the both scaling methods the most important ranking criterion is Number of Papers Indexed by SCI, SSCI and AHCII. Moreover, academicians think the least important criterion is Number of Graduate of Member Who Has Nobel Prize. In both scaling methods the order of university ranking criteria does not change.

Keywords: Scaling, Thurstone, Bradley-Terry, Paired Comparison, Thurstone's III. Case

1. Giriş

Yükseköğretim kurumlarının yürüttüğü bilimsel araştırma-geliştirme faaliyetleri ile ülkelerin uluslararası arenadaki gücü ve toplumlarının refah seviyesi arasında oldukça sıkı bir bağ vardır. Bu bağlamda her ulus bilimsel etkinlik yürüten kurumlarının niceliğini arttırmanın yanında niteliklerini de geliştirmenin yollarını aramaktadır. Bu noktada küresel veya yerel lig tabloları yükseköğretim kurumu seçme durumundaki öğrencilere yol gösterirken yükseköğretim kurumu yöneticilerine ve yasa yapıcılara iyileştirilmesi gereken noktaları belirlemelerinde yardımcı olan önemli bir araç haline gelmiştir (Dill ve Soo, 2005; Lo, 2014, s. 41; Saka ve Yaman, 2011). Günümüzde yüksek öğretim kurumları sıralamaları üniversiteler, özel kuruluş veya devlet kurumlarına sponsor olan iş dünyası, potansiyel öğrenciler ve ebeveynleri ve halk tarafından dikkatle takip edilmektedir (Kaycheng, 2015). Üniversitelerin ve ülkelerin bu sıralamaları ciddiyetle takip etme sebeplerinden biri hiç kuşkusuz uluslararası eğitim sağlamanın, bir ülkenin veya üniversitenin ekonomisine yaptığı önemli katkıdır. OECD raporlarına göre 1975'te 0.6 milyon olan uluslararası öğrenci sayısı 2006 yılına gelindiğinde 2.9 milyona ulaşmıştır (Marconi ve Ritzen, 2015). Bu öğrenciler her yıl ev sahibi ülkelerde eğitim ücreti ve yaşam gideri olarak 30 milyar dolar harcamaktadır. Bu küresel marketin yükseköğretim kurumları üzerindeki baskısı ile küresel sıralamalar, üniversitelerin, ülkelerin ve öğrencilerin davranışlarını etkileyen ve akademik kalitenin bir kriteri olarak görülen önemli bir etken haline almıştır (Dill ve Soo, 2005; Hazelkorn, 2013; Jarocka, 2015). Bu listelerin üst sıralarında kendi kurumlarına yer açmaya çalışan birçok ülke yükseköğretim kurumlarında gerek ekonomik gerek yapısal bazı revizyonlarda bulunmuşlardır. Örneğin, Tayvan hükümeti beş yıl içerisinde farklı alanlarda özelleşmiş elit araştırma merkezleri kurmak ve bazı üniversitelerini 10 yıl içinde dünyanın en iyi 100 üniversitesi listelerine yerleştirebilmek için "beş yıl-beş milyar" programını hayata geçirmiştir. Bu program ile Tayvan yükseköğretiminin küreselleşmesi ve üst düzey araştırmalarının teşvik edilmesi amaçlanmıştır (Lo, 2014, s. 82). Benzer bir biçimde Fransa, Almanya, Rusya, İspanya, Çin, Güney Kore, Malezya, Finlandiya, Hindistan ve Japonya gibi birçok ülke bazı sıralama kuruluşları tarafından üstün başarıyı tanımlayan birtakım göstergeleri baz alarak dünya çapında üniversiteler oluşturma yoluna gitmişlerdir (Hazelkorn, 2013).

Günümüzde uzmanlık alanlarını ve bölgesel farklılıkları dikkate alan 60'tan fazla yerel sıralama listesi ve sekiz tane önemli küresel sıralama listesi vardır (Hazelkorn, 2013). Saka ve Yaman (2011) bazı uluslararası sıralamaları şu şekilde listelemişlerdir:

- Times Dergisi Yüksek Öğretim Sıralaması (Times Higher Education Supplement, THES)
- Şanghai Jiaotong Üniversitesi Dünya Üniversiteleri Akademik Sıralaması (Academic Ranking of World Universities)
- Newsweek Dergisinin Dünyanın En İyi 100 Üniversitesi Sıralaması (Top 100 Global Universities)

- Cybermetric Laboratuvarları Webometrics: Webde Dünya Üniversitelerinin Sıralaması (Webometrics: World Universities' Ranking on the Web)
- Google Search Uluslararası Üniversiteler Sıralaması (G-Factor International University Rankings)
- MINES Paris Teknolojileri Dünya Üniversiteleri Meslek Sıralaması (Professional Ranking)
- HEETACT Dünya Üniversiteleri Bilimsel Çalışmalar Performans Sıralaması (Performance Ranking of Scientific Papers for World Universities)
- Çin Wuhan Üniversitesi Küresel Üniversite Sıralaması (Global University Ranking)

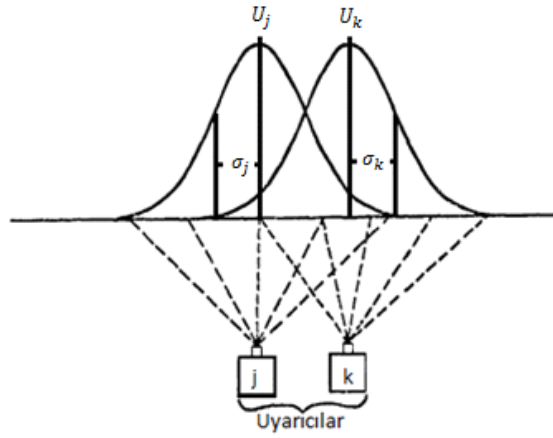
Üniversite başarı sıralaması listelerinden Shanghai Üniversitesi'nin yayımladığı Dünya Üniversiteleri Akademik Sıralaması ilk defa 2003'te yayımlanmıştır. Bu listeyi Times tarafından yayımlanan Dünya Üniversiteler Sıralaması (Times Higher Education Supplement, THES) takip etmiştir. Aslında Anglo-Amerikan ülkelerinde üniversite sıralamalarına ilişkin tabloların yayımlanması daha eski tarihlere dayanmaktadır. Yalnız bu sıralamalar sadece yerel ölçekte yapılmış olup uluslararası karşılaştırmalara yer verilmemiştir. Örneğin, ilk lisansüstü eğitim programları değerlendirmesi 1920'lerde başlamış; 1983'te ABD kolejleri sıralaması yayımlanmış ve 1990'larda ise bu akım Birleşik Krallık'ı etkisi altına almıştır. Sadece İsviçre, Avusturya ve Almanya gibi Almanca konuşan topluluklara ait yükseköğretim kurumlarının sıralaması olan Centrum für Hochschulentwicklung sıralaması (CHE) da yine ilk olarak 1998 yılında yayımlanmıştır. Tüm bunlara rağmen üniversite sıralamalarına dünya genelinde ilginin çekilmesi Shanghai sıralaması ile ilk olarak 2003 yılında olmuştur (Erkkilä, 2013; Hazelkorn, 2013). Türkiye'de 2009'da Orta Doğu Teknik Üniversitesi'nin öncülüğünde yeni bir yerel sıralama sistemi gündeme gelmiş ve daha sonra hayata geçirilmiştir (Alaşehir, 2010).

Literatürde üniversite başarı sıralamalarının birçok yönü eleştirilmektedir. Yükseköğretim kurumundan toplanan verilerin güvenilir olmama sorunu (Kaycheng, 2015; Hazelkorn, 2013), İngilizcenin resmi dil olarak kullanıldığı ülkelerin avantajlı olması (Kaycheng, 2015), öğretim yerine araştırmacının ödüllendirilmesi (Marconi ve Ritzen, 2015), sonuca odaklı araştırmaların teşvik edilmesi (Lo, 2014), hükümetler tarafından bazı üniversitelere yüksek kaynak aktarılırken diğerlerinin daha az pay alması (Hazelkorn, 2013), sıralamada kullanılan kriterlerin ağırlıklandırılması ve toplam puan elde edilmesi (Kaycheng, 2015; Rauhvarger, 2011) üniversitelerin sadece %1-%3'nün değerlendirilmesi (Rauhvarger, 2011), yoğun olarak doğa ve sağlık bilimleri çalışmaları yapan üniversitelerin sıralamalarda avantajlı olması (Rauhvarger, 2011) eleştirilen noktalardan bir kaçısıdır.

Ölçekleme kavramı, temelleri ilk olarak Fechner tarafından 19. yüzyıl sonlarında atılan psikofizik bilim dalında kullanılmıştır (Espinoza ve Conner, 1982; Brown ve Peterson, 2009). Psikofizik bizleri kuşatan çevreden veya bedenimizden kaynaklı fiziksel uyarıcıların algılanan psikolojik boyuttaki şiddeti, ayırt edilebilirliği ve ölçeklenebilirliği üzerine çalışmaların ve bu çalışmalardan yola çıkarak oluşturulan kanunların bütünü olarak tanımlanabilir (Turgut ve Baykul, 1992; Borg, 1990; Irtel, 2005). Psikofizikte ölçeklemede uyarıcıların fiziksel boyutta olduğu ve bu fiziksel uyarıcıların psikolojik boyutta algılandığı varsayımı vardır. Thurstone, psikolojik boyuttaki a ve b gibi iki uyarıcı arasındaki farkın yine psikolojik boyuttaki c ve d uyarıcıları arasındaki farka eşitlenebileceğini varsayarak psikofiziksel ölçeklemeye yeni bir bakış açısı getirmiştir (Gilbert, Fiske ve Lindzey, 1998).

Ölçekleme söz konusu olduğunda ayırt etme sürecinde çoğunlukla sıralama işin içine girer. Bir çift yargıdan araştırılan özelliğe hangisinin daha fazla sahip olduğunun sorgulanması işlemi ikili karşılaştırma olarak tanımlanır (Irtel, 2005). Thurstone ikili karşılaştırma yargılarına dayandırdığı "Karşılaştırmalı Yargı Kanunu" isimli çalışmasında psikolojik ölçeklemenin temelini hipotetik bir "ayırtıcı sürece" dayandırmıştır. Ona göre ikili karşılaştırma yargılarından her biri bireydeki psikolojik ölçekte bir noktaya denk gelir. Bu iki yargının psikolojik ölçekteki yeri tanımlanamayan bir ayırtıcı süreç tarafından belirlenir. İnsanoğlunun algısal durumunun karmaşık doğası gereği bu ayırtıcı süreç her defasında farklı şekilde meydana gelir. Bu durumda aynı ikili karşılaştırma yargısı bireye tekrar tekrar sunulduğunda, bunlar algılanan boyutta normal

bir dağılım sergilerler. Bu dağılımın ortalaması standart sapmanın birim olarak kullanıldığı aralık ölçeğinde o yargının ölçek değeri olarak kabul edilir. Şekil 1’de j ve k uyarıcılarının psikolojik boyuttaki hipotetik dağılımı ve ölçek değeri görülmektedir (Espinoza ve Conner, 1982).



Şekil 1. Uyarıcıların Psikolojik Boyutta Dağılımı

Thurstone (1927) j ve k gibi iki uyarıcının psikolojik boyuttaki ölçek değerleri arasındaki farkı Eşitlik 1 yardımı ile ortaya koymuştur. Bu denkleme *karşılaştırmalı yargı kanunu* adını vermiştir. Bulunan bu farkın olasılık değeri normal dağılım eğrisinde iki değer arasında kalan alanın hesaplanması ile elde edilir.

$$S_j - S_k = Z_{jk} \sqrt{\sigma_j^2 + \sigma_k^2 - 2r\sigma_j\sigma_k} \quad (1)$$

Eşitlik 1’de S_j ve S_k karşılaştırılan iki uyarıcının psikolojik ölçek değerini, Z_{jk} $P_{j>k}$ yargılarının oranının toplamını ve σ_j, σ_k uyarıcıların standart sapmasını ve r sembolü de uyarıcıların standart sapmaları arasındaki korelasyonu temsil eder. Karşılaştırmalı yargı kanununun farklı varsayımlar altında kullanılan beş farklı hali olduğu vurgulanmıştır. Uyarıcılar arasında dikkat çekici bir uyarıcı olmadığı durumlarda yani uyarıcılar homojen olduklarında bu uyarıcılara ait standart sapmalar arasındaki ilişki sıfıra çok yakın değerler alır. Korelasyonun sıfır kabul edildiği durumda, III. hal denklemleri olarak adlandırılan Eşitlik 2’deki formüle ulaşılır.

$$S_j - S_k = Z_{jk} \sqrt{\sigma_j^2 + \sigma_k^2} \quad (2)$$

Bu denklemin matematiksel işlemlerinin yoğunluğu nedeniyle Thurstone (1927) psikolojik ölçekleme için neredeyse kullanılamaz olduğunu vurgulasa da, Burros (1951, akt. Turgut ve Baykul, 1992) bu hesaplamaları yapmanın yöntemini geliştirmiştir. Uyarıcıların standart sapmalarının eşit olduğu varsayımı altında karşılaştırmalı yargı kanununun V. haline ulaşılır. V. hal denklemleri Eşitlik 3’te görüldüğü gibidir:

$$S_j - S_k = Z_{jk} = Z_{jk} \quad (3)$$

İkili karşılaştırma yargılarına dayalı ölçekleme tekniklerinden bir diğeri psikolojik deneyler, spor müsabakaları ve genetik gibi geniş bir kullanım alanı olan Bradley-Terry modelidir (Turner ve Firth, 2010; Chetrite, Diel ve Lerasle, 2015). Bu model ilk olarak Ralph Bradley ve Milton Terry’nin 1952’de

yayınladıkları *Eksik Blok Tasarımında Sıra Analizi: İkili Karşılaştırma Yöntemi* isimli çalışmalarında tanıtılmıştır (Hallinan 2005; Tsukida ve Gupta, 2011).

Bu yaklaşım da Thurstone modeli gibi ayırıcı süreç varsayımına dayanmaktadır. Yalnız bu modelde yargıların psikolojik boyuttaki ölçek değerlerini tespit etmek için normal toplamli dağılım fonksiyonu yerine lojistik toplamli dağılım fonksiyonu kullanılmaktadır (Brown ve Peterson, 2009; Agresti, 1996). Modelin genel hali Eşitlik 4'te gösterilmiştir.

$$P_{jk} = \frac{\pi_j}{\pi_j + \pi_k} \quad (4)$$

Burada P_{jk} , j yargısının k yargısına tercih edilme olasılığını, π_j ve π_k karşılaştırma cümlelerinin gücünü temsil etmektedir. Dikkat edildiğinde π_j ve π_k toplamının bire eşit olduğu görülebilir. Bu modelde j yargısı k yargısı ile karşılaştırıldığında psikolojik boyutta gözlenemeyen ve diğer yargılardan bağımsız olan bir S_i puanının meydana geldiği varsayılır. S_i değerinin $\log \pi_j$ parametresine uygun bir dağılımı takip ettiği kabul edilir. Bu durumda S_i 'nin toplamli dağılım fonksiyonu olan $F_j(s)$ Eşitlik 5'te gösterilen şekli alır (Hallinan 2005).

$$F_j(s) = \exp(-e^{-(s - \log \pi_j)}) \quad (5)$$

Ardından iki yargıya ait psikolojik ölçek değeri arasındaki farkın ($S_j - S_k$) ortalaması olan lojistik dağılımdan Eşitlik 6 yardımı ile hesaplanır (Hallinan 2005).

$$S_j - S_k \sim F_j(s) = \frac{1}{1 + e^{-(s - (\log \pi_j - \log \pi_k))}} \quad (6)$$

Kökenini psikofizikten alan Thurstone'nun karşılaştırmalı yargı kanununun V. hal denklemi ile tüketici tercihleri ve deneysel desen çalışmalarından geliştirilmiş Bradley-Terry modelinin benzer sıralamalar sunduğu bilinen bir durumdur. Benzer sonuçlar üretmelerine karşın Bradley-Terry modelinin istatistiksel prosedürünün daha kolay olduğu ve matematiksel olarak daha gelişmiş olduğu savunulmuştur (Handley, 2004). Alanyazında V. hal denklemi ile Bradley-Terry modelinin karşılaştırıldığı çalışmalara rastlamak mümkündür. Alanyazında veri-model uyumsuzluğu durumunda V. hal denkleminin alternatif olarak kullanılan III. hal denklemi ile Bradley-Terry modelinin sonuçlarının karşılaştırıldığı bir çalışmaya rastlanmamıştır. Bu iki modelin ortaya koyduğu sonuçların tartışılması açısından bu çalışma önemlidir. Bu çalışmada öğretim elemanlarının görüşlerine göre üniversite sıralama kriterlerinin III. hal denklemi ve Bradley-Terry ölçekleme modeli ile elde edilen kriter sıralamalarının karşılaştırılması amaçlanmıştır. Bu doğrultuda aşağıdaki üç probleme yanıt aranmıştır.

- Üniversite sıralama kriterlerinin III. Hal denkleminin dayalı ölçekleme yöntemine göre elde edilen kriter sıralaması nasıldır?
- Üniversite sıralama kriterlerinin Bradley-Terry modeline dayalı ölçekleme yöntemine göre elde edilen kriter sıralaması nasıldır?
- Üniversite sıralama kriterlerinin III. Hal denkleminin dayalı ölçekleme yöntemi ve Bradley-Terry modeline dayalı ölçekleme yöntemine göre elde edilen kriter sıralaması arasında nasıl bir ilişki vardır?

2. Yöntem

2.1. Araştırma Deseni

Bu çalışmada iki farklı ikili karşılaştırmaya dayalı ölçekleme yöntemi ile ölçekleme yapılmış ve bu iki modelden elde edilen kriter sıralamalarının tutarlılığı incelenmiştir. Çalışma nicel araştırma tekniklerinin kullanıldığı temel araştırma modelinde desenlenmiştir (Karasar, 2002).

2.2. Çalışma Grubu

Araştırmanın çalışma grubunu ölçek formuna internet ortamında yanıt veren Türkiye'deki kamu veya özel/vakıf yükseköğrenim kurumlarında çalışmakta olan 247 akademisyen oluşturmaktadır. Katılımcılara ilişkin demografik bilgiler Tablo 1'deki gibidir:

Tablo 1. Katılımcılara İlişkin Demografik Değişkenler

Değişken	Düzye	n	%
Cinsiyet	Kadın	78	31,6
	Erkek	169	68,4
Unvan	Prof. Dr.	24	9,7
	Doç. Dr.	34	13,8
	Yrd. Doç. Dr.	74	30
	Arş. Gör. Dr./Öğr. Gör. Dr.	15	6,1
	Arş. Gör.	73	29,6
	Öğr. Gör. /Okt./ Uzman	27	10,9
Çalışma Süresi	1-5 Yıl Arası	99	40,1
	6-10 Yıl Arası	51	20,6
	11-15 Yıl Arası	29	11,7
	15 Yıldan Fazla	68	27,5

Tablo 1 incelendiğinde katılımcıların büyük bir bölümünün erkek olduğu ve genel olarak unvan arttıkça katılımcı sayısının azaldığı görülebilir. Benzer bir biçimde araştırmaya en fazla 1-5 yıl arasında kıdeme sahip akademisyenler katılırken, 11-15 yıl arası çalışma süresi olan öğretim elemanları katılımcılar arasındaki en küçük gruptur.

2.3. Veri Toplama Aracı

Araştırmada ölçme aracı olarak üniversite sıralamalarında kullanılan yedi kriterden elde edilmiş 21 ikili karşılaştırma yargısı içeren bir ölçek kullanılmıştır. Ölçeğin geliştirilmesinde öncelikle literatür taraması yapılarak önde gelen üniversite sıralama kuruluşlarının kriterleri belirlenmiştir. Ardından uzman görüşleri dikkate alınarak, bu kuruluşların kullandığı ortak kriterlerden yedisinin (U1: üniversitesin Nobel sahibi üye veya mezun sayısı, U2: Nature ve Science dergilerinde yayımlanan makale sayısı, U3: SCI, SSCI ve AHCI indekslerinde taranan makale sayısı, U4: öğretim üyesi başına düşen atıf sayısı, U5: öğretim üyesi başına düşen öğrenci sayısı, U6: uluslararası öğrenci sayısı, U7: uluslararası öğretim üyesi sayısı) kullanılmasına karar verilmiştir. Son olarak, bu yedi kriterin kendi aralarında ikili olarak karşılaştırılmalarından elde edilen 21 ikili karşılaştırma yargısı oluşturulmuştur. Karşılaştırma yargılarından çevrimiçi ölçek formu hazırlanmıştır.

2.4. İşlem

Kriterler ilk olarak Thurstone'nun ikili karşılaştırma yargılarına dayalı ölçekleme yöntemlerinden V. hal denklemi ile ölçeklenmiş ve verilerin bu modele uyum sağlamadığı tespit edilmiştir ($\chi^2 = 34.17, p < .05$). Bu durumda Guilford (1954) tarafından tavsiye edilen III. hal denklemi kullanılmıştır.

III. hal denkleminde dayalı ölçekleme işleminde öncelikle frekans tablosundan Z birim matrisi oluşturulmuştur. Ardından Z birim matrisinin sütun toplamları ($\sum Z_j$) hesaplanmıştır. Daha sonra Z birim normal sapmalar karesi matrisinden sütunların toplamı ($\sum Z_j^2$) bulunmuş ve uyarıcı sayısı ile çarpılmıştır ($K\sum Z_j^2$). Sütun toplamlarının kareleri alınarak $(\sum Z_j)^2$ hesaplanmıştır. Ardından sütun elemanlarının standart sapması (KV_j) hesaplanarak $(\sum 1/KV_j)$ değeri bulunmuştur. $\sum 1/KV_j$ değerlerinin toplamından K.C sabit değeri elde edilmiştir.

$$K.C = \frac{2.K}{\sum_{j=1}^K \frac{1}{KV_j}} \quad (7)$$

$$\sigma_j = \frac{K.C}{KV_j} - 1 \quad (8)$$

Eşitlik 7 ve Eşitlik 8 yardımıyla ayırt etme yargılarının standart sapma değeri hesaplanmıştır. Bulunan standart sapmaların kareleri alınarak varyans değerlerine ulaşılmıştır. Varyansların ikişer ikişer toplamı hesaplanarak varyans toplamları matrisi elde edilmiştir. Bu matrisin elemanlarının karekökü alındıktan sonra her biri Z birim matrisinin elemanları ile çarpılarak S matrisi elde edilmiştir. S matrisinin sütun toplamları elde edilerek $\sum Z_j$ değerleri bulunmuş ardından bu değerler K uyarıcı sayısına bölünerek S_j değerlerine ulaşılmıştır. Son olarak en küçük S_j değerini sıfır yapan değer tüm S_j değerlerine eklenerek ölçek değerlerine (S_c) ulaşılmıştır. S_c ölçek değerleri küçükten büyüğe doğru sıralanarak kriterlerin önem sırası tespit edilmiştir (Güler ve Anıl, 2009). Bu hesaplamaların tümünde Excel programı kullanılmıştır. Bradley-Terry modeline dayalı ölçekleme analizinde Microsoft Excel programında eklenti olarak çalışan Xlstat yazılımı kullanılmıştır.

3. Bulgular

3.1. Birinci Probleme İlişkin Bulgular

Katılımcılardan üniversite sıralama kriterlerinden her defasında iki tanesini önem düzeyine göre karşılaştırmaları istenmiş ve Tablo 2’deki gibi frekanslar matrisi elde edilmiştir. Ham puanlar matrisinde satırlardaki uyarıcıların sütundaki uyarıcılara tercih edilme frekansı gösterilmektedir. Örneğin, U1 kriterinin U2 kriterinden daha önemli olduğunu düşünen katılımcı sayısı 57 iken, U2 uyarıcısının daha önemli olduğunu düşünen 190 kişi vardır. Bir uyarıcının kendisi ile karşılaştırması yapılmadığından köşegenlerin boş olduğu görülebilir.

Tablo 2. Yedi Kritere İlişkin Ham Puanlar Matrisi (F)

	U1	U2	U3	U4	U5	U6	U7
U1		57	46	51	112	91	66
U2	190		47	82	157	152	135
U3	201	200		137	202	200	191
U4	196	165	110		215	205	187
U5	135	90	45	32		127	101
U6	156	95	47	42	120		67
U7	181	112	56	60	146	180	
Toplam	5187						

Ham puanlar matrisinin her bir elemanı katılımcı sayısına (247) bölünerek oranlar matrisi elde edilmiştir. Tablo 3’te oranlar matrisi görülmektedir.

Tablo 3. *Yedi Kriteria İlişkin Oranlar Matrisi (P)*

	U1	U2	U3	U4	U5	U6	U7
U1		0,23	0,19	0,21	0,45	0,37	0,27
U2	0,77		0,19	0,33	0,64	0,62	0,55
U3	0,81	0,81		0,55	0,82	0,81	0,77
U4	0,79	0,67	0,45		0,87	0,83	0,76
U5	0,55	0,36	0,18	0,13		0,51	0,41
U6	0,63	0,38	0,19	0,17	0,49		0,27
U7	0,73	0,45	0,23	0,24	0,59	0,73	

Tablo 3'teki değerler incelenecek olursa karşılıklı oranlar toplamının bire eşit olduğu görülebilir. Bu karşılıklı değerler üzerinden matrise yanlış veri girilip girilmediği kontrol edilebilir. Oranlar matrisindeki her bir değer için birim normal dağılımının z değeri hesaplanarak Tablo 3'te görülen birim normal standart sapmalar matrisi elde edilmiştir. Bu hesaplamalar yardımı ile ayırt etme yargılarının standart sapmaları ve varyanslarına ulaşılmıştır.

Tablo 4. *Birim Normal Sapmalar (Z) Matrisi ve Gözlemci Varyanslarının Kestirilmesi*

	U1	U2	U3	U4	U5	U6	U7
U1		-0,74	-0,89	-0,82	-0,12	-0,34	-0,62
U2	0,74		-0,88	-0,43	0,35	0,29	0,12
U3	0,89	0,88		0,14	0,91	0,88	0,75
U4	0,82	0,43	-0,14		1,13	0,95	0,70
U5	0,12	-0,35	-0,91	-1,13		0,04	-0,23
U6	0,34	-0,29	-0,88	-0,95	-0,04		-0,61
U7	0,62	-0,12	-0,75	-0,70	0,23	0,61	
$\sum Z_j$	3,52	-0,18	-4,44	-3,90	2,46	2,43	0,10
$\sum Z_j^2$	2,52	1,72	3,74	3,55	2,28	2,25	1,87
$K\sum Z_j^2$	17,64	12,04	26,16	24,83	15,99	15,75	13,10
$(\sum Z_j)^2$	12,40	0,03	19,71	15,17	6,05	5,92	0,01
KV_j	2,29	3,46	2,54	3,11	3,15	3,14	3,62
$\sum 1/KV_j$	0,44	0,29	0,39	0,32	0,32	0,32	0,28

Tablo 5'te ayırt etme yargılarının standart sapma ve varyans değerleri görülmektedir.

Tablo 5. *Ayırt Etme Yargılarının Standart Sapması ve Varyans Değerleri*

	U1	U2	U3	U4	U5	U6	U7
σ_j	1,60	0,72	1,34	0,91	0,89	0,90	0,64
σ_j^2	2,55	0,51	1,80	0,84	0,79	0,80	0,42

Bu matris herhangi iki uyarıcının psikolojik boyuttaki değerlerinin farkını ($S_j - S_k$) göstermektedir.

Thurstone ölçekleme yönteminde uyarıcılar ölçek değeri en küçük olandan en büyük olana doğru sıralanırlar. Örneğin bu çalışmada en önemli bulunan uyarıcı ölçek değeri sıfır olan U3 uyarıcısı, ikinci en önemli uyarıcı U4 uyarıcısıdır.

Tablo 6. *S* Matrisi ($S_j - S_k = Z_{jk}$)

	U1	U2	U3	U4	U5	U6	U7
U1		-1,29	-1,86	-1,51	-0,21	-0,62	-1,07
U2	1,29		-1,33	-0,50	0,40	0,34	0,11
U3	1,86	1,33		0,22	1,46	1,42	1,12
U4	1,51	0,50	-0,22		1,44	1,22	0,78
U5	0,21	-0,40	-1,46	-1,44		0,04	-0,25
U6	0,62	-0,34	-1,42	-1,22	-0,04		-0,67
U7	1,07	-0,11	-1,12	-0,78	0,25	0,67	
Toplam	6,56	-0,29	-7,41	-5,23	3,29	3,08	0,01
S_j	1,09	-0,04	-1,06	-0,75	0,47	0,44	0,00
S_c	2,15	1,02	0,00	0,31	1,53	1,50	1,06

Tablo 7. *III. Hal Denklemi ile Yapılan Ölçeklemeden Elde Edilen Ölçek Değerleri ve Kriterlerin Önem Sırası*

Sıra	Ölçek Değeri	Uyarıcılar
1	0,00	U3: SCI, SSCI ve AHCI İndekslerinde Taranan Makale Sayısı
2	0,31	U4: Öğretim Üyesi Başına Düşen Atıf Sayısı
3	1,02	U2: Nature ve Science Dergilerinde Yayımlanan Makale Sayısı
4	1,06	U7: Uluslararası Öğretim Üyesi Sayısı
5	1,50	U6: Uluslararası Öğrenci Sayısı
6	1,53	U5: Öğretim Üyesi Başına Düşen Öğrenci Sayısı
7	2,15	U1: Üniversitenin Nobel Sahibi Üye veya Mezun Sayısı

Akademisyenlere göre üniversite sıralama kuruluşlarının kullandığı kriterlerin önem düzeyine göre sıralanmış hali ve bu kriterlerin ölçek değerleri Tablo 7'de görüldüğü gibidir. Buna göre akademisyenler SCI, SSCI ve AHCI indekslerinde taranan makale sayısının üniversite sıralaması için en önemli kriter olması gerektiğini düşünürken, üniversitenin Nobel ödülü alan üye veya mezun sayısını en önemsiz kriter olarak görmekte dirler.

3.2. İkinci Probleme İlişkin Bulgular

Bradley-Terry modeli ile yapılan ikili karşılaştırma yargılarına dayalı ölçekleme sonucunda uyarıcıların tercih edilme durumlarına ilişkin bazı bilgiler Tablo 8'de özetlenmiştir.

Tablo 8. *Bradley-Terry Modeli ile Yapılan Ölçeklemede Uyarıcılara İlişkin Özet*

Uyarıcı	Tercih Edilme Sayısı	Tercih Edilmeme Sayısı	Tercih Edilme Oranı	Tercih Edilmeme Oranı
U2	763,00	719,00	51,48	48,52
U1	423,00	1059,00	28,54	71,46
U3	1131,00	351,00	76,32	23,68
U6	561,00	921,00	37,85	62,15
U4	1078,00	404,00	72,74	27,26
U5	530,00	952,00	35,76	64,24
U7	701,00	781,00	47,30	52,70

Tablo 8'de ikili karşılaştırmalar sonucunda hangi uyarıcının katılımcılar tarafından kaç kez daha önemli bulunduğu, kaç kez daha az önemli bulunduğu ve bunların oranlarına ilişkin değerler görülmektedir. Örneğin, U2 kriterinin ikili karşılaştırmalarda 763 kez daha önemli olduğu düşünülürken; 719 kez karşılaştırıldığı kriterin daha önemli olduğu belirtilmiştir. Tablo 9'da Bradley-Terry modeli ile elde edilen ölçek değerleri ve bu değerlere ait standart sapmalar ve güven aralıkları sunulmuştur.

Tablo 9. *Bradley-Terry Modeli ile Yapılan Ölçeklemeden Elde Edilen Ölçek Değerleri, Standart Sapmalar ve Güven Aralıkları*

Uyarıcılar	Ölçek Değeri	Standart Sapma	Güven Aralığı Alt Sınır	Güven Aralığı Üst Sınır
U2	0,77	0,00	0,77	0,77
U1	0,32	0,01	0,31	0,33
U3	2,23	0,03	2,18	2,29
U6	0,49	0,00	0,48	0,49
U4	1,99	0,02	1,95	2,02
U5	0,47	0,00	0,46	0,47
U7	0,74	0,00	0,45	0,46

Tablo 9'daki ölçek değerleri Thurstone modelindeki ölçek değerleri gibi yorumlanmaz. Burada Thurstone modelinin aksine değerlerin büyük olması uyarıcıların önem sırasını yükselten bir durumdur. Örneğin analiz sonucuna göre en önemli uyarıcı en büyük değere (2,23) sahip olan U3 uyarıcısı iken, en önemsiz bulunan uyarıcı 0,32 ölçek değerine sahip U1 uyarıcısıdır. Tablo 10'da üniversite sıralama kriterlerinin önem düzeylerine göre sıralanmış hali ve ölçek değerleri görülmektedir.

Tablo 10. *Bradley-Terry Modeli ile Yapılan Ölçeklemeden Elde Edilen Ölçek Değerleri ve Kriterlerin Önem Sırası*

Sıra	Ölçek Değeri	Uyarıcılar
1	2,23	U3: SCI, SSCI ve AHCI İndekslerinde Taranan Makale Sayısı
2	1,99	U4: Öğretim Üyesi Başına Düşen Atıf Sayısı
3	0,77	U2: Nature ve Science Dergilerinde Yayımlanan Makale Sayısı
4	0,74	U7: Uluslararası Öğretim Üyesi Sayısı
5	0,49	U6: Uluslararası Öğrenci Sayısı
6	0,47	U5: Öğretim Üyesi Başına Düşen Öğrenci Sayısı
7	0,32	U1: Üniversitesin Nobel Sahibi Üye veya Mezun Sayısı

Tablo 10'da görüldüğü üzere akademisyenlere göre üniversite sıralamalarında kullanılması gereken en önemli kriter SCI, SSCI ve AHCI indekslerinde taranan makale sayısı iken, en önemsiz bulunan kriter üniversitesin Nobel ödülü alan üye veya mezun sayısıdır.

3.3. Üçüncü Probleme İlişkin Bulgular

Tablo 11'de iki yöntemden elde edilen kriterlerin önem düzeyine göre sıralanmış hali ve ölçek değerleri görülmektedir.

Tablo 11. *III. Hal Denklemi ve Bradley-Terry Modeli ile Ölçekleme Sonuçlarının Karşılaştırılması*

Uyarıcı Sıraları			Ölçek Değerleri	
III. Hal	Bradley-Terry	Uyarıcılar	III. Hal	Bradley-Terry
1	1	U3: SCI, SSCI ve AHCI İndekslerinde Taranan Makale Sayısı	0,00	2,23
2	2	U4: Öğretim Üyesi Başına Düşen Atıf Sayısı	0,31	1,99
3	3	U2: Nature ve Science Dergilerinde Yayımlanan Makale Sayısı	1,02	0,77
4	4	U7: Uluslararası Öğretim Üyesi Sayısı	1,06	0,74
5	5	U6: Uluslararası Öğrenci Sayısı	1,5	0,49
6	6	U5: Öğretim Üyesi Başına Düşen Öğrenci Sayısı	1,53	0,47
7	7	U1: Üniversitesin Nobel Sahibi Üye veya Mezun Sayısı	2,15	0,32

Tablo 11’de iki yöntemin birebir aynı sıralama sonuçlarını meydana getirdiği görülmektedir. İki yöntemin birbirine alternatif olarak kullanılabilmesi ve sıralamaların aynı çıkmasında iki yöntemin benzer dağılımlara dayalı hesaplamalarının etkili olduğu söylenebilir.

4. Tartışma ve Sonuç

Urban, Culler, Thurstone ve Stevens’in katkıları ile ölçeklemenin kapsamı psikofiziğin ilgi alanı olan duyumsal eşik değerleri gibi konuların ötesine geçerek psikolojinin konusu olan algısal büyüklüklere kadar genişlemiştir. Günümüzde temel amacı bireydeki duyuşsal değerler, inançlar, tutumlar, kişilik özellikleri ve önyargılar gibi psikolojik yapıların ölçülmesi olan psikolojik ölçekleme, psikofiziksel ölçeklemeden bağımsız bir bağlama oturmuştur (Rajamanickam, 2002). Bu çalışmada ikili karşılaştırma yargılarına dayalı iki ölçekleme yönteminin, uyarıcıların psikolojik boyuttaki sıralaması üzerindeki etkisi incelenmiştir.

Her iki ölçekleme yönteminde de üniversite sıralama kriterlerinin sıralarının aynı olduğu belirlenmiştir. Akademisyenler SCI, SSCI ve AHCI indekslerinde taranan makale sayısının en önemli kriter olduğunu düşünürken, üniversitelerin Nobel Ödülü alan üye veya mezun sayısını en önemsiz kriter olarak görmektedir.

Bu bulgular ışığında kullanılan ölçekleme yönteminin uyarıcıların psikolojik boyuttaki sırasını bir farklılık yaratmadığı söylenebilir. Alanyazında V. hal denklemi ile Bradley-Terry modelinin çok yakın sonuçlar meydana getirdiğine ilişkin araştırmalar vardır. Bu çalışma ile bu durumun III. hal denklemi için de geçerli olduğu tespit edilmiştir. Bu durumda matematiksel olarak gelişmiş olan Bradley-Terry yöntemi (Handley, 2004) ile yaygın olarak kullanılan Thurstone yöntemi benzer durumlarda aynı amaç için kullanılabilir. Yalnız istatistik yazılımları sayesinde kolayca uygulanabilmesi Bradley-Terry yönteminin önemli bir avantajı olarak görülebilir.

5. Öneriler

Ölçekleme çalışması yapacak araştırmacılara Bradley-Terry yönteminde eksik veri matrisi ile analizler yapıp Thrustone modeli ile karşılaştırmaları önerilebilir. Ayrıca Bradley-Terry yöntemi ile sıralama yargılarına dayalı ölçekleme yöntemleri karşılaştırılabilir.

Kaynakça

- Alaşehir, O. (2010). *University ranking by academic performance: A scientometrics study for ranking world universities* (Doctoral dissertation). Middle East Technical University, Ankara.
- Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work, Environment & Health*, 16, 55–58.
- Brown, T. C., & Peterson, G. L. (2009). An enquiry into the method of paired comparison. *US Department of Agriculture, General Technical Reports*. CO, USA Fort Collins,.
- Chetrite, R., Diel, R., & Lerasle, M. (2015). Bradley-Terry model in random environment: Does the best always win? <https://hal.archives-ouvertes.fr/hal-01204517> adresinden erişilmiştir.
- Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, 49(4), 495–533. doi:10.1007/s10734-004-1746-8.
- Erkkilä, T. (2013). *Global university rankings: Challenges for european higher education*. Basingstoke, UK: Palgrave Macmillan.
- Guilford, J. P. (1954). *Psychometrics methods*. New York: Mc Graw-Hill Book Co.
- Güler, N. ve Anıl, D. (2009). Scaling through pair-wise comparison method in required characteristics of students applying for post graduate programs. *International Journal of Human Sciences*, 6(1), 627-639.

- Hallinan, S. E. (2005). *Paired comparison models for ranking national soccer teams* (Doctoral dissertation). Worcester Polytechnic Institute.
- Handley, J. C. (Ed). (2001). Final program and proceedings of IS&T's PICS Conference bildirileri: Image Processing, Image Quality, Image Capture Systems Conference. Montréal, Quebec, Canada.
- Hazelkorn, E. (2013) How Rankings are reshaping higher education. Climent, V., Michavila, F. and Ripolles, M. (eds), *Los rankings univertarios*. Mitos y Realidades.
- Irtel, H. (2005). Psychophysical scaling. B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1628–1632). Chichester, U.K.: Wiley.
- Karasar, N. (2002). *Bilimsel araştırma yöntemi* (2. Baskı). Ankara: Nobel Yayınları.
- Kaycheng, S. (2015). Multicollinearity and indicator redundancy problem in world university rankings: an example using times higher education world university ranking 2013-2014 data. *Higher Education Quarterly*, 69(2), 158–174. doi:10.1111/hequ.12058
- Lindzey, G., Gilbert, D., & Fiske, S. T. (1998). *The handbook of social psychology*. UK: Oxford University Press.
- Lo, W. Y. W. (2014). *University rankings: Implications for higher education in Taiwan*. Singapore: Springer.
- Marconi, G., & Ritzen, J. (2015). Determinants of international university rankings scores. *Applied Economics*, 47(57), 6211–6227. doi:10.1080/00036846.2015.1068921
- Rajamanickam, M. (2002). *Modern psychophysical and scaling methods and experimentation*. New Delhi: Concept Pub. Co.
- Rauhvargers, A. (2011). *Global university rankings and their impact. EUA report on rankings*. Brussels: European University Association.
- Thurstone, L. L. (1927). Psychophysical analysis. *The American journal of psychology*, 38(3), 368-389.
- Tsukida, K., & Gupta, M. R. (2011). *How to analyze paired comparison data*. Washington University Seattle Department of Electrical Engineering. Seattle, WA.
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM Yayınları.
- Turner, H., & Firth, D. (2010). Bradley-Terry models in R: the BradleyTerry2 package. *Journal of Statistical Software*, 48(9).