TÜRKİYE CUMHURİYETİ'NİN YÜZÜNCÜ YILI

Veriden Kavrama: Eğitimde Ölçme Değerlendirme

Editör
Gökhan Kayır
HüsemKorkmaz

BİDGE

BİDGE Yayınları

Veriden Kavrama: Eğitimde Ölçme Değerlendirme

Editör: Doç. Dr. Gökhan KAYIR, Dr. Hüsem KORKMAZ

ISBN: 978-625-372-525-9

1. Baskı
Sayfa Düzeni: Gözde YÜCEL
Yayınlama Tarihi: 25.12.2024
BİDGE Yayınları

Bu eserin bütün hakları saklıdır. Kaynak gösterilerek tanıtım için yapılacak kısa alıntılar dışında yayıncının ve editörün yazılı izni olmaksızın hiçbir yolla çoğaltılamaz.

Sertifika No: 71374
Yayın hakları © BİDGE Yayınları
www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com
Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.
Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya / Ankara

İçindekiler

Test Eşitleme: Kuramsal Bir Çerçeve 4
Burcu DEMİRÖZ 4
Ölçek Geliştirme 43
Emrah BÜYÜKATAK 43
Teoriden Uygulamaya Bibliyometrik Analiz 60
Burcu DEMİRÖZ 60
Çoktan Seçmeli Test Geliştirme Sürecinde Madde Analizi ve Madde Yanlılığının İncelenmesi 122
Yusuf KASAP 122
Üniversite Öğrencilerinin Ekolojik Sorunlara İlişkin Görüşlerinin İkili Karşılaştırma Metodu İle İncelenmesi 141
Sümeyra SOYSAL 141
Bilgisayar Ortamında Bireye Uyarlanan Test Uygulamalarında Kullanıcı Deneyimine Etki Eden Faktörler 163
Alper TOSUN 163
Alper SİNAN 163
Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Test Sonlandırma Kuralları 205
Dilek GÜR ALP 205
Alper SİNAN 205

BÖLÜM I

Test Eşitleme: Kuramsal Bir Çerçeve

Burcu DEMİRÖZ[1]

Giriş

Testlerden alınan puanlar, akademik programlara kabul, başarıdaki ilerlemeyi izleme, belirli bir görevdeki yeterlilikleri belirleme, mesleğe yerleşme vb. amaçlarla kullanılır. Bu amaçlar doğrultusunda puanlar bireylerin yetenek düzeyi hakkında bilgi parçası olarak kullanılır. Böylece bireylerin puanları ile ölçülen belirli özelliklere göre yetenek düzeyleri karşılaştırılır. Bireylere aynı test maddeleri rutin olarak uygulanırsa, maddeler hakkında başkalarını bilgilendirebilirler. Birden fazla kez sınava giren kişilere aynı maddeler uygulanabilir. Bu gibi güvenlik sorunlarının önüne geçmek amacıyla aynı özelliği ölçmeyi amaçlayan farklı test formlarının kullanılması yaygındır. Aynı özelliği ölçmeyi amaçlayan farklı formlar oluşturulurken içerik ve istatistiksel

[1]Dr., Türkiye Cumhuriyeti Milli Eğitim Bakanlığı, Türkiye, ORCID: 0000-0002-8326-4671, e-posta: bd14.olcme@gmail.com

özellikler açısından formların birbirine olabildiğince benzer olması gerekir. İçerik ve istatistiksel özellikler açısından benzer formlar oluşturulmaya çalışılsa da madde güçlüğü açısından tamamen eşdeğer olan formlar oluşturmak neredeyse imkansızdır. Formlar arasındaki güçlük farklılıklarından dolayı daha kolay maddeleri yanıtlayan bireyler zor maddeleri yanıtlayan bireylere karşı avantaj kazanmış olur. Bu durum adaletsizliğe neden olur. Adaletsizliğin yanı sıra iki form uygulandığında ortalamaları (başlangıç noktaları) ve standart sapmaları (birimleri) farklı olduğundan bu formlardan alınan puanlar birbirleri ile doğrudan karşılaştırılamaz (Keleci-oğlu, 2024). Livingston'a (2004) göre bir grup için iki formdaki puanlar aynı göreceli konumu temsil ediyorsa eşdeğer olarak ele alınmalıdır. Başlangıç noktası ve birimleri farklı olan ölçümlerde aynı yetenek düzeylerindeki bireyler farklı konumlarda yer alır. Bu nedenle puanlar birbirleri ile doğrudan karşılaştırılamaz. Test eşitleme, madde güçlüğü farklarını ayarlamayı ve formlardan alınan puanların aynı konuma getirilerek karşılaştırılabilir ve birbirinin yerine kullanılabilir olmasını sağlamayı amaçlar. Eşitleme, aynı amaca yönelik olarak hazır-lanmış, içerik ve zorluk açısından benzer formlardan alınan puanların birbiri yerine kullanılabilmesini sağlamak amacıyla yapılan bir istatistiksel işlemdir (Kolen & Brennan, 2004). Angoff'a (1984) göre eşitleme, bir formun birim sistemini diğer bir formunun birim sistemine dönüştürmektir. Bir başka ifade ile puanların diğer formdan elde edilen puanlarla ilişkilendirilmesine veya dönüştürülmesine eşitleme denir (Hambleton & Swaminathan, 1985).

Eşitleme Koşulları

Testlerden alınan puanlarının eşitlenebilmesi için belirli koşulların sağlaması gerekir (Angoff, 1971; Harris & Crouse, 1993;

Holland & Dorans, 2006; Lord, 1980; Petersen vd., 1989). Bu koşullar şunlardır:

a) Simetriklik

Formlardan hangisinin X hangisinin Y olarak adlandırıldığından bağımsız olarak eşitlemenin aynı sonucu vermesi gerekmektedir (Lord, 1980). Yani Form X'ten alınan bir puanı Form Y'den alınan puana eşitlemek için kullanılan eşitleme fonksiyonu, Form Y'deki bir puanı Form X'ten alınan puana eşitlemek için kullanılan fonksiyonun tersi olmalıdır. Buna göre birinci forma ait x_i puanı ikinci forma eşitlendiğinde elde edilen eşitlenmiş puan y_i ise, y_i puanı birinci forma eşitlendiğinde x_i puanı elde edilmelidir. Örneğin, Form X'ten alınan 10 puan Form Y'den alınan 12 puana eşitlenmiş ise Form Y'den alınan 12 puanın Form X'ten alınan 10 puana eşitlenmesi gerekmektedir.

b) Aynı Yapıyı Ölçme

Angoff'a (1970) göre, formların eşitlenebilmesi için öncelikli şart, bu formların aynı yapıyı, özelliği veya yeteneği ölçmesidir. Formlar eşitlenecekse bu durum göz önünde bulundurularak oluşturulmalıdır. Aksi takdirde, kullanılan istatistiksel prosedürlerden bağımsız olarak eşitleme gerçekleştirilemez ve puanlar birbirinin yerine kullanılamaz. Örneğin matematik testinden alınan puan fen bilimleri testinden alınan puana eşitlenemez. İstatistiksel olarak testlerden alınan puanlar birbirine eşitlense bile aynı anlamı ifade etmeyecektir.

c) Eşitlik

y_i gerçek puan için, gözlenen puanların dağılımı, X formunda eşitlenmiş puan dağılımı ile aynı olmalıdır (Kolen & Brennan, 2004). Eşitlik koşulu sağlandığında bireylerin hangi test formunu

aldığı önemsizleşir. Bu koşulun sağlanması için formların aynı içeriği ölçmesi ve benzer psikometrik özelliklerde olması gerekir. Bu durumda formların birebir aynı olmasını gerektirir. Formlar aynı olduğunda test eşitlemeye gerek kalmayacaktır. Eşitlemeye gerek kalmaması nedeniyle eşitlik koşulu esnetilerek zayıf eşitlik koşulunun sağlanmasının yeterli olacağı belirtilmiştir (Morris, 1982). Birinci sıralı eşitlik için Y formundan elde edilen gözlenen puan dağılımlarının ve eşitlenmiş X formu puanlarının ortalamasının eşit olmasıdır. İkinci sıralı eşitlik için Y formundan elde edilen gözlenen puan dağılımlarının ve eşitlenmiş X formu puanlarının standart hataları eşit olmalıdır.

d) Eşitlik

y_i gerçek puan için, gözlenen puanların dağılımı, X formunda eşitlenmiş puan dağılımı ile aynı olmalıdır (Kolen & Brennan, 2004). Eşitlik koşulu sağlandığında bireylerin hangi test formunu aldığı önemsizleşir. Bu koşulun sağlanması için formların aynı içeriği ölçmesi ve benzer psikometrik özelliklerde olması gerekir. Bu durumda formların birebir aynı olmasını gerektirir. Formlar aynı olduğunda test eşitlemeye gerek kalmayacaktır. Eşitlemeye gerek kalmaması nedeniyle eşitlik koşulu esnetilerek zayıf eşitlik koşulunun sağlanmasının yeterli olacağı belirtilmiştir (Morris, 1982). Birinci sıralı eşitlik için Y formundan elde edilen gözlenen puan dağılımlarının ve eşitlenmiş X formu puanlarının ortalamasının eşit olmasıdır. İkinci sıralı eşitlik için Y formundan elde edilen gözlenen puan dağılımlarının ve eşitlenmiş X formu puanlarının standart hataları eşit olmalıdır.

d) Eşit Güvenirlik

İki formun güvenilirlikleri birbirine yakın olmalıdır (Angof, 1971; Crocker & Algina, 2006; Tanguma 2000). Formların güvenirlikleri arttıkça eşitleme işlemleri daha az hatalı sonuçlar verir.

e) Grup Değişmezliği

Eşitleme fonksiyonu, testin uygulandığı bireylerden bağımsızdır. Eşitleme fonksiyonu evrenden seçilen farklı alt gruplardan etkilenmeden aynı sonucu vermelidir (Holland, 2007). Yani eşitlemenin yapıldığı evrende yer alan tüm alt gruplar için aynı eşitleme fonksiyonu geçerli olmalıdır. Örneğin grup değişmezliği koşulu sağlandıysa, kadınlar ve erkekler için aynı eşitleme fonksiyonu bulunacaktır. Lord ve Wingersky (1984), eşitlemenin gözlenen puan özelliklerine dayanan yöntemlerin kesinlikle grup değişmezliği olmadığını belirtmişlerdir. Van der Linden (2000) belirli teorik koşullar altında, gerçek puan eşitleme yöntemlerinde grup değişmezliği olduğunu göstermiştir. Dorans ve Holland (2000), grup değişmezliğini araştırmak için iki istatistik önermektedir. Standartlaştırılmış Kök Ortalama Kare Fark (**R**oot **M**ean **S**quare **D**ifference/RMSD) belirli bir X puanıyla ilişkilidir. Standartlaştırılmış Kök Beklenen Ortalama Kare Farkı (Standardized Root Expected Mean Square Difference/REMSD) tüm popülasyon için genel farkları özetlemektedir.

Eşitleme koşulları eşitleme işlemi gerçekleştirildikten sonra kontrol edilebilmektedir. Bu nedenle koşulların sağlanması için eşitleme işlemleri doğru bir şekilde yapılmalıdır. Eşitleme süreci aşağıdaki adımlardan oluşmaktadır:

1. Eşitleme deseninin seçilmesi

2 .Eşitleme desenine göre test formlarının uygulanması

3. Eşitleme yönteminin seçilmesi

4. Eşitleme işlemlerinin gerçekleştirilmesi

5. Eşitleme sonuçlarının değerlendirilmesi

Eşitleme Desenleri

Eşitleme işleminin yapılabilmesi için ilk olarak eşitleme deseni belirlenmelidir. Eşitlemede veri toplama yani testlerin gruplara uygulanması "eşitleme deseni" olarak ifade edilir. (Kolen & Brennen, 2004). Desenler grup sayısı, testlerin uygulanma sayısı ve sırasına göre farklılaşmaktadır. Her desenin avantaj ve dezavantajları bulunmaktadır. Test eşitleme süreçleri desenlere göre farklılaşmaktadır.

a) Tek Grup Deseni

Tek grup deseninde, evrenden seçkisiz olarak tek bir örneklem alınır. Bireyler iki testide alır. Bireylere iki testte uygulandığı için formlar arasındaki güçlük farklılığının etkisi ortadan kalkmaktadır. Çoğu durumunda, bireylere birden fazla formun uygulanması uzun zaman almaktadır. Gruba önce X formu sonra Y formu uygulandığında, Y formunda yorgunluk etkisi görülebilir. Bireylerin yetenek düzeylerindeki değişim gözlenebilmektedir. X formunda yer alan maddeler bireylerde öğrenmeye neden olabilir. Bu nedenle Y formunda öğrenme etkisi gözlenebilir. İlk test formu yorgunluk ve gibi nedenler ile ikinci test formundan elde edilen sonuçları etkileyebilmektedir. Bu etkiye sıra etkisi denir. Tek grup deseni, küçük örneklemlerde daha doğru eşitleme sonuçları sağlayabilmektedir. İki test formunu da alan grup aynı olduğu için, puanlara örneklemeden karışan hata miktarı azalır.

b) Random Grup Deseni

Evrenden gelen bireyler seçkisiz olarak iki gruba ayrılır. Grupların yetenek dağılımları farklılaşabilir. Bu durum eşitlemeye yanlılık karışmasına neden olabilir (Hambleton & Swaminathan, 1985). Gruplardan birine X formu diğerine Y formu uygulanır. Her bireye tek bir form uygulanmaktadır. Bu nedenle uygulama süresi bireylerin birden fazla form aldığı bir tasarıma göre kısalır. Grupların performanslarındaki farklar, formların zorluk farkını gösterir (Kolen & Brennan, 2014). Tek grup tasarımında ortaya çıkabilecek olası sıra etkileri sorununu ortadan kaldırmaktadır. Sıra etkisi gözlenmediği için örneklem büyük olduğunda random grup deseni genellikle iyi bir seçim olarak kabul edilmektedir (von Davier, 2011). Büyük ve heterojen gruplar seçilmelidir. Random grup deseni simetri özelliğini sağlamaktadır.

c) Denelenmiş Tek Grup Deseni

Tek grup desenin sınırlılığı olan sıra etkisini engellemek için geliştirilmiştir. Evrenden seçkisiz olarak tek bir örneklem alı-nır. Alınan örneklem seçkisiz olarak iki gruba ayrılır. Grupla-rın, yetenek dağılımlarının benzer olması gerekir. Bir gruba önce X sonra Y formu uygulanır. Diğer gruba ise önce Y sonra X formu uygulanır. İlk X formunun alınmasının Y formuna etkisi, ilk Y formunun alınmasının X formuna etkisine eşit ka-bul edilir (Kolen & Brennan, 2014). Küçük örneklemlerde kul-lanışlıdır fakat bireyler iki formu da aldığı için uygulama süre-si uzun olabilir (González & Wiberg, 2017).

d) Denk Olmayan Gruplarda Ortak Madde Deseni

Evrenden gelen bireyleri seçkisiz olarak iki gruba ayırmak her zaman mümkün olmayabilir. Farklı evrenlerden gelen iki gruba

formlar uygulanır. Grupların yetenek düzeyleri farklı olduğu için iki formdan alınan puanlardaki farklılığın formlardan mı bireylerden mi kaynaklandığı anlaşılamaz. Bu nedenle gruplara ortak maddeler uygulanır. Ortak maddeler iki grubun yetenek düzeyleri arasındaki farkı düzenler. Gruplar test formlarından birini ve ortak maddeleri alır. Ortak maddeler gruplara uygulanan formların içine eklemişse ve ortak maddelerden alınan puanlar bireylerin puanına ekleniyorsa iç ortak test, formlardan ayrı bir form olarak uygulanıyorsa dış ortak test olarak adlandırılır (Crocker & Algina, 2006). Eşitleme fonksiyonu ortak maddeler üzerinden elde edilir. Ortak maddeler ve testlerin ölçecekleri içerik aynı olmalıdır (Cook & Petersen, 1987). Ortak maddelerin ve testlerin ortalama güçlükleri benzer olmalıdır (Petersen vd., 1982). Angoff 'a (1971) göre testlerde yer alan madde sayısı 40 ve altında olduğunda ortak madde oranı %20; 40'ın üzerinde olduğunda ise %30 olmalıdır. Spence (1996) ortak madde sayısının testteki tüm madde sayısının %20 ile %30'u arasında değişmesi gerektiğini belirtir. Çok boyutlu testlerde bu oran %50'dir (Peterson, 2014).

Test Eşitleme

Test eşitleme deseni seçildikten sonra formlar gruplara uygulanır. Formlar uygulandıktan elde edilen puanlar ile eşitleme işlemi gerçekleştirilir. Eşitleme işleminin gerçekleştirilmesinde ilk olarak hangi eşitleme yönteminin kullanılacağı belirlenmelidir. Eşitleme yönteminin seçilebilmesi için verilerin boyutluluğu incelenmelidir.

Boyutluluk

Test maddeleri tarafından ölçülen minimum gizil özellik sayısı boyut olarak tanımlanmaktadır (Tindal & Haladyna, 2002). Kline'e

(2005) göre ise soyut, gözlenemeyen, dolaylı yollarla ölçülebilen; gizil nitelik taşıyan özelliklere "boyut" adı verilir. Bir n test veya n ikili madde kümesi, bir gizli özellik içeren ortak bir boyut modeline uyuyorsa tek boyutludur (McDonald, 1981). Goldstein ve Wood (1989), her iki formda madde yanıtlarının altında yatan gizil özelliklerin aynı lineer bileşimi gözleniyorsa boyutluluğun ihmal edilebilir olduğunu belirtmiştir. Fakat gizil özelliklerin aynı lineer bileşimi göstermesi her zaman mümkün olmayabilir.

Çoğu istatistiksel yöntem tek boyutluluk varsayımına dayanır. Tek boyutluluk varsayımı ihlal edildiğinde güvenirliğin gerçek güvenirliği doğru bir şekilde temsil edip etmediği şüphelidir. Çok boyutlu testlerde tek boyutlu eşitleme yöntemlerinin kullanılması eşitleme hatasının miktarında artışa neden olabilir. Livingston vd. (1990) çok boyutluluğun derecesi yüksek olduğunda eşitlenmiş puanların çok büyük ya da çok küçük kestirildiğini göstermiştir.

Standartlar'a (1999) göre boyutluluğun incelenmesi yapı geçerliği kanıtlarının toplanmasıdır. Geçerliği etkileyecek testin ölçmek istediği yapı ile ilgisi olmayan gizil değişkenler belirlenmelidir. Testin boyutluluğunu incelemede faktör analizi, paralel analiz, yamaç birikinti grafiği, Stout DIMTEST ve POLYTEST vb. yöntemler kullanılmaktadır. İncelemeler gerçekleştirilirken birden fazla yöntem ile yapıya dair kanıtlar desteklenmelidir. Boyutluluk incelemeleri gerçekleştirildikten sonra kullanılacak eşitleme yöntemi belirlenir.

Eşitleme Yöntemleri

a) Tek Boyutlu Eşitleme (TBE)

Bir gizil değişkeni, yeteneği veya özelliği ölçen testler kullanıldığı zaman tek boyutlu test eşitleme yöntemleri tercih edil-

melidir. Tek boyutlu testlerde klasik test kuramına (KTK) ve madde tepki kuramına (MTK) dayalı eşitleme yöntemleri kullanılmaktadır.

a.1. Klasik test kuramına dayalı test eşitleme yöntemleri

Klasik test kuramına dayalı eşitleme yöntemleri; ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme olarak sınıflandırılmaktadır. Her yöntemin kendine özgü varsayımları bulunur ve kullanılan eşitleme desenine bağlı olarak eşitleme fonksiyonları farklılık gösterir.

Ortalama eşitleme yöntemi X ile Y formundan elde edilen puanlar arasındaki farkların testi alan her bireyde aynı miktarda olduğu varsayımı üzerine kurulmuştur (Kolen & Brennan, 2004). Yani, iki formun madde güçlükleri arasındaki farkın tüm yetenek ölçeği boyunca eşit olduğu varsayılır. Ortalama eşitleme yönteminde, puanların ortalamaları arasındaki fark kullanılarak eşitleme yapılır (Livingston, 2004). X formu Y formuna eşitlenirken, X formundan elde edilen puandan X formunun ortalaması çıkarılır. Bulunan sonuca Y formunun ortalaması eklenir. X formundan elde edilen puan x_i, X formundan elde edilen puanları ortalaması η_X, Y formundan elde edilen puanları ortalaması η_Y ve x_i puanının Y formuna eşitlenmesi ile elde edilen eşitlenmiş puan $Y(x_i)$ olmak üzere eşitleme fonksiyonu aşağıdaki gibidir (Kolen & Brennan, 2004);

$$Y(x_i) = x_i - \eta_X + \eta_Y$$

Elde edilen eşitlenmiş puanların aritmetik ortalaması Y formunun ortalamasına eşittir (Livingston, 2014). Formlardan alınan puanların varyanslarının eşit olduğu varsayıldığında ortalama eşitleme yöntemi kullanılabilir (González & Wiberg, 2017).

Doğrusal eşitleme yöntemi X ile Y formundan elde edilen puanlara ait dağılımlarının aynı olduğunu varsayar. Puanların aritmetik ortalamaları ve standart sapmalarının eşit olduğu varsayılır. x_i puanının y_i puanına eşit olabilmesi için hem $Y(x_i)$'nin hem de y_i'nin ortalamalarına olan uzaklıklarının z puanlarının aynı olması gerekmektedir (Livingston, 2014). X formuna ait puanların standart sapması σ_X ve Y formuna ait puanların standart sapması σ_Y olmak üzere eşitleme fonksiyonu aşağıdaki gibidir (Angoff, 1982);

$$Y(x_i) = \eta_Y + \frac{\sigma_Y}{\sigma_X}(x_i - \eta_X)$$

Eşitleme işleminden sonra, eşitlenmiş puanlar Y formundan alınacak maksimum ya da minimum puanı geçebilmektedir. Eşitlenmiş puan, maksimum puandan yüksek ise maksimum değere sabitlenir. Minimum değerden küçük ise, minimum değere sabitlenir (Kolen & Brennan, 1995). Bu işleme *budama (truncation)* denir. Yöntem puanların dağılımını temel aldığı için eşitlenmiş puanlar formların uygulandığı gruplardan etkilenmektedir. Eşitlenmiş puanların gruptan etkilendiği durumlarda farklı güçlük düzeylerine sahip iki formdan alınan puanlar eşitlendiğinde, yetenek düzeyi yüksek bir grup ile düşük bir grupta eşitleme farklılık gösterebilir. Yetenek düzeyi yüksek gruba ait eşitleme fonksiyonun eğimi ($\frac{\sigma_Y}{\sigma_X}$) az iken düşük gruba ait eşitleme fonksiyonun eğimi daha yüksek olacaktır. Bu nedenle büyük ve heterojen örneklemlerde doğrusal eşitleme yöntemi kullanılmalıdır.

Gruplar, aynı yetenek düzeyinde değilse yani denk olmayan gruplarda ortak madde deseni kullanılarak veriler toplandıysa farklı doğrusal eşitleme yöntemlerinin kullanılması önerilir. Bu yöntemlerde farklı evrenlerden gelen gruplar evrenlerin ağırlıkları göz önünde bulundurularak birleştirilir. Oluşan yeni evrene *sentetik evren* denir. X formunun uygulandığı gruba ait evrenin ağırlığı w_X,

Y formunun uygulandığı gruba ait evrenin ağırlığı w_Y olmak üzere ağırlıklar aşağıdaki koşulları sağlar;

$$w_X + w_Y = 1$$

$$w_X, w_Y > 0$$

İlk grup (evren 1) X formunu aldığında elde edilen puanlar X_1, ikinci grup (evren 2) Y formunu aldığında elde edilen puanlar Y_2, ortak maddelerden alınan puanlar V olarak adlandırılır. X_1 puanları Y_2 puanlarına eşitlendiğinde X_2 puanları; Y_2 puanları X_1 puanlarına eşitlendiğinde Y_1 puanları elde edilecektir. Tucker ve Levine eşitleme yöntemleri çeşitli varsayımlar yardımı ile bu değerleri kestirir. Tucker yönteminde, toplam puanların ortak maddelerden alınan puanlar ile regresyonu iki evrende eşittir. Toplam puanların ortak maddelerden alınan puanlar ile koşullu varyansı iki evrende eşittir. Tucker yöntemi eşitleme yapılırken gözlenen puanları kullanır. Levine yöntemi ise X ve Y formlarının güvenirlikleri eşit olduğunda gözlenen; güvenirlikler eşit olmadığında ise gerçek puanları kullanır (Levine, 1955). Varsayımları iki puan türü içinde aynıdır. Gerçek puanlar (T_X ve T_Y) ile ortak maddelerden ait gerçek puanlar (T_V) arasında iki evrede de mükemmel ilişki vardır (r = 1). Gerçek puanlar (T_X ve T_Y) ile ortak maddelerden ait gerçek puanlar (T_V) ile regresyonu iki evrende eşittir. X, Y ve V'nin hata varyansları iki evrende eşittir. Hata varyansları gözlenen ve gerçek puan varyanslarının farkına eşittir.

Eşit yüzdelikli eşitleme formlar aynı puan dağılımına sahip olmadığında önerilir (Zhu, 1998). Aynı yüzdelik sıraya denk gelen puanlar eşit kabul edilir. Bunu belirlemek için iki formun yığılmalı frekans dağılımları oluşturulur. Yığılmalı frekanslarda aynı yüzdelik sıraya denk gelen puanlar eşitlenir. X formuna ait kümülatif dağılım

fonksiyonu F, Y formuna ait kümülatif dağılım fonksiyonu G, X formundan elden edilen puanları Y formuna eşitlemek için kullanılan simetrik fonksiyon e_Y olmak üzere eşitleme fonksiyonu aşağıdaki gibidir;

$$e_Y = G^{-1}[F(x_i)]$$

Örneğin X formunda %50'de yer alan puan 45, Y formunda %50'de yer alan puan 52 ise x_i = 47 için eşitlenmiş puan 52'dir. Fakat her zaman aynı yüzdelikte yer alan puanları bulmak kolay değildir. Küçük gruplarla yapılan eşitlemelerde, örneklemeden kaynaklanan hata düzensizlik problemine neden olmaktadır (Kolen, 1988). Bunun nedeni evrenden seçilen gruplarda testten alınabilecek her puanın gözlenememesidir. Yığılmalı frekans grafiğinde süreklilik gözlenmez. Grupların evrenleri daha iyi temsil etmesi için büyük örneklemlerde eşitleme gerçekleştirilmeli ve ya grafiklerin analitik yöntemler kullanılarak sürekli hale getirilmesi gerekmektedir. Dağılımı sürekli hale getirmek için düzgünleştirme yöntemleri kullanılır. Düzgünleştirme yöntemlerinin kullanılması eşitlenmenin güvenirliği arttırır. Kolen (2007), ön düzgünleştirme (presmoothing), son düzgünleştirme (post-smoothing), log doğrusal ön düzgünleştirme (log-linear pre-smoothing) ve kübik spline son düzgünleştirme (cubic spline post-smoothing) yöntemlerini önermektedir. Ön düzgünleştirme yöntemi, eşitleme işlemi gerçekleştirilmeden önce uygulanır. Son düzgünleştirme yöntemi, eşitlenmiş puanlara uygulanır. Log doğrusal ön düzgünleştirme, puanlarının polinom fonksiyonu cinsinden beklenen oranlarını ifade eder ve eşitleme işlemi gerçekleştirilmeden önce uygulanır. Kübik spline son düzgünleştirme, sürekli bir kübik spline eğrisi kullanarak veri noktalarını ile birleştir. Dağılımdaki düzensizliklerin yanı sıra eşit yüzdelikli eşitleme yapabilmek için eşitleme fonksiyonun

maksimum ve minimum puanları geçmemesi gerekir. Eşit yüzdelikli eşitlemede bir diğer önemli nokta ise X formundan elde edilen puanların yüzdelik sırasının Y formundaki yüzdelik sıralarla eşleşmesidir. Sırlarda eşleşme gözlenmez ise öteleme gerçekleştirilir. Eşitlemede puanların bir miktar arttırılması veya azaltılmasına öteleme denir.

a.2. Madde tepki kuramına dayalı test eşitleme yöntemleri

Madde tepki kuramı (MTK) bireylerin yetenek düzeyleri ile doğru cevap olasılıkları arasında ilişki olduğu varsayar. Bireylerin gözlenen puanları (doğru cevap olasılıklarının toplamı) madde ve yetenek parametrelerinin bir fonksiyonudur. Kolen ve Brennan (2014) kestirilen madde ve yetenek parametrelerinin karşılaştırılabilir olması ve gözlenen puanların eşitlenebilmesi için ortak bir ölçekte yer alması gerektiğini belirtmiştir. Tek grup, random grup ve dengelenmiş grup deseni kullanılarak veri toplandığında gruplar seçkisiz olarak oluşturulmaktadır. Bu nedenle bu eşitleme desenleri kullanılarak veri toplandığında parametrelerin aynı ölçekte olduğu kabul edilir. Denk olmayan gruplarda ortak madde deseni kullanıldığında parametreler örneklemler denk olmadığı için farklı ölçekler üzerindedir. Parametrelerin aynı ölçekte yer alması için kalibrasyon gerçekleştirilmelidir (Han, 2008). Parametrelerin doğrusal bir dönüşümle aynı ölçeğe getirilmesine kalibrasyon denir.

Ayrı kalibrasyon ve eş zamanlı kalibrasyon olmak üzere iki kalibrasyon yöntemi bulunmaktadır. Ayrı kalibrasyon yönteminde ölçek kalibrasyonunda X (yeni) formundan kestirilen madde ve yetenek parametreleri, ortak maddelerin parametreleri aracılığıyla, Y (eski) formun ölçeğine getirilir. Ayrı kalibrasyon yöntemleri

kestirilen madde ve yetenek parametrelerini aynı ölçeğe dönüştürmek orijin (translation indeterminacy) ve ölçü birimi belirsizliğini (dilation/axis indeterminacy) giderir. Yetenek parametrelerinin konumları orijin belirsizliğine neden olur. Orijin belirsizliği, formda yetenek parametresinin sıfır değerinin nerede olduğuyla ilgilidir. Ölçü birimi belirsizliği ise farklı formlarda kullanılan ölçü birimlerinin farklı olmasından kaynaklanır. Bir formdan elde edilen puanlardaki bir birimlik artışın yetenek düzeyinde kaç birimlik artışa neden olduğu anlamına gelir. Farklı formlarda farklı artışlar gözlenebilir. Bu durumda, bir formdan alınan puanın başka bir forma göre daha büyük ya da daha küçük bir yetenek değişimini temsil etmesi olasıdır (Kolen & Brennan, 2014). Belirsizlikleri gidermek için doğrusal (lineer) bir dönüşüm fonksiyonu kullanılır. Doğrusal dönüşüm fonksiyonu elde etmek için eğim (A) ve kesişim (B) katsayıları kullanılır. i maddesinin Y (eski) formundan kestirilen ayırt edicilik parametresi a_{Yi}, i maddesinin Y formundan kestirilen güçlük parametresi b_{Yi}, i maddesinin X (yeni) formundan kestirilen ayırt edicilik parametresi a_{Xi} ve i maddesinin X formundan kestirilen güçlük parametresi b_{Xi} olmak üzere madde parametreleri için doğrusal dönüşüm fonksiyonları aşağıdaki gibidir;

$$a_{Yi} = \frac{a_{Xi}}{A}$$

$$b_{Yi} = A\, b_{Xi} + B$$

$$c_{Yi} = c_{Xi}$$

X formunu alan j kişisinin yetenek düzeyi θ_{Xj} olmak üzere madde parametreleri için doğrusal dönüşüm fonksiyonları aşağıdaki gibidir (Ukşul, 2024);

$$\theta_{Yj} = A\,\theta_{Xj} + B$$

Ayrı kalibrasyon yöntemleri A ve B katsayılarını elde etme yöntemlerine göre farklılaşmaktadır (Embretson & Reise, 2000). Ayrı kalibrasyon yöntemleri momente ve karakteristik yöntemleri olmak üzere ikiye ayrılır. Momente yöntemleri ortalama-ortalama ve ortalama-standart sapma; karakteristik eğriye yöntemleri Haebara ile Stocking ve Lord yöntemleridir.

Ortalama-ortalama yönteminde madde ayırt edicilik ve madde güçlük parametrelerinin ortalaması kullanılır. Y formuna ait madde ayırt ediciliklerinin ortalaması $\eta_Y(a_{Yi})$, X formuna ait madde ayırt ediciliklerinin ortalaması $\eta_X(a_{Xi})$, Y formuna ait madde ayırt güçlüklerinin ortalaması $\eta_Y(b_{Yi})$ ve X formuna ait madde ayırt güçlüklerinin ortalaması $\eta_X(b_{Xi})$ olmak üzere A ve B katsayıları aşağıdaki gibidir;

$$A = \frac{\eta_Y(a_{Yi})}{\eta_X(a_{Xi})}$$

$$B = \eta_X(b_{Xi}) - A\,\eta_Y(b_{Yi})$$

Ortalama-sigma yönteminde madde güçlük parametrelerinin standart sapması kullanılmaktadır. Y formuna ait madde güçlüklerinin standart sapması $\sigma_Y(b_{Yi})$ ve X formuna ait madde güçlüklerinin standart sapması $\sigma_X(b_{Xi})$ olmak üzere A ve B katsayıları aşağıdaki gibidir;

$$A = \frac{\sigma_X(b_{Xi})}{\sigma_Y(b_{Yi})}$$

$$B = \sigma_X(b_{Xi}) - A\,\sigma_Y(b_{Yi})$$

Haebara yönteminde belirli bir yetenek düzeyinde her bir madde için madde karakteristik eğrileri arasındaki farkların karelerinin toplamını alarak karakteristik eğriler arasındaki farkı hesaplar (Hdiff(θ_i)). Stocking ve Lord (1983) yönteminde ise belirli bir yetenek düzeyinde her bir madde için madde karakteristik eğrileri arasındaki farkların toplamının karesini alarak karakteristik eğriler arasındaki farkı hesaplar (SLdiff(θ_i)).

Eş zamanlı kalibrasyon yönteminde iki formdaki ortak maddelerin aynı madde parametrelerine sahip olduğu varsayılarak madde ve yetenek parametrelerini aynı anda ve ortak bir veri seti kullanarak kestirir. Kestirim gerçekleştirilirken grupların yetenek dağılımları arasındaki fark göz önünde bulundurularak madde parametreleri kestirildiği için, kestirilen parametreler aynı ölçek üzerinde yer almaktadır. Eş zamanlı kalibrasyonda ortak maddelerin yanı sıra ortak bireylerde kullanılabilmektedir. Eş zamanlı kalibrasyonda tüm maddeler ve bireyler için parametre kestirimi tek bir analizle yapılır. Bu da veri tutarlılığını artırır.

Formlardan kestirilen madde ve yetenek parametreleri aynı ölçeğe dönüştürüldükten sonra eşitleme gerçekleştirilebilir. Madde tepki kuramına dayalı yöntemler gerçek ve gözlenen puan eşitleme olmak üzere ikiye ayrılır. Madde tepki kuramında bireylerin gerçek puanları, her maddenin doğru yanıtlanma olasılıklarının toplanması ile bulunur. Madde tepki kuramı gerçek puan eşitleme işleminde, belirli bir θ_i için gerçek puanlar $\tau_X(\theta_i)$ ve $\tau_Y(\theta_i)$ eşdeğer kabul edilir. Eşdeğer gerçek puanı bulmak için X formundaki gerçek puan $\tau_X(\theta_i)$ belirlenir. Daha sonra bu gerçek puana karşılık gelen yetenek düzeyi Newton-Raphson metodu ile bulunur. Y formunda bu

yetenek düzeyine karşılık gelen gerçek puan bulunur. Bu işlem, X formundan elde edilen tüm gerçek puan değerleri için tekrarlanır.

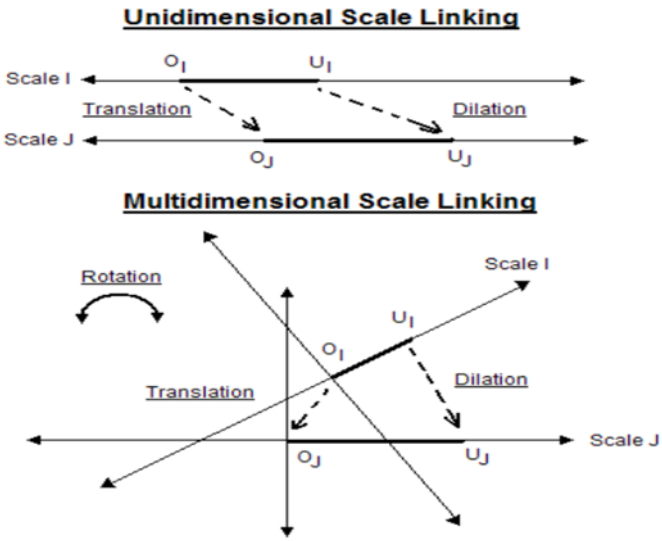
Gözlenen puan eşitleme yönteminde her bir formun koşullu gözlenen puan dağılımı özyinelemeli Lord-Wingersky algoritması ile kestirilir. Kestirilen koşullu gözlenen puan dağılımları yetenek yoğunluğu ($\psi(\theta)$) ile çarpılır. Çarpımlar marjinal gözlenen puan dağılımını belirlemek amacıyla tüm yetenek düzeyleri boyunca toplanır. Böylece marjinal gözlenen puan dağılımları kestirilir. Marjinal toplam gözlenen puan dağılımlarını eşitlemek için eşit yüzdelikli eşitleme yapılır.

b) Çok Boyutlu Eşitleme (ÇBE)

Birden fazla gizil değişkeni, yeteneği veya özelliği ölçen testler kullanıldığı zaman çok boyutlu test eşitleme yöntemleri tercih edilmelidir. Çok boyutlu eşitlemenin yapılabilmesi için madde ve yetenek parametrelerinin aynı ölçekte olması gereklidir. Kullanılan eşitleme desenine göre çeşitli kalibrasyon yöntemleri kullanılmaktadır. Dengelenmiş tek grup ve random grup deseni kullanılarak veri toplandığında tek boyutlu eşitleme ile benzer olarak orijin ve ölçü birimlerinin aynı olduğu varsayılarak orijin ve ölçü birimi belirsizliğini giderilmiş olur. Fakat rotasyonel/dönme belirsizlik (rotation indeterminacy), korelasyon belirsizliğinin (correlation indeterminacy) giderilmesi gerekmektedir. Rotasyonel belirsizlik madde ayırt edicilik ve yetenek parametrelerinin dönme belirsizliğidir. Rotasyonel belirsizliğinde, doğru cevaplama olasılığı değişmeden sonsuz sayıda madde ayırt edicilik ve yetenek parametresi elde edilebilir (Hirsch, 1989). Korelasyon belirsizliği boyutlar arasındaki korelasyonun, formlardaki benzersiz maddelerden dolayı farklı kestirilebilmesidir. Boyutlar arasındaki

korelasyonların farklı kestirilmesi her form için kestirilen parametrelerin, vektör bileşimlerinin farklı olduğu anlamını taşır. Korelasyon belirsizliği sorunu genellikle, gizil yeteneklerin çok değişkenli bir normal dağılım izlediğini ve yetenek uzayında eksenlerin karşılıklı olarak birbirine dik olduğu varsayılarak giderilir. Her iki form farklı koordinat sistemlerinde yer alsa da marjinal gözlenen puan dağılımları, çok boyutlu madde tepki kuramı modellerinin değişmezlik özelliğinden dolayı koordinat sistemi seçiminden etkilenmez. Bu rotasyonel belirsizliğini gidermiş olur. Brossman ve Lee'de (2013), random grup deseni altında ölçek kalibrasyonunun gerekli olmadığını belirtmiştir.

Denk olmayan gruplarda ortak madde deseni kullanılarak veri toplandığında ayrı ve eş zamanlı kalibrasyon yöntemleri uygulanabilir. Formlardan madde ve yetenek parametreleri kestirildikten sonra ayrı kalibrasyon yöntemleri, rotasyonel/dönme belirsizlik (rotation indeterminacy), korelasyon belirsizliği (correlation indeterminacy), orijin (translation indeterminacy) ve ölçü birimi belirsizliği (dilation/axis indeterminacy) gidererek parametre kestirimlerini aynı ölçeğe getirir. Rotasyonel belirsizliği giderilirse, orijin ve ölçü birimi belirsizlikleri tek boyutlu eşitleme kalibrasyon yöntemleri kullanılarak çözülebilir (Li & Lissitz ,1998). Rotasyonel belirsizliği giderilmez ise ölçek (orijin ve ölçü birimi belirsizliği) belirsizliğinin giderilmesi için, her boyut için yetenek parametrelerinin kestirimleri, ortalaması sıfır ve varyansı bir olacak şekilde standartlaştırılabilir (Li & Lissitz, 2000).



Şekil 1: Çok boyutlu ölçek kalibrasyonu

Kaynak: Brossman, 2010

Denk olmayan gruplarda ortak madde deseninde ayrı kalibrasyon yöntemleri rotasyonel, korelasyon, orijin ve ölçü birimi belirsizliklerini gidermek için skalerler ve matrislerden faydalanır. Kullanılan kalibrasyon yöntemine göre skaler ve matrislerin kestirim formülleri değişmektedir. m testin boyut sayısı, A rotasyon ve ölçü birimi belirsizliğini hesaba katan ve boyutlar (θ yetenek düzeyleri) için varyans ve kovaryansları düzenleyen m x m boyutunda matris (döndürme matrisi/rotation matrix), β orijin belirsizliğini hesaba katan mx1 boyutunda vektör (öteleme vektörü/translation vector), X hedef ölçeği ve Y referans ölçek olmak üzere genel kalibrasyon formülleri aşağıdaki gibidir (Davey vd., 2000);

$$\theta_{Yi} = (A\theta_{Xi}) + \beta$$

$$\alpha_{Yi}^{T} = (A^{-1})^{T}\, a_{Xi}$$

$$d_{Yi} = d_{Xi} - \alpha_{Xi}^{T} A^{-1}\beta$$

$$c_{Yi} = c_{Xi}$$

Davey vd. (2000) tek boyutlu eşitleme için kullanılan kalibrasyon yöntemlerini çok boyutlu eşitlemeye genişleterek dört yöntem önermişlerdir. Bu yöntemlerden ilki formlar arasında ortak madde parametre kestirimleri arasındaki kare farkını en aza indiren Divgi (1985) yönteminin uzantısı olan doğru yöntemidir (The Direct Method). "Eşitlenmiş Fonksiyon yöntemi" olarak yöntemi ortalama-ortalama yönteminin bir uzantısıdır (The Equated/Matching Scale Function Method).Madde karakteristik Fonsiyonu yöntemş Haebara madde karakteristik eğri yönteminin uzantısıdır (The Item Characteristic Function Method). Çok boyutlu madde karakteristik yüzeylerinin farkını minimum yapan dönüşüm matrisleri bulunur. Son yöntem olan Eşleşen Test Yanıt Fonksiyonları veya Yüzeyleri yöntemi Stocking ve Lord'un (1983) test karakteristik eğri yönteminin bir uzantısıdır (Matching Test Responce Function/Surface). Çok boyutlu test karakteristik yüzeylerinin farkını minimum yapan dönüşüm matrisleri bulunur. Boyutlar arasındaki korelasyonlar iki formda benzer ise ortogonal rotasyon matrisi; farklı ise eğik rotasyon matrisi kullanılır (Simon, 2008).

Li ve Lissitz (2000), önerilen dört yöntemde kullanılan rotasyon matrisinin ortogonal olmadığı için birden fazla formu olabileceğini öne sürmektedir. Bu nedenle ortogonal Procrustes dönüşümünü, çeviri vektörünü ve tek bir genişleme parametresini içeren yöntemler önermişlerdir. k genişletme parametresi (dilation parameter), T ortogonal matris (döndürme matrisi/rotation matrix),

β vektörü (öteleme vektörü/translation vector) olmak üzere kalibrasyon formülleri aşağıdaki gibidir (Li & Lissitz,2000);

$$a_{jY} = k\, a_{jX}^{T}\, T$$

$$\theta_{jY} = \frac{1}{k}\,(T^{-1}\theta_{jX} - \beta)$$

$$d_{jY} = d_{jX} + \left(a_{jX}^{T}\, T\right)\beta$$

T ortogonal matrisi A matrisi ile benzerdir. β vektörleri yöntemler için aynıdır. A matrisi, yetenek ve madde ayırt edicilik parametrelerini döndürür ve genişletirken T ortogonal matrisi sadece döndürme yapar. Genişletme için k parametresi eklenmiştir. T ortogonal matrisi ve ölçekleme katsayıları (k genişletme parametresi ve β vektörü) ayrı kestirilir. T ortogonal matrisi formlardan kestirilen madde ayırt edicilik parametrelerinin karelerinin farklarının toplam değerini minimize edecek şekilde ortogonal Procrustes rotasyonu (Schönemann, 1966) kullanılarak kestirilir.

Ölçekleme katsayılarının üç farklı yöntem ile kestirilebilir. Ölçekleme katsayılarının eşzamanlı kestirimi için test karakteristik yüzeyleri eşleştirme yöntemi (Matching the Test Response Surface) Stocking Lord prosedürünün uzantısıdır. β vektörü için k parametresini kestirmek için özdeğerlerin oranı ile birleştirilmiş en küçük kareler tahmini ve β vektörü için k parametresini kestirmek için izin oranı ile birleştirilmiş en küçük kareler tahmini yöntemleri de önerilmektedir.

Denk olmayan gruplarda ortak madde deseninde eş zamanlı kalibrasyonda, X ve Y formları tek bir testmiş gibi ele alınır. Tüm madde ve yetenek parametreleri aynı anda kestirir. Eşzamanlı

kalibrasyonda, üç alt testten bulunur: 1) Y (referans/base grup) formunda yer alan benzersiz maddeler, 2) X (hedef/odak/target grup) formunda yer alan benzersiz maddeler ve 3) iki formda yer alan ortak maddeler. Her formunun benzersiz maddeleri diğer formu için kayıp veri olarak tanımlanır. Formlardan elde edilen yanıt matrisleri birleştirilerek tek bir matris oluşturulur. Kayıp veriler ile kestirimler yapılır. X formunu alan hedef grubun yetenek parametreleri, Y formunu alan referans gruba göre kestirilir.

Formlardan kestirilen madde ve yetenek parametreleri aynı ölçeğe dönüştürüldükten sonra eşitleme gerçekleştirilebilir. Basit yapılı çok boyutlu madde tepki kuramı (ÇB-MTK) gözlenen puan eşitleme (SMO) (Lee & Brossman, 2012), basit yapılı ÇB-MTK gerçek puan eşitleme (SMT) (Kim vd., 2019), tam ÇB-MTK gözlenen puan eşitleme (MOSE), ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE), ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşım (ATSE) (Brossman, 2010), bi-faktör ÇB-MTK gözlenen puan (Lee & Lee, 2016), bi-faktör ÇB-MTK gerçek puan (Lee vd., 2015), madde takımı yanıt modeli ÇB-MTK gözlenen puan ve madde takımı yanıt modeli ÇB-MTK gerçek puan (Tao & Cao, 2016) yöntemlerinin çok boyutlu test eşitleme çalışmalarında kullanıldığı belirlenmiştir.

b.1. Basit yapılı ÇB-MTK gözlenen puan yöntemi

Basit yapılı ÇB-MTK gözlenen puan yönteminde temel varsayımlar bulunmaktadır: (a) Karma testlerdeki her madde, belirli bir boyutu ölçer ve boyutlar arasında korelasyon olmasına izin verilir; (b) aynı boyutla ilişkili olan maddeler, Tek Boyutlu Madde Tepki Kuramı modeli ile modellenebilir; (c) eski ve yeni formu alan gruplarındaki bireyler, boyutların temsil ettiği yetenek düzeyleri

açısından seçkisiz olarak denk kabul edilir. Basit yapılı ÇB-MTK gözlenen puan yönteminde her boyut için koşullu gözlenen puan dağılımı hesaplanır. Koşullu toplam gözlenen puan dağılımı, boyutların bir fonksiyonudur. Her boyuttaki yetenek düzeylerinin kombinasyonuna bağlı olarak toplam puanlar üretir. Boyutlar için hesaplanan gözlenen puan dağılımları hesaplandıktan sonra her form için m tane boyut üzerinden koşullu toplam gözlenen puan dağılımını hesaplanır. Birinci boyut için koşullu gözlenen puan dağılımı $f_1(x_1|\theta_1)$, ikinci boyut için koşullu gözlenen puan dağılımı $f_2(x_2|\theta_2)$, m'inci boyut için koşullu gözlenen puan dağılımı $f_m(x_m|\theta_m)$, birinci boyutun ağırlığı w_1 , ikinci boyutun ağırlığı w_2 ve m'inci boyutun ağırlığı w_m olmak üzere koşullu yerel madde bağımsızlığı varsayımı altında koşullu toplam gözlenen puan dağılımı aşağıdaki gibidir;

$$f(x|\theta_1, \theta_2, \dots \theta_m) = \sum_{X=w_1x_1+w_2x_2+\cdots+w_mx_m} f_1(x_1|\theta_1)f_2(x_2|\theta_2) \; .. f_m(x_m|\theta_m)$$

Kestirilen koşullu gözlenen puan dağılımları çok değişkenli bir yetenek dağılımı $g(\theta_1, \theta_2, \dots \theta_m)$ ile çarpılır. Birleşik m boyutlu yeterlilik dağılımı, Mislevy'nin (1984) prosedürü kullanılarak tahmin edilebilir. Çarpımlar marjinal gözlenen puan dağılımını belirlemek amacıyla tüm yetenek düzeyleri boyunca toplanır. Marjinal toplam gözlenen puan dağılımı aşağıdaki gibidir;

$$f(x) = \sum_{\theta_1}\sum_{\theta_2} f(x|\theta_1, \theta_2, \dots \theta_m)g(\theta_1, \theta_2, \dots \theta_m)$$

Marjinal toplam gözlenen puan dağılımlarını eşitlemek için eşit yüzdelikli eşitleme yapılır. Basit yapılı ÇB-MTK gözlenen puan Tek Boyutlu Madde Tepki Kuramı gözlenen puan eşitleme yöntemine benzemektedir. Ancak yetenek yoğunluğu ($\psi(\theta)$) yerine

çok değişkenli bir yetenek dağılımı $g(\theta_1, \theta_2, \ldots \theta_m)$ ile çarpılır. Ayrıca bileşik toplam gözlenen puan dağılımı kullanılmaktadır.

b.2. Basit yapılı ÇB-MTK gerçek puan yöntemi

Basit yapılı ÇB-MTK gerçek puan eşitleme yönteminde her boyut için madde ve yetenek parametreleri gerçekleştirildikten sonra her boyut için Tek Boyutlu Madde Tepki Kuramı gerçek puan eşitleme ayrı ayrı gerçekleştirilir. Elde edilen eşitleme sonuçları toplam test puanı için kullanılır. Her boyut puanı kombinasyonu için eşitlenmiş bileşik puan boyut ağırlıkları ile çarpılarak bulunur. Boyut sayısı m, boyutun ağırlığı w_m olmak üzere bileşik aşağıdaki gibidir (Kim vd., 2019);

$$eq(x|x_1, x_2 \ldots, x_m) = w_1 eq(x_1) + w_2 eq(x_2) \ldots + w_m eq(x_m)$$

Bileşke puan elde edildikten sonra çok değişkenli gözlenen puan dağılımı $f(x_1, x_2 \ldots, x_m)$ belirlenir. Çok değişkenli gözlenen puan dağılımı, m tane boyutun gözlenen puan dağılımının çarpımı ile bulunur (Kim vd., 2019);

$$f(x_1, x_2, \ldots, x_m|\theta_1, \theta_2, \ldots, \theta_m) = f(x_1|\theta_1) f(x_2|\theta_2) \cdots f(x_m|\theta_m),$$

Boyutların gözlenen puan dağılımları Lord-Wingersky algoritması ile bulunabilir. Çok değişkenli gözlenen puan dağılımı çok değişkenli gizli yetenek dağılımları $g(\theta_1, \theta_2, \ldots \theta_m)$ ile çarpılarak çok değişkenli dağılım elde edilir. Çok değişkenli dağılım tanımlandıktan sonra, her bir bileşik puana yol açan bölüm puanları için frekans ağırlıkları aşağıdaki fonksiyon ile bulunur (Kim vd., 2019);

$$f(x_1, \ldots, x_m|x) = \frac{f(x_1, \ldots, x_m)}{\sum_{X=int(\sum_{i=1}^{m} w_i X_i)} f(x_1, \ldots, x_m)},$$

Eşitlenmiş bileşik puanların ağırlıklı toplamı ile nihai eşitlenmiş puan aşağıdaki fonksiyon ile kestirilir (Kim vd., 2019);

$$f(x_1, \ldots, x_m|x) = \frac{f(x_1, \ldots, x_m)}{\sum_{X=int(\sum_{i=1}^{m} w_i X_i)} f(x_1, \ldots, x_m)},$$

b.3. Tam ÇB-MTK gözlenen puan yöntemi

Tam ÇB-MTK gözlenen puan eşitleme yönteminde bireylerin farklı yetenek düzeylerini dikkate alarak, puan eşitleme işlemini gerçekleştirir. Yetenek düzeyleri vektörünü kullanan Lord-Wingersky algoritması ile yetenek düzeyleri kombinasyonları için koşullu gözlenen puan dağılımları kestirilir. Boyut sayısı m, boyut vektörü θ ve madde indeksi r olmak üzere koşullu gözlenen puan dağılımı, aşağıdaki gibidir (Peterson, 2014);

$$\begin{aligned} f_r(x|\boldsymbol{\theta}_j) &= f_{r-1}(x|\boldsymbol{\theta}_j)(1-P_r), && x=0 \\ &= f_{r-1}(x|\boldsymbol{\theta}_j)(1-P_r) + f_{r-1}(x-1|\boldsymbol{\theta}_j)P_r, && 0<x<r, \\ &= f_{r-1}(x-1|\boldsymbol{\theta}_j)P_r, && x=r. \end{aligned}$$

Koşullu gözlenen puan dağılımları çok değişkenli yetenek yoğunluğu ile çarpılır. Tüm yetenek düzeyleri boyunca toplanarak marjinal gözlenen puan dağılımı elde edilir. Çok değişkenli yetenek yoğunluğu $\psi(\theta_m)$ olmak üzere marjinal gözlenen puan dağılımı aşağıdaki gibidir (Peterson, 2014);

,

$$f(x) = \int_{\theta_1} \int_{\theta_2} \dots \int_{\theta_m} f(x|\boldsymbol{\theta})\psi(\boldsymbol{\theta})\, d(\boldsymbol{\theta}).$$

Marjinal toplam gözlenen puan dağılımlarını eşitlemek için eşit yüzdelikli eşitleme yapılır.

b.4. ÇB-MTK tek boyutlu yaklaşımlar eşitleme yöntemi

Zhang ve Stout'a (1999) çok boyutlu yapının çeşitli yöntemler ile tek boyutlu yapıya indirgenebileceğini ve tek boyutlu yöntemlerin kullanılarak eşitleme yapılabileceğini belirtilmiştir. ÇB-MTK tek boyutlu yaklaşımlar Zhang'in (1996) çalışmalarını temel alarak ortaya çıkmıştır. Yöntemde ilk olarak çok boyutlu yeteneklerin doğrusal bileşimi olan tek boyutlu bir bileşik yetenek parametresi kestirilir. Daha sonra tek boyutlu bileşik yetenek parametresine karşılık gelen tek boyutlu madde parametreleri elde edilir. Doğrusal bileşim elde edilirken ağırlıklandırma yapılır. Ağırlık vektörünü kestirmek için en çok bilgi veren yetenek düzeyleri belirlenir. Boyut indeksi k, i maddesi için k. boyuta ait madde ayırt ediciliği $\hat{a}_{ik}$, toplam boyut sayısı m, k. boyutun doğrusal bileşik katsayılar vektöründe standartlaştırılmış katsayı $\hat{w}_k$ ($\underline{w}$), testte yer alan toplam madde sayısı N (n) olmak üzere en çok bilgi veren test seviyesi yönüne karşılık gelen ağırlık vektörü her aşağıdaki gibidir (Lee, 2013);

$$\hat{w}_k = \frac{\sum_{i=1}^{N} \hat{a}_{ik}}{\sqrt{\sum_{k=1}^{m} (\sum_{i=1}^{n} \hat{a}_{ik})^2}}$$

Ağırlık vektörü her boyut için kestirilir. Tek boyutlu parametreler kestirilirken ağırlık vektöründen yararlanılır. Çok boyutlu yetenek kovaryans matrisi (mxm birim matris) $\hat{\Sigma}$, ağırlık katsayısı $\underline{\hat{w}}$ ($\hat{w}_k$), madde ayırıcılık parametresi koşullu kovaryans matrisi $\hat{\sigma}_{iw}^2$, gizil vektör θ_w'nin bir bileşimi $\underline{\theta}$, çok boyutlu zorluk parametresi ile ilgili bir indeks $\widehat{d_{iw}}$ ve $\underline{\hat{a}}_i^T$ çok boyutlu madde ayırıcılık vektörü olmak üzere tek boyutlu madde ve yetenek parametreleri aşağıdaki gibi kestirilir (Lee, 2013; Brossman & Lee, 2013);

$$\hat{a}_{iw} = (1 + \hat{\sigma}_{iw}^2)^{-\frac{1}{2}} \underline{\hat{a}}_i^T \hat{\Sigma} \underline{\hat{w}}$$

$$\hat{b}_{iw} = -\frac{\hat{d}_{iw}}{\hat{a}_{iw}}$$

$$\hat{d}_{iw} = (1 + \hat{\sigma}_{iw}{}^2)^{-\frac{1}{2}} . \hat{d}_i$$

$$\hat{c}_{iw} = \hat{c}_i$$

$$\hat{\sigma}_{iw}{}^2 = \underline{\hat{a}}_i^T \hat{\Sigma} \hat{a}_i - (\underline{\hat{a}}_i^T \hat{\Sigma} \underline{\hat{w}})^2$$

$$\underline{\theta} = \hat{w}_k^T \theta_w$$

Çok boyutlu madde ve yetenek parametreleri tek boyuta indirgendikten sonra ÇB-MTK gerçek veya gözlenen puan eşitleme tek boyutlu yaklaşımları uygulanabilir. ÇB-MTK gerçek puan eşitleme tek boyutlu yaklaşımı tek boyutlu gerçek puan eşitleme yöntemi ile aynıdır. X formunda karşılığı bulunmak istenen bileşik gerçek puan (τ_{Yw}) belirlenerek Y formunda bu puana karşılık gelen

tek boyutlu bileşik yetenek ($\underline{\theta}$) düzeyi aşağıdaki formül ile bulunur (Lee, 2013);

$$f(\theta_{jw}) = \tau_{Yw} - \sum_{i \in Y} p_{ij}(\theta_{jw}; a_{iw}, b_{iw})$$

Y formunda bileşik yetenek düzeyine karşılık gelen puan bulunur. Bileşik yetenek düzeyiyle ilişkili bileşik gerçek puan, her bileşik yetenek düzeyindeki madde doğru yanıt olasılıklarının toplamıdır.

ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşımı tek boyutlu gözlenen puan eşitleme yöntemi ile aynıdır. Koşullu gözlenen puan dağılımı bütün bileşik yetenek düzeyleri için Lord-Wingersky formülü ile kestirilir. Her form için gözlenen marjinal dağılımı belirlemek için, koşullu dağılım popülasyondan kestirilen tek boyutlu yetenek yoğunluğu ($\psi(\theta_{jw})$) ile çarpılır. Çarpımlar marjinal gözlenen puan dağılımını kestirmek amacıyla ve kestirilen tek boyutlu yetenek uzayı boyunca toplanır. Marjinal toplam gözlenen puan dağılımlarını eşitlemek için eşit yüzdelikli eşitleme yapılır.

b.5. Bi-faktör ÇB-MTK gözlenen puan eşitleme yöntemi

Bi-faktör ÇB-MTK gözlenen puan eşitleme yönteminde genel ve özel boyutlar dikkate alarak, puan eşitleme işlemini gerçekleştirir. Madde ve yetenek parametreleri bi-faktör modele göre kestirilir. Koşullu gözlenen puan dağılımları ikili puanlanan maddeler için Lord-Wingersky ve çoklu puanlanan maddeler için Hanson (1994) formülü ile belirlenir. Bi-faktör model için koşullu gözlenen puan

dağılımları elde edilirken algoritmalara genel faktör eklenir (Choi, 2019). Koşullu gözlenen puan dağılımları için marjinal gözlenen puan dağılımları elde edildikten sonra eşit yüzdelikli eşitleme yapılır.

Eşitleme Hatası

Eşitleme sonucunda elde edilen puanların doğruluğu hata miktarı ile değerlendirilir. Eşitlemede random (tesadüfi) ve sistematik hata olmak üzere iki tür hata vardır. Örneklemin evreni temsil etme düzeyinden kaynaklı hatalar random hatalardır. Eşitlemenin standart hatası (SEE) olarak adlandırılır. SEE eşitlemenin tekrarlanması sonucunda elde edilen hataların standart sapmasına eşittir. Örneklemin evreni temsil etme düzeyi arttıkça standart hata azalır. Örneklem küçüldükçe evreni temsil etme düzeyi azalır. Bunun sonucunda SEE artar. Uygun eşitleme deseni büyük örneklemlere uygulandığında SEE'nin sıfıra yakın değerler alması beklenir. Kullanılan yöntemlerden kaynaklanan yanlılık sistematik hatadır. Eşitlemenin yanlılığı (BIAS) olarak adlandırılır ve yanlılık çalışmalarıyla belirlenir. Varsayımların ve eşitleme koşullarının sağlanmamasından kaynaklanır (Zheng, 1991).

Eşitleme hatasını kestirmek için bootstrap yöntemi kullanılabilir (Kolen & Brennan, 2004). Bootstrap yöntemi, X formundan elde edilen puanların Y formuna ait ölçek üzerinde tekrarlı eşitlenmiş puan kestirimini elde eder. X formundan elde edilen puan x_i, tekrar (replikasyon) sayısı R, x_i için r. tekrardan elde edilen eşitlenmiş puan $\widehat{t_{yr}}(x_i)$ olmak üzere tekrarlardan elde edilen eşitlenmiş puanların ortalaması $\bar{t}(x_i)$ aşağıdaki gibidir(Lee, 2013);

$$\bar{t}(x_i) = \frac{1}{R}\sum_{r=1}^{R} \widehat{t_{yr}}\ (x_i)$$

Eşitlemenin standart hatası (SEE) aşağıdaki gibidir;

$$SEE(x_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R} (\widehat{t_{yr}}(x_i) - \bar{t}(x_i))^2}$$

BIAS ve RMSE aşağıdaki gibidir;

$$BIAS(x_i) = \bar{t}(x_i) - t_y(x_i)$$

$$RMSE(x_i) = \sqrt{(SEE(x_i))^2 + (BIAS(x_i))^2}$$

Negatif ve pozitif yanlılık değerlerinin birbirini nötrlemesini önlemek için ortalamalar hesaplanırken yanlılık değerlerinin mutlak değeri alınabilir (Öztürk Gübeş, 2014; Stone, 1992).

Kaynaklar

American Educational Research Association (AERA), American Psychological Association(APA), & National Council on Measurement in Education (NCME). (2014).*Standards for educational and psychological testing*. American EducationalResearch Association.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 508–600). American Council on Education.

Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 55–70). New York, NY: Academic Press.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory* (Doctoral dissertation). University of Iowa, USA.

Brossman, B. G., & Lee, W. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement, 37*(6), 460-481.

Choi, J. (2019). *Comparison of MIRT observed score equating methods under the commonitem nonequivalent groups design* (Doctoral Dissertation). University of Iowa, USA.

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.

Davey, T., Oshima, T. C., & Lee, K. (2000). Multidimensional linking: Four practical approaches. *Journal of Educational Measurement, 37* (4), 357-373.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Han, K. T. (2008). *Impact of item parameter drift on test equating and proficiency estimates* (Doctoral Dissertation). University of Massachusetts, USA.

Holland, P.W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187–220).Westport, CT: American Council on Education and Praeger.

Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology, 42*, 139-167.

González, J., & Wiberg, M. (2017). *Applying test equating methods: Using R* (1st ed.). Cham, Switzerland: Springer International Publishing.

Hanson, B. A. (1994). An investigation of item response theory models for polytomously scored items. *Journal of Educational Measurement, 31*(2), 155–170.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Harris, D. J., & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education, 6*(3), 195–240.

Hirsch, T. (1989). Multidimensional equating. *Journal of Educational Measurement, 26*, 337-349.

Holland, P., & Rubin, D. (1982). *Test equating*. New York: Academic Press.

Holland, P. W. (2007). A framework and history for score linking. N. J. Dorans, M. Pommerich, ve P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5-30). New York, NY: Springer-Verlag

Kelecioğlu, H. (2024, May 2). Test eşitleme. Retrieved October 29, 2024, from Epodder.org website: https://kitap.epodder.org/test-esitleme/index.html

Kim, S. Y., Lee, W., & Kolen, M. J. (2019). Simple-structure multidimensional item response theory equating for multidimensional test. *Educational and Psychological Measurement, 80*(1), 91-125.

Kline R. B. (2005). *Principles and practice of structural equation modeling.* New York, NY: The Guilford Press.

Kolen, M. J. (1988). Traditional equating methodology. *Educational measurement: Issues and practice, 7*(4), 29-37.

Kolen M.J. (2007). Data Collection Designs and Linking. N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and*

aligning scores and scales (pp. 5-30). New York, NY: Springer-Verlag.

Kolen, M., & Brennan, R. (1995). *Test equating: methods and practices.* New York: Springer-Verlag.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (1st ed.). New York, NY: Springer Science and Business Media.

Kolen, Michael J., & Brennan, R. L. (2014). *Test equating, scaling, and linking* (3rd ed.). New York, NY: Springer Science and Business Media.

Lee, W., & Brossman, B. G. (2012). Observed score equating for mixed-format tests using a simple-structure multidimensional IRT framework. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating* (vol 2.2, s. 115-142). Center for Advanced Studies in Measurement and Assessment.

Lee, E. (2013). *Equating multidimensional test under a random groups design: A comparison of various equating procedures* (Doctoral Dissertation). University of Iowa, USA.

Lee, G., Lee, W., Kolen, M. J., Park, I. Y., Kim, D. I., & Yang, J. S. (2015). Bi-factor MIRT true-score equating for testlet-based tests. *Journal of Educational Evaluation, 28*, 681-700.

Lee, G., & Lee, W. (2016). Bi-factor MIRT observed-score equating for mixed-format tests. *Applied Measurement in Education, 29*, 224-241.

Levine, R. (1955). Equating the score scales of alternate forms administered to samples of different ability. *ETS Research Bulletin Series, 1955*(2), i-118.

Li, Y. H., & Lissitz, R. W. (1998). *An evaluation of multidimensional ırt equating methods by assessing the accuracy of transforming parameters onto a target test metric*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, America, 12-16 April, 1998.

Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement, 24*(2), 115-138.

Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best?. *Applied Measurement in Education*, *3*(1), 73–95.

Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: ETS.

Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd ed.). Princeton, NJ: Educational Testing Service.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed score equatings. *Applied Psychological Measurement*, *8*, 452–461.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359–381.

McDonald, R. P. (1981). The dimensionality of tests and items. *The British Journal of Mathematical and Statistical Psychology*, *34*(1), 100–117. doi:10.1111/j.2044-8317.1981.tb00621.x

Morris, C. N. (1982). On the foundations of test equating. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 169–191). New York: Academic.

Öztürk Gübeş, N. (2014). *Testlerin boyutluluğunun, ortak madde formatının, yetenek dağılımının ve ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesine etkisi* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests* (Doctoral Dissertation), University of Iowa, USA.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. *Educational Measurement, 3*, 221–262.

Petersen, N. S., Marco, G. L., & Stewart, E. E. (1982). A test of the adequacy of linear score equating method. In P. W. Holland & D. B. Rubin (Ed.), *Test equating* (pp. 71–135). New York: Academic Press.

Simon, M. L. D. (2008). *Comparıson of concurrent and separate multıdımensıonal ırt lınkıng of ıtem parameters* (Doctoral Dissertation). University of Minnesota, USA.

Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika, 31*(1), 1-10.

Spence, P. D. (1996). *The effect of multidimensionality on unidimensional equating with item response theory* [Unpublished Doctoral Dissertation]. University of Florida, USA.

Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the twoparameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement, 16*, 1-16.

Tao, W., & Cao, Y. (2016). An extension of IRT-based equating to the dichotomous testlet response theory model. *Applied Measurement in Education, 29*, 108-121.

Tindal, G., & Haladyna, T. M. (2002). *Large scale assessment programs for all students: Validity, technical adequacy and implementation.* Lawrence Erlbaum Associates.

Ukşul, E. (2024). *Madde takımı tabanlı testlerde ölçek dönüştürme hatalarının incelenmesi* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Van der Linden,W. J. (2000). A test-theoretic approach to observed-score equating. *Psychometrika, 65*, 437–456.

Zhang, L., & Stout, W. (1999). A comparison of item response theory and classical test theory using a multidimensional item response model. *Journal of Educational Measurement, 36*(2), 139-160.

Zeng, L. (1991). *Standard errors of linear equating for the single-group design*. ACT Research Report Series. 91-4.

Zhang, J. (1996). *Some fundamental issues in item response theory with applications* [Unpublished Doctoral Dissertation]. Illinois Üniversitesi.

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

BÖLÜM II

Ölçek Geliştirme

Emrah BÜYÜKATAK[1]

Giriş

Etkili ölçüm, bilimsel araştırmanın temel taşıdır. Ölçümleri geliştirmek ve iyileştirmek için çeşitli stratejiler vardır. Belirli bir stratejinin önemi, hangi tür bilimsel olgunun ölçüldüğüne bağlı olacaktır. Örgütsel davranıştaki araştırmanın odak noktası genellikle yapılar arasındaki ilişkilerdir. Yapı, doğrudan gözlemlenemeyen teorik olarak ilgi duyulan bir olguyu tanımlamak için kullanılan kavramsal bir terimdir (Edwards ve Greenberg, 2003). Yapılara verilen örnekler arasında tutum, memnuniyet, güven, bağlılık vb. bulunur. Araştırmacılar genellikle bir dizi öğenin içeriğinden ziyade, altta yatan gizli bir yapıyı ölçmeyi amaçlarlar. Yapılar genellikle kolayca gözlemlenemez, diğer bir ifadeyle ilgilendiğimiz yapıların çoğunu göremeyiz veya dokunamayız. Bu nedenle, gizli yapıları

[1] Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Bölümü, Ankara/Türkiye, Orcid: 0000-0002-5341, ebuyukatak2014@gmail.com

ölçmek için özel geliştirilen ölçme araçları, ölçekler veya bir dizi tezahür ettirilmiş gözlem kullanırız.

Ölçü, bir katılımcının yapının var olduğuna, hissedildiğine veya ifade edildiğine inanma derecesinin niceliksel bir değerlendirmesidir. Bu yapıların dolaylı değerlendirmesi, yapıyı ölçmek için birden fazla öğenin veya göstergenin kullanıldığı ölçümler aracılığıyla gerçekleştirilir, yani bir yapı "ölçeklenir".

Son yıllarda, çeşitli bilimsel uygulamalarda tutumları, davranışları ve müdahaleleri değerlendirmek için çok sayıda ölçek geliştirilmiştir. Ölçek algı, tutum, davranış, beceri, yetenek, gibi gözlemlenip sayısal olarak ifade edilemeyen yapıların ölçülüp sayısal olarak ifade edilebilmesi amacıyla oluşturulan ölçme araçlarına verilen genel isimdir. Birçok tanımlaması olan ölçek, bir tek ve tanımlı psikolojik özelliği ölçmek amacıyla geliştirilen ölçme araçlarıdır (Erkuş, 2012). Ölçekler, doğrudan gözlemlenemeyen gizli kavramları bir grup somut ifadeyle yakalamaya çalışır. Ölçekler aynı zamanda doğrudan yollarla kolayca gözlemlenemeyen teorik değişkenlerin seviyelerini ortaya çıkarmak için tasarlanmış bileşik bir puanda birleştirilmiş madde koleksiyonlarıdır (DeVellis, 2012). Ölçekler ayrıca ölçmeyi kolaylaştırmakla birlikte elde edilen sonuçların niteliğinin de belirlenmesini sağlamaktadır. Bilimsel gelişme ölçmeye dayanır, duyarlı ölçüm vasıtasıyla yapılan ölçmeler de bu gelişmeyi artırmaktadır (Tavşancıl, 2014).

Ölçek geliştirme süreci, çok sayıda teorik ve metodolojik titizlik gerektiren karmaşık ve sistematik prosedürler ile istatistiksel yeterlilik içermektedir. Ölçek geliştirmede bireylerin sahip oldukları psikolojik özelliğin ne olduğunu ortaya çıkaracak maddeler oluşturulur (Erkuş, 2012). Ölçek geliştirme süreci sonunda ölçekleri

oluşturacak öğelerin oluşturulmasıyla başlar. Buradaki en önemli konu ölçeklerin bir örneklem popülasyonundan veri toplamak için kullanılmadan önce değerlendirilmesidir. Ölçek geliştirme aşamasında dikkate alınması gereken bir dizi önemli faktör ve ölçek geliştirmeye yardımcı olmak için izlenmesi gereken belirli adımlar vardır. Ölçek geliştirme adımlarında ölçek geliştiricilere göre farklılıklar olsa da, izlenen adımların içerikleri büyük ölçüde aynıdır (DeVellis, 2012; Erkuş, 2012). Bu kapsamda 8 adımdan oluşan ölçek geliştirme süreci açıklanmıştır.

Ölçeğin Amaç ve Kapsamının Belirlenmesi

Bir ölçek geliştirmeden çok önce ölçeğin amacının ne olduğu ve gelecekte hangi uygulamalara sahip olabileceği belirlenmelidir. Bu, uzunluk ve yanıt formatı gibi alınması gereken kararlarda yardımcı olacaktır. Ölçek geliştirme çalışmasında ölçülmek istenen yapının özelliklerinin belirlenmesi birinci olarak yapılması gereken işlemdir. Ölçek geliştirme çalışmalarında alan uzmanlarının yardımıyla ve literatür taranarak ölçülmek istenen kavramın yapısal özellikler belirlenir. Ölçek geliştiriciler kendilerine rehberlik edecek kuram belirleyerek kavramsal yapıyı oluşturmalı ve ölçülmek istenen yapının diğer yapılardan ne derecede farklı olduğunu göstermelidirler (DeVellis, 2014).

Ölçek geliştirme sürecindeki en temel ve önemli aşama ölçmek istenilen yapının belirlenmesi aşamasıdır. Zor ve uzun bir süreç olan bu aşamadaki oluşan eksiklikler sonraki aşamaları da etkileyeceğinden ölçek geliştiricilerin en çok yoğunlaşması ve zaman harcaması gereken aşama ölçeğin amaç ve kapsamının belirlenmesi aşamasıdır (Erkuş, 2012).

Bu aşamada neden bir ölçek geliştirileceğine dair gerekçeler ortaya konur. Bu gerekçeler mevcut bir ölçeğin eksiklerinin olması, yeni bir değişken ile ilgili araştırma yapılması, yeni bir değişkene dair araştırma olabilir. Ölçeğin hangi amaçla geliştirileceği aşamasında yeni bir ölçme aracına ihtiyaç olup olmadığı, ölçeğin neyi ölçmeye çalışacağı, hedef grubun kim olduğu, uygulama koşullarının neler olduğu, testin ideal formunun ne olduğu ve uygulama süresinin önemli olup olmadığı gibi bazı ön soruların yanıtlanması gerekmektedir (Yalçın, 2021).

Ölçülecek Yapının Tanımlanması

Ölçülecek değişkenin kuramsal açıklaması belirlenerek tanımlanır. Ardından kuramsal çerçevede, belirtilen yapıya ilişkin davranışsal göstergeler tespit edilir. Ölçülmesi istenen yapıya karar verilirken öncelikle ölçülecek yapının kavramsal ve kuramsal tanımlaması yapılmalıdır (Tezbaşaran, 2008). Bu yapıların davranışsal göstergelerinin de belirlenmesi önemlidir. İçerik analizi bu davranışsal göstergeleri belirlemede kullanılabilir. Ayrıca ilgili literatür incelenebilir, uzman görüşüne başvurulabilir veya gözlem de yapılabilir (Crocker ve Algina, 1986).

Ölçülecek kavramın/yapının kuramsal temeline göre, bu kavramın veya yapının alt boyutları olması durumunda alt boyutların da kurama uygun olarak belirlenmesine ihtiyaç duyulur. Birkaç boyuta sahip olduğu düşünülen yapılarda, ölçek geliştiricisi ölçekteki her boyutu temsil edecek bir dizi öğeyi göz önünde bulundurmalıdır. Bu açıdan, çok boyutlu ölçekler yalnızca bir boyuta sahip olanlardan daha uzun olmaktadır. Literatürde benzer konularda geliştirilen geçerli ve güvenilir ölçeklere ait yapıların incelenmesi bu

konuda fikir verebilir. Madde yazımından önce, ölçeğin alt boyutları mutlaka tanımlanmalıdır.

Yapının boyutsal yapısını haritalamak için de bir literatür taraması gereklidir. Çünkü araştırmacıların her boyuta ilişkin teorik anlayışlarını yansıtan öğeleri oluşturmaları gerekir. Soyutluk (genişlik) arttıkça, yapının birden fazla boyuttan oluşması beklenebilir. Örneğin, lisansüstü öğrenci danışman desteğine ilişkin bir literatür taraması, üç ayrı türde (veya boyutta) öğretim üyesi danışman desteğini ortaya çıkarabilir: Psikososyal, araştırma ve kariyer.

Ölçek Geliştirme Tekniğinin Belirlenmesi

Ölçmek istenilen özelliği ölçmede çok sayıda yanıt verme biçimi bulunmaktadır. Hangi tekniğin kullanılacağına karar vermede ölçülen özellik belirleyicidir ve bu süreç madde oluşturma ile beraber gerçekleştirilmelidir (DeVellis, 2014). Ölçek geliştirici formatın ne olacağını erken bir aşamada düşünmelidir. Bu adım, öğelerin oluşturulmasıyla aynı anda gerçekleşmelidir. Örneğin, seçilen yanıt formatı tek kelimelik öğelerden oluşan bir kontrol listesiyse, uzun bir beyan ifadeleri listesi oluşturmak zaman kaybı olabilir. Genel olarak, bir süreklilikte puanlanabilen ve bir ölçek puanı oluşturmak için toplanan öğelerden oluşan ölçekler teorik yönelimle en uyumlu olanlardır. Elde edilen verilerin matematiksel özellikleri bakımında format sıralama, sınıflama, eşit aralıklı ve oran ölçeği şeklinde olabilir. Ayrıca alan yazında Bogardus, Guttman, Thurstone, Likert ve Osgood tipi özel ölçek geliştirme teknikleri de bulunmaktadır. Ancak çoğunlukla Guttman, Thurstone, Likert tipi ölçekler kullanılmaktadır.

Guttman Ölçeği

Guttman ölçeği, üç tek boyutlu ölçekten biridir. iğer ikisi ise Likert Ölçeği ve Thurstone Ölçeği'dir. Guttman ölçeği, kümülatif ölçekleme olarak da adlandırılır ve hiyerarşik bir şekilde sıralanabilen öğelerle oluşturulur. Katılımcıların konuyla ilgili aşırı "tutumlarını", yani son derece olumlu veya olumsuz tutumlarını temsil eder. Bu ölçek, araştırmacılar tarafından tek boyutlu bir ölçeğin gerekli olduğu durumlarda kullanılır. Tek boyutlu ölçek, yanıt seçeneklerinin yalnızca bir ölçüm parametresine sahip olduğunu, yani ölçekle bir dizi sayının ilişkilendirilebileceğini gösterir. Örnek olarak havayolu hizmetindeki memnuniyet derecesi 0-10 arasında bir puana göre ölçülmek istendiğinde, bu durum tek boyutlu yanıt seçenekleriyle gösterilebilir.

Guttman ölçeğinin bir ifade listesi vardır. Bu listenin sonuna yerleştirilen ifadeyi kabul eden katılımcıların, sonuncusunun üstündeki diğer tüm ifadeleri de kabul edecekleri sonucuna da varılır. Her ifadenin ilişkili bir ağırlığı vardır. Katılımcı geri bildirimlerine göre ağırlığın toplanması, araştırmacıların katılımcılar için kabul edilebilir ifadelerin sayısını tahmin etmelerine yardımcı olacaktır. Örneğin, 5 ölçekli bir Guttman ölçeğinde, bir katılımcı 3 puan alırsa, ölçeğin ilk 3 ifadesini kabul ettiği anlamına gelir. Farklı bir katılımcı 5 puan alırsa, kümülatif ölçekteki tüm ifadeleri kabul ettiği anlamına gelir.

Thurstone Ölçeği

Thurstone ölçeği, katılımcının bir konuya karşı davranışını, tutumunu veya hissini izlemek için kullanılan tek boyutlu bir ölçek olarak tanımlanır. Bu ölçek, her bir ifadenin katılımcının konuya karşı tutumunu olumlu veya olumsuz olarak gösteren sayısal bir

değere sahip olduğu belirli bir konu hakkındaki ifadelerden oluşur. Katılımcılar katıldıkları ifadeleri belirtir ve bir ortalama hesaplanır.

Bu ölçek, Robert Thurstone tarafından eşit görünen aralık seviyelerinde ölçümü yaklaşık olarak belirlemek için geliştirilmiştir. Thurstone ölçeği, Likert ölçeğinin temelleri üzerine inşa edilmiştir, ancak bu tutum ölçeği oluşturma yöntemi, nihai tutum puanını değerlendirirken yalnızca her bir öğenin değerini dikkate almakla kalmaz, aynı zamanda nötr öğelere de hitap eder. Guttman ölçeği ve Bogardus sosyal mesafe ölçeği de öğelerin hiyerarşik bir şekilde sıralanabileceği tek boyutlu ölçeklerin varyasyonlarıdır. Thurstone ölçek tipinde üç ölçek bulunmaktadır, ancak en yaygın kullanılan yöntem eşit görünen aralıklardır ve bu nedenle ölçek eşit görünen aralık ölçeği de olarak da adlandırılmaktadır. Diğer ikisinin geliştirilmesi biraz daha karmaşıktır, ancak yine de aynı katılıyorum/katılmıyorum sınav sorusu türüyle sonuçlanır; bunlar ardışık aralıklar yöntemi ve eşleştirilmiş karşılaştırmalar yöntemidir.

Likert Ölçeği

Likert Ölçeği, 1932'de yaygın olarak kullanılan 5 puanlık ölçek biçiminde ortaya çıkmıştır. Bu ölçekler, genel konulardan oluşan bir gruptan, katılımcılardan onay veya inanç düzeylerini belirtmelerini isteyen en spesifik olanlara kadar uzanmaktadır. Likert Ölçeği, araştırmacıların katılımcıların tutumlarını ve görüşlerini toplamak için kullandıkları tek boyutlu bir ölçektir. Araştırmacılar genellikle bu psikometrik ölçeği bir markaya, ürüne veya hedef pazara yönelik görüşleri ve bakış açılarını anlamak için kullanırlar.

Psikolog Rensis Likert, bir grup maddeye (belki 8 veya daha fazla) verilen yanıtların bir koleksiyonundan oluşan bir ölçek ile bir

ayrım yapmıştır. İnsanların bir konuya yönelik tutumlarını ölçerek bir dizi ifadeye cevap vermelerini ve her biri hakkında ne hissettiğini değerlendirmelerini sağlayarak bu yöntemi geliştirmiştir. Likert Ölçeği soruları, anketler ve araştırmalar için temeldir ve katılımcıların çeşitli seçenekler arasında tutumlarını, görüşlerini veya davranışlarını ifade etmelerine olanak tanımaktadır.

Uyarıcıların ve Uygun Tepki Kategorilerinin Yazılması (Madde Yazımı)

Maddelerin yazımı planlanırken en temelde düşünülmesi gereken husus, maddelerin sade ve anlaşılır olması, bir maddenin birden fazla yargı ve düşünce ifadesine sahip olmamasıdır (Naresh, 2006). Bir madde havuzu, bir ölçeğin ortaya çıkabileceği zengin bir kaynak olmalıdır. İlgi duyulan içerikle alakalı çok sayıda madde içermelidir. İçerikle ilgili çok sayıda madde olması bir yükümlülük değil, bir ihtiyaçtır. Bu, aynı zamanda geçerliliğin temeli olan iç tutarlılık güvenilirliğinin de temelidir. Olumlu ve olumsuz ifade edilen maddeler havuza dahil olsun veya olmasın, ifadeleri yerleşik dilbilgisi kurallarına uymalıdır. Bu durum belirsizlik kaynaklarından bazılarının önlenmesine yardımcı olacaktır. Ölçek geliştirme sürecinde iki seçenekli, çoktan seçmeli, dereceleme türü, senaryo tipi ve simülasyon tipi maddeler kullanılmaktadır.

Madde yazılırken mutlaka literatür taranmalıdır. Bununla birlikte alan uzmanlarının görüşlerinden ve ölçeğin geliştirileceği düşünülen hedef gruptan da faydalanılabilir. Maddeler yazılırken dikkat edilmesi gereken hususlar aşağıda belirtilmiştir (Gültekin, 2012; Tekin, 2000):

- Maddelerdeki ifadeler ölçülecek olan nitelikle uygun olmalıdır. Bir tutum ölçeği geliştiriliyorsa belirlenen

tutumun düşünsel, duyuşsal ve davranışsal unsurlarını içeren maddeler oluşturulmalıdır (Bailey, 1994; Likert, 1967; Akt. Tavşancıl, 2014).

- Bir madde ile birden çok özelliğin ölçülmemesi gerekir.
- Maddeler mümkün olduğuca kısa ve anlaşılır ifade edilmelidir.
- Rastgele cevaplama yapanları ayırmada kullanılan yarısı olumlu diğer yarısı olumsuz maddelerin yazımı, ters olarak belirtilen maddeleri anlamada sorun yaşatabilir.
- Seçenek sayısı ve seçeneklerin ifadeleri önemlidir.
- Aynı maddede iki tane olumsuz ifadenin olmaması gerekir.
- Madde sayısının ne kadar olacağını önceden belirlemek zor olduğundan olabildiğince fazla sayıda madde hazırlamak hedeflenmelidir. Madde yazımında ölçekte kullanılması planlanan madde sayısının çoğunlukla üç katı kadar madde yazılması yeterlidir. Böylelikle amaca hizmet etmeyen ve ölçekten çıkarılabilecek maddeler olabilir (Baykul, 2010; Tezbaşaran, 2008).

Maddelerin Gözden Geçirilmesi ve Düzeltilmesi

Maddelerin yazımı tamamlandıktan sonra maddelerin ölçülmek istenilen davranışı ölçüp ölçmediği incelenir. Dil bakımından maddelerin anlaşılırlık düzeyi, bilimsel olarak hata olup

olmadığı, maddelerin dilbilgisi hatalarının olup olmadığı, oluşturulan maddelerin uygulanacak gruba uygunluğu detaylı olarak incelenir. Maddeler, ölçek yapısına göre uzmanlar tarafından kontrol edilir ve gerekli düzenlemeler ile düzeltmeler yapılır (Atılgan, Kan ve Doğan, 2013).

Oluşturulan bu madde havuzunda uzman görüşü en önemli adımlardan birisidir. Uzmanlar tarafından maddelerin gerekliliği, açık ve özgüllüğünün anlaşılırlığının incelenmesi yapılır. Yapılan inceleme sonucunda bazı ifadelerin çıkarılması, eklenmesi veya değiştirilmesi yönünde tavsiyeler oluşturulur. Ölçek hazırlayan bu tavsiyeleri kabul ederek gerekli düzeltmeleri ve düzenlemeleri yapabilir. Ancak bazı durumlarda ölçek hazırlayanlar tavsiyelerin tamamını dikkate almazlar ve kendi yazdıkları maddeleri kullanırlar.

Ölçek maddelerinin değerlendirilmesindeki uzman görüş aşamasında formlar geliştirilebilir ve uzmanlar tarafından bu formlar kullanılabilir. Uzmanların burada yapacağı esas değerlendirme ölçeğin amacına uygun olarak maddelerin yazılıp yazılmadığıdır. Maddenin uygunluk derecesi (hiç uygun değil, kısmen uygun, uygun) belirtilir.

Ön Deneme Uygulamasının Yapılması

Bir ölçeğe ön deneme uygulamasının yapılması, ölçeğin ölçülmek istenen hedef kitleye yakın benzerlikte küçük grup üzerinde uygulanarak, maddelerin gözden geçirilmesine imkan tanımaktadır. Bu aşama madde yazımından sonra ölçek geliştirme sürecinin özü olarak kabul edilir. Ön deneme uygulaması maddeler, yorumlar ve öneriler hakkında istatistiksel bilgi elde etmek için ölçeği hedef popülasyondan temsili bir örneklemde test etmeyi içermektedir. Bu ön deneme esnasında maddelerin okunabilirliği,

ölçeğin eğer gerekiyorsa cevaplama süresinin belirlenmesi ve ölçeğin uygulanmasında ortaya çıkabilecek anlaşılmayan bölümlerin belirlenmesi sağlanır.

Örneklem grubuna ön denemenin uygulanması ölçeğin geçerliğinin ve güvenirliğinin belirlenmesi açısından da çok önemlidir. Ancak burada esas olan örneklem grubunun evreni temsil etmesidir. Bu sebeple amaca uygun örneklem grubu belirlenmelidir ve örneklem grubunun sayısı tespit edilmelidir. Ön deneme uygulamasındaki örneklem büyüklüğü ile ilgili olarak farklı görüşler bulunmaktadır. Evci ve Aylar (2017) ön denemenin hedef kitlenin yaklaşık %5'lik kısmı ile yapılmasını ifade ederken, Şeker ve Gençdoğan (2014) 30 ila 50 arasında katılımcı seçilerek uygulama yapılmasının yeterli olacağını ifade etmiştir

Ön deneme uygulmasının yapılmasının ardından elde edilen veriler kapsamında istatistikler elde edilir ve ön denemesi yapılan ölçek madde istatistikleri incelenir. Hesaplanan bu istatistikler madde toplam korelasyonları, maddelerin ortalamaları, madde varyansları, KMO, faktör analizleri ve iç tutarlılık katsayılarıdır. Madde istatistiklerinin belirlenmesinden sonra istenen özelliği ölçmeyen veya kriterleri sağlamayan maddelerin ve aşırı benzer maddelerin çıkarılması ve düzeltilmesi gibi işlemler ile ölçeğin ideal yapı ve uzunluğa gelmesi sağlanır.

Deneme Uygulamasının Yapılması

Ön deneme uygulaması sonrası yapılan düzeltmeler ve uzman görüşleri de dikkate göz önünde bulundurularak maddeler, ölçek yönergesi, ölçekleme yöntemi ve ölçek süresi belirlenerek ölçeğin son hali verilir. Deneme uygulamasının amacı, ölçülmesi istenen özelliğe en uygun yapının belirlenmesi için gerekli bilgilerin

sağlanmasıdır. Deneme uygulamasında seçilecek örneklem, ölçülmek istenen özellik bakımından heterojen olmalıdır.

Deneme uygulamasındaki katılımcı sayısı ile ilgili kesinlik yoktur. Bazı akademisyenler uzman incelemesinden sonra 20 maddelik bir ölçek için 300 gibi büyük bir örneklemi önerirken (DeVellis, 2014), Nunually (1978) bir ölçek için ortalama 300 kişilik bir grubun yeterli olduğunu, Comrey ve Lee (1992) 100 kişilik bir örneklemin zayıf, 200 kişilik örneklemin orta, 300 kişilik örneklemin iyi, 500 kişilik örneklemin çok iyi ve 1000 kişilik örneklemin ise mükemmel olacağını belirtmiştir. Bazı araştırmalarda ise örneklem grubu ile madde sayısı arasında ilişki kurularak sayının belirlenmesinin uygun olacağı vurgulanmaktadır. Örneklem sayısı madde sayısının en az 5, 10 veya 15 katı olması yönünde görüşler bulunmaktadır (Delice ve Ergene, 2015). Çok az sayıda örneklem grubu kullanarak deneme yapılması maddeler arasındaki kovaryans örüntüsünün kararlı olmamasına sebep olabilir. Evrenin örneklem tarafından temsil edilememesi de oluşabilecek diğer bir olumsuzluktur (DeVellis, 2012).

Ölçeğin deneme formunun hazırlanmasında konu birliğinin sağlanması, yönergede olması gereken bilgiler kapsamında yönergenin yazılması ve gözden geçirilmesi, ölçek deneme formunun baskı için hazırlanması ve çoğaltılması aşamaları, deneme uygulamasının yer, zaman, uygulayıcılar açısından planlanması ve sonunda deneme uygulamasının belirtilen şartlarda gerçekleştirilmesi önemlidir.

İstatistiksel Analizlerin Yapılması ve Ölçeğe Son Halinin Verilmesi

Deneme uygulamasından sonra maddelerin ve ölçeğin istatistikleri yapılır. Yapılan istatistiksel analizlerle ölçeğin geçerlik ve güvenirliğinin kanıtlanması sağlanır. Ölçek için hazırlanan maddelerin uygunluğunun veya uygun olmayan maddelerin belirlenmesi için hem madde istatistikleri hem de ölçek analizleri yapılır. İstatistiksel analiz aşamasında madde güçlükleri ve madde ayırt edicilikleri hesaplanarak düzeltilecek maddeler belirlenir.

Madde istatistiklerinin incelenmesi aşamasında ölçekte kullanılan maddeler birbiriyle yüksek düzeyde ilişkili olmalıdır. Bu amaçla korelasyon matrisi incelenmesi yapılır ve inceleme sonucunda korelasyon değerleri düşük olan maddelerin çıkarılması veya değiştirilmesi sağlanır. Madde analizlerinde madde-toplam korelasyonları incelenerek de madde istatistiklerine bakılabilir. Her bir madde diğer maddelerle veya ölçeğin toplam puanıyla yüksek düzeyde ilişkili olmalıdır. Bu incelemede maddelerin kendisi dışındaki maddelerin toplam puanlarıyla korelasyonuna, her bir maddenin o madde dâhil tüm ölçek puanıyla ilişki katsayısına bakılır. Ayrıca madde varyansları ve madde ortalamaları da maddelerin hakkında önemli bilgiler vermektedir. Eğer bir maddenin varyansı sıfıra yakınsa, tüm katılımcıların bu maddeye benzer cevaplar verdiği sonucu ortaya çıkmaktadır ve bu durum istenilen bir özellik değildir. İdeal olan ise maddelerin orta değere yakın ortalamalar vermesidir. Uç değerlere yakın ortalamalar gösteren bir maddenin istenilen yapıyı ölçemediğine karar verilir.

Ölçekle ilgili olarak geçerlik ve güvenirlik kanıtları da açıkça ortaya konulmalıdır. Güvenirlik kanıtları için tüm ölçek ve alt boyutlar için ayrı ayrı hesaplanmak üzere iç tutarlılık katsayısı olan

Cronbach Alfa katsayısı kullanılır. Çok boyutlu ölçeklerde buna ek olarak Bileşik Güvenirlik katsayısının da sunulması gereklidir. Bazı durumlarda mümkünse test tekrar test yöntemi ile 30-50 kişilik bir örneklemde güvenirlik katsayısı hesaplanmalıdır. Ancak ölçülmesi istenen özellik kısa sürede değişkenlik gösterecekse, bu yöntemin kullanılmaması gerekir.

Geçerlik kanıtları için ölçekle ilgili olarak kapsam ve yapı geçerliği incelenir. Kapsam geçerliğinde uzman görüşleri alınarak ölçeğin kapsam geçerliği kanıtları ortaya konur. Yapı geçerliğinde ise açımlayıcı faktör analizi veya temel bileşenler analizi kullanılır. Bu analizler tutumun alt boyutları ve sayıları hakkında bilgi verirler (Tavşancıl, 2014). Faktör sayısının belirlenmesinde hangi yöntemin kullanıldığı, çok boyutlu ölçeklerde ise hangi döndürme yönteminin belirlenme sebebi açıklanır. Burada ideal olan madde faktör yüklerinin 0.40'ın üzerinde olmasıdır. Eğer bir maddenin faktör yük değeri kabul edilebilir sınırda ancak düşükse (0.31 ile .39 arası) ve bu madde çok önemli veya kritik bir madde değilse, bu madde ölçekten çıkarılabilir. Ayrıca açımlayıcı faktör analizi yapılan ölçeğin, başka bir grupta uygulanması sağlanır ve doğrulayıcı faktör analizi de yapılır. Böylelikle ölçülmek istenen yapının başka grupta da doğrulandığı kanıtı sağlanır.

Türkiye'de geliştirilmiş bir ölçme aracının kullanılması durumunda ölçme aracının geçerlik ve güvenirlik değerleri incelendikten sonra araştırmacı/lardan izin alınması gerekmektedir (Yalçın, 2021). Bununla birlikte her ölçme aracının geliştirildiği yaş ve gruplarda geçerli ve güvenilir olduğu dikkate alınmalıdır. Eğer ölçeğin standardizasyon çalışması yapılmamışsa, farklı örneklemden veri toplanması durumunda (aynı yaş grubunda için

kullanılsa dahi) ölçeğin geçerlik ve güvenirlik kanıtları tekrar sunulmalıdır.

Sonuç

Ölçek geliştirme çalışmaları uzun literatür taraması gerektiren, emek ve zaman harcanarak yapılan çalışmalardır. Birbiriyle tutarlı, hataları olduğunca en asgari seviyeye indirilmiş, geçerli veriler üretmek ve çözüm geliştirmek adına doğru tespitler yapabilmek için ölçülmek istenen özelliğe duyarlı ölçek kullanabilme imkanlarının devamlı olarak araştırılması bilime olan katkısını daha da artıracaktır.

Kaynaklar

Atılgan, H., Kan, A ve Doğan, N. (2013). *Eğitimde ölçme ve değerlendirme* (6.Baskı). Ankara: Anı Yayıncılık.

Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2.Baskı). Ankara: Pegem Akademi.

Comrey, A. L., ve Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Crocker, L., ve Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Delice, A., ve Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Eğitim Bilimleri Dergisi*, *3*(1), 60-75.

DeVellis, R. F. (2012). Scale development. *Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

DeVellis, R. F. (2014). *Ölçek geliştirme kuram ve uygulamalar* (T. Totan, Çev.). Ankara: Nobel Yayıncılık.

Edwards, J. R.ve Greenberg, I. J. (2003). *Construct validation in organizational behavior research.* Organ. Behav. A Manag. Chall., 2nd ed., Lawrence Erlbaum, Mahwah, NJ, 311-354.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi Yayınları.

Evci, N., ve Aylar, F. (2017). Ölçek geliştirme çalışmalarında doğrulayıcı faktör analizinin kullanımı. *Sosyal Bilimler Dergisi, 4*(10), 389-412.

Gültekin, S. (2017). Testlerde kullanılacak madde türleri, hazırlama ilkeleri ve puanlanması. RN Demirtaşlı. *Eğitimde ölçme ve değerlendirme içinde*, 145-222.

Naresh, K. M. (2006). Questionnaire design and scale development. *MVR Grover, The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, 83-94.

Nunnually, J.C. (1978). *Psychometric theory.* NewYork: McGraw-Hill Companies.

Şeker, H. ve Gençdoğan, B. (2014). *Psikolojide ve eğitimde ölçme aracı geliştirme* (2. Basım). Ankara: Nobel Yayıncılık.

Tavşancıl, E. (2014). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayıncılık.

Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları.

Tezbaşaran, A. (2008). *Likert tipi ölçek hazırlama kılavuzu.* Ankara: Türk Psikologlar Derneği Yayınları.

Yalçın, S. (2021). Ölçek geliştirme ve uyarlama süreci bilgi notu.

BÖLÜM III

Teoriden Uygulamaya Bibliyometrik Analiz

Burcu DEMİRÖZ[1]

Giriş

Literatürde yer alan araştırmaların sistematik olarak incelenmesi 1910'lara dayanmaktadır. Bu anlamda ilk araştırma, Cole ve Eales tarafından 1917'de yapılmıştır. Araştırmayı, İngiliz Patent Ofisi eski kütüphanecisi tarafından yapılan bilim tarihi üzerine bir istatistiksel analiz izlemiştir (Hulme, 1923). Lotka (1926), bilimsel üretkenliğin frekans dağılımını analiz etmiş ve bu araştırması, Lotka Kanunu'nun geliştirilmesine yol açmıştır. Lotka kanunu, bibliyometrinin en yaygın kullanılan kanunlarından biridir ve yazar üretkenliğindeki desenleri değerlendirir. 1927'de Gross ve Gross kimya dergisindeki makalelere eklenen atıfları saymış ve analiz etmiştir. Bir diğer önemli araştırma ise Bradford'un (1934) dergilerde yayımlanan makalelerin frekans dağılımını incelediği

[1]Dr., Türkiye Cumhuriyeti Milli Eğitim Bakanlığı, Türkiye, ORCID: 0000-0002-8326-4671, e-posta: bd14.olcme@gmail.com

araştırmadır. Ortaya çıkan Bradford Kanunu, bibliyometrik analizde dergi üretkenliğini incelemek için yaygın olarak kullanılmaktadır. Zaman içinde, literatür incelemeleri daha kapsamlı ve sistematik hale gelerek, günümüz araştırmalarının temel yapı taşlarından oluşturmuştur. Bu alandaki araştırmalara rağmen, 1969 yılına kadar "bibliyometri" terimi kullanılmamıştır. "Bibiliyometri" kavramı Pritchard tarafından 1969 yılında ortaya atılmıştır. Pritchard (1969) bibliyometriyi, yazılı iletişimin süreçlerini ve bir disiplinin doğasını ve gelişim sürecini, yazılı iletişimin çeşitli yönlerini sayarak ve analiz ederek aydınlatmak için matematiksel ve istatistiksel yöntemlerin uygulanması olarak tanımlamıştır. Pritchard'ın bu tanımı, bibliyometrik analizin teorik çerçevesini belirleyerek alanın akademik bir disiplin olarak tanınmasını sağlamıştır. Bibliyometri bir literatür bütününün bibliyografik özelliklerinin nicel analizidir (Hawkins, 1977).

Bibliyometrik analiz, literatürde yer alan araştırmaları nicel olarak değerlendirmede ve araştırma alanlarını anlamada önemli bir araçtır. Bu yöntem, bilimsel üretkenliği; yazarlar, kurumlar ve ülkeler arasındaki iş birliğini ölçmek için matematiksel ve istatistiksel teknikler kullanarak araştırmacılara belirli bir alandaki eğilimleri, önemli kaynakları ve işbirlikği ağlarını tanımlama imkânı sunar (Aria ve Cuccurullo, 2017). Bibliyometrik analiz, yayınların etki düzeylerini belirlemek ve zaman içindeki gelişimini izlemekle kalmaz, aynı zamanda politikaların şekillendirilmesinde, finansman stratejilerinin belirlenmesinde ve bilimsel alanlardaki boşlukların belirlenmesinde de kritik bir rol oynar (Leydesdorff & Persson, 2010). Bu sayede akademik araştırmaların kalitesi artırılabilir, kavramsal bağlantılar daha net bir şekilde ortaya konulabilir ve gelecekteki araştırmalar için yol haritaları çizilebilir (Kessler, 1963).

Biliyometrik analiz, bilimsel literatürdeki yenilikleri izleme, önemli bilim insanlarını tanıma ve daha etkili araştırma stratejileri geliştirme konusunda araştırmacılara önemli bilgiler sağlar.

Bibliyometrik analiz gerçekleştirilirken araştırma tasarımı, veri toplama, veri analizi-haritalandırma (görselleştirme) ve yorumlama aşamaları araştırmanın geçerliğini ve güvenirliğini artırmak için dikkatlice planlanmalı ve uygulanmalıdır.

1.Araştırma tasarımı

Bibliometrik analiz araştırmasının tasarım aşamasında, öncelikle araştırmanın amacı ve kapsamı netleştirilir. Amaçlar belirlenir. Amaçlar belirlendikten sonra bu amaçlara ulaşmayı sağlayacak açık, anlaşılır ve net araştırma soruları belirlenmelidir. Bu aşama araştırma sorularını ve araştırmanın kapsamını derinlemesine anlamayı gerektirmektedir.

2.Veri toplama

Bir bibliyometrik analizde veriler, belirli bir alandaki yayınları derleyen bir veritabanından veya disiplinler arası yayınları biraraya getiren bir veritabanından elde edilebilir. Tarama yapılacak veri tabanı seçilirken akademik literatürü ne ölçüde kapsadığına dikkat edilmelidir. Veri tabanının kapsamın ülkelere, dillere, bilim alanlarına veya yayınevlerine karşı önyargılı olmaması gerekir (Neuhaus & Daniel, 2008). Veri tabanında yer alan bibliyografik verilerde yazar adlarının doğru yazılmış ve tutarlı olmasına, yayınların yayımlandığı dergilerin isimlerinin ve yazarların kurumsal bilgilerinin standartlaştırılmış olmasına dikkat edilmelidir. En çok kullanılan veri tabanları Web of Science, Scopus, Google Scholar, PubMed, ProQues, CrossRef, ERIC, ArXiv, CiteSeerX ve Dimensions en çok kullanılan veri tabanlarıdır.

İlk olarak 1964 yılında Eugene Garfield tarafından kurulan Institute for Scientific Information (ISI) tarafından geliştirilen Web of Science (WoS), günümüzde Clarivate Analytics tarafından işletilmektedir. WoS, dünya genelinde en yüksek etkiye sahip lider disiplinler arası atıf veritabanlarından biridir. 21,000'den fazla yüksek kaliteli akademik dergiyi indeksler. Ücretli bir veri tabanıdır. Erişim üniversiteler veya kurumlar aracılığıyla sağlanır. WoS, fen ve sağlık bilimleri alanlarında daha kapsamlı bir tarama imkânı sunarken sosyal ve beşeri bilimler gibi alanlarda diğer veri tabanlarına göre daha sınırlı kapsama sahiptir.

Scopus veri tabanı 2004 yılından itibaren, Elsevier tarafından kullanıma açılmıştır. Scopus, 40.000'den fazla hakemli dergi, konferans bildirileri, ticari yayınlar, kitap serileri ve patentleri kapsadığı için en büyük veritabanıdır. Scopus'un en önemli özelliği, disipliler arası en kapsamlı veri tabanı olmasıdır. Avrupa, Latin Amerika ve Asya Pasifik bölgelerinde gerçekleştirilen araştırmaları da bünyesinde barındırmaktadır. Henüz yayımlanmamış makalelere de erişim sağlanabilmektedir. Ücretli bir veri tabanıdır.

Google tarafından geliştirilen ve 2004 yılından itibaren kullanılabilen diğer bir veri tabanı ise Google Scholar'dır. Veri tabanı, akademik literatür için geniş bir tarama yapmanın basit bir yolunu sunar ve disiplinler arasıdır. Google Scholar, tarama sırasında makalelerin tam metnini, yazarını, makalenin yayımlandığı yayını ve makalenin diğer akademik literatürde ne kadar alıntılandığını dikkate alarak sonuçları sıralar. Böylece anahtar kelimeler ile en ilişkili sonuçlar her zaman ilk sırada görünür. Ücretsiz olması ve kullanım kolaylığı popüler olmuş bir araçtır.

PubMed, biyomedikal ve sağlık bilimleri alanlarında en kapsamlı ve güvenilir veri tabanlarından biridir. Amerika Birleşik Devletleri Ulusal Tıp Kütüphanesi (NLM) tarafından geliştirilmiş ve 1996 yılında kullanıma sunulmuştur. Özellikle MEDLINE veri tabanındaki makalelere erişim sağlar. Yaklaşık 6.600 açık erişimli dergi, kitap bölümü ve PubMed Central (PMC) aracılığıyla tam metin kaynak sunar. PubMed özet (abstract) ve başlık bilgilerini ücretsiz olarak sunar. PMC tarafından sunulan tam metinler ücretsizken farklı dergilerde yayımlanan makalelerin tam metnine erişebilmek için ücret ödemesi gerekebilir

Araştırmanın kapsamına uygun veri tabanları belirlendikten sonra anahtar kelimeler belirlenerek tarama yapılır. Tarama yapılırken Boolean mantığı kullanılır. Boolean mantığı 19. yüzyılda matematikçi George Boole tarafından geliştirilmiştir (Boole, 1854). Boolean mantığı kullanılarak gerçekleştirilen bir taramada anahtar kelimeler arasındaki ilişkileri tanımlamak için AND, OR ve NOT gibi mantıksal operatörlerden faydalanılır. İki anahtar kelime arasına AND yazılırsa her iki anahtar kelimeyi de içeren sonuçlar elde edilir. Kelimeler arasına OR yazılırsa iki kelimeden birini veya ikisini içeren sonuçlar elde edilir. Anahtar kelimeden önce NOT yazılırsa anahtar kelimeyi içermeyen sonuçlar elde edilir.

WoS veri tabanında tarama yapılırken “Advance Search” kullanılabilir. Etkili bir arama yapabilmek için etiketlerden faydalanılır. TS etiketi bir makalenin başlık, özet ve anahtar kelimelerinde geçen terimleri aramak için kullanılır. TI etiketi yalnızca makale başlıklarında; AB etiketi yalnızca özet bölümünde arama yapar. AU ve AI etiketleri, belirli bir yazarın araştırmalarını bulmak için kullanılır. Fakat AU etikietinde yazar isimleri aratılırken AI etiketinde ResearcherID veya ORCID aratılır. AK etiketi, bir

makalenin yazarları tarafından belirlenen ve araştırmanın ana konularını özetleyen anahtar kelimelerde arama yapmak için kullanılır. SO etiketi, belirli bir dergi veya kaynağı hedefler. PY etiketi, belirli bir yayın yılına göre arama yapmayı mümkün kılar. AD etiketi, yazarın bağlı bulunduğu kurumu ya da adres bilgisini tarar.

Scopus veri tabanında "Advance Search" bölümünde tarama yapılırken AUTHOR etiketi belirli bir yazarın yayınlarını, TITLE, ABS ve KEY etiketleri sırasıyla başlık, özet ve anahtar kelimelerdeki terimleri arar. TITLE-ABS-KEY etiketi başlık-özet-anahtar kelime alanlarını birlikte tarar. AFFIL etiketi kurum bazlı aramalar yaparken, PUBYEAR etiketi yayımlanma yılına göre filtreleme sağlar. Ayrıca, DOCTYPE etiketi doküman türünü, SOURCE belirli bir dergiyi hedefler.

Araştırmamızdaki belgelerin gerçekten araştırma alanını temsil ettiğinden emin olmamız gerektiği unutulmamalıdır. Bunun için tarama sonuçlarında elde edilen araştırmalar incelenmelidir. Araştırmaların özet bölümleri okunarak araştırma alanını temsil etmeyen araştırmalar veri setinden çıkarılmalıdır. İncelemeler sonucunda elde edilen nihai veri seti uygun format seçilerek kaydedilir. Farklı veri tabanlarından indirilen veri setlerinin tek bir dosyada birleştirilmesi gereklidir. Ortak olan araştırmalar bir defa yazılmalıdır.

3.Veri analizi ve yorumlama

Araştırma sorularına göre veri analizinde kullanılanacak yöntemler belirlenir. Yöntemler performans analizi ve haritalandırma olarak iki gruba ayrılmaktadır. Performans analizi toplam yayınlar, yazar katkıları ve atıfla ilgili göstergeler gibi

ölçütleri kullanarak yazarların, kurumların ve ülkelerin konu alanına etkisini değerlendirir. Yayınlar ile ilgili ölçümlerin başında toplam yayın sayısı gelir. Toplam yayın sayısı bir yazarın, kurumun veya ülkenin bilimsel üretkenliğini değerlendirmede önemli bir ölçüt olarak öne çıkar. Toplam yayın sayısı, bir yazar kurum veya ülke tarafından üretilen toplam yayın sayısıdır. Tek yazarlı yayınlar, bir yazar tarafından yazılan ve bireysel araştırma katkılarını gösteren yayınlardır. Ortak yazarlı yayınlar birden fazla yazar tarafından yazılan yayınlardır ve iş birliğini yansıtır. Yayın yapılan aktif yıl sayısı bir yazarın veya kurumun aktif olarak yayın yaptığı yılları belirtir. Aktif yılı başına üretkenlik aktif yıl başına üretilen ortalama yayın sayısıdır. Katkıda bulunan yazar sayısı bir araştırma grubuna katkıda bulunan benzersiz yazarların toplam sayısını ifade eder. Atıflar ile ilgili ölçümler de ise toplam atıf sayısı incelenerek başlanır. Toplam atıf sayısı bir yazarın veya kurumun tüm yayınlarına aldığı toplam atıf sayısıdır. Ortalama atıf sayısı yayın başına ortalama atıf sayısıdır ve yayının konu alana etkisini gösterir. Yayınları ve atıfları ölçen bileşik ölçümler araştırma sonuçlarının özelliklerine ve literatürde kabulüne ilişkin kapsamlı bilgiler verir. Atıf yapılan yayın başına atıf sayısı, atıf yapılan yayın başına ortalama atıf sayısını gösterir. Atıf yapılan yayın sayısı bir yazarın veya kurumun en az bir kez atıf yapılan yayınlarının toplam sayısıdır. Atıf alan yayınların oranı toplam yayın sayısı içerisinde atıf alan yayınların oranını beliritir.

Yayın ve atıf indeksleri, bireysel veya kurumsal akademik performansı ölçmekte etkili araçlardır fakat yazarlar arasındaki iş birliğinin kapsamını değerlendirmek için farklı ölçütlere ihtiyaç duyulmaktadır. Bu bağlamda iş birliği indeksleri önemli bir gösterge olarak öne çıkar. İşbirliği katsayısı araştırmanın iş birliği derecesini

gösterir. g-indeksi, bir yazarın en çok atıf alan "g" sayıda makalesinin, toplamda en az g^2 atıf aldığı durumlarda g sayısını temsil eder. Örneğin, bir yazarın ilk 4 makalesi sırasıyla 16, 9, 4 ve 1 atıf aldıysa, g = 3 olacaktır. Çünkü ilk üç makale toplam 29 atıfa sahiptir (en az $3^2 = 9$). h-indeksi, bir yazarın en az h sayıda makalesinin her birinin en az h atıf aldığı durumlarda h sayısını temsil eder. Örneğin bir yazarın 5 yayını en az 5 atıf aldıysa bu yazarın h-indeksi 5'tir. h-indeksi hem yayın sayısını hem de bu yayınların aldığı atıf sayısını dikkate alır. m-indeksi yazarın h-indeks değerinin, yazarın yayınladığı ilk yayın yılından itibaren geçen yıl süresine oranıdır. g, h ve m-indeksleri dışında i-indeksi de kullanılmaktadır. i-10, i-100 ve i-200 olmak üzere üç çeşit i indeksi bulunur.i-10 en az 10, i-100 en az 100 ve i-200 en az 200 atıf alan yayınların sayısıdır. Örneğin, bir yazarın i-indeksi 10 ise, en az 10 makalesi 10 veya daha fazla atıf almıştır. Bu indeks, özellikle daha yeni yayınların etkisini vurgulamak için kullanılır ve genellikle diğer ölçümlerle birlikte değerlendirilir.

Haritalandırma araştırmanın bileşenleri arasındaki entelektüel etkileşimleri ve yapısal bağlantıları ortaya çıkarır. Haritalandırma yöntemleri arasında atıf analizi, ortak atıf analizi, bibliyografik eşleştirme, ortak kelime analizi ve ortak/eş yazarlık analizi yer alır. Bu tür yöntemler, ağ analiziyle birleştirildiğinde, araştırma alanının bibliyometrik yapısını ve entelektüel yapısını sunmada etkilidir (Baker vd., 2020; Donthu vd., 2021; Tunger ve Eulerich, 2018).

3.1. Atıf analizi (Citation)

Atıf analizi, yazarlar veya dergiler arasındaki ilişkileri belirleyen bir yöntemdir. Atıf analizinde, atıf ile referans kavramlarını karıştırmamak gereklidir. Bir atıf, yazarın literatürde

yer alan araştırmadan bahsetmesidir. Referans ise bir araştırmada kullanılan tüm kaynakların detaylı biçimde listelendiği bölümdür. Atıf ve referans kavramları birlikte çalışır. Atıflar, metinde belirli bir bilgiyi işaret ederken referanslar, bu bilgilerin kaynağını tam olarak tanımlar.Yazarlar kendileri ile benzer konulardaki literatüre atıf yapar. Bu nedenle atıf yapan ve yapılan araştırmalar arasında bir ilişki olduğu düşünülür. Ancak, bu ilişki atıfta bulunulan belgede verilen diğer referansları da içermektedir. Çünkü bu referanslar üçüncü bir yayının dayandığı araştırmanın temelini oluşturmaktır. Sonuç olarak, bir atıf analizi, bir referans veya atıfla bağlantılı yayınlar arasındaki bu karmaşık ilişkileri inceler. Bir yazar, araştırma veya derginin aldığı atıf sayısı ne kadar fazlaysa, alanında bilinirliği o kadar fazla olur. Ayrıca, yazarların, araştırmaların veya dergilerin bilimsel topluluklara etkisi atıf analizi yoluyla belirlenebilir. Fakat bir araştırmanın kalitesi yalnızca atıf sayısına bağlı değildir. Bu bulgu diğer analiz sonuçlarıyla desteklenmelidir.

Atıf analizleri, bilimsel literatürün hangi yönlerde büyüdüğünü ve nasıl geliştiğini farklı düzeylerde anlamaya yardımcı olur. Mikro düzey analiz, bireysel araştırmalar veya yazarlar arasındaki bağlantılara odaklanır. Bu, bir yazarın alana katkısını veya bir araştırmanın diğer araştırmalar üzerindeki etkisini incelemek için kullanılır. Örneğin, bir makalenin kaç kez atıf aldığı, bireysel bir araştırmanın etki düzeyini ortaya koyabilir.

Makro düzey analiz ise, daha geniş ölçekli bir perspektif sunar ve alanın genel dinamiklerini anlamaya çalışır. Burada, belirli bir disiplindeki eğilimler, alanların zaman içindeki büyümesi veya farklı disiplinler arasındaki etkileşimler incelenir. Bu düzeydeki analizler, alanın genel etkisini veya uluslararası düzeydeki iş birliği kalıplarını anlamak için kullanılır.

Mezzo düzey analiz, mikro ve makro düzey arasında bir yere konumlanır. Bu düzeydeki analizler, araştırma grupları, kurumlar veya ülkeler arasındaki bağlantıları inceler. Örneğin, bir üniversitenin alandaki etkinliği veya iki ülke arasındaki bilimsel iş birliği bu kapsamda değerlendirilebilir. Bu üç düzeyde yapılan atıf analizleri, alanın büyümesini ve bilgi yayılımını detaylı bir şekilde anlamaya olanak sağlar Aynı zamanda politikaların şekillendirilmesi, finansman kaynaklarının tahsisi ve araştırma önceliklerinin belirlenmesi gibi karar süreçlerine de katkıda bulunur.

Atıf analizlerinde kendi kendine atıf yapan araştırmalara dikkat edilmelidir. Kendi kendine atıf, yazarların önceki araştırmalarına atıfta bulunmasıdır. Aksnes'e (2003) göre atıfta bulunan makalenin yazarları ile atıf yapılan makalenin yazarlarının en az bir ortak yazarı paylaştığı durumlara denir. Bu noktada atıflar dikkatli incelenmelidir. Çünkü aynı ada sahip farklı kişilerbu analizin güvenilirliğini etkileyebilir ve kendi kendine atıf sayılarının hatalı hesaplanmasına neden olabilir. Bunun dışında yazar adlarındaki yazım farklılıkları veya yanlış yazımlar, kendi kendine atıfların belirlenemesine neden olabilir (Glänzel vd., 2004).

3.2. Ortak atıf analizi (Co-citation)

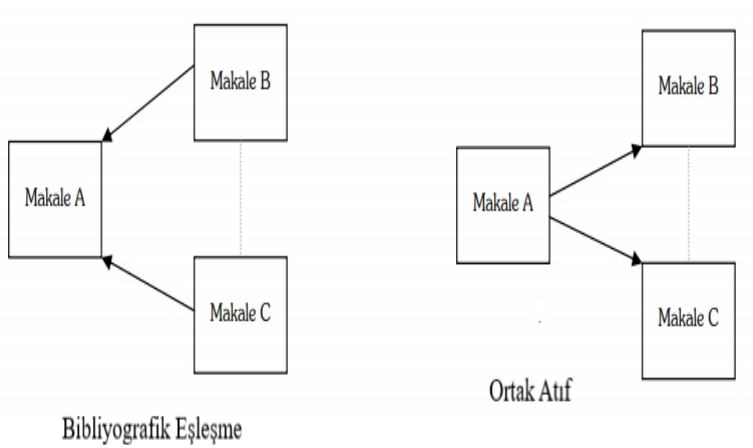
Ortak atıf analizi, iki yazara aynı belgede atıfta bulunulmasını ifade eder. Bir başka ifadeyle iki araştırma başka bir araştırmanın referans listesinde birlikte yer alıyorsa bu iki araştırma birbiriyle bağlantılıdır. Aynı araştırmada iki referansın birlikte yer alması, bu referansların sadece araştırmayla değil, birbiriyle de bir ortak yanı olduğunu gösterir. Ortak atıfların içerik açısından yakınlığın bir göstergesidir. Ortal atıf analizinde her atıfın birinci yazarı dikkate alınır. Diğer ortak yazarların katkılarını göz önünde bulundurmaz.

Ortak atıf analizinde farklı istatistiksel yöntemlere kullanılabilir. Kümeleme analizi, faktör analizi ve çok boyutlu ölçekleme en sık kullanılan yöntemlerdir. Bu yöntemler, yazarların en sıklıkla birlikte atıf aldığını gösteren sıklık sayısını kullanarak yazar ilişkilerini belirler. Kümeleme analizinde en sık ortak atıf yapılan yazar kümeleri veya grupları belirlenir. Her bir yazar belirli bir kümede yer alır. Aynı yazar birden fazla kümeye dahil edilemez. Bu sorunu çözmek için, bir yazarın birden fazla faktöre ait olmasına olanak tanıyan faktör analizi yöntemi kullanılabilir. Faktör analizinde ilişkili olan yazar kümelerini/gruplarını belirlemek için korelasyon değerleri incelenir. Gruplar, faktörler olarak adlandırılır. Her bir faktör, ortak bir temayı veya araştırma alanını temsil eder. Her bir yazarın belirli bir faktörle ne derece ilişkili olduğunu gösteren faktör yükleri hesaplanır. Yazarın en çok yüklendiği faktör, en çok ilişkili olduğu temayı ifade eder. Eğer yazar birden fazla faktörde yüksek yüke sahipse, araştırmaları birden fazla temayla bağlantılıdır. Çok boyutlu ölçeklemede (ÇBÖ) ise yazarlar veya makaleler arasındaki benzerlikler bir uzaklık matrisi ile ölçülür. Yakınlık-uzaklık matrisi ilişkileri iki veya üç boyutlu bir düzleme indirger. ÇBÖ, sık atıf alan yazarları haritada birbirine yakın konumlandırarak ilişkileri görselleştirme imkanı sunar. Görselleştirme, karmaşık ilişkileri daha sade ve anlaşılır bir şekle dönüştürür. Benzer temalar üzerinde çalışan yazar gruplarını veya disiplinler arası farkları keşfetmeye olanak tanır. Bu nedenle, ÇBÖ özellikle ortak atıf analizinde gizil ilişkileri ortaya çıkarmada güçlü bir keşifsel yöntemdir.

3.3. Bibliyografik eşleştirme (Bibliographic coupling)

Farklı iki kaynakta aynı yayına atıf yapılması bibliyografik eşleştirmedir. İki araştırma, aynı yayınları ne kadar çok alıntılarsa,

kaynakçaları o kadar çok benzeşir ve ilişkileri o kadar güçlüdür (Öztürk ve Güler, 2022). Bibliyografik eşleştirme ortak atıf analizinin tam tersidir. Ortak atıf analizinde A makalesi B ve C makalelerine atıf yapar. Bibliyografik eşleştirme de ise B ve C makaleleri A makalesine atıf yapar.



Şekil 1: Bibliyografik eşleşme ve ortak atıf analizi

Kaynak: Garfield, 1988

3.4. Ortak kelime analizi (Co-occurrence/Co-word)

Ortak kelime analizinin ilk çıkış noktası, bir konunun kavramsal alanını temsil eden bir temalar ağı ve bunların ilişkilerinin belirlenmesidir (Zupic ve Cater, 2015). Bir başka ifade ile ortak kelime analizi, anahtar kelimelerin birlikte kullanılma sıklığına bağlı olarak konunun kavramsal yapısı ve alt konuları belirlemek için kullanılır. Bir konunun kavramsal yapısı konu alanını temsil ederken alt konu alanları arasındaki ilişkiler haritalandırılabilir (Cho, 2014). Böylece konun gizil yapıları ortaya çıkarılır.

Ortak kelime analizinde birimler kullanılan anahatar kelimeler olur. Anahtar kelimeler aracılığı ile araştırmanın içeriği incelenir. Ortak kelime analizinin iki varsayımı bulunur. İlk varsayımı, bir araştırmada kullanılan anahtar kelimeler, araştırmanın anlamını ve ilişkili olduğu kavramları yansıttığıdır. Yani anahtar kelimeler 'makroskobik düzeylerde içerik tanımlayıcıları' olarak kabul edilir (Chen vd., 2008). İkinci varsayım ise farklı araştırmalarda iki kelimenin birlikte kullanılmasının, iki araştırma arasında bir ilişki olduğunu göstermesidir (Feng vd., 2017). Analizde anahtar kelimelerin dışında başlık, özet veya tam metinlerde en sık kullanılan kelimelerde birbirine bağlanabilmektedir.

3.4. Ortak/eş yazarlık analizi (Co-authorship)

Ortak yazarlık analizi, bir konudaki yazarlar arasındaki işbirliklerini inceler. Yazarlar arasındaki sosyal ağlar oluşturulabilmektedir. Buna ek olarak bibliyografik veriler yazarların kurumsal bağlantıları ve ülkeleri hakkında bilgiler içermektedir. Ortak yazar analizi kurumlar ve ülkeler arasındaki iş birliğini inceleme fırsatı sunar (Zupic ve Cater, 2015). Araştırma konusunda öne çıkan bilim insanlarıyla iş birliği yapmak isteyen yeni bilim insanlarına değerli bilgiler sunmak açısından önemli bir analizdir.

Bibliyometrik analiz yöntemleri, bilimsel üretkenlik ve etkileşimlerin yapısal özelliklerini anlamamıza olanak sağlarken, bu analizlerin temellerini oluşturan bibliyometrik yasalar, bilimsel iletişimin genel dinamiklerini kavramamıza rehberlik eder. Bibliyometrik yasalar, bilimsel iletişimde yayın dağılımı, yazar üretkenliği ve atıf yoğunluğu gibi unsurların genel kalıplarını

açıklayan matematiksel ve istatistiksel prensipler sunar. Bu nedenle, analiz yöntemlerinin ortaya koyduğu bulgular, bibliyometrik yasalarla desteklendiğinde daha kapsamlı ve anlamlı hale gelir.

3.5. Brandford yasası

Bibliyometrik analizde en üretken dergiler hakkında tanımlayıcı analizler yapılabilmektedir. Bu bağlamda, belirli bir alandaki dergi üretkenliğini anlamak için Bradford Yasası kullanılmaktadır. Bradford Yasası, 1934 yılında Samuel C. Bradford tarafından bilimsel literatürün dergilerdeki dağılımını incelemek amacıyla formüle edilmiştir. Bradford 1931 ile 1933 yılları arasında yer bilimi alanında gerçekleştirilmiş ve bu alanda bulunan tüm makaleler toplanmış ve bu makalelerin yayımlandığı dergileri analiz etmiştir. Analiz sonuçlarına göre belirli bir düzen olduğu belirlenmiştir: Bir konu alanında yayımlanan makale sayısı ile bu makalelerin yayımlandığı dergi sayısı arasında ters bir ilişki olduğu gözlenmiştir. Belirli bir konu alanında az sayıda dergi toplam yayınların büyük bir kısmını oluştururken, daha fazla sayıda dergi ise o alanla ilgili daha az sayıda makale yayımlamaktadır. Başka bir ifade ile üretilen toplam makale sayısı, dergilerin azalan üretkenlik sırasına göre düzenlenmesi durumunda, dergilerin sayısının logaritması ile orantılıdır.

Yasa uygulanırken dergiler, yayımladıkları makale sayısına göre sıralanır. Dergiler her grup eşit sayıda makale içericek şekilde üç veya daha fazla gruba (zone) ayrılır. Bradford Yasası, dergileri üretkenliklerine göre sıralayarak her bir gruptaki dergi sayısının artarken makale sayısının sabit kaldığını varsayar. Makaleleri yayımlayan dergi sayısı her bölgede aynı olmayacaktır; çünkü bazı dergiler diğerlerinden daha üretkendir. Gruplar, Bradford bölgeleri

olarak adlandırılır. İlk bölge, "çekirdek dergiler" olarak adlandırılan en üretken dergileri içerir. Çekirdek dergiler belirli bir konudaki makalelerin önemli bir bölümünü yayımlar (Tonta ve Al, 2007). Üçüncü ya da en son bölgede ise konuyla ilgili daha az üretken olan dergiler bulunur (Yazıcı, 2023). Çekirdek ve ardışık bölgelerdeki dergi sayıları $1:k:k^2$ oranını takip eder. Bradford sabiti/çarkanı/katsayısı k, en üretken derginin yayımladığı makale sayısı (maksimum makale sayısı) y_m, bölge sayısı p ve Euler sayısı e^{γ} (γ = 0.5772) olmak üzere aşağdaki formül ile hesaplanır (Egghe ve Rousseau, 1990);

$$k = (e^{\gamma}.y_m)^{1/p}$$

Bradford katsayısı elde edildikten sonra ilk bölgede yer alması beklenen dergisi sayısı bulunur. İlk gruptaki dergi sayısı r_o, toplam dergi sayısı T olmak üzere aşağıdaki gibidir;

$$r_o = \frac{T.(k-1)}{k^p - 1}$$

Bölgeler boyunca bulunması beklenen dergi sayıları r_p aşağıdaki gibi bulunur;

$$r_1 = r_o * k$$
$$r_2 = r_o * k^2$$
$$r_p = r_o * k^{p-1}$$

Bradford Yasası'nın beklenen dağılımı ile gerçek verilerden elde edilen gözlenen dağılımın ne derece uyumlu olduğunu test etmek mümkündür. Bölgelerdeki beklenen dergi sayısı ile gözlenen dergi sayıları karşılaştırılır. Hata yüzdesi aşağıdaki gibi hesaplanır (Kumar ve Mohindra, 2015);

$$e = \frac{|T - (r_o + r_1 + \cdots + r_P|}{T} * 100$$

3.6. Price yasası

Derek J. de Solla Price, bilimle ilgili özel bir üssel büyüme yasasını formüle eden ilk bilim insanı olmuştur. Bu yasa Price Yasası olarak bilinmektedir. Öncü yazar sayısının toplam yazar sayısına göre oldukça az olduğu belirtilmiştir. Yasaya göre bir konudaki araştırmaların yarısı toplam yazar sayısının karekökü kadar yazar tarafından yazılmaktadır. Buna ek olarak toplam yazar sayısının karekökü tane dergi makale sayısının yarısı kadar araştırma içerir (Senqupta, 1992). Örneğin toplam 500 makale var ise ve toplam 100 yazarın oluşturduğu varsayılırsa, Price Yasasına göre, 250 makale (500/2= 250), 100 yazarın karekökü olan ($\sqrt{100}$ = 10), 10 yazar tarafından yazılmalıdır.

3.7. Lotka yasası

Lotka (1926), bilimsel üretkenliğin frekans dağılımını analiz etmiş ve bu araştırması, Lotka kanununun geliştirilmesine yol açmıştır. Yazar üretkenliğini inceler. Yazarların konu alanına katkıda bulunduğu araştırma sayısı n ise, bir araştırma ile katkıda bulunan yazar sayısının n ile katkıda bulunan yazar sayısına oranı $1/n^2$'dir. Örneğin konu alanına bir yayın ile katkısı olan yazar sayısı iki yayın ile katkı sağlayan yazarlara oranı $1/2^2 = 1/4$ 'tür. Bir başka ifade ile Lotka Yasası bir alanda, yazarların %60'ının bir yayın ile, %15'inin 2 yayın ile, %7'sinin 3 yayın ile katkıda bulunduğunu öngörürmektedir. (Rowlands, 2005).

3.9. Pareto (80/20 kuralı) yasası

Pareto yasası, İtalyan ekonomist Vilfredo Pareto tarafından 1896 yılında geliştirilmiştir. Başlangıçta gelir dağılımındaki eşitsizlikleri açıklamak için kullanılan bu yasa, genellikle "80/20

kuralı" olarak bilinir. Yayın ve atıf dağılımlarını anlamak için kullanılmaktadır. Pareto, kaynakların (örneğin, yayınların) küçük bir yüzdesinin (genellikle %20) toplam çıktının (örneğin, atıfların) büyük bir kısmını (%80) ürettiğini öne sürmüştür. Bibliometrik analizde bu, belirli bir alanda az sayıda makalenin büyük ölçüde atıf aldığı, geri kalan çoğunluğun ise nispeten az atıf aldığı anlamına gelir. Pareto dağılımı, bilimsel üretkenliğin ve eşit olmayan etki düzeyi dağılımını açıklamada kullanılarak, hangi yayınların veya yazarların bir alandaki bilgi üretimine en fazla katkıda bulunduğunu ortaya koymada önemli yasadır.

3.10. Zipf yasası

1932 yılında George Kingsley Zipf dilbilimsel bir teoriye dayanak oluşturacak kalıpları bulmak amacıyla kelime kullanımına dair büyük miktarda veri analiz etmiştir. Bunun sonucunda Zipf yasasını geliştirmiştir. Zipf yasası, metin içerisinde yer alan kelimeler kullanım sıklığına göre sıralandığında (f), kelimelerin sıklıkları sıra numaralarıyla (r) çarpılırsa yaklaşık olarak sabit bir değer elde edilir ($r \times f = C$).

Bibliyometrik analizi gerçekleştirmek için çeşitli yazılımlar kullanılmaktadır. BibExcel, Pajek, Gephi, SciMat, Sci2 ve UCINET programları bulunsa da CiteSpace, R programlama dili ve VOSviewer en çok kullanılan programlardır. CiteSpace (Chen, 2006), performans analizi ve haritalandırma için tasarlanmış, ücretsiz bir Java uygulamasıdır. Kullanıcılar, veritabanlarından elde edilen bibliyografik verileri analiz ederek, atıf ağlarını, co-citation (birlikte atıf) ilişkilerini ve anahtar kelime ağlarını görselleştirebilir. CiteSpace, zaman dilimlerine dayalı görselleştirmeler yaparak, belirli bir alanın gelişim dinamiklerini ve araştırma öncülerini ortaya

koyar. Ayrıca, küme analizi, patlama algılama (burst detection) ve kavramsal yapı analizi gibi özellikleri sayesinde bilimsel iletişim ağlarını detaylı bir şekilde analiz etme imkânı sunar.

R programlama dilinde Bibliometrix (Aria ve Cuccurullo, 2017) paketi ile bibliyometrik analizler gerçekleştirilmektedir. Bibliometrix paketi kullanıcılara Shiny uygulaması üzerinden, WoS, Scopus ve Dimensions veritabanlarından elde edilen bibliyografik verileri analiz etme ve görselleştirme sağlar. Plain text ve BibTeX dosya formatları kullanılabilmektedir. Bibliometrix paketi yalnızca haritalandırmaya odaklanmaz bunun yanı sıra sonuçların doğruluğu da önem verir. Kavramsal yapıyı incelemek için Faktör Analizi (FA) kullanan bir yaklaşımda önermektedir.

VOSviewer, 2009 yılında, Leiden Üniversitesi'nden Nees Jan van Eck ve Ludo Waltman tarafından geliştirilen ücretsiz bir programdır. Tüm işletim sistemlerinde çalışabilmektedir. Bibliografik veri, ağ verisi ve metin verisi ile analizler gerçekleştirmektedir. Fakat verilerin düzenlenmesi için menüler içermez.

Her programın olumlu ve olumsuz yönleri vardır. Örneğin, Pajek ve UCINET Gephi ve R programlama diline göre daha yavaştır. Oluşturulan haritaların esnekliği bir diğer olumsuz yanıdır. BibExcel eş-atıf ağları oluşturmak için kullanılabilmektedir. Fakat BibExcel grafik oluşturmada kullanılamaz. VOSviewer programı ile oluşturulan bir haritada, aynı kelimelerin farklı biçimleri birleştirilemezken, Gephi programında bu kelimeler birleştirilebilir. Bu gibi sorunları çözmek için programlar birlikte kullanmaktır. Örneğin, VOSviewer yalnızca haritaları oluşturmak için kullanılırken R programlama dilinde önerilen farklı analizler (faktör

analizi) gerçekleştiriliebilir. Hangi programların kullanılacağına programların özellikleri ve haritaların esnekliği ile ilgili hususlar dikkate alınarak karar verilmelidir (Donthu vd., 2021).

Analizler gerçekleştirildikten sonra elde edilen bulgular açık ve net bir şekilde ifade edilmelidir. Mevcut bilgi birikimi ortaya konmalıdır. Konunun araştırılmamış bölümleri vurgulanmalıdır. Araştırmacılara yeni yollar göstermelidir. Konunun gelişimi hakkında bilgi verildikten sonra ilgili alanyazın ve kuramlar ile bulgular desteklenebilir.

Araştırmanın bu bölümüne kadar bibliyometrik analiz kavramının ne olduğu, analizde kullanılan yöntemler, temel teorik yasalar ve analiz sürecinde dikkat edilmesi gereken noktalar detaylı bir şekilde ele alınmıştır. Bu bilgiler, bibliyometrik analizin hem teorik altyapısını hem de uygulama pratiğini kavramayı kolaylaştırmak adına sunulmuştur. Şimdi, bibliyometrik analizin daha somut ve anlaşılır hale gelebilmesi için, bu teorik bilgilerin ışığında örnek bir uygulama gerçekleştirilecektir. Bu uygulama, bibliyometrik analiz sürecinin aşamalarını ve elde edilebilecek sonuçları adım adım göstermek amacıyla hazırlanmıştır.

Örnek Bir Uygulama

Bibliyometrik analizin daha iyi anlaşılabilmesi amacıyla örnek bir uygulama gerçekleştirilmiştir. İlk olarak araştırmanın amacı belirlenmiştir. Bu araştırmanın amacı, matematik başarısı hakkında yapılan araştırmaların genel görünümünü haritalamak ve literatürdeki boşlukları belirlemektir. Bu amaç doğrultulsunda aşağıdaki araştırma sorularına yanıt aranmıştır:

Araştırma sorusu 1: Matematik başarısı konulu makalelerin yıllara göre durumu nedir?

Araştırma sorusu 2: Matematik başarısı konusunda en üretken dergiler hangileridir?

Araştırma sorusu 3: Matematik başarısı konusunda en üretken yazarlar kimlerdir?

Araştırma sorusu 4: Matematik başarısı konulu makalelerde sorumlu yazarların ülkelere göre dağılımı nasıldır?

Araştırma sorusu 5: En sık kullanılan yazar anahtar kelimeleri nelerdir?

Araştırmanın amacı ve araştırma soruları belirlendikten sonra örnek veri seti oluşturulmuştur. Örnek veri seti oluşturulurken Scopus ve Web of Science (WoS) veri tabanlarında tarama yapılmış ve Boolean mantığı kullanılmıştır. Tarama yapılırken Scopus veri tababında TITLE-ABS-KEY; WoS veri tabanında TS etiketi kullanılmıştır. "*math success*", "*math achievement*" , "*mathematic achievement*", "*mathematic success*", "*math performance*" ve "*mathematic performance*" anahtar kelimeleri kullanılmıştır. Anahtar kelimelerden herhangi birini içeren araştırmalara ulaşılmak hedeflendiği için kelimeler arasında OR operatörü kullnılmıştır. Tarama sonucunda Scopus veri tabanında 2793 ve WoS veri tabanında 2781 araştırmaya ulaşılmıştır. Elde edilen araştırmaların kapsamını sınırlandırmak için filitreleme yapılmıştır. Scopus veri tabanında Subject Area "Social Sciences" ve WoS veri tabanında Research Area "Education Educational Research" filitresi uygulanarak sadece eğitim alanındaki araştırmalara ulaşmak hedeflenmiştir. Her iki veri tabanında da Document Type "Article" filitresi uygulanmıştır. Böylece sadece makalelerden oluşan bir veri seti elde edilecektir. İki veri tabanından elde edilen makalelerin başlangıç yıllarının aynı olması için Scopus veri tabanında 1977 ve

2025 yılları arasında olacak şekilde yıl filitresi uygulanmıştır. Filitreleme sonucunda Scopus veri tabanında 1610 ve WoS veri tabanında 812 makaleye ulaşılmıştır (Tarama 5 Aralık 2024 tarihinde gerçekleştirilmiştir). Veri setlerinde ortak olan makaleler çıkarıldığında nihai veri seti 1806 makaleden oluşmaktadır.

Örnek uygulamada Veri analizleri R Studio (R Core Team, 2024) yazılımı "bibliometrix" paketi (Aria ve Cuccurullo, 2017) kullanılarak gerçekleştirilmiştir. Bu nedenle veri tabanları taranarak ulaşılan bibliyografik veriler BibTeX dosya formatı kullanılarak indirilmiştir. Scopus veri tabanında tarama sonuçlarının yer aldığı sayfada "All" kutucuğu işaretlenir. Seçilen veriler üç nokta işaretinde yer alan "Save to list" kısmından listeye kaydedilir. Daha sonra profilimizde bulunan "Saved lists" bölümden kaydedilen liste açılır. "Export" basılır ve açılan pencere BibTeX seçilir. Seçim yapıldıktan sonra indirilecek verilerde bulunması istenen bilgiler işaretlenir. Bu noktada "Include references" kutucuğunun işaretlendiğine dikkat edilmelidir. WoS veri tabanında tarama sonuçlarının yer aldığı sayfada "Export" basılır ve açılan pencere BibTeX seçilir. WoS veri tabanında ulaşılan araştırmalar beşyüzerli listeler halinde kaydedilebilmektedir. Bu nedenle beş yüzden fazla sonuç elde edildiğinde birden fazla dosya indirilir. İndirilecek verilerde bulunması istenen bilgiler "Record Content" başlığı altında yer alan "Custom selection" veya tüm bilgiler doğrudan indirilmek isteniyorsa "Full Record and Cited References" seçeneği işaretlenebilir. Farklı veri tabanları veya WoS veri tabanından birden fazla dosya indirildiğinde dosyaların birleştirilmesi gerekmektedir. Dosyaları birleştirmek için R Studio'da aşağıda yer alan kodlar ile gerçekleştirilir.

```
getwd()
setwd()

library(bibliometrix)
library(bibtex)


WOS1 = convert2df("wos1.bib", dbsource = "wos", format = "bibtex")

WOS2 = convert2df("wos2.bib", dbsource = "wos", format = "bibtex")

SCOPUS = convert2df("scopus.bib", dbsource = "scopus",format = "bibtex")

veri = mergeDbSources(WOS1, WOS2,SCOPUS ,remove.duplicated = TRUE)

write.xlsx(veri, file = "veri.xlsx")

biblioshiny()
```

Şekil 2: Bibliyografik eşleşme ve ortak atıf analizi

R Studio'da getwd() fonksiyonu, R çalışma ortamında yer alan mevcut çalışma dizinini (working directory) döndüren bir komuttur. Çalışma dizini, R'nin dosya okuma ve yazma işlemleri sırasında varsayılan olarak kullandığı dizindir. Örneğin, bir dosya açmaya veya bir veri setini kaydetmeye çalıştığınızda, işlemler bu dizinde gerçekleştirilir. R Studio'da analizlere başlamadan önce çalışma dizni kontrol edilmelidir. Veri setlerimizin yer aldığı klasörün çalışma dizni olarak ayarlanması için setwd() komutu kullanılır. Komutun içerisinde çalışma dizini olarak ayarlamak istediğiniz yol tırnak işareti içerisinde belirtilir. Yolları belirtirken, Windows kullanıcıları ters eğik çizgi \ yerine ileri eğik çizgi / veya \\ kullanmalıdır.

Çalışma dizini ayarlandıktan sonra indirilen veri setlerinin R Studio ile okutulması gerekmektedir. BibTeX formatındaki dosyaların işlenebilmesi için "bibliometrix" ve "bibtex" paketlerinin yüklenmesi önemlidir. Bu paketler daha önce yüklenmişse, yalnızca aktif edilerek kullanılabilir. "convert2df()" fonksiyonu, "bibliometrix" paketinde yer alan ve veri tabanlarından indirilen bibliyografik verileri analiz için uygun bir veri çerçevesine

dönüştürmek amacıyla kullanılan bir fonksiyondur. Bu fonksiyon, file, dbsource ve format olmak üzere üç temel argümanla çalışır. File, veri dosyasının yolu veya adını ifade ederken, dbsource veri dosyasının indirildiği kaynağı belirtir. Format ise dosyanın biçimini tanımlar. Uygun argümanlar belirtildikten sonra kod çalıştırıldığında, veri dosyaları R Studio'nun Environment bölümünde görünür hale gelir. Tüm veri dosyalarının başarıyla okutulmasının ardından, mergeDbSources() fonksiyonu kullanılarak bu dosyalar birleştirilebilir. Bu fonksiyonun ilk argümanında veri dosyalarının isimleri belirtilirken, remove.duplicated = T argümanı tekrar eden yayınların yalnızca bir defa yazılmasını sağlar. Veri setlerinin birleştirilmesiyle oluşturulan nihai veri seti write.xlsx () fonksiyonu kullanılarak çalışma dizinine excel dosya formatında yazdırılırç

biblioshiny() fonksiyonu, R'de bibliometrix paketinin bir parçasıdır ve kullanıcı dostu bir grafiksel arayüz sağlayarak bibliyometrik analizlerin interaktif bir şekilde gerçekleştirilmesini mümkün kılar. Bu fonksiyon, herhangi bir programlama bilgisi gerektirmeden, bibliyografik veri setlerini analiz etmenize ve sonuçları görselleştirmenize olanak tanır. biblioshiny() fonksiyonunu çalıştırdığınızda, varsayılan web tarayıcınızda bir arayüz açılır. Açılan web sayfasında data menüsünden veriler ara yüze yüklenir. "Import Raw Files" seçeneği, çeşitli veri tabanlarından indirilen ham bibliyografik dosyaların analiz için yüklenmesinde kullanılır. Belirli formatlardaki dosyaları destekler ve her veri tabanı için özel bir yapı sunar. Örneğin, WoS veri tabanından indirilen dosyalar genellikle .bib (BibTeX) formatında olur ve doğrudan analiz edilebilir. Scopus veri tabanı, hem .csv hem de .ris formatlarını desteklerken, PubMed için düz metin (.txt)

formatı tercih edilir. Dimensions, Lens.org ve CrossRef gibi diğer veri tabanları ise genellikle .csv veya .bib formatlarında dosya sağlar.

"Load bibliometrix files daha önce bibliometrix paketi kullanılarak işlenmiş ve çalışma dizinine yazdırılmış dosyaları yeniden yüklemek için kullanılan bir özelliktir. Bu seçenek, genellikle .RData formatında saklanan işlenmiş bibliyografik verileri hızlı bir şekilde analiz sürecine dahil ederek, sıfırdan veri yükleme ve işleme ihtiyacını ortadan kaldırır. Daha önce yapılan analizleri yeniden yükleme olanağı sunarak, zamandan tasarruf sağlar ve analizlere kaldığınız yerden devam etmenizi mümkün kılar.

"Use a sample collection" bibliyometrik analizlere başlamak için örnek bir veri seti kullanmanızı sağlayan bir özelliktir. Örnek veri seti, bibliometrik analizlerin nasıl yapıldığını öğrenmek ve arayüzün sunduğu özellikleri keşfetmek için standart bir kaynak sağlar. İçerisinde genellikle bilimsel yayınlara ait temel bibliyografik bilgiler bulunur ve analizler, veri temizleme, görselleştirme ve diğer bibliometrik yöntemlerin test edilmesi için uygundur. Bu özellik, kullanıcıların arayüzü keşfetmesini kolaylaştırırken, gerçek veri setlerine uygulamadan önce pratik yapma imkanı sunar. Özellikle yeni kullanıcılar veya hazır bir veri setine erişimi olmayanlar için idealdir.

Veri setleri yüklendikten sonra, Start butonuna basarak analiz sürecini başlatılır. Bu işlem, tüm analiz adımlarını tek bir seferde başlatmanızı sağlar ve işlem adımlarını hızlıca tamamlar. Eğer veriler veri tabanlarından indirilirken herhangibir filitreleme gerçekleştirilmediyse biblioshiny Filters (filitreler) menüsünden

filitreleme yapılabilir. Bu menü, kullanıcıların belirli kriterlere göre veri setini daraltmasına olanak tanır. Örneğin, yalnızca belirli bir tarih aralığındaki yayınları, belirli bir yazar veya dergide yer alan makaleleri, ya da belirli anahtar kelimeleri içeren araştırmaları seçmek mümkündür.

Overview (genel bakış) nenüsü, bibliyometrik analizlerin genel bir özetini sunarak kullanıcıya verilerin ve analizlerin genel yapısını gösterir. Main information (asıl bilgiler) alt menüsünde yer alan

*Timespan bibliyometrik analizlerde belirli bir zaman aralığını ifade eder ve analiz edilen verilerin hangi dönemi kapsadığını;

*Sources toplam kaynak (dergi, konferans, kitap, vb.) sayısını (her bir kaynak bir defa sayılır);

*Documents yayın sayısını;

*Annual Growth Rate (Yıllık Büyüme Oranı) bir yıl içerisindeki büyüme oranını;

*Documents Average Age (Belgelerin Ortalama Yaşı) bir veri setindeki yayınların yaş ortalamasını,

*Average Citations per Document (Belge Başına Ortalama Atıf Sayısı) her bir yayının aldığı ortalama atıf sayısını,

*References (Referanslar) veri setinde yer alan yayınların atıfta bulunduğu toplam kaynakça sayısını,

*Affilations (Kurumlar) veri setinde yer alan araştırmacıların bağlı olduğu kurum sayısını,

*Keywords Plus anahtar kelime sayısını,

*Authors' Keywords (Yazar Anahtar Kelimeleri) yazar anahtar kelime sayısını,

*Authors (yazarlar) farklı yazar sayısını;

*Authors of single-authored docs tek yazarlı yayınlardaki farklı yazar sayısı

*Single-authored doc tek yazarlı yayın sayısını;

*Co-Authors per Document (Belge Başına Ortak Yazar Sayısı), bibliyometrik analizlerde, yayın başına düşen ortalama yazar sayısını;

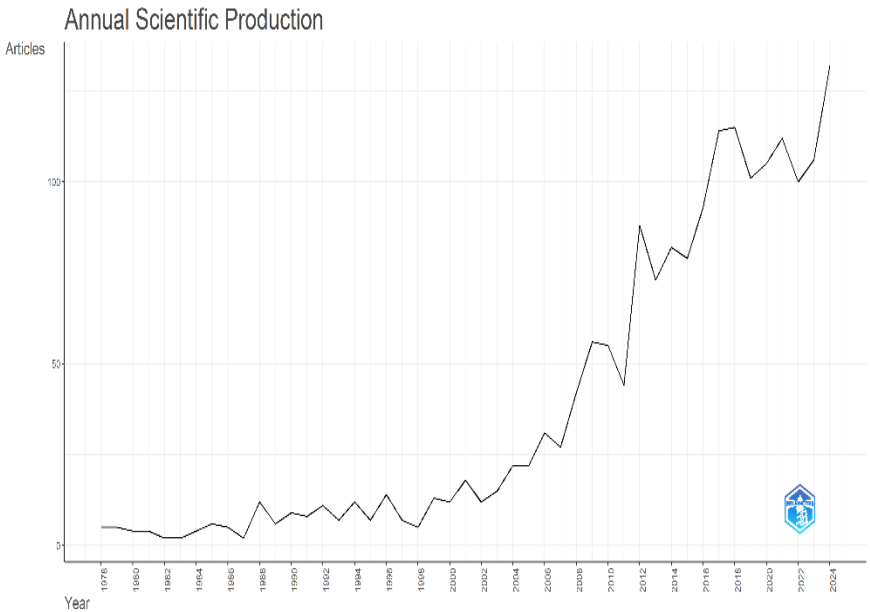
*International Co-authorship (Uluslararası Ortak Yazarlık), ortak yazarlık ilişkilerinin ne kadarının farklı ülkelerden gelen yazarlar arasında gerçekleştiğinin oranı gösterir.

Örnek veri setine ait betimse istatistikler Tablo 1'de yer almaktadır.

Tablo 1: Betimsel istatistikler

Description	Results
Main Information About Data	
Timespan	1978:2024
Sources (Journals)	524
Documents	1806
Annual Growth Rate %	7.38
Doc Average Age	10.1
Average citations per doc	39.46
References	15819
Affiliations	1763
Document Contents	
Keywords Plus (ID)	2353
Author's Keywords (DE)	3357
Authors	
Authors	4230
Authors of single-authored docs	204
Authors Collaboration	
Single-authored doc	248
Co-Authors per doc	3.15
International co-authorships %	6.977

Annual Scientific Production (yıllık bilimsel üretim) alt menüsünde her yıl yayımlanan yayın sayısını gösteren grafik yer almaktadır. Yıllık bazda yayın sayısındaki artış ya da azalış trendleri izlenebilir. Üretkenliğin zaman içinde nasıl değiştiğini gösterir. Artan bir üretkenlik eğilimi, araştırma alanına ilginin arttığını gösterirken, azalan bir üretkenlik eğilimi araştırma alanına ilginin azaldığını gösterir. Örnek veri setine ait yıllık bilimsel üretim grafiği Şekil 3'te yer almaktadır.

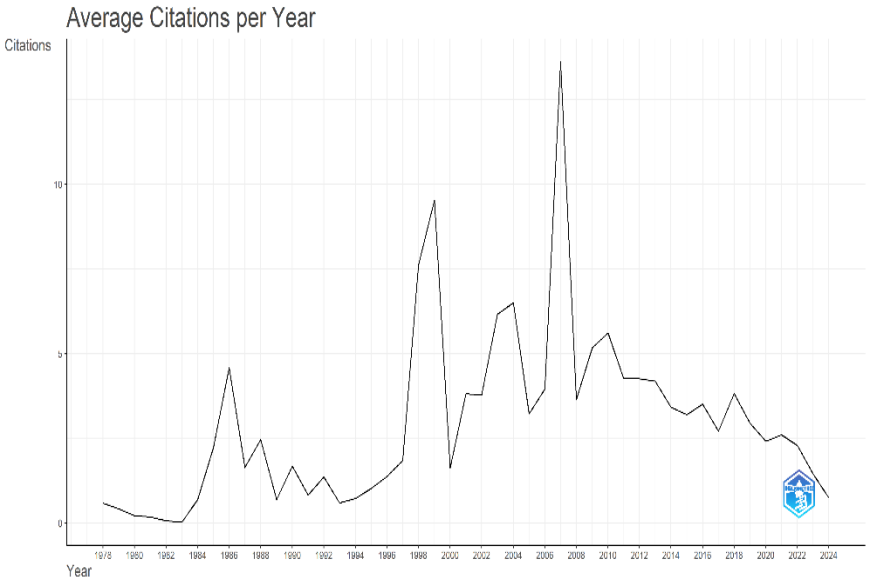


Şekil 3: Yıllık bilimsel üretim

Şekil 3 incelendiğinde 1978 ve 2005 yılları arasında yayımlanan makale sayısında genel olarak dalgalı bir artış gözlenmiştir. Yıllık yayımlanan makale sayısı 2 ile 22 arasında değişmektedir. 2005 yılından sonra makale sayısı istikrarlı bir şekilde artmaktadır. 2005 yılından sonra grafikte gözlemlenen keskin artışlar dikkat çekicidir. Keskin artışlar, her yıl bir önceki yıla kıyasla makale sayısında nispeten daha büyük bir artış olduğunu göstermektedir. En çok makale 2024 yılında yayımlanmıştır (n = 132). Makale sayısındaki artış konuya olan ilginin arttığını göstermektedir.

Average Citations Per Year (yıllık ortalama atıf sayısı) alt menüsünde yıllık bazda ortalama atıf sayısını gösteren grafik yer alır. Bu grafik ile yıllık atıf trendleri izlenebilir. Artan bir atıf eğilimi, yayınların etkisinin arttığını gösterirken, azalan bir atıf eğilimi

yayınların etkisinin azaldığını gösterir. Örnek veri setine ait yıllık ortalama atıf sayısı grafiği Şekil 4'te yer almaktadır.

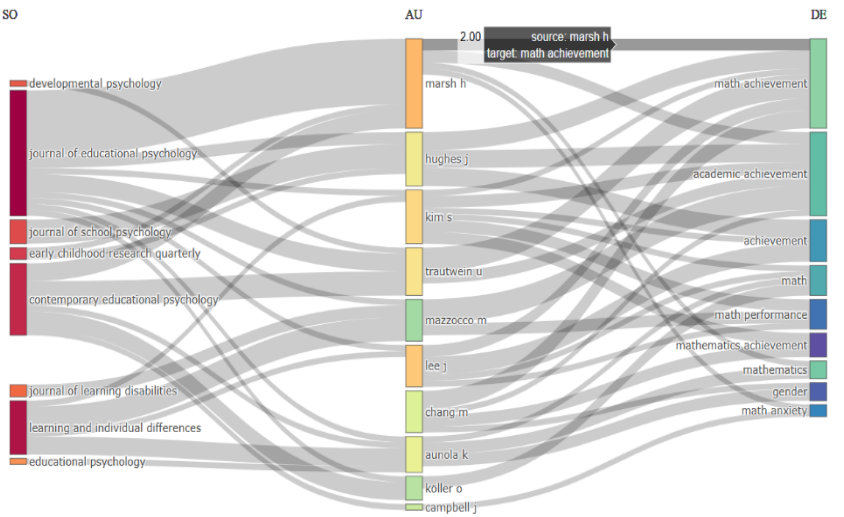


Şekil 4: Yıllık ortalama atıf sayısı

Şekil 4 incelendiğinde yıllık ortalama atıf sayıları 0.02 ile 13.62 arasında değişmektedir. 1978-1983 yılları arasında ortalama atıflar düşüktür. Bu dönemdeki makaleler fazla ilgi görmemiştir. 1998 yılında ortalama atıflarda önemli bir artış olmuş ve ilk zirve 1999 yılında gözlenmiştir. Bu dönemde yayımlanan makaleler akademik camiada daha fazla ilgi görmüş ve atıf almıştır. 2000-2006 yılları arasında ortalama atıflarda bir düşüş gözlenmiştir. İkinci zirve 2007 yılında gözlenmiştir. 1998 ve sonrasındaki makaleler alanda daha etkili olmuştur. En düşük ortalama atıf 1982, en yüksek ortalama atıf 2007 yılında gözlenmiştir.

Three-Field Plot (üç alan grafiği) alt menüsünde üç temel bileşen arasındaki ilişkiyi keşfetmek için kullanılan Sankey

diyagramıyer almaktadır. Temel bileşenler yazarlar, dergiler, ülkeler, anahtar kelimeler ve ya kurumlar seçilebilir. Üç alan grafiğinde, kutu boyutları sıklıkla ve düğümler arasındaki çizgilerin genişliği ise bağlantı sayısıyla orantılıdır (Koo, 2021). Örnek veri setine ait üç alan grafiği (yazar- anahtar kelime-dergi) Şekil 5'te yer almaktadır.

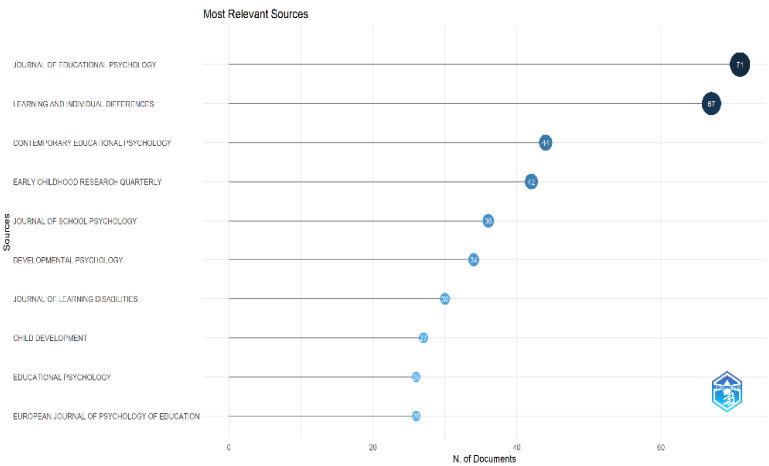


Şekil 5: Üç alan grafiği

Şekil 5'te dergi alanındaki en büyük kutu "Journal of Educational Psychology" dergisidir. "Comtemporary Educational Psychology" dergisi ikinci sırada yer almıştır. Etkili yazarların makaleleri bu dergilerde yayımlanmıştır. Marsh, H. ve Hughes, J., yazar alanındaki en büyük kutulardır. Anahtar kelime tarafında, en büyük kutunun en sık kullanılan anahtar kelimeleri içermektedir. *Math achievement*, *academic achievement* ve *achievement* anahtar kelimeleri en büyük kutularda yer almaktadır ve anahtar kelime alanında en fazla bağlantıya sahip kelimelerdir. Bu anahtar kelimeler

birkaç yazar tarafından (örn. Marsh, H. ve Hughes, J.) makalelerde kullanılmıştır. “Journal of Educational Psychology” dergisi en fazla bağlantıya sahip dergidir (n=8). Ardından “Comtemporary Educational Psychology” (n=5) ve “Learning and İndividual Differences” (n=4) gelmektedir. Kim, S. en fazla bağlantıya sahip yazardır (n=8). Ardından Marsh, H. ve Aunola, K. gelmektedir (n=7).

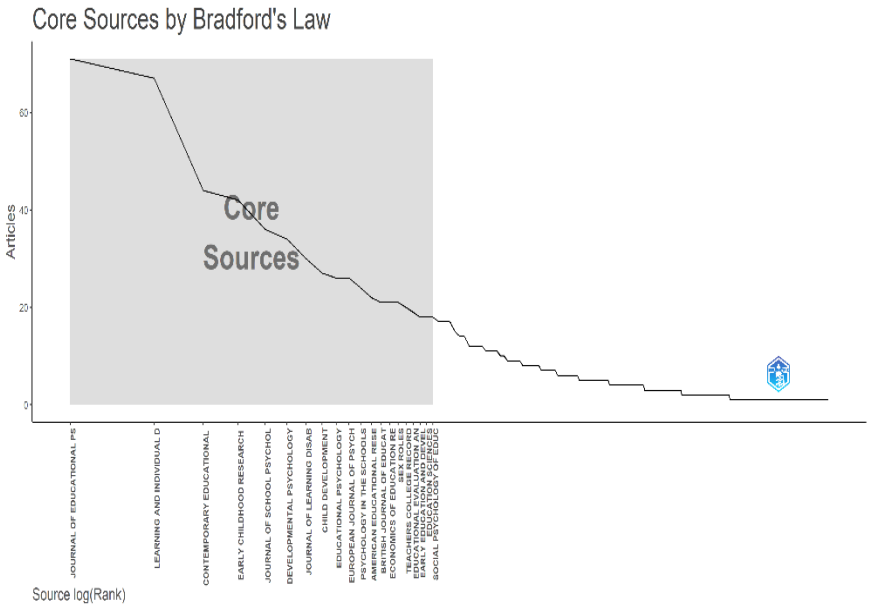
Sources (kaynaklar) menüsü, veri setindeki yayınların yayımlandığı dergilerin, konferansların ve diğer yayın kaynaklarının analizini yapmaya yönelik bir menüsüdür. Bu menü, bibliyometrik analizlerde kaynakların sayısını, üretkenliklerini, etkileşimlerini ve önemlerini anlamak için kullanılır. Most Relevant Sources (en ilgili kaynaklar) alt menüsünde konu hakkındaki yayınların en çok hangi kaynakta yayımlandığını gösteren grafik yer alır. Table (tablo) sekmesinden tüm kaynakların yayımladıkları yayın sayısına göre sıralamasına ulaşılabilir. Örnek veri setine ait grafik Şekil 6’da yer almaktadır.



Şekil 6: En ilgili kaynaklar

Araştırmada değerlendirilen 1806 makale arasında 524 farklı dergi yer almıştır (bkz. Tablo 1). İlk on dergi toplam 403 makale yayınlamış olup makalelerin %22.31'ini oluşturmaktadır. En çok makale (f=71, %3.93) "Journal Of Educational Psychology" dergisinde yayımlanmıştır. "Learning and İndividual Differences" dergisi ikinci sırada yer almıştır (f=67, %3.70). "Contemporary Educational Psychology" dergisi üçüncü sırada yer almış olup "Learning and İndividual Differences" dergisini takip etmiştir (f=44, %2.43). En az yirmi makale yayımlayan on altı dergi vardır.

Bradford's law (Bradford yasası) alt menüsünde Bradford eğrisi yer almaktadır. Bu eğri, Bradford Yasası'nın mantığını görsel olarak temsil eder ve kaynakların yayın sayılarına göre nasıl sınıflandırıldığını/sıralandığını anlamak için kullanılır. X (yatay) eksen dergi isimlerini; Y (dikey) eksen toplam yayın sayısını gösterir. Çekirdek kaynaklar solda başlar ve daha az katkı sağlayan kaynaklar sağa doğru sıralanır. Eğri, dik bir şekilde yükselirse bu, çekirdek kaynakların az sayıda olduğu fakat yoğun bir şekilde yayın yaptığı gösterir. Eğri yataylaşır ve yavaş bir artış gösterirse bu, çok sayıda kaynağın az miktarda yayın içerdiğini gösterir. Hangi kaynakların literatürde baskın olduğunu anlamaya yardımcı olur. Çekirdek kaynakları belirleyerek literatür taramalarını ve araştırma planlamalarını daha verimli hale getirir. Yazarlara, aratırmalarını hangi kaynaklarda yayımlamayı hedefleyeceklerini gösterebilir. Örnek veri setine ait Bradford eğrisi Şekil 7'de yer almaktadır.



Şekil 7: Bradford eğrisi

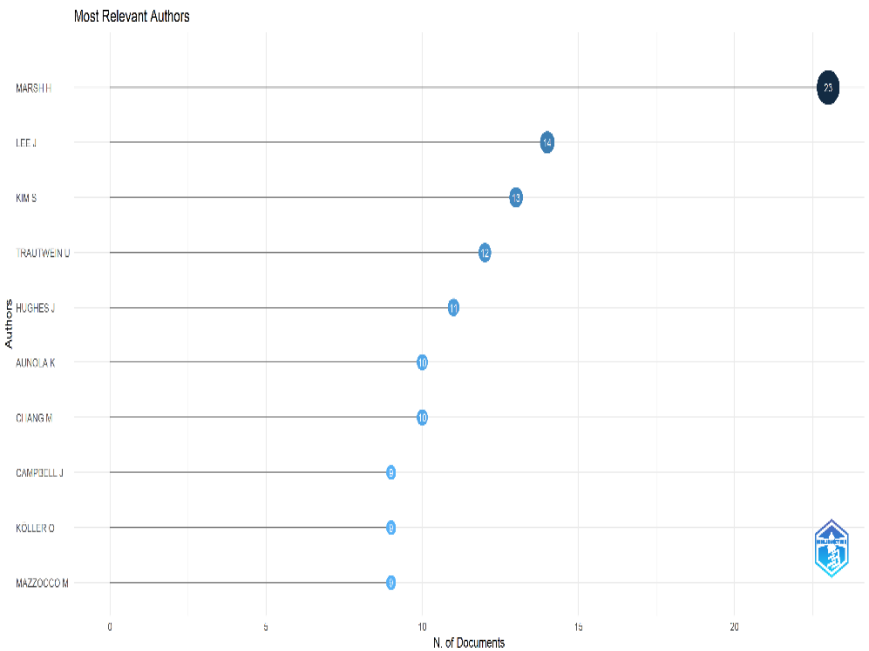
Şekil 7'de dergiler makale sayısına göre çoktan aza doğru sıralanmış ve üç bölgeye ayrılmıştır (p=3). Her bölge ideal olarak ortalama 602 (y_0) makaleden oluşmaktadır (31806 ÷ 3 = 602). En fazla makale sayısı (y_m=71) ile Bradford çarpanı (k) 5.01 olarak hesaplanmıştır. Birinci gruptaki çekirdek dergi sayısının 17 (r_0); ikinci grupta 84 (r_1); üçüncü, daha az üretken grupta ise 423 (r_2) olması beklenmektedir. Örnek veri setinde, birinci gruptaki çekirdek dergi sayısı 20 iken, ikinci grupta 88 dergi, üçüncü, daha az verimli grupta ise 416 dergi yer almaktadır. Bu bulgu, makalelerin dergiler arasındaki dağılımının Bradford Yasası ile uyumlu olmadığını göstermektedir. Birinci (çekirdek) ve ikinci bölgeler, Bradford Yasası tarafından öngörülen değerden büyüktür. Üçüncü bölge ise öngörülenden daha küçüktür.

Çekirdek bölgedeki dergi sayısının ardışık bölgelerdeki dergi sayısına oranı 17: 17 (5.01): $17(5.01)^2$ idi. Bu nedenle, üç bölgedeki beklenen toplam makale sayısı 524 olmuştur. Gözlenen ve beklenen toplam makale sayısı eşittir. Hata değeri sıfır bulunmuştur. Bu nedenle, her bölgedeki dergi sayısı için hata değerleri incelenmiştir. Birinci bölge için %15, ikinci bölge için %4.54 ve üçüncü bölge için %1 hata değerleri gözlenmiştir.

Sources' Local Impact (kaynakların yerel etkisi) alt menüsü, veri setindeki kaynakların etkisini anlamak ve analiz etmek için h-indeksi, g-indeksi, m-indeksi ve toplam atıf sayısı gibi ölçümleri kullanarak kaynakları sıralar. Bir kaynağın yerel düzeydeki (yani veri seti içindeki) etkisini değerlendirmeyi sağlar. Sources' Production Over Time (kaynakların zaman içindeki üretimi) alt menüsünde bir kaynağın yayımladığı yayınların zaman içindeki dağılımını yıllık bazda veya kümülatif olarak gösteren grafik yer alır.

Authors (yazarlar) menüsü, veri setindeki yayınların yazarlarının analizini yapmaya yönelik bir menüdür. Bu menü, bibliyometrik analizlerde yazarların sayısını, üretkenliklerini etkilerini, iş birliği ağlarını ve önemlerini anlamak için kullanılır. Most Relevant Authors (en ilgili yazarlar) alt menüsünde konu hakkında en çok yayını olan yazarların sırasını gösteren grafik yer alır. Table (tablo) sekmesinden tüm yazarların yayımladıkları yayın sayısına göre sıralamasına ulaşılabilir. Sıralama ölçütü yayın sayısı (N. of documents), yayın yüzdesi (percentage) veya paylaştırılmış katkı (fractionalized factory) olarak seçilebilir. Paylaştırılmış katkı her bir yazarın veya katkıda bulunan birimin, bir yayın üzerindeki katkısını toplam katkıya oranlayarak hesaplayan bir ölçüm yöntemidir. Eğer bir yayında n yazar varsa, her bir yazarın katkısı $1/n$ olacaktır. Örneğin, üç yazarlı bir makalede her bir yazarın

katkısı 1/3 olur. Yazarın birden fazla yayını var ise paylaştırılmış katkı bu yayınlardaki katkıların toplamına eşit olur. Bu ölçüt, bilimsel çıktıların paylaşımlı olarak değerlendirilmesine olanak tanır. Örnek veri setine ait grafik Şekil 8'de yer almaktadır.

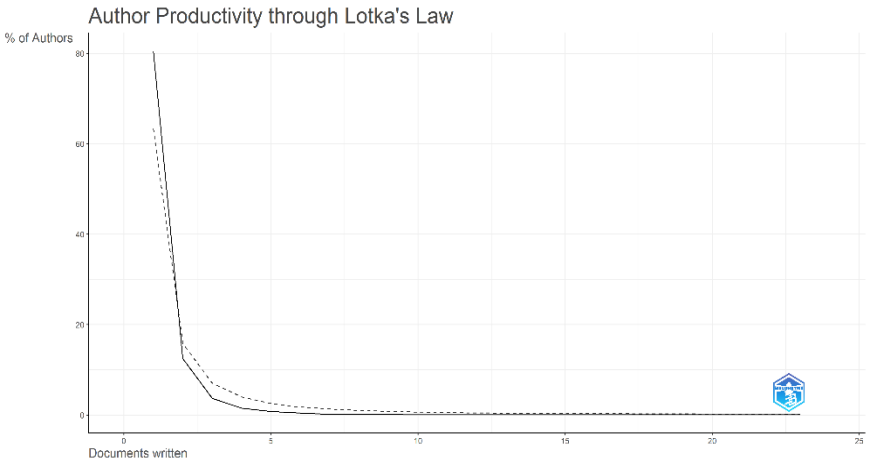


Şekil 8: En ilgili yazarlar

Araştırmada değerlendirilen 1806 makale arasında 4230 farklı yazar yer almıştır (bkz. Tablo 1). İlk on yazar toplam 120 makale yayımlamış olup makalelerin %6.64'ünü oluşturmaktadır. En çok makale (f=23, %3.93) Marsh, H. tarafından yayımlanmıştır. Lee, J. (f=14, %0.77) ikinci ve Kim, S. (f=13, %0.71) üçüncü sırada yer almıştır. En az on makale yayımlayan yedi yazar vardır.

Authors' Productivity Over Time (yazarların zaman içindeki üretkenliği) alt menüsünde bir yazarın yayımladığı yayınların zaman

içindeki dağılımını yıllık bazda gösteren grafik yer alır. Lotka's law (Lotka yasası) alt menüsünde Lotka eğrisi yer almaktadır. Bu eğri, Lotka Yasası'nın mantığını görsel olarak temsil eder ve yayın sayılarına göre yazar yüzdesini gösterir. X (yatay) eksen yayın sayısını; Y (dikey) eksen yazar yüzdesini gösterir. Örnek veri setine ait Lotka eğrisi Şekil 9'da yer almaktadır.

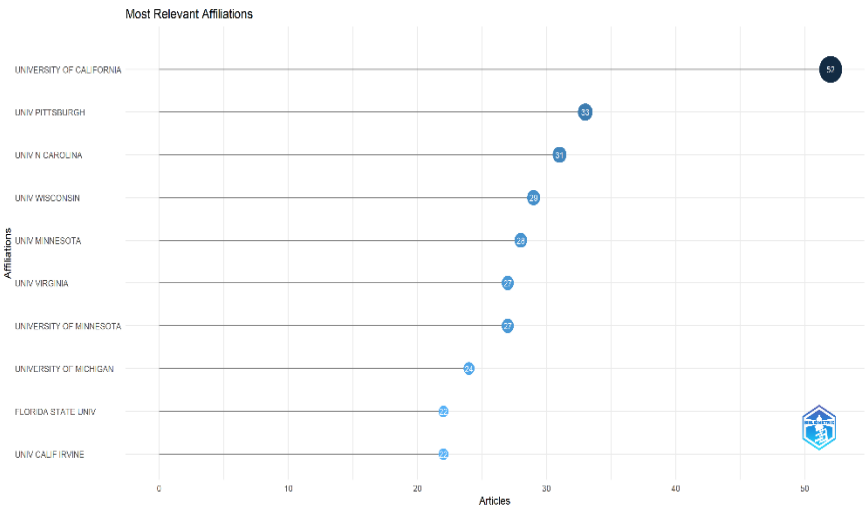


Şekil 9: Lotka eğrisi

Şekil 9 incelendiğinde 4230 farklı yazarın %80'inin tek makale yayımladığı görülmektedir (n=3407). Yazarların %12.5'i (n=528) iki ve %3.7'si (n=157) üç makale yayımlamıştır. İki makale yazan yazar sayısı bir makale yazan yazar sayısının (n =2) $1/n^2$ = ¼'ü olan 851 sayısını sağlamamıştır. Benzer durum üç makale yazan yazar sayısı içinde geçerlidir. Bu bulgu, makalelerin yazarlar arasındaki dağılımının Lotka yasası ile tutarlı olmadığını göstermektedir.

Authors' Local Impact (yazarların yerel etkisi) alt menüsü, veri setindeki yazarların etkisini anlamak ve analiz etmek için h-indeksi, g-indeksi, m-indeksi ve toplam atıf sayısı gibi ölçümleri kullanarak

kaynakları sıralar.h-indeksine göre sıralama yapıldığında h-indeksi sıfır olan yazarlar tabloda yer almaz. Bir yazarın yerel düzeydeki (yani veri seti içindeki) etkisini değerlendirmeyi sağlar. Most Relevant Affiliations (en ilgili kurumlar) alt menüsünde konu hakkındaki yayınların en çok hangi kurumlardaki yazarlarca yayımlandığını gösteren grafik yer alır. Grafik oluşturulurken Affiliation Name Disambiguation (kuruluş adı belirsizliğinin giderilmesi) evet olarak işaretlenmesine dikkat edilemelidir. Böylece kurumun adının, farklı yayınlarda ve veri tabanlarında farklı biçimlerde yazılması nedeniyle ortaya çıkan belirsizlikler ortadan kaldırılır. Örnek veri setine ait grafik Şekil 10'da yer almaktadır.

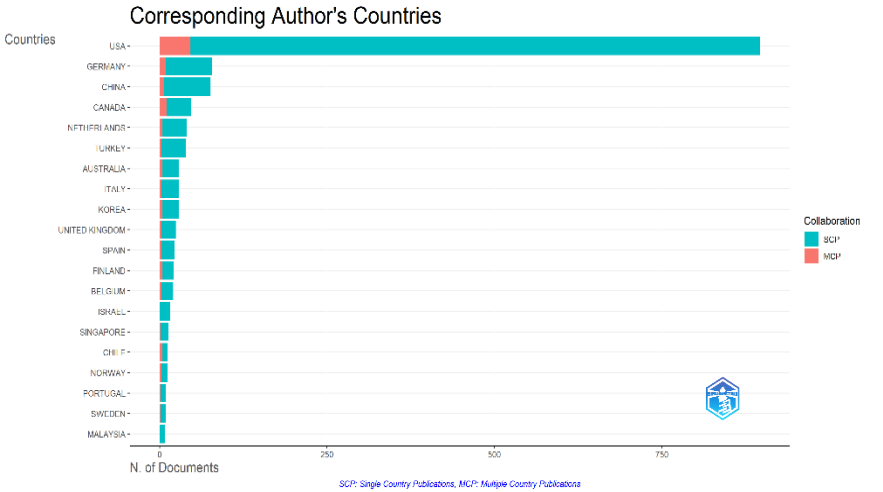


Şekil 10: En ilgili kurumlar

Araştırmada değerlendirilen 1806 makale arasında 1763 farklı kurum yer almıştır (bkz. Tablo 1). İlk on krum toplam 295 makale yayınlamış olup makalelerin %16.33'ünü oluşturmaktadır. En çok makale (f=52, %2.87) "University of California" tarafından yayımlanmıştır. "University of Pittsburgh" ikinci sırada yer almıştır

(f=33, %1.82). “University of Carolina” üçüncü sırada yer almış olup “University of Pittsburgh” takip etmiştir (f=31, %1.71). En az yirmi makale yayımlayan on altı kurum vardır.

Affiliations' Production over Time (kurumların zaman içindeki üretimi) alt menüsünde kurumların yayımladığı yayınların zaman içindeki dağılımını yıllık bazda gösteren grafik yer alır. Corresponding Author's Countries (sorumlu yazar ülkeleri) alt menüsü yayınlarda sorumlu yazar olarak belirlenen kişilerin ülkelerini ve ilgili ülkede üretilen yayınların toplam sayısını belirtmektedir. Bu sayede, en üretken ülkeler, uluslararası etkileşim ve iş birliği düzeyleri, farklı ülkelerden gelen katkılar ve hangi ülkelerin akademik işbirliklerinde daha etkin olduğu ve ülke bazlı eğilimler hakkında önemli ipuçları sunar. SCP yalnızca bir ülkenin araştırmacıları tarafından yayımlanan yayınları ifade eder. SCP yayınlar, tek bir ülke içinde yapılan araştırmaların ne kadar etkili ve yerel işbirliklerinin ne kadar güçlü olduğunu gösterir. SCP sayısı, bir ülkenin bilimsel üretkenliğini ve kendi içindeki akademik etkileşimleri yansıtan önemli bir göstergedir. Örneğin, yalnızca Belçika'daki araştırmacılar tarafından yazılan ve Belçika’daki dergilerde yayımlanan bir makale, SCP yayın olarak sınıflandırılır. MCP ise birden fazla ülkenin araştırmacıları tarafından birlikte yayımlanan yayınları ifade eder. MCP yayınlar, uluslararası işbirliklerinin ve küresel akademik etkileşimin gücünü gösterir. MCP sayısı, uluslararası iş birliği düzeyini anlamak için kullanılır. Örneğin, Kanada, Amerika ve İngilitere'den araştırmacıların birlikte yayımladığı bir makale, MCP yayın olarak sınıflandırılır. Örnek veri setine ait grafik Şekil 11’de yer almaktadır.



Şekil 11: Sorumlu yazar ülkeleri

Şekil 11, yazarlar yirmi ülke arasında dağılmaktadır. Sorumlu yazarların çoğunlukla ABD, Almanya, Çin, Kanada ve Hollanda ile bağlantılı olduğunu göstermektedir. SCP makalelerin yoğunlukta olduğu gözlenmiştir. Bu durum, ulusal bilimsel kapasitenin güçlü olduğunu ancak uluslararası işbirliklerin sınırlı olabileceğini göstermektedir. En fazla yayın 896 makale ile (%49.6) Amerika'da yapılmıştır. Araştırmalar ülkelere göre sıralandığında ikinci sırada 78 (%4.3) araştırma ile Almanya ve üçüncü sırada 24 araştırma ile (% 4.2) Çin yer almaktadır.

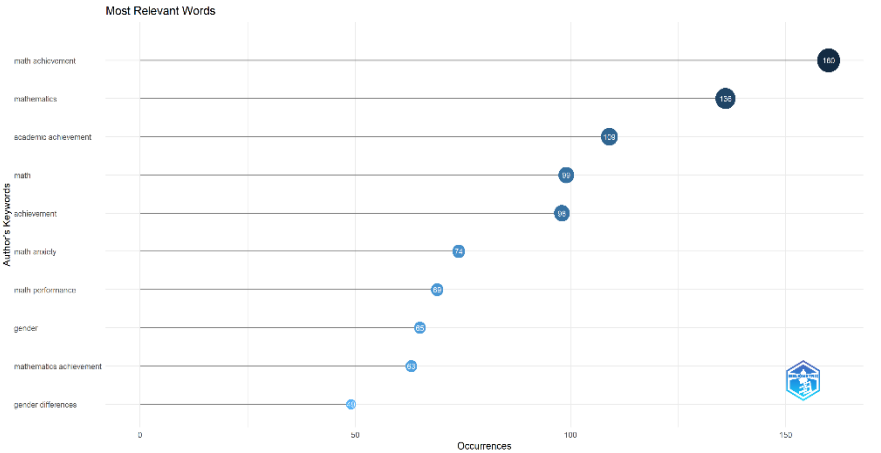
Countries' Scientific Production (ülkelerin bilimsel üretkenliği) alt menüsünde bilimsel üretkenlik dünya haritası yer almaktadır. Haritada mavi ve gri renkler kullanılır. Mavi rengin tonları üretkenlik düzeyini göstermektedir. Koyu mavi renkteki ülkeler en fazla katkıda bulunan ülkeledir. Rengin açılması üretilen yayın sayısının azaldığını gösterir. Gri renk o ülkeye ait bir çalışmasının bulunmadığı anlamına gelir. Countries' Production

over Time (ülkelerin zaman içindeki üretimleri) alt menüsünde ülkelerin yayımladığı yayınların zaman içindeki dağılımını yıllık bazda gösteren grafik yer alır. Most Cited Countries (en çok atıf alan ülkeler) ülkelerin yayınlarının aldığı toplam veya ortalama atıf sayısı gösterilir.

Documents (yayınlar) menüsü, veri setindeki yayınların analizini yapmaya yönelik bir menüdür. Bu menü, bibliyometrik analizlerde yayınların türleri, özellikleri ve bilimsel katkılarına ilişkin özet bilgiler sağlar. Most Global Cited Documents (en çok küresel alıntı yapılan belgeler) alt menüsü küresel düzeyde (global) toplam atıf sayısı ve yıllık bazda ortalama atıf ölçümlerini kullanarak yayınları sıralar. Alandaki en etkili çalışmaların tanımlanması ve temel referansların belirlenmesi için kullanılır. Most Local Cited References (en çok yerel alıntı yapılan kaynaklar) alt menüsü yerel düzeydeki toplam atıf sayısı ölçümlerini kullanarak yayınları sıralar. Yani veri setinde yer alan ve veri setindeki yayınların kaynakçalarında yer alan yayınları sıralar. Veri setindeki yayınlar arasındaki bağlantıları anlamak amacıyla kullanılır. Reference spectroscopy (referans yayın yılı spektroskopisi) araştırmanın tarihsel köklerini ve önemli makaleleri gözler belirlemeyi sağlar. Araştırma alanındaki tüm yayınların atıfta bulunduğu referansların yayın yıllarının analizi (referans yayın yılları/RYY), RYY'lerin eşit dağılmadığını göstermektedir. Bazı yayın yılları, diğerlerine göre daha sık yer alır. Bu yıllar, referans yayın yıllarının dağılımında (RYY spektrogramında) belirgin zirveler olarak gözlenir. Zirveler genellikle, diğer yayınlara kıyasla çok daha fazla alıntılanmış yayınlara dayanır. Bu yayınlar, alanın kökenlerini ve entelektüel temellerini temsil eder. Eğer bir yayın yılı t için zirve çok yüksek pozitif bir değere sahipse kullanılan veri setinden atıf yapılan birçok

referans, t yılında yayınlanmıştır. t − 2 ile t + 2 arasındaki yayın yılı dönemi baz alınarak atıf sayılarının medyanı hesaplanır (Anand vd., 2020). Medyan incelemesi her yılın komşu yıllarla karşılaştırılmasını sağlar. Zirveler, atıf sayısına dayalı grafiğe kıyasla daha kolay tespit edilebilir.

Most Frequent Words (en sık kullanılan kelimeler) alt menüsünde yazar anahtar kelimeleri, başlıklar, özet, Keyword Plus ve konu kategorilerinde en sık kullanılan kelimeleri ve terimleri belirlenir. Keyword Plus, Web of Science gibi indekslerde kullanılan bir anahtar kelime sistemidir. Yayınların başlıklarında veya yazar tarafından sağlanan anahtar kelimelerde bulunmayan ancak makalelerin atıf yaptığı kaynaklardan türetilen anahtar kelimelerdir. Araştırmanın bağlamsal anlamını genişletir. En sık kullanılan kelimeler, veri setinde hangi konuların öne çıktığını görselleştirir. Ayrıca, kelimelerin kullanım sıklığına göre bir sıralama elde edilir. Araştırmacıların literatürdeki temel konuları ve kavramları keşfetmesine, konular ve kavramlar arasındaki ilişkilerin anlaşılmasına ve araştırmanın genel içeriğini ortaya koymaya yardımcı olur. Örnek veri setine ait en sık kullanılan kelimeler Şekil 12'de yer almaktadır.



Şekil 12: En sık kullanılan yazar anahtar kelimeleri

Analizlere dahil edilen 1806 makalede 3357 farklı yazar anahtar kelimesi kullanılmıştır. En çok "math achievement", "mathematics", "academic achievement", "math" ve "achievement" yazar anahtar kelimeleri kullanılmıştır. En az on defa kullanılan 70 kelimeye ulaşılmıştır. Makale başına ortalama yaklaşık 2 anahtar kelime düşmektedir.

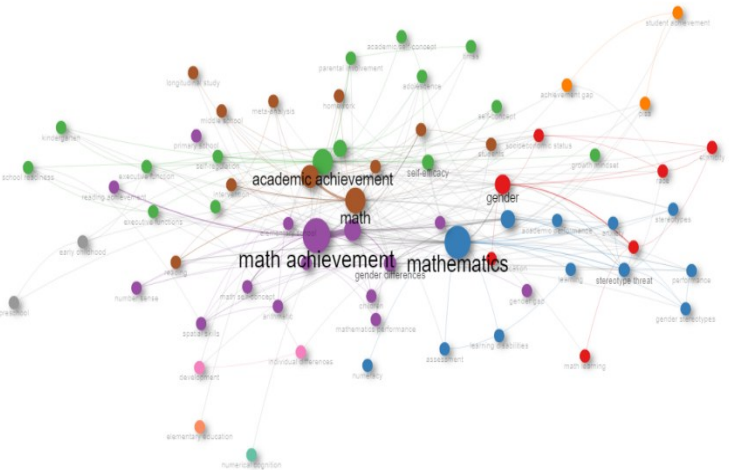
Makalelerde en çok kullanılan kelimeler ile Words Cloud (kelime bulutu) oluşturulabilmektedir. Kelimelerin kullanım sıklığına göre farklı boyutlarda ve renklerde gösterildiği bir görselleştirme yöntemidir. Kelime bulutunda merkeze yakınlık ve kelimelerin boyutlarının büyüklüğü sırasıyla ilgili kelimelerin konu alanına özgü kelimeler olduklarını ve çok sık kullanılan kelimeler olduklarını gösterirken, merkeze uzaklık ve kelimelerin boyutunun küçülmesi ise o kelimelerin daha az tercih dildiğini göstermektedir (Aslancı, 2022) . TreeMap (kelime ağacı) ise kelimelerin kullanım sıklığı ve bütün kelimeler içindeki yüzdelik dilimi gösterilmektedir. Ağaçta, her bir kelimeyi gösteren dikdörtgenlerin boyutu kelimenin

kullanım sıklığına göre belirlenmektedir. Sık kullanılan kelimelerin dikdörtgenleri büyüktür. Words' Frequency over Time (zaman içinde kelimelerin frekansı) alt menüsünde en sık kullanılan kelimelerin zaman içindeki kullanım sıklığını yıllık bazda veya kümülatif olarak gösteren grafik yer alır. Zaman içinde anahtar kelimelerin gelişiminin izlenebilmesini sağlamaktadır (Aydın İnci, 2022). Trend Topics (trend konular) alt menüsünde en sık kullanılan kelimelerin yıllara göre hangi sıklıkla kullanıldığını gösteren grafik yer alır. Her yıl hangi kelimenin popüler olduğu gözlenir. Kelime çizgisinde bulunan balonlar, kelimenin yılda kaç kere kullanıldığını gösterir. Kullanım sıklığı arttıkça balon büyür.

Conceptual Structure (kavramsal yapı) menüsü bir araştırma alanındaki temel terimler veya anahtar kelimeler arasındaki etkileşimleri, anlamlı ilişkileri, kümelenmeleri ve ortak kullanım sıklıklarını analiz etmeye yarar. Kavramsal yapılar, araştırmacılara önemli temaları, kavramları ve gelişim süreçlerini görselleştirme ve analiz etme imkanı sunar. Bu, özellikle yeni veya az çalışılmış alanlarda literatür taraması yaparken faydalı olmaktadır. Bu menüde Network Approach (ağ yaklaşımı) ve Factorial Approach (faktöriyel yaklaşım) başlıkları altında alt menüler bulunmaktadır. Ağ yaklaşımı, yayınlar arasındaki bağlantıları, yazarlar, anahtar kelimeler, dergiler veya ülkeler gibi araştırma bileşenlerinin çeşitli düzeylerdeki ilişkileri ağ tabanlı bir yaklaşımla gösterir. Oluşturulan ağları değerlendirirken çeşitli ölçümler kullanulmaktadır. Bir ağdaki ilişkisel bağlarının sayısını merkezilik derecesi ifade eder. Örneğin, ortak yazarlık ağındaki bir yazar dört farklı yazarla bağlantılı ise o yazarın merkezilik (centrality) derecesi dörttür. Toplam ilişkisel bağ sayısının her bağın gücüyle çarpılmasıyla ağırlıklandırılmış merkezilik (weighted centrality) derecesi bulunur. Aradalık

merkeziliği (Betweenness centrality), bir düğümün, bağlantısız düğüm grupları arasında bilgi taşıma yeteneğini ifade eder. Araştırma bileşeninin bir ağda oynadığı rol hakkında bilgi sunar. aradalık merkeziliği, belirli bir düğümden geçen en kısa yolların toplam sayısının tüm ağdaki en kısa yolların toplam sayısına bölünmesiyle bulunur. Özvektör merkeziliği (eigenvector centrality) değerinin, yüksek olması bilgiyi diğer yüksek bağlantılı düğümlere iletmekten sorumlu olan ağdaki düğümün önemini gösterir. Yakınlık merkeziliği, düğümlerin ağdaki diğer düğümlere daha yakın olarak bilgiyi etkili bir şekilde taşıma kabiliyetini ifade eder. Bu tür düğümlerin ağdaki diğer düğümlerden uzaklıklarının toplamı, bu düğümlerin bilgiyi etkili bir şekilde taşımasının göreceli kolaylığını gösterir.

Co-occurrence Network (ortak kelime ağı) alt menüsünde ortak kelime ağı yer almaktadır. Ortak kelime ağı kavramsal alanı temsil eden anahtar kelimeler ilişkilerini ve oluşturdukları temaları gösterir. Ortak kelime ağında, her düğüm bir anahtar kelimeyi gösterir. Anahtar kelimenin kullanım sıklığı arttıkça, düğüm büyür. Anahtar kelimeler arasındaki bağlantı çizgileri kelimelerin birlikte kullanılma sıklıklarını gösterir. Kelimelerin birlikte kullanılma sıklıkları arttıkça, düğümler arasındaki bağlantı çizgileri kalınlaşır. Kelimeler arasındaki mesafe genellikle iki kelimenin bir arada ne kadar sık geçtiğiyle ters orantılıdır. Yani, birlikte sık kullanılan kelimeler daha yakın konumlandırılırken, birlikte daha az geçen kelimeler daha uzak mesafede yer alır. Birlikte kullanılan kelimeler aynı kümelerde gruplandırılır. Her küme, alandaki temel temaları ve odak noktalarını vurgular. Her grup farklı bir renkle temsil edilir. Örnek veri setine ait ortak kelime ağı Şekil 13’te verilmiştir.



Şekil 13:Ortak kelime ağı

Örnek veri setinde on kesme noktası olarak belirlendiğinde sık kullanılan yetmiş yazar anahtar kelime belirlenmiştir. Başka bir deyişle, yetmiş yazar anahtar kelimesi on veya daha fazla kez kullanılmıştır. Şekil 13 incelendiğinde ortak kelime ağının, beş büyük kümenin (kırmızı, mavi, yeşil, mor ve kahverengi) yanı sıra beş nispeten daha küçük kümeden oluştuğu belirlenmiştir. Kırmızı renk ile gösterilen birinci küme gender, stem, mathematics education, socioeconomic status, race, math learning ve ethnicity kelimelerini içermektedir. Gender (cinsiyet), race (ırk), ethnicity (etnisite) ve socioeconomic status (sosyoekonomik durum), bireylerin başarılarını etkileyen önemli demografik faktörlerdir. Mathematics education (matematik eğitimi) ve math learning (matematik öğrenimi), bu faktörlerin matematik eğitimi ve öğrenme süreçlerindeki etkilerini incelemeyi ifade eder. Anahtar kelimeler,

eğitimde fırsat eşitliği, farklı grupların başarı düzeyleri, matematik öğreniminde karşılaşılan engeller ve bu engellerin aşılması için uygulanan politikalar ile ilgilidir. Bu nedenle, bu anahtar kelimeler matematik eğitiminde demografik faktörlerin rolü temasına odaklanmıştır.

Mavi renk ile gösterilen ikinci küme mathematics (matematik), math performance (matematik performansı), stereotype threat (önyargı tehtidi), academic performance (akademik performans), anxiety (kaygı), assessment (değerlendirme), performance (performans), stereotypes (önyargı), learning (öğrenme), gender stereotypes (cinsiyet önyargısı), learning disabilities (öğrenme güçlükleri) ve numeracy (sayısal yetenek) anahtar kelimelerini içermektedir. Kümedeki anahtar kelimeler, bireylerin matematik öğrenme süreçleri, performansları ve bu süreçlerde karşılaştıkları sorunlarla ilgilidir. Bu nedenle, bu kelimeler matematik performansını etkileyen sosyal ve psikolojik faktörler teması altında incelenebilir.

En büyük küme 15 anahtar kelimeden oluşan üçüncü kümedir (yeşil). Academic achievement (akademik başarı) ve mathematics achievement (matematik başarısı) , bireylerin okulda ve matematikte gösterdikleri performansı ifade ederken, self-efficacy (öz yeterlilik) ve academic self-concept (akademik öz-kavram), bireylerin öğrenme süreçlerindeki özgüvenini ve kendilerini nasıl algıladığını açıklar. Executive functions (yönetici işlevler) ve self-regulation (kendini düzenleme), bireylerin dikkat, planlama ve davranış kontrolü gibi bilişsel becerileri yönetme yetilerini ifade ederek, bu becerilerin okul başarısına etkisini vurgular. TIMSS, uluslararası düzeyde matematik ve fen başarısını değerlendiren bir çalışma olarak bu bağlamda önemli bir referans noktasıdır. Parental

involvement (ebeveyn katılımı), ebeveyn desteğinin ve katılımının çocukların akademik başarısına olan etkisini incelerken, growth mindset (büyüme zihniyeti), öğrenmeye yönelik olumlu bir yaklaşımı ve gelişime açık bir tutumu yansıtır. Adolescence (ergenlik) ve kindergarten (anaokulu), gelişim dönemleri olarak bu süreçlerin başlangıç ve ilerleyiş aşamalarını temsil eder. Longitudinal (boylamsal) çalışmalar, bu faktörlerin uzun vadeli etkilerini izlerken, school readiness (okula hazır olma), çocukların okula başlamadan önce gerekli becerilere sahip olmasını ifade eder. Bu anahtar kelimeler çocuk ve ergenlerde eğitimsel gelişim: bilişsel, duygusal ve sosyal faktörler teması altında incelenebilir.

Mor renk ile gösterilen dördüncü kümede math anxiety (matematik kaygısı), öğrencilerin matematikte yaşadığı stres ve korkuyu ifade ederken, gender differences (cinsiyet farklılıkları) ve gender gap (cinsiyet açığı), matematik başarısındaki cinsiyet temelli farklılıklara odaklanır. Working memory (çalışma belleği) ve spatial skills (mekansal beceriler) gibi bilişsel beceriler, öğrencilerin aritmetik ve sayı algısı gibi matematiksel görevlerdeki başarılarını etkiler. Math self-concept (matematik öz-kavramı), bireyin matematikteki yetkinliği hakkındaki algısını tanımlarken, reading achievement (okuma başarısı) ve matematik arasındaki ilişki de önem taşır. Anahtar kelimeler, çocukların matematiksel başarılarını etkileyen çalışma belleği, mekansal beceriler, cinsiyet farklılıkları, öz-kavram, matematik kaygısı ve okuma başarısı gibi unsurları kapsamaktadır. Bu nedenle çocukluk döneminde matematik başarısı ve bilişsel faktörler teması altında yer alır.

Turuncu ile gösterilen beşinci küme incelendiğinde, kümede yer alan student achievement (öğrenci başarısı) anahtar kelimesi

diğer anahtar kelimelerden uzakta konumlandırılmıştır. Bu, kelimenin farklı bağlamlarda sıklıkla kullanıldığını göstermektedir

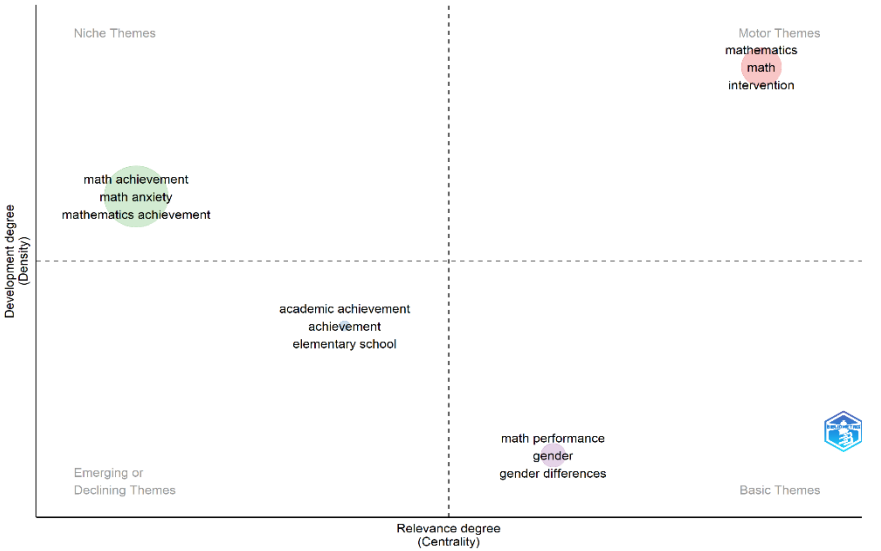
Kahverengi ile gösterilen altıncı kümede öğrencilerin akademik başarıları ve motivasyonlarını etkileyen çeşitli faktörlerle ilgilidir. Özellikle "math" (matematik) ve "reading" (okuma), eğitim araştırmalarında yaygın olarak değerlendirilen dersleri ifade eder. "Achievement" (başarı), öğrencilerin bu alanlardaki performanslarını vurgular. "Motivation" (motivasyon), öğrencilerin çabalarını etkileyen içsel dürtüleri ele alırken, "intervention" (müdahale), performansı artırmayı amaçlayan stratejileri ifade eder. "High school" (lise) ve "middle school" (ortaokul) terimleri, incelenen eğitim seviyelerini gösterirken, "students" (öğrenciler) ise bu çalışmaların hedef grubudur. "Homework" (ödev), öğrencilerin ders dışındaki görevlerinin başarıları üzerindeki etkisini araştıran bir değişkendir. "Meta-analysis" (meta-analiz), birden fazla çalışmanın birleştirilerek genel eğilimlerin değerlendirildiği bir araştırma yöntemini ifade ederken, "longitudinal study" (uzunlamasına çalışma), öğrencilerin akademik sonuçlarındaki değişimleri anlamak için uzun bir süre boyunca izlenen araştırmayı tanımlar. Bu anahtar kelimeler, "öğrencilerin akademik başarıları ve motivasyonu" teması altında toplanabilir.

Thematic Map (tematik harita) alt menüsünde tematik harita yer alır. Tematik harita araştırmaların hangi temalar etrafında yoğunlaştığını, hangi temaların eski hangi temaların yeni olduğunu belirlemede kullanılır. Ortaya çıkan her bir tema, en sık kullanılan kelimelere göre oluşturulmaktadır. Kelimelerin kullanım sıklığı ile orantılı olarak temalar boyutlandırılır (Aria vd., 2020). Tematik harita, iki boyutlu ve dört gruptan/çeyrekten oluşan bir diyagramdır. Temaların eksenlere göre konumları merkezilik ve yoğunluklarına

göre belirlenir. Diyagramda yer alan X ekseni bir ağın, merkeziyeti ifade etmektedir. Merkeziyet derecesi, kümenin diğer kümelerle etkileşim derecesinin yani dışsal uyumunun ve araştırma alanının gelişiminde öneminin ölçüsüdür. Y ekseni ise yoğunluğu ifade eder. Yoğunluk bir küme ağının iç gücünü, içsel uyumunu ölçer ve temanın gelişimi hakkında fikir verir (Kurnaz, 2021).

Temalar yerleştirildikleri gruba/çeyreğe göre analiz edilir. Sağ üst çeyrekte yer alan temalar yüksek merkeziyete ve yüksek yoğunluğa sahiptir. Bu temalara motor temalar denir. Motor temalar, konunun çekirdek ve destekleyici unsurlarını oluşturur. Literatürün belirli bir dönemdeki evrimini, popüler konuları, anahtar kelimeleri, temel teorik ve metodolojik çerçeveleri yansıtarak bir alanın gelişim sürecindeki önemli dönüm noktalarını da gösterebilir. Motor temadaki kelimeler hem çok kullanılmakta hem de birlikte kullanımları yaygın olan kelimelerdir (Tereci ve Yüksel, 2024). Sağ alt çeyrekte yer alan temalar yüksek merkeziyete ve düşük yoğunluğa sahiptir. Bu temalara temel ve çapraz temalar denir. Temel temalar, bir araştırma alanının temel yapı taşlarını ve ana konularını ifade eder. Çapraz temalar, farklı araştırma alanları veya konular arasında kesişen, birleşen veya ortak paydada buluşan multidisipliner temalardır. Çeşitli disiplinler arasında bilgi aktarımı sağlar ve farklı konuları bir araya getirerek daha geniş bir perspektif oluştururlar. Sol üst çeyrekte yer alan temalar düşük merkeziyete ve yüksek yoğunluğa sahiptir. Bu temalara niş temalar denir. Niş tema içerisindeki kelimeler ağın merkezinde olmayan ama birbiriyle sık kullanılan kelimelerdir (Tereci ve Yüksel, 2024). Sol alt çeyrekte yer alan temalar düşük merkeziyete ve düşük yoğunğa sahiptir. Yeni ortaya çıkan veya kaybolan temalar bu çeyrekte yer alır. Tematik haritada sağ üst ve sağ alt kadranlar incelenmelidir (Cobo vd., 2012).

Örnek veri setine ait sık kullanılan yetmişyazar anahtar kelimeden oluşturulan tematik harita Şekil 14'te verilmiştir.



Şekil 14:Tematik harita

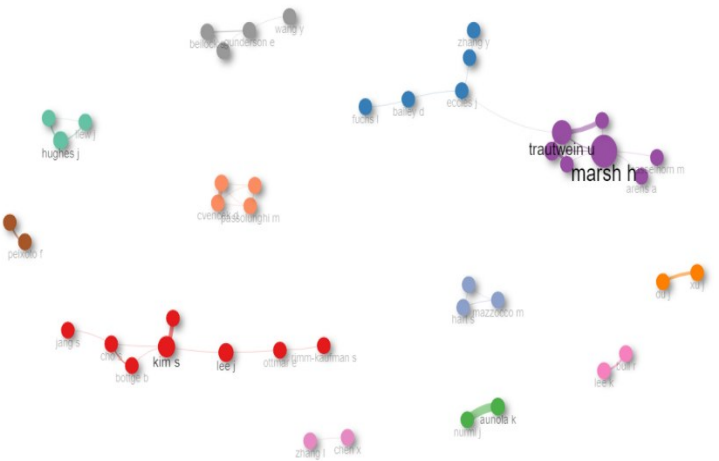
Şekik 14 incelendiğinde sağ üst çeyrekte mathematics (matematik), math (matematik) ve intervention (müdahale) anahtar kelimelerinden oluşan küme yer almaktadır. Anahtar kelimeler tematik haritada motor tema olarak belirlenmiştir. Bu kavramların araştırma alanında merkezi ve etkili bir rol oynadığını göstermektedir. Matematik ve müdahale çalışmaları, eğitimde hem teorik hem de uygulamalı yönleriyle araştırmaların odak noktası olarak kritik bir işlev görmektedir. Bu tema, diğer konularla güçlü bağlantılar kurarak alanın bilimsel ağını şekillendirmektedir. Sağ alt çeyrekte math performance (matematik başarısı), gender (cinsiyet) ve gender differences (cinsiyet farklılıkları) anahtar kelimelerinden oluşan küme yer almaktadır. Anahtar kelimeler tematik haritada temel ve çapraz tema olarak belirlenmiştir. Kümenin merkeziyetinin

düşük yoğunluğunun nispeten yüksek olduğu gözlenmiştir. Bu durum konuların belirli bir uzmanlık alanında derinlemesine çalışıldığını ancak daha geniş araştırma ağıyla sınırlı bağlantılar kurduğunu göstermektedir. Sol üst çeyrekte math achievement (matematik başarısı), math anxiety (matematik kaygısı) ve mathematic achievement (matematik başarısı) kelimelerinden oluşan küme yer almaktadır. Anahtar kelimeler tematik haritada niş tema olarak belirlenmiştir. Matematik başarısı ve kaygısı üzerine yapılan araştırmaların özgül bir akademik topluluk içinde önem taşıdığını, ancak daha geniş bilimsel bağlamlarda henüz yeterince ilişkilendirilmediğini işaret eder. Sol alt çeyrekte academic achievement (akademik başarı), elementary school (ilkokul) ve achievement (başarı) kelimelerinden oluşan küme yer almaktadır. Anahtar kelimeler tematik haritada yeni ortaya çıkan veya kaybolan tema olarak belirlenmiştir. Kümenin düşük merkezilik düşük yoğunlukta olduğu gözlenmiştir. Bu durum konuların literatürde yeni tartışılmaya başlandığını veya araştırmacılar arasında farkındalığın arttığını göstermektedir. Kümenin gelecekte daha yoğun bir şekilde araştırılabileceği veya daha fazla bağlantı kurarak merkeziliğini artırabileceği öngörülmektedir.

Thematic Evolution (tematik evrim) alt menüsünde belirli zaman aralıklarında temaların bağlantılarının nasıl olduğunu ve temaların zaman içerinde nasıl geliştiğini, değiştiğini gösterem Sankey diyagramı yer alır. Dikdörtgenler temaları gösterir. Dikdörtgenlerin boyutu, temanın kullanım sıklığıyla orantılı olarak büyür veya küçülür. Dikdörtgenler arasında bulunan bağlantılar, temanın zaman içerisindeki değişimini gösterir. İki tema arasındaki ilişki arttıkça, düğümler arasındaki bağlantı çizgileri kalınlaşır. Zaman içerisinde temalar yok olabilmektedir. Farklı temalar

birleşerek yeni tamalar oluşturabilmektedir. Bazı temalar ise değişmez. Factorial Analysis (faktöriyel analiz/ FA) anahtar kelimeler arasındaki ilişkilere göre kelimeleri gruplara ayırır.

Social Structure (sosyal yapı) menüsü bilim insanları, kurumlar veya ülkeler arasındaki iş birliği ağlarını görselleştirme ve analiz etme imkanı sunar. Yazarların belirli bir alandaki merkezi konumları ve işbirlikleri, o alandaki önde gelen araştırmacıları gösterebilir ve işbirliklerinin doğası hakkında fikir verebilir. Collaboration Network (iş birliği ağı) alt menüsünde ortak yazar ağı yer almaktadır. İşbirliği ağı alana katkıda bulunan yazarların birbirleriyle ilişki durumlarını gösterir. İşbirliği ağında, her düğüm bir yazarı/araştırmacıyı gösterir. Yazarın yayın sayısı arttıkça, düğüm büyür. Yazarlar arasındaki bağlantı çizgileri iş birliği sıklıklarını gösterir. Ortak yazılan makale sayısı arttıkça, düğümler arasındaki bağlantı çizgileri kalınlaşır. Yazarlar arasındaki mesafe genellikle iki yazarın birlikte ne kadar sık yayın yaptığıyla ters orantılıdır. Yani, birlikte sık sık yayın yapan yazarlar daha yakın konumlandırılırken, birlikte daha az yayın yapan yazarlar daha uzak mesafede yer alır. Birlikte çalışan araştırmacılar aynı kümelere gruplandırılır. Her grup farklı bir renkle temsil edilir. Örnek veri setine ait iş birliği ağı Şekil 15'te verilmiştir.

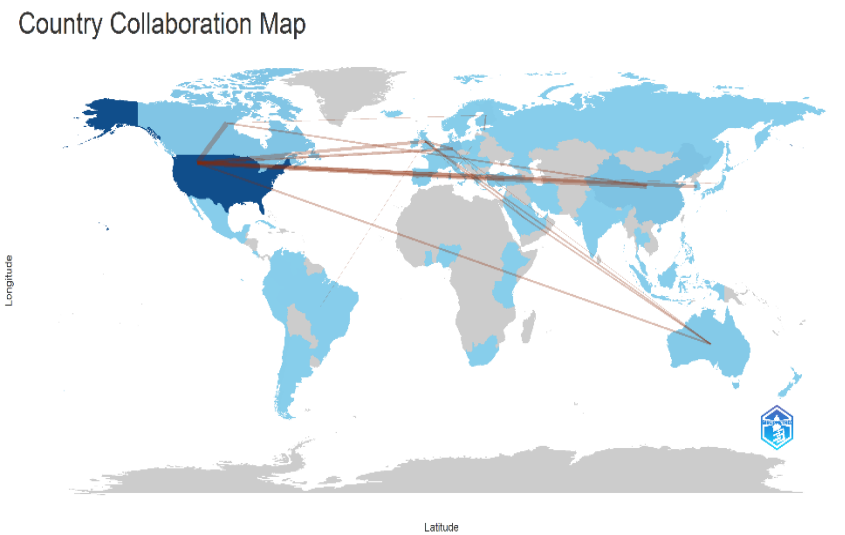


Şekil 15:İş birliği ağı

Örnek veri setinde beş kesme noktası olarak belirlendiğinde sık kullanılan yetmiş beş yazar belirlenmiştir. Başka bir deyişle, yetmiş beş beş veya daha fazla makale yayınlamıştır. Şekil 15 incelendiğinde iş birliği ağının on iki kümeden oluştuğu belirlenmiştir. Kırmızı renkle gösterilen birinci küme, Şekil 15'te sol alt köşede yer almaktadır. En büyük küme olan birinci küme sekiz yazardan oluşmaktadır. Kim, S. birinci kümenin merkezinde yer almaktadır. En büyük ikinci küme dördüncü kümedir. Mor renkle gösterilen dördüncü küme, Şekil 15'te sağ üst köşede yer almaktadır. Dördüncü kümede yedi yazar yer almaktadır. Marsh, H. dördüncü kümenin merkezindedir. Beşten fazla yazarın bulunduğu üç küme vardır. Bununla birlikte, çoğu küme yalnızca iki veya üç yazar arasındaki ilişkilerden oluşuyordu.

Her makaleye katkıda bulunan yazar sayısı bir ile on yedi arasında değişmektedir. İki yüz kırk sekiz (%13.73) makalenin tek yazarı, sekiz yüz on dokuz (%45.34) makalenin iki yazarı ve bin iki yüz yirmi sekiz (%67.99) makalenin üç yazarı vardır. Makale başına düşen ortalama ortak yazar sayısı 3.15'dir. Uluslararası ortak yazarların yüzdesi ise 6.977'dir. İşbirliği Endeksi (CI), çok yazarlı makalelerin toplam yazar sayısının çok yazarlı makalelerin toplam sayısına bölünmesiyle ölçülür (Elango ve Rajendran, 2012). İşbirliği endeksi (CI) 3.49'dur. Başka bir deyişle, çok yazarlı makale başına düşen ortalama ortak yazar sayısı 3.49'dur. Genel olarak, iş birliği ağı incelendiğinde, ağın seyrek olduğu görülmüştür. Yazarlar arasında sınırlı iş birliği vardır. Etkili yazarlar, izole edilmiş silolarda çalışma eğilimindedir. Örneğin, birinci kümenin merkezinde yer alan etkili yazarlardan Kim, S., dördüncü kümenin merkezinde yer alan Marsh, H. ile iş birliği yapmamıştır.

Ülke iş birliği haritası, ülkelerin işrbirliği ilişkilerini ve uluslararası iş birliği yoğunluklarını göstermek için kullanılmaktadır. Haritada mavi ve gri renkler kullanılır. Mavi rengin tonları ülkelerin iş birliğini sayısını göstermektedir. Koyu mavi renkteki ülkeler en fazla iş birliği yapan ülkeledir. Rengin açılması iş birliği sayısının azaldığını gösterir. Gri renk o ülkenin ait işbirliği yapmadığı gelir. İş birliği yapan ülkeler arasındaki bağlantı, çizgiler ile gösterilir. İş birliği sayısı arttıkça, ülkeler arasındaki bağlantı çizgileri kalınlaşır. Örnek veri setine ait ülkelerin iş birliği ağı Şekil 16'da verilmiştir.



Şekil 16: Ülkelerin iş birliği haritası

Şekil 16' da en çok uluslararası iş birliği ABD - Kanada (17) ve ABD - Çin (16) arasında gerçekleşmiştir. ABD'dan sonra en fazla iş birliği İngitere ve Almanya (5) arasındadır. ABD toplam 86 iş birliği ile en çok iş birliği gerçekleştiren ülke olmuştur. ABD'yi Çin 23 ve Almanya 21 iş birliği ile takip etmektedir.

Sonuç

Bibliometrik analiz, literatürün nicel olarak değerlendirilmesini sağlayan, disiplinler arası bağlantıları gösteren ve giderek daha fazla önem kazanan bir yöntemdir. Bu yöntem, yayınların nicel ve ortak kelime analizi gibi nitel unsurları bir araya getirerek, bir araştırma alanının genel görünümünü ve evrimini gözler önüne serer. Yayınların, atıfların, anahtar kelimelerin ve işbirliklerinin incelenmesi yoluyla, belirli bir alandaki eğilimleri, etkili kaynakları ve yazarları, önemli yayınları, güncel temaları ve yeni ortaya çıkan araştırma konularını detaylı bir şekilde anlamamızı

sağlar. Bu tür analizler, araştırmacıların hem de kurumların, hangi konularla yakından ilgilendiklerini gösterir.

Bibliometrik analiz, yalnızca bilimsel üretkenlik veya araştırma alanlarının haritalanması ile sınırlı kalmaz. Bunun yanı sıra üretilen bilgilerin akışını ve ilişkileri anlamada da güçlü bir araçtır. Araştırmaların zaman içindeki gelişimini görselleştirme, hangi yıllarda hangi temaların ilgi çektiğini belirleme ve araştırmacılar arasındaki işbirliklerini analiz etme gibi imkanlar sunar. Literatürdeki boşlukları ve araştırma gereksinimlerini belirler. Böylece yapılan araştırmalardan yola çıkarak gelecekteki araştırmalara yol gösterir. Özellikle hızlı gelişen alanlarda, araştırmacıların hangi yöne ilerlemesi gerektiğine dair değerli yol haritaları sunar. Bibliyometrik analiz bilimsel ilerlemeyi destekleyen temel bir yöntemdir.

Bu araştırmada gerçekleştirilen örnek bibliometrik analiz uygulaması, matematik başarısı konusunda literatürün kapsamlı bir değerlendirmesini sunmuştur. Analiz sonuçlarına göre matematik başarısı konusunda yapılan araştırmaların sayısı 1978 - 2024 yılları arasında artmıştır. Konuya olan ilgi son yıllarda artış göstermiştir. En üretken dergiler olarak “Journal Of Educational Psychology”, “Learning and İndividual Differences” ve “Contemporary Educational Psychology” belirlenmiştir. En etkili yazarlar Marsh, H., Lee, J. ve Kim, S. idi. Bu durum, matemamatik başarısı konsunda yapılacak bir araştırma için ilk olarak bu dergilerin ve yazarların incelenebileceğini göstermektedir.

İş birliği ağları birlikte çalışmanın önemini göstermektedir. Ülkeler ve yazarlar arasındaki iş birliği zayıf kalmıştır. Etkili yazarlar yalnızca kendi kümelerinde çalışma eğilimindedir. Tek bir

ülke içinde yapılan araştırmaların uluslararası araştırmalardan fazla olduğu görülmüştür. Yazarlar ve ülkeler tarafından yapılan iş birliklikleri konunun ilerlemesi için büyük önem taşımaktadır. Bu nedenle, daha geniş ölçekte iş birlikleri teşvik edilmelidir.

Örnek uygulamada bazı sınırlıklıkar da vardı. Sadece Web of Science ve Scopus veri tabanlarındaki makaleler incelenmiştir. Araştırmalarda konferanslar, kitaplar ve benzeri kaynaklar gibi diğer referans türlerini de dahil ederek veri seti genişletilmelidir. Konu alanının ilerlemesini ve evrimini yakından takip etmek için bu tür analizler belirli zaman aralıklarında yenilenmelidir.

Kaynaklar

Aksnes, D. W. (2003). A macro study of self-citations. *Scientometrics, 56*(2), 235–246.

Anand, K. J. S., Roue, J.-M., Rovnaghi, C. R., Marx, W., & Bornmann, L. (2020). Historical roots of pain management in infants: A bibliometric analysis using reference publication year spectroscopy. *Paediatric and Neonatal Pain, 2*(2), 22–32.

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959-975.

Aria, M., Misuraca, M. and Spano, M. (2020). Mapping the evolution of social research and data science on 30 years of social indicators research. *Social Indicators Research*, 149(3),803-831,

Aslanci, S. (2022). Araştırma sorgulamaya dayalı öğrenme : Bibliyometrik bir analiz. *Scientific Educational Studies*, *6*(1). doi:10.31798/ses.1068633

Aydın İnci, G. (2022). *Bilimin doğası öğretimi üzerine bibliyometrik bir analiz* (Yüksek Lisans Tezi). Alanya Alaaddin Keykubat Üniversitesi, Alanya.

Baker, H. K., Kumar, S., & Pandey, N. (2020). A bibliometric analysis of Managerial Finance: A retrospective. *Managerial Finance, 46*(11), 1495–1517.

Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering, 137*, 85–86.

Boole, G. (1854). An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities. Macmillan and Co.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J*ournal of the American Society for Information Science and Technology, 57*(3), 359–377.

Chen, C., Song, I. Y., Yuan, X., & Zhang, J. (2008). The thematic and citation landscape of Data and Knowledge Engineering (1985–2007). *Data and Knowledge Engineering, 67*(2), 234–259.

Cho, J. (2014). Intellectual structure of the institutional repository field: A co-word analysis. *Journal of Information Science, 40*(3), 386-397.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology, 63*(8), 1609-1630.

Cole, F. J., & Eales, N. B. (1917). The history of comparative anatomy. Part I: A statistical analysis of the literature. *Science Progress (London), 11*(4), 578–596.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 136*, 252-261.

Egghe, L. and Rousseau, R. (1990) Introduction to ınformetrics: Quantitative methods in library, documentation and information science. Amsterdam: Elsevier.

Elango, B., & Rajendran, P. (2012). Authorship trends and collaboration pattern in the Marine Sciences literature: A scientometric study. *International Journal of Information Dissemination and Technology, 2*(3), 166–169.

Feng, J., Zhang, Y. Q. ve Zhang, H. (2017). Improving the co-word analysis method based on semantic distance. *Scientometrics, 111* (3), 1521-1531.

Glänzel, W., Thus, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics, 59*(1), 63–77.

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science, 66*, 1229–1234.

Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. In *Incentives and performance: Governance of research organizations* (pp. 121–139). Cham: Springer.

Hawkins, D. T. (1977). Unconventional uses of online information retrieval systems: Online bibliometric studies. *Journal of the American Society for Information Science, 28*(1), 13–18.

Hulme, E. W. (1923). Statistical bibliography in relation to the growth of modern civilization. London: Grafton.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25.

Koo, M. (2021). Systemic lupus erythematosus research: a bibliometric analysis over a 50-year period. *International Journal of Environmental Research and Public Health, 18*(13).

Kumar, A. & Mohindra, R. (2015). Bibliometric analysis on knowledge management research. *International Journal of Information Dissemination and Technology, 5*(2), 106-113.

Kurnaz, A. (2021). Etnosentrizm ile ilgili çalışmaların bibliyometrik analizi. *Beykoz Akademi Dergisi, 9*(2), 98-118.

Leydesdorff, L., & Persson, O. (2010). The top 100: The influence of the Nobel Prize on the structure of science. *Journal of the American Society for Information Science and Technology, 61*(2), 366-374

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(12), 317–323.

Neuhaus, C., & Daniel, H. D. (2008). Data sources for performing citation analysis: an overview. *Journal of Documentation, 64*(2), 193–210.

Öztürk, O., & Gürler, G. (Ed.). (2022). *Bir literatür incelemesi olarak bibliyometrik analiz*. Ankara: Nobel Yayıncılık.

Pareto, V. (1896). *Cours d'économie politique*. Lausanne: F. Rouge.

Pritchard, A. (1969). Statistical bibliography or bibliometrics?. *Journal of Documentation, 25*(4), 348–349.

R Development Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rowlands, I. (2005). Emerald authorship data, Lotka's law and research productivity. Aslib Proceedings: New Information *Perspectives, 57*(1), 5-10.

Sengupta, A. (1992). The value of information in scientific communication: A bibliometric study. *Libri, 42*(2), 75-84.

Tereci, A. & Sezen-Yüksel, N. (2024). Tesadüfi öğrenmenin tematik çerçevesinin belirlenmesi. *Türk Eğitim Bilimleri Dergisi, 22*(2),1524-1550

Tonta, Y., & Al, U. (2007). Türkiye'nin bilimsel yayın haritası: Türkiye'de dergi yayıncılığı üzerine bibliyometrik bir araştırma (TÜBİTAK Projesi No. SOBAG-105K088). Hacettepe Üniversitesi.

Tunger, D., & Eulerich, M. (2018). Bibliometric analysis of corporate governance research in German-speaking countries: Applying bibliometrics to business research using a custom-made database. *Scientometrics, 117*(3), 2041–2059.

Yazıcı, A. M. (2023). A bibliometric analysis of research on climate and change business enterprises. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 12* (3), 2056 - 2077.

Zipf, G. K. (1932). Selected studies of the principle of relative frequency in language. London: Oxford University Press.

Zupic, I. & Cater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429-472.

BÖLÜM IV

Çoktan Seçmeli Test Geliştirme Sürecinde Madde Analizi ve Madde Yanlılığının İncelenmesi

Yusuf KASAP[1]

Giriş

Çoktan seçmeli test maddesi seçenekler ve madde kökünden oluşur. Seçeneklerden bir tanesi doğru cevap diğerleri ise çeldiricilerdir. Çoktan seçmeli testlerde yanıtlama özgürlüğü maddelerin seçenek sayısı ile sınırlıdır (Özçelik, 2010). Diğer sınavlardan farkı doğru cevabı kendi içinde bulundurmasıdır. Çoktan seçmeli madde yazma sürecinde madde yazarı, madde yazma niteliklerine ve tecrübesine sahip olup madde yazma tekniklerini en iyi biçimde uygularsa hazırlayacağı maddelerin kalitesi artar (Turgut & Baykul, 2011). Madde yazma teknikleri tecrübe ile öğrenilebilir. Bir madde yazarı uzun yıllar ders anlatıp ders anlattığı öğrencilerin

[1] Dr. Yusuf KASAP., Hacettepe Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, Ankara/Türkiye, Orcid: 0000-0002-5114-1175, yusufkasapolc@gmail.com

yaşadıkları kavram yanılgılarına hakim olursa çeldiricileri bu doğrultuda hazırlayıp yazılan maddelerin niteliğini artıracaktır (Doğan, 2020). Bununla birlikte madde yazarının testin hazırlanacağı konu alanına hakim olması gerekir. Burada söz edilen hakimiyet ifadesi, konuya aşinalık ya da konuyu ana hatlarıyla bilmek değil konuyu derinlemesine üst düzeyde bilme anlamındadır. Madde yazarı, hedef davranışların aşamalı olarak sınıflandırılması, bu sınıflamada her sınıfın özelliklerini bilmelidir. Eğer maddde yazarı bu özelliklere sahip değilse bir ölçme uzmanıyla birlikte çalışması veya yazılan maddeler için uzman görüşü alması gerekmektedir. Madde yazarının sahip olması gereken diğer bir özellik ise Türkçe'ye hakim olmasıdır. Yazarın madde kökünü açık ve anlaşılır, dilbilgisi yönünden hatasız olarak yazması gerekir. Madde yazarının bunu başarabilmesi için dili iyi kullanması gerekmektedir. Madde yazarı yazmış olduğu maddeleri tecrübesi arttıkça eleştirirse ve başkalarının yazmış oluğu maddelerin güçlü ve zayıf yönlerini ifade ederse istenen deneyimi kazanmış olur (Turgut & Baykul, 2011). Çoktan seçmeli testler üzerine bir çok tartışma vardır. Bunlardan bir tanesi bu araçla sadece alt düzey davranışların ölçülebilip üst düzey davranışların tamamının ölçülemeyeceği, diğeri soruların cevaplarının şansla bulunabileceğidir (Tan, 2011).

Çoktan seçmeli testlerde sentez basamağı hariç taksonominin her basamağındaki davranışlar ölçülebilir (Turgut & Baykul, 2011). Ancak madde yazarı tarafından konu alanının etraflıca ve derinliğine bilinmesi ve soru yazmada teknik beceriye sahip olması gerekir. Soruların, cevaplarını bilmeyen öğrenciler tarafından doğru cevaplandırılabilmesi durumunun bazı önlemlerle azaltılabilmesi, ancak tamamen ortadan kaldırılmaması çoktan seçmeli testlerin yok edilemeyen olumsuz özellikleridir. Bunu

önlemek için madde yazarı çeldiricileri doğru cevaba anlamca yaklaştırarak sorunun güçlük düzeyini ayarlayıp doğru cevap seçeneğini ölçülen kazanıma sahip olanları çekecek biçimde çeldiricileri ise ölçülen kazanıma sahip olmayanları çekecek biçimde yazar (Haladyna & Downing, 2011). Çeldirici yazmak ustalık gerektirmektedir. Öğrencilerin diğer sınavlarda yaptığı yanlışlıklar, iyi öğrenilmemiş konular çeldiriciler için fikir verebilir. Ayrıca, maddede sorulan problemin hatalı çözümleri analiz edilerek iyi çeldiriciler yazmak mümkündür. Eğer madde yazarı çeldirici bulamıyorsa madde formunu değiştirme yoluna gitmelidir. Çeldirici yazmada yararlanılacak kaynaklardan biri konu alanıdır. Bu bakımdan maddeler yazılmadan önce konunun değişik kaynaklardan okunması faydalı olacaktır. Çeldirici bulmada diğer kaynak, öğrencilerin sınıfta yaptıkları yanlışlardır. Çeldirici bulmada başka bir kaynak ise teste konulacak soruların daha önce kısa yanıtlı olarak öğrencilere sorulmasıdır. Bu sınavda öğrencilerin verdiği yanıtlar çeldirici yazmak için fikir verecektir (Turgut & Baykul, 2011). Diğer bir tartışma konusu ise çoktan seçmeli testlerde öğrencinin şans ile doğru yanıtı bulabilmesidir. Çoktan seçmeli testlerde şans ile doğru yanıtı bulmadan kaynaklı olarak ölçme sonuçlarına karışan hatalar düzeltme formülü (yanlışlar ile doğruların götürüldüğü formül) ile ortadan kaldırılır (Tekindal, 2010).

Çoktan Seçmeli Testlerin Özellikleri:

- Diğer soru tiplerinden farkı doğru cevabın madde içinde verilmesidir.
- Çok sayıda soru sorabilme imkanı olduğundan Kapsam geçerliği yüksektir.
- Güvenirliği, Geçerliği en yüksek ölçme aracıdır.

- Yazı güzelliği, kompozisyon yeteneği gibi değişkenlerin işe karışmaması geçerliğini yükseltir.
- Puanlama tamamen objektif ve kolaydır.
- Şans başarısı vardır. Bu durum güvenirliği düşürür. Ancak seçenek sayısının çokluğu şans başarısını azaltır.
- Üst düzey davranışların ölçülmesinde uygun değildir.
- Uygulama ve puanlama yönünden kullanışlıdır. Çünkü uygulaması ve puanlaması kolaydır. İlköğretim 3. sınıftan itibaren uygulanabilir.
- Hazırlanması zor ve uzmanlık gerektirdiği için hazırlama yönünden kullanışlılığı düşüktür.
- Önceden güçlükleri, güvenirlikleri, geçerlikleri tahmin edilebilir (Doğan, 2020).

Çoktan Seçmeli Test Geliştirmede Adımlar

1. Testin amacını belirleme: Testle hangi özelliklerin ölçüleceği ve hangi amaç için kullanılacağının belirlenmesidir. Örneğin: Personel seçmek için adayların sayısal yeteneklerinin ölçülmesi

2. Ölçülecek nitelikleri saptama: Bu adımda ölçülecek kapsam ve bu kapsamdaki davranışlar belirlenir. Örneğin; Personel seçmek için adayların sayısal yeteneklerinin ölçülmesi için matematik alanının hangi konu ve ünitesinden soruların sorulacağı belirlenir.

3. Maddelerin Yazılması: Ölçülecek özellikle ilgili davranışlara uygun maddeler yazılır ve soruların sayısı konuların ağırlığına bağlı olarak belirlenir.

Çoktan Seçmeli Madde Yazmada Genel İlkeler:

- Her madde bir davranışı yoklamalıdır.
- Madde kökü olabildiğince açık ve net olmalıdır.
- Seçenekler arasında büyüklük sırası varsa bunlar sıralı olarak verilmelidir.
- Maddeler birbirinden bağımsız cevaplandırılacak şekilde yazılmalıdır.
- Seçenekler birbirinden bağımsız olmalı, biri diğerini içermemeli, ipucu olmamalıdır.
- Böyle bir zorunluluk varsa bileşik cevap gerektiren soru tipi kullanılmalıdır.
- Testteki tüm maddelerin seçenek sayıları aynı olmalıdır.
- İyi bir test maddesinin madde kökü okunduğunda seçeneklere bakılmaksızın davranışa sahip olan cevaplayıcı doğru cevabı verebilmelidir.
- Madde kökünde gereksiz ifade ve açıklamalardan kaçınılmalıdır.
- Madde kökünde seçeneklerden herhangi birine ipucu veren ifadeler kullanılmamalıdır.
- Seçeneklerde tekrar eden kelimeler madde köküne alınmalı ve tekrar edilmemelidir.

- Yazı büyüklüğü uygulanan kişilerin düzeyine uygun olmalıdır.
- Maddelerdeki ifadeler bilinen bir kaynaktan olduğu gibi alınmamalıdır.
- Her test maddesi önemli bir davranışı yoklamalıdır.
- Çeldiriciler davranışa sahip olmayanları yanıltmalı fakat sahip olanları yanıltmamalıdır.
- Birbirinin zıddı olan seçeneklerin kullanılmaması gerekir.
- Madde noktalama ve yazım kurallarına uygun yazılmalıdır.
- Seçeneklerin uzunlukları birbirine yakın olmalıdır.
- Doğru cevaplar test içine belli bir örüntüde yerleştirilmemelidir.
- Doğru cevaplar seçeneklere yaklaşık olarak eşit dağıtılmalıdır.
- Olumsuz maddeler olumlu algılanma riskine karşı, olumsuzluk ifadesinin altını çizme, koyu yazma vb. yoluyla belirginleştirilmelidir. (Doğan, 2020).

4. Maddelerin gözden geçirilmesi (Redaksiyon): Sorular yazıldıktan sonra, bilimsel veya dil hataları, açık ve anlaşılırlık ve ölçülmek istenen davranışı ölçebilme yönlerinden incelenerek deneme uygulaması hazırlanmadan önce gerekli düzeltmeler yapılır.

5. Deneme formunun hazırlaması: Maddeler düzenlendikten sonra test deneme formu haline getirilir.

6. Deneme uygulamasının yapılması: Deneme formunu uygulamak için asıl gruba benzer bir grup seçilip pilot uygulama yapılır.

7. Deneme uygulaması cevap kağıtlarının puanlanması, madde analizi ve madde seçimi: Deneme uygulaması yapıldıktan sonra cevaplayıcıların cevapları için puanlama yapılır. Puanlama 1-0 ölçme kuralına göre yapılır. Puanlama yapıldıktan sonra madde istatistikleri hesaplanır. Daha sonra elde edilen test ve madde istatistiklerine bağlı olarak işleyen ve işlemeyen maddeler belirlenip hangi maddelerin seçileceğine, hangi maddelerde ne tür düzeltmeler yapılacağına, nihai teste hangi maddelerin alınacağına karar verilir.

Madde Analizi

Madde analizi genellikle maddenin güçlük indeksinin ve ayırtedicilik indeksinin hesaplandığı analizdir. Madde analizi sonucunda istenilen güçlük düzeyine sahip ve ayırtedici maddeler seçilerek nihai test oluşturulur.

Madde Güçlük İndeksi (P_j)

Bir maddeye doğru cevap verenlerin oranı yada yüzdesidir. Madde güçlük indeksi, maddenin zor mu yoksa kolay mı olduğunu gösteren sayısal değerlerdir. Madde güçlük indeksi 0,00 ile 1,00 arasında değer alır. Eğer bir maddeyi maddenin uygulandığı hiçbir birey doğru cevaplayamazsa maddenin güçlük indeksi ‘’0’’ olur. Eğer bir maddeyi maddenin uygulandığı tüm bireyler doğru cevaplamışsa maddenin güçlük indeksi ‘’1’’ olur (Tekin, 2010). Madde güçlük indeksinin değeri 0’a yaklaştıkça zorlaşır, 1’e yaklaştıkça kolaylaşır. Sorunun bilenle bilmeyeni en iyi ayırt ettiği durumda beklenen güçlük düzeyi ortadır. Ancak orta güçlük düzeyi

maddenin ayırtediciliğini garanti edememektedir (Murpy & Davishofer, 2005).

Tablo 1. Madde Güçlük İndeksinin Yorumlanması

Madde güçlük indeksi(P_j)	**Yapılacak yorum**
0,00 ile 0,19 arasında çıkarsa	Çok zor
0,20 ile 0,39 arasında çıkarsa	Zor
0,40 ile 0,60 arasında çıkarsa	Orta güçlükte
0,61 ile 0,80 arasında çıkarsa	Kolay
0,81 ile 1,00 arasında çıkarsa	Çok kolay

Crocker & Algina, 1986

$$P_j = \frac{D}{n}$$

D: Soruyu doğru cevaplayan öğrenci sayısı

n: Soruyu cevaplandırmış (doğru ve yanlış)Toplam öğrenci sayısı

Örnek1:

50 kişinin katıldığı (n= 50) sınavdaki bir maddeye; 40'ı doğru (D=40); 10'u yanlış cevap vermiş olsun, Bu maddenin güçlük indeksinin değeri;

$P_j = \frac{40}{50} = 0.80$ olarak bulunur.

Örnek2:

On soruluk bir sınavda her soruya doğru cevaba “1”, yanlış cevaba “0” olarak puanlama yapılmıştır. Uygulama sonucu aşağıdaki tablo ile verilmiştir. Cevaplanmayan soru bulunmamaktadır.

		Sorular									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Öğrenciler	**A**	1	1	1	0	1	0	1	0	1	0
	B	1	0	0	0	1	1	1	1	1	0
	C	0	0	1	0	0	0	1	0	0	0
	D	1	0	0	0	1	0	1	1	0	0
	E	1	0	1	0	0	1	1	0	1	1
	F	1	0	0	0	1	0	1	1	1	0
	G	0	0	1	0	0	0	1	0	0	0
	H	0	0	1	0	1	0	1	1	0	0
	I	1	0	0	0	1	0	1	1	0	0
	J	0	0	0	0	1	1	1	0	0	1

Tabloda verilen bilgilere göre madde ve test istatistikleri hesaplanacak olursa:

Madde güçlük indeksleri (P_j)

1. madde için, $p_1 = \frac{6}{10} = 0{,}60$

2. madde için, $p_2 = \frac{1}{10} = 0{,}10$

3. madde için, $p_3 = \frac{5}{10} = 0{,}50$

4. madde için, $p_4 = \frac{0}{10} = 0$

6. madde için,

$p_6 = \frac{3}{10} = 0{,}30$

7. madde için,

$p_7 = \frac{10}{10} = 1$

8. madde için,

$p_8 = \frac{5}{10} = 0{,}50$

9. madde için,

$p_9 = \frac{4}{10} = 0{,}40$

5. madde için, $p_5 = \frac{7}{10} = 0{,}70$ 10. madde için, $p_{10} = \frac{2}{10} = 0{,}20$

Yorum: Orta güçlük düzeyinde olan 1, 3, 8 ve 9. maddedir. Bu maddeler orta güçlük düzeyinde oldukları için standart sapmaları ve varyansları yüksek çıkar. En zor madde hiçbir bireyin doğru yanıt veremediği 4. madde iken en kolay madde tüm bireylerin doğru yanıtladığı 7. Maddedir. Bu maddeler bireysel farkları ortaya koyamadıklarından standart sapmaları ve varyansları sıfırdır.

Madde Ayırt Edicilik İndeks (r_{jx})

Madde ayırt edicilik gücü indeksi, maddenin kalitesini gösteren sayısal değeridir. Maddenin kalitesi bilen öğrenci ile bilmeyen öğrenciyi birbirinden ayırt edebilmesine bağlıdır. Bir maddenin teste alınıp alınmayacağına, madde ayırt edicilik gücü indeksine bakılarak karar verilir. Madde ayırt edicilik gücü indeksi, -1,00 ile +1,00 arasında değer alır. -1,00'e yaklaştıkça ayırtedicilik düşer. + 1,00'e yaklaştıkça ayırtedicilik yönde yükselir (Crocker & Algina, 1986).

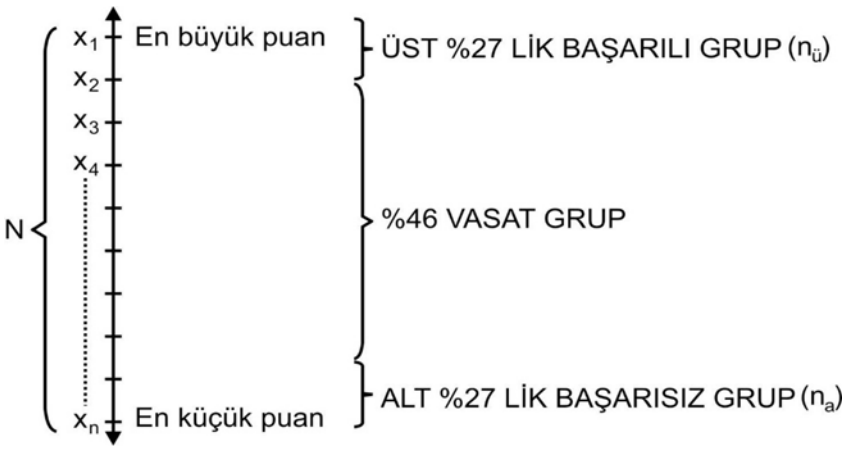
Tablo 2. Madde Ayırtedicilik İndeksinin Yorumlanması

Ayırt edicilik indeksi (r_{jx})	Yorum
$r_{jx} < 0.20$	(düşük) ise madde testten atılır.
$0.20 - 0.29$	(orta) arası düzeltilebilir.
$0.30 - 0.39$	(iyi) arası teste alınır (geliştirilebilir)
$r_{jx} \geq 0{,}40$	(mükemmel) ideal madde

Crocker & Algina, 1986

Alt ve Üst Grup tekniği ile Madde Analizi

Alt ve üst grup tekniğinde öğrencilere uygulanan test sonucunda alınan puanlar elde edildikten sonra puanlar en büyük olanda en küçük olana doğru sıralanır. En başarılı %27'lik kısımdaki bireyler üst grubu, En başarısız %27'lik kısımdaki bireyler alt grubu oluşturur. %46'lık orta düzey başarı gösteren vasat grup analize alınmaz (Doğan, 2020). Çünkü orta kesim maddenin ayırtediciliğini belirlemede etkili değildir. Bu teknik analiz için grubun heterojenliğini sağlayan tekniktir.



Örnek: Bir testteki 18. maddenin alt ve üst grup tekniğine göre madde analizi uygulaması için seçenek analiz tablosu aşağıdaki gibidir.

18.Madde	A	B	C	D*	E	Boş	Toplam
Üst Grup	2	35	8	45	2	8	100
Alt Grup	20	23	20	20	4	13	100

18. maddenin doğru cevabı D seçeneğidir.

Tabloda verilen bilgilere göre madde istatistiklerinin bulunuşu ve çeldirici analizi

$$p_j = \frac{Dü + Da}{N}$$

Dü: Üst grupta maddeye doğru cevap verenlerin sayısı

Da: Alt grupta maddeye doğru cevap verenlerin sayısı

N: Analize alınan toplam kişi sayısı

$p_{18} = \frac{45+20}{200} = 0{,}32$ olduğundan 18.madde gruba zor gelmiştir.

$$r_{jx} = \frac{Dü - Da}{n}$$

r_{jx} : j.maddenin ayırtedicilik indeksi

Dü : Üst grupta maddeye doğru cevap verenlerin sayısı

Da : Alt grupta maddeye doğru cevap verenlerin sayısı

n: Üst veya alt gruptaki kişi sayısı (Toplamın yarısı)

Yukarıdaki seçenek analiz tablosuna göre 18. maddenin ayırtediciliği hesaplanacak olursa;

$r_{18} = \frac{45-20}{100} = 0{,}25$ maddenin ayırtediciliği orta düzey olduğu için maddenin düzeltilmesi gerekir.

Maddenin Düzeltilmesi için Gereklilikler:

Doğru cevap nasıl olmalıdır?

Bir maddenin doğru cevabı üst gruptakileri (bilenler.) çekmeli, alt gruptakileri (bilmeyenleri) çekmemelidir.

Çeldiriciler Nasıl Olmalıdır?

Çeldiriciler üst gruptakileri çekmemeli, alt gruptakileri ise çekmelidir. Aksi durumda, çeldirici doğru cevap gibi davranarak zıt yönde çalışacağından bilenleri doğru cevaba göndermez. Bu durumda ayırtedicilik düşer. Ayrıca soruya yanlış cevap verenlerin üst grupta da alt grupta da çeldiricilere dağılım oranlarının mümkün

olduğunca eşit olması demektir. Yani çeldiriciler çeldiricilere düşen toplam kişi sayısını dengeli paylaşmalıdır. Aksi durumda, zayıf ve kuvvetli çeldiricilerin meydana gelmesine neden olur.

Zıt yönde çalışan çeldirici:

Üst grupta alt gruptakinden daha fazla kişiyi kendisine çeken çeldiricidir. Bu çeldiriciler bilenleri doğru cevaba göndermeyeceğinden ayırtedicilik düşer.

Zayıf çeldiriciler:

Alt ve üst grupta toplamda az sayıda kişiyi kendisine çeken çeldiricilerdir. Zayıf çeldiriciler maddeyi kolaylaştıran ve yeterince iyi çalışmayan çeldiricilerdir. Bu çeldiriciler bilmeyenleri kendisine çekemediğinden bilmeyenlerin doğru cevaba gitme ihtimalini artırarak ayırt ediciliği düşürebilirler. Ayrıca zayıf çeldiriciler cevaplayıcı tarafından kolayca eleneceğinden şans başarısını artırırlar.

Kuvvetli (etkili) çeldiriciler:

Alt ve üst grupta toplamda çok sayıda kişiyi kendisine çekerek maddeyi zorlaştıran çeldiricilerdir. Alt grupta çok sayıda kişiyi çeken kuvvetli çeldiriciler iyi çalışan çeldiricilerdir. Eğer kuvvetli çeldirici üst grupta çok sayıda kişiyi çekerse bilenler doğru cevaba gitmeyeceğinden ayırtedicilik düşer.

Tablodaki Seçeneklerin Analizi

18.madde	A	B	C	D*	E	Boş	Toplam
Üst Grup	2	35	8	45	2	8	100
Alt Grup	20	23	20	20	4	13	100

D doğru cevap seçeneği: Üst grupta yeterli sayıda bireyi çekememiş alt grupta ise gereğinden fazla bireyi çekmiştir.

A çeldiricisi: Bu çeldirici üst grupta az, alt grupta çok sayıda bireyi çektiğinden en iyi çalışan çeldiricidir. Ayırtediciliği en olumlu etkileyen çeldiricidir.

B çeldiricisi: Üst grupta alt gruptan daha fazla bireyi çektiğinden zıt yönde çalışan ve doğru cevap gibi davranmış olan çeldiricidir. Ve zıt yönde çalışan kuvvetli çeldiricidir. En kötü çalışan çeldirici olduğundan ayırtediciliği en olumsuz etkileyen çeldiricidir.

C çeldiricisi: Bu çeldirici üst grupta az, alt grupta çok sayıda bireyi çektiğinden iyi çalışmış bir çeldiricidir. Üst grupta A çeldiricisine göre daha fazla bireyi çektiğinden A çeldiricisi kadar iyi çalışmamıştır. Ayırtediciliği olumlu etkileyen çeldiricidir.

E çeldiricisi: Zayıf çeldirici olduğundan iyi çalışmamıştır. Zayıf çeldirici olduğundan ayırtediciliği olumsuz etkileyen çeldiricidir.

Çeldiricilerin en iyi çalışandan en kötü çalışana sıralaması: A,C,E,B biçimindedir.

Seçenek analizi yapılan 18. maddenin ayırtediciliğini artırmak için yapılması gerekenler:

- Öncelikle zıt yönde çalışan B çeldiricisi üst gruptaki bireyleri çekmeyecek, alt gruptakileri çekecek şekilde düzeltilmelidir. Düzeltilemiyorsa çıkarılıp alt gruptaki bireyleri çekecek biçimde iyi bir çeldirici yazılmalıdır.
- Sonra zayıf çeldirici olan E çeldiricisi alt gruptakileri çekecek şekilde kuvvetlendirilmelidir.

- Doğru cevap seçeneği üst gruptaki bireyleri çekecek biçimde düzeltilmelidir.

8. Son düzeltmelerin yapılması: Madde analizi sonunda istenilen güçlük düzeyine sahip ve ayırtedici maddeler seçildikten sonra açık anlaşılır olması ve bilimsel hataların olmaması için son kontroller yapılır.

9. Nihai testin oluşturulması ve madde istatistiklerinin kestirilmesi: Deneme uygulaması sonucunda seçilip son düzeltme yapılan maddeler nihai formu oluşturur (Turgut & Baykul, 2011).

Madde Yanlılığının İncelenmesi

Genellikle, eğitim ve psikoloji alanında kullanılan testlerdeki maddelerin yansız ve maddelerle oluşturulan testlerin yordama geçerliğinin yüksek olduğu varsayılır. Örneğin, üniversite giriş sınavlarında cinsiyet veya bölge faktörlerine bağlı olarak öğrenci grupları arasında madde puanı ortalamaları yönünden farklılık çıktığında, grupların gerçek başarılarının böyle olduğu varsayılır. Ancak bu bilgi gerçeği her zaman yansıtmaz. Maddenin yanlı ölçüm yapması, farklı grupların ortalamaları arasında fark çıkmasına neden olabilir. Farklı grupların ortalamaları arasındaki farkın gerçek kaynağını ortaya koyabilmek testte yer alan herbir maddenin yanlılık çalışması yapılmalıdır. Bir testten ölçme sürecinde bir özelliğe yanlı ölçümler elde ediliyorsa, bu test ile yapılan ölçümlerin geçerliğinden şüphe etmek gerekir. Test geliştirme süreci, çok sayıda işlem ve adımlardan oluşur. Test maddelerinin geçerliğini belirlemek bu işlem ve adımların en önemlisidir. Geçerli bir test geçerli maddelerden oluşmaktadır. Madde yanlılığını belirlemek ise testin geçerliğini belirleme teknikleri arasında yer alır. Eğitim ve psikolojide kullanılan testlerdeki maddelerin yanlılık çalışmaları,

testlerin geçerliği ile ilgili kanıtlar elde etmenin önemli tekniklerindendir (Doğan & Öğretmen 2016)

Madde yanlılığı, bir test maddesine aynı yetenek düzeyinde fakat farklı alt gruplardan gelen bireylerin doğru cevap verme olasılıklarının farklı olması olarak tanımlanmaktadır. Yanlı bir madde değişen madde fonksiyonu (Differential Item Function - DIF) verir. DIF aynı yetenek düzeyine sahip farklı gruplardan yanıtlayıcıların, bir maddedeki doğru yanıtlama olasılıklarının farklılaşması sonucu ortaya çıkan madde yanlılığı olarakta ortaya çıkar. Ancak maddenin değişen madde fonksiyonu vermesi kendi başına madde yanlılığının bir kanıtı değildir (Doğan & Öğretmen, 2016). Bir maddeye gelen cevaplardaki farklılık, cevaplayıcı grupların yetenek düzeylerindeki farklılık olduğu zaman beklenen bir durumdur. Genellikle farklı yetenek düzeyindeki grupların madde cevaplarındaki farklılık madde yanlılığı değildir; tam tersine maddenin geçerliği ve yansızlığıdır (Raju, 1988). Diğer yandan, bu analizler psikolojik yapı ölçümlerinin gruplara göre farklılık gösterip göstermediğini incelemek için yapılmaktadır (Mellenberg, 1983).

Madde yanlılığını incelemenin iki amacı vardır. Birincisi, madde puanlarının değişik gruplardan gelen farklı değişkenlik kaynaklarından etkilenip etkilenmediğinin araştırılması; ikincisi ise bu ilişkisiz kaynakların bazı alt gruplar için bir avantaj sağlayıp sağlamadığıdır. (Crocker ve Algina, 1986). Eğer madde, ilgilenilen farklı alt gruplar için adil ise maddenin farklı alt gruplardaki aynı değişkenlik kaynağından etkilendiği söylenebilir. Bu durumda cevaplayıcılar, aynı maddenin ölçtüğü psikolojik yapı açısından benzer düzeyde olacaklardır. "İlişkisiz değişkenlik kaynağının dağılımı bütün alt gruplar için benzer olacaktır" (Crocker ve Algina, 1986). Diğer yandan yetenek düzeyi aynı gruplar arasında madde

performansı bakımından fark olduğu zaman maddenin yanlı olduğu söylenir. Eğer bir test çok heterojen bir evren için kullanılacaksa, yanlılık analizleri madde seçimi sürecinin en önemli unsuru haline gelir. Çünkü, yapılan ölçme işleminde test uzmanı için en önemli kriter farklı alt gruplardan gelen ve testi alan bireyler için en adil ve en doğru sonuçları elde etmek olacaktır (Doğan & Öğretmen, 2016).

KAYNAKÇA

Crocker, L. and Algina, J., (1986). Introduction to Classical and Modern Test Theory. Orlando: Rinehart and Winston, Inc.

Doğan, N., & Öğretmen, T. (2016). Test ve Madde Yanlılığı. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 5(1).

Doğan, N. (2020). Eğitimde Ölçme ve Değerlerdirme (2. baskı, s. 114-120). Ankara: Pegem Yayın Dağıtım.

Haladyna, T. M., & Downing, S. M. (2011). Handbook of Test Development. Taylor & Francis. Prentice Hall.

Mellenberg, G. J., (1983). Conditional Item Bias Methods. In S. H. Irvine and W. J. Barry (Eds), Human Assesment and Cultural Factors (pp. 293-302), Newyork: Plenum Pres.

Murpy, R., K. & Davidshofer, O., C. (2005). Psychological Testing: principles and aplications. Pearson/Prentice Hall.

Özçelik, D.A. (2010). Okullarda Ölçme ve Değerlendirme: Öğretmen El Kitabı (1. Baskı, s. 137-144). Ankara: Pegem Yayın Dağıtım.

Raju, N. S., (1988). "The Area Between Two Item Characteristic Curves." Psychometrica, 53,495-502.

Tan, Ş. (2011). Öğretimde Ölçme ve Değerlendirme: KPSS El Kitabı (6. baskı, s. 176-190). Ankara: Pegem Yayın Dağıtım.

Tekindal, S. (2010). Eğitimde Ölçme ve Değerlendirme (2. baskı, s. 94-97). Ankara: Pegem Yayın Dağıtım.

Turgut, M. F., & Baykul, Y. (2011). Eğitimde Ölçme ve Değerlerdirme (2. baskı, s. 122-139). Ankara: Pegem Yayın Dağıtım.

BÖLÜM V

Üniversite Öğrencilerinin Ekolojik Sorunlara İlişkin Görüşlerinin İkili Karşılaştırma Metodu İle İncelenmesi

Sümeyra SOYSAL[1]

Giriş

Ekoloji, canlıların çevreleri ile uyum içinde yaşamalarını sürdürmelerini inceleyen bir bilim dalı olarak tanımlanırken ekosistem, insan ve diğer canlıların birlikte, uyum ve denge içinde varlıklarını ve gelişmelerini sürdürebilmeleri için var olan şartların tamamını içeren bir kavram olarak karşımıza çıkmaktadır. Ekosistemde bütün canlıların çevreleriyle dengeli bir şekilde yaşaması söz konusudur. Bu dengenin bozulması ekolojik dengeye dışardan müdahalelerin olması ile olur. Bu durum ekolojik

[1] Doç. Dr., Necmettin Erbakan Üniversitesi, Ahmet Keleşoğlu Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Konya/Türkiye, Orcid: 0000-0002-7304-1722, sumeyrasoysal@erbakan.edu.tr

sorunların veya çevre sorunlarının ortaya çıkarmasına neden olmaktadır (Görmez, 2003).

21. yüzyılda teknoloji, sanayi ve kentleşmenin hızlı gelişimi, artan enerji ihtiyacı ve değişen tüketim alışkanlıkları çevre sorunlarını hızla büyütmüştür. Hava, su ve toprak kirliliği gibi çevre kirliliği insan hayatını tehdit eder hale gelmiş, küreselleşme ve hızlı kentleşme insan ihtiyaçlarını hızla değiştirmiştir. Bu durum, çevrenin aşırı tüketilmesine ve doğal kaynakların hızla tükenmesine yol açarak, tüm canlıların yaşamını tehlikeye atmaktadır (Baykal & Baykal, 2008; Akyüz, 2015).

Ekoloji ve çevre sorunları yalnıca ülkelerin değil 2024 itibari ile yaklaşık 8 milyar olan (United Nations Population Fund-UNPA, 2024) tüm insanlığın ortak bir sorunudur. Artan nüfusla günümüz dünyasında, çevre sorunları giderek daha karmaşık hale gelmekte ve ekolojik denge üzerindeki tahribat, insanoğlunun geleceğini tehdit eder boyutlara ulaşmaktadır. Hava kirliliği, su kaynaklarının kirlenmesi, doğal kaynakların tükenmesi, ormansızlaşma, iklim değişikliği, biyolojik çeşitliliğin kaybı ve atık yönetimi gibi sorunlar, sadece çevreyi değil, aynı zamanda sosyal, ekonomik ve kültürel yapıları da derinden etkilemektedir (UNEP, 2020). Bu bağlamda, çevre bilinci ve ekolojik farkındalık, toplumların sürdürülebilir bir geleceğe doğru ilerlemesi için kritik öneme sahiptir.

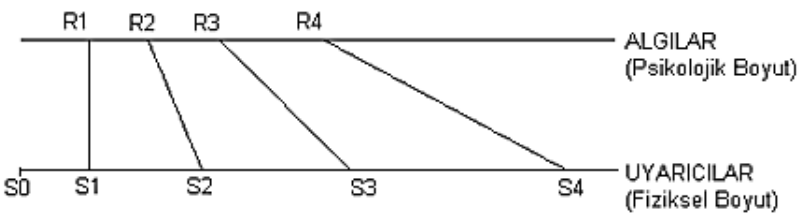
Gençler, özellikle üniversite öğrencileri, bu sürecin en dinamik ve bilinçli aktörlerinden biri olarak, çevre sorunlarına ilişkin görüş ve tutumlarıyla önemli bir rol oynamaktadır. Üniversite öğrencileri, çevre bilincinin şekillendiği, sosyal sorumluluğun arttığı ve kolektif eylemlerin önemli bir güç oluşturduğu toplumsal bir gruptur (Schultz, 2002). Bununla birlikte, üniversite öğrencilerinin

çevreye dair görüş ve tutumları, bireysel davranışlarla sınırlı kalmayıp, toplumsal düzeyde daha büyük etkiler yaratma potansiyeline sahiptir. Genç nüfus, çevresel değişimlere karşı duyarlı ve aktif bir yaklaşım benimseme potansiyeline sahip olup, gelecekteki politika ve çözüm süreçlerinde etkili olabilecek bir güçtür.

Bu çalışma ile üniversite öğrencilerinin ekolojik sorunlara karşı var olan ilgilerini arttırmak, farkındalıklarını pekiştirmek ve ekolojik sorunları önem düzeylerine göre nasıl sıraladıklarını görmek açısından alan yazına katkı sağlamak hedeflenmektedir.

İkili karşılaştırmalar metodu ile ölçekleme

Ölçekleme işlemleri, psikofizik adı verilen bir bilim dalında ele alınır ve bu alan, bireylerin fiziksel büyüklükleri nasıl algıladığını incelemeye odaklanır. Psikofizik, çevresel uyarıcıların fiziksel özellikleri ile bu uyarıcıların psikolojik algıları arasındaki ilişkiyi ortaya koyan yasaları araştıran bir disiplindir (Turgut & Baykul, 1992). Şekil 1'de uyarıcıların ölçülen özellikleri ile algılanan özellikleri arasındaki ilişki şematize edilmiştir.



Şekil 1. Uyarıcıların ölçülen özellikleri ile algılanan özellikleri arasındaki ilişki

(Kaynak: Turgut ve Baykul, 1992)

Ölçekleme, değişkenler arasındaki farkları daha belirgin hale getirmek ve ölçümler üzerinde matematiksel işlemleri çeşitlendirmek amacıyla, sıralama ve sınıflama düzeyindeki verilerin eşit aralıklı ölçek düzeyine dönüştürülmesidir. Bu süreçte kullanılan teknikler, deneklere dayalı yaklaşımlar ve yargıcı kararlarına dayalı yaklaşımlar olmak üzere iki ana grupta toplanabilir.

Deneklere dayalı yaklaşımlar, cevaplayıcının odaklandığı bir yöntem olup, bireylerin verdikleri tepkiler doğrultusunda ölçek üzerinde farklı noktalara yerleştirilmesi esasına dayanır (Torgerson, 1958; Crocker & Algina, 1986). Bu yaklaşımlar, dereceleme toplamlarıyla ölçekleme ve çok boyutlu ölçekleme gibi teknikleri içerir (Turgut & Baykul, 1992). Diğer yandan, yargıcı kararlarına dayalı yaklaşımlar, denek yerine uyarıcıya odaklanan bir yaklaşımdır ve burada her bir uyarıcının psikolojik ölçek üzerindeki yeri gözlemcilerin değerlendirmelerine dayanarak belirlenir (Arık & Kutlu, 2013). Yargıcı kararlarına dayalı yaklaşımlar, ikili karşılaştırmalarla ölçekleme, sınıflama yargılarıyla ölçekleme, sıralama yargılarıyla ölçekleme ve mutlak yargılarla ölçekleme gibi dört ana başlık altında ele alınır (Kan, 2008). Bu çalışmada ise ikili karşılaştırmalarla ölçekleme yöntemi kullanılmıştır.

İkili karşılaştırmalar yoluyla ölçekleme metodunun temeli Thurstone'un 1927 ve 1959'daki çalışmalarına dayanmaktadır (Heldsinger & Humphry, 2010). İkili karşılaştırmalarla ölçekleme metodu, ilk olarak tutum cümlelerinin ölçeklendirilmesinde kullanılmıştır. Cevaplayıcıların uyarıcıları ikişerli olarak değerlendireceği her durumda kullanılabileceği için bu yöntemin geniş bir uygulama alanı bulunmaktadır. Özellikle, duyuşsal

alandaki pek çok davranışın ölçeklenmesinde kullanıldığı (Anıl ve Güler, 2006) gibi pazar araştırmalarında ve karar verme süreçlerinde de kullanılmaktadır. Herhangi bir iş grubundaki bireylerin mesleki sorunlarının tespiti (Anıl, Ertuna & Uysal, 2015), standart akademik performans kararlarının alınması (Bramley, Bell & Pollitt, 1998), bireylerin sosyal aktivite tercihlerinin belirlenmesi (Polat & Göksel, 2014), çeşitli istatistiksel yöntemlerin karşılaştırılması (Ömür & Erkuş, 2013; Güler, İlhan & Taşdelen Teker, 2018) gibi cevaplayıcıların uyarıcıları ikişerli olarak değerlendirebileceği her durumda ikili karşılaştırmalarla ölçekleme yöntemi kullanılabilmektedir.

Çalışmanın amacı

Bu araştırmanın amacı, üniversite öğrencilerinin insanlığın belli başlı ekolojik sorunlarını en önemliden en önemsize doğru sıralanmasını ikili karşılaştırmalar metodu ile incelemektir. Bu bakımdan araştırmanın alt amaçları aşağıdaki şekilde ifade edilmiştir:

1. Üniversite öğrencileri ekoloji sorunları en önemliden en önemsize doğru nasıl sıralamaktadırlar?

2. Eğitim Fakültesi, Edebiyat Fakültesi, Mühendislik Fakültesi ve İktisadi ve İdari Bilimler Fakültesi'nde okuyan öğrenciler ekoloji sorunları en önemliden en önemsize doğru nasıl sıralamaktadırlar?

3. Her bir uygulamaya ait ölçekleme sonuçlarının ölçek değerlerinin iç tutarlılıkları nasıldır?

Yöntem

Araştırmanın örneklemi

Bu araştırmanın örneklemini, Cumhuriyet Üniversitesi'nin Eğitim Fakültesi, Mühendislik Fakültesi, İktisadi ve İdari Bilimler Fakültesi ve Edebiyat Fakültesi'nde okuyan toplam 482 öğrenci oluşturmaktadır. Öğrencilerin yaklaşık %51'i kadındır ve fakültelere göre dağılımları Tablo'1 de özetlenmiştir.

Tablo 1: Örneklemin dağılımı

Fakülte	**N**	**%**
Eğitim	166	34,44
Edebiyat	177	36,72
Mühendislik	74	15,35
İİB	57	11,83
Kayıp	8	1,66
Toplam	482	100,00

Ölçme aracının hazırlanması ve uygulanması

Türkmen (2010) çalışmasında insanlığın belli başlı ekolojik sorunlarını ekolojik sorunları sekiz başlıkta ele almıştır: Nüfus artışı ve mekan sorunu, enerji sorunu, tarımsal sorunlar, çölleşme, çoraklaşma ve kuraklık, ormansızlaşma, Canlı türlerinin yok olması ve küresel ısınma. Çalışmada veri toplama amacıyla bu ekolojik sorunları ifade eden bir anket formu kullanılmıştır. Kullanılan bu ankette, bu sekiz ekolojik sorun ikili gruplar hâlinde eşleştirilmiştir. Öğrencilerden her bir grupta verilen iki sorundan en önemli buldukları sorunu işaretlemeleri istenmiştir. Anket yirmi sekiz çift eşleştirmeden oluşmakta olup sekiz ekolojik soruna ait bütün ikili eşlemeleri içermektedir.

Verilerin analizi

Veriler, ekolojik sorunların önemine göre sıralanmasına ilişkin fakülte ayrımı yapmaksızın tüm üniversite öğrencilerinin tepkileri ve farklı fakültelerde (Eğitim Fakültesi, İktisadi ve İdari Bilimler Fakültesi, Mühendislik Fakültesi ve Edebiyat Fakültesi) okuyan öğrencilerin tepkileri olmak üzere iki kategoride değerlendirilmiştir. Her bir kategori için işlem basamakları şu şekildedir (Anıl & Güler, 2006; Güler & Anıl, 2009):

Frekans (F) matrisinin oluşturulması: Öncelikle, cevaplayıcıların yaptıkları sıralamalardan her bir ifadenin bir diğerine göre kaç defa daha öncelikli sıralandığı bulunmuştur. Veri toplama aracında toplam sekiz sorun bulunduğundan toplam 56 [yani 8x(8-1)/2] tane frekans elde edilmiştir. Söz gelimi F matrisinin fij elemanı i sorununun j sorunundan daha öncelikli görülme frekansını göstermektedir. Dolayısıyla katılımcı sayısı n olmak üzere fji=n–fij ile hesaplanarak 8x8'lik matris oluşturulmuştur. Bir ifade kendisi ile karşılaştırılamayacağından hesaplarda kolaylık sağlaması amacıyla fii=n/2 olarak alınmıştır.

Oranlar (P) matrisinin oluşturulması: F matrisinin her bir elemanı cevaplayıcı sayısına bölünerek oranlar matrisi oluşturulmuştur. P matrisinin elemanları pij=fij/n şeklindedir.

Standart oranlar (Z) matrisinin oluşturulması: Oranlar matrisinin elemanları standart Z puanlarına dönüştürülerek yeni bir matris (Z matrisi) oluşturulmuştur. Z matrisine, elemanları sütun elemanlarının ortalaması olan bir satır eklenmiştir. Bu satır elemanları en küçük eleman 0'a gelecek şekilde sayı doğrusu üzerinde ötelenmiştir. Böylece ifadelerin ikili karşılaştırma ile

sıralanmaları sayı doğrusu üzerinde elde edilmiş ve sıfıra daha yakın olan sorun daha öncelikli sorun olarak değerlendirilmiştir.

Tutarlılık Ölçüsü: Ölçek değerlerinin iç tutarlığının hesaplanması, gözlenen Pjk oranlarının ölçek değerlerinden elde edilen Pjk değerleriyle ne kadar uyumlu olduğu hesaplanarak yapılmaktadır. Bu amaçla, uygulama verilerinin analizinden elde edilen ölçek değerlerinden hareket edilerek bir Z' birim normal sapmalar matrisi ve bu matristen de teorik oranlar matrisi elde edilir. Teorik oranlar ile gözlenen oranlar arasındaki uygunluğa bakılır. Bu durumda öncelikle ölçek değerlerinin ortalama hatası hesaplanır. Elde edilecek küçük bir ortalama hata gözlemci yargılarının güvenilir olduğuna, büyük bir hata ise gözlemci yargılarının güvenilir olmadığına işaret etmektedir. Büyük bir hata değeri ya modelin varsayımlarının sağlanamadığını ya gözlemci yargılarının güvenilir olmadığını ya da her iki hata kaynağının da etkili olduğunu göstermektedir (Turgut & Baykul, 1992). Bir başka deyişle, ortalama hata değeri gözlenen değerlerle teorik değerler arasındaki uygunluğun ortalama değerini, yani uyumun bir ölçüsünü verir. Bu uygunluk derecesinin anlamlı olup olmadığı ise ki kare (chi-square) testi ile yoklanır. Bu çerçevede tüm hesaplamalar Microsoft Office Excel programı ile yapılmıştır.

Birinci alt problemin çözümünde anket formuna cevap veren tüm öğrencilerin verileri (N=482) kullanılırken ikinci alt problemin çözümünde kayıp veriler çıkarıldıktan sonra kalan toplam 474 öğrencinin cevapları kullanılmıştır. Bulgularda yer alan matrislerin satır ve sütunlarında yer alan hafler araştırmada ele alınan ekolojik sorunları temsil etmektedir. Kodlamalar şu şekildedir: A-Nüfus artışı ve mekan sorunu, B- Enerji sorunu, C- Tarımsal Sorunlar D-

çölleşme, E- Çoraklaşma, F-Ormansızlaşma, G-Canlı türlerinin yok olması, H- Küresel ısınma.

Bulgular

Üniversite öğrencilerinin ekolojik sorunları sıralamasına ilişkin bulgular

Sekiz ekolojik sorunun sıralanması sonucu oluşturulan 28 tane ikili karşılaştırmaya ait frekans matrisi Tablo 2'de özetlenmiştir.

Tablo 2: Frekans matrisi

F MATRİSİ								
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A		246	297	193	162	181	223	151
B	236		289	220	196	216	165	135
C	185	193		183	171	209	142	125
D	289	262	299		197	245	178	134
E	320	286	311	285		279	207	164
F	301	266	273	237	203		190	139
G	259	317	340	304	275	292		78
H	331	347	357	348	318	343	304	

Not: A-Nüfus artışı ve mekan sorunu, B- Enerji sorunu, C- Tarımsal Sorunlar D- çölleşme, E- Çoraklaşma, F-Ormansızlaşma, G-Canlı türlerinin yok olması, H- Küresel ısınma.

Tablo 2'deki satır, sütun isimleri verilerin Excel Programına işlendiği şekliyle yazılmıştır. Matrisin *i*.satır *j*.sütun elemanı (*x*ij); satırdaki sorunun sütundaki soruna göre kaç defa daha öncelikli gösterildiğini belirtmektedir. Örneğin, $x_{A,B} = 246$ değeri, 246 öğrencinin A sorununu B sorunundan daha önemli veya daha öncelikli bulduğunu gösterir. Dolayısıyla $x_{B,A}$ değeri cevaplayıcı sayısından $x_{A,B}$ değerinin çıkarılmasıyla elde edilmektedir ($x_{A,B}$ = *N*–246 ve *xij* + *xji* = 482).

F matrisi esas köşegene göre simetriktir ve simetrik elemanların toplamı, toplam gözlemci sayısını verir ki burada toplam kişi sayısı olan 482'ye eşittir. Tablo 2'deki F matrisinin her bir elemanı N=482'ye bölünerek Tablo 3'deki P oranlar matrisi hesaplanmıştır.

Tablo 3: Oranlar matrisi

P MATRİSİ								
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A		0,51	0,62	0,40	0,34	0,38	0,46	0,31
B	0,49		0,60	0,46	0,41	0,45	0,34	0,28
C	0,38	0,40		0,38	0,36	0,43	0,30	0,26
D	0,60	0,54	0,62		0,41	0,51	0,37	0,28
E	0,66	0,59	0,65	0,59		0,58	0,43	0,34
F	0,62	0,55	0,57	0,49	0,42		0,39	0,29
G	0,54	0,66	0,71	0,63	0,57	0,61		0,16
H	0,69	0,72	0,74	0,72	0,66	0,71	0,63	

Tablo 3 incelenirse, P matrisinin esas köşegene göre simetrik olan elemanlarının toplamı birdir. Bir başka ifadeyle, P matrisinin elemanları pij+pji=1 eşitliğini sağlamaktadır. Bir sonraki aşamada P matrisindeki her bir p değerinin birim normal dağılımın z değeri bulunmuş ve Tablo 4'te verilen birim normal sapmalar matrisi oluşturulmuştur.

Tablo 4: Birim normal sapmalar matrisi

Z MATRİSİ								
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A		0,03	0,30	-0,25	-0,42	-0,32	-0,09	-0,49
B	-0,03		0,25	-0,11	-0,24	-0,13	-0,41	-0,58
C	-0,30	-0,25		-0,31	-0,37	-0,17	-0,54	-0,65
D	0,25	0,11	0,31		-0,23	0,02	-0,33	-0,59
E	0,42	0,24	0,37	0,23		0,20	-0,18	-0,41
F	0,32	0,13	0,17	-0,02	-0,20		-0,27	-0,56
G	0,09	0,41	0,54	0,33	0,18	0,27		-0,99
H	0,49	0,58	0,65	0,59	0,41	0,56	0,33	
ΣZjk	1,25	1,24	2,58	0,46	-0,87	0,43	-1,49	-4,26
Sj	0,16	0,16	0,32	0,06	-0,11	0,05	-0,19	-0,53
Sc	0,69	0,69	0,86	0,59	0,42	0,59	0,35	0,00

Tablo 4 incelendiğinde, Z matrisindeki esas köşegene göre elemanlar birbirinin ters işaretlisidir ve mutlak değerce birbirlerine eşittirler. Matrisin sonunda her bir sütuna ait değerlerin toplamını (ΣZjk) gösteren bir satır oluşturulmuş ve bu satırdaki her bir z hücre

değerlerinin sütunlar boyunca ortalamaları alınarak yani sütunların eleman sayısı olan 8'e bölünerek ölçek değerleri (Sj) hesaplanmıştır. Tablo 4'ün alttan ikinci satırında görüldüğü gibi, Z değerlerinin en küçüğü H sütununa ait olan -0,533'tür. Eksenin başlangıcı (0 noktası) doğrusal dönüştürme ile yani bulunan en küçük değerin tüm uyarıcı değerlerine eklenmesi ile sıfır noktasına kaydırılarak Sc değerleri bulunmuştur. Sc değerleri en küçükten büyüğe doğru sıralanarak elde edilen Sc ölçek değerleri Tablo 5'te özetlenmiştir.

Tablo 5: Ölçek değerlerinin sıralı olarak gösterimi

	Küresel ısınma	**Canlı türlerinin yok olması**	**Çoraklaşma ve Kuraklık**	**Ormansızlaşma**	**Çölleşme**	**Enerji sorunu**	**Nüfus artışı ve mekan sorunu**	**Tarımsal sorunlar**
Sıra	1	2	3	4	5	6	7	8
Sc	0	0,35	0,42	0,59	0,59	0,69	0,69	0,86

Ekolojik Sorunlar Anketi'nin üniversite öğrencilerinden elde edilen verilerinin V. Hal denklemi ile ölçeklenmesi sonucu ekolojik sorunların önem sıraları, Tablo 5'te verildiği gibi gözlenmiştir. Tablo 5 incelendiğinde, öğrencilere göre en önemli ekolojik sorun (en düşük ölçek değerine sahip) küresel ısınma iken sekiz ekolojik sorun içinden diğerlerin en az önemli olan (en düşük ölçek değerine sahip) tarımsal sorunlar olduğu gözlenmiştir.

Eğitim Fakültesi, Edebiyat Fakültesi, Mühendislik Fakültesi ve İktisadi ve İdari Bilimler Fakültesi'nde okuyan öğrencilerinin ekolojik sorunları sıralamasına ilişkin bulgular

Eğitim Fakültesi, Edebiyat Fakültesi, Mühendislik Fakültesi ve İktisadi ve İdari Bilimler Fakültesi'nde okuyan öğrencilerin görüşlerine göre ekolojik sorunları önem düzeylerine göre sıralamalarına ilişkin frekans matrisi, oranlar matrisi ve birim sapmalar matrisi yukarıda birinci alt probleminin analizinde gösterildiği gibi hesaplanmıştır.

Tablo 6: Ölçek değerlerinin uygulama yapılan fakültelere göre sıralı olarak gösterimi

Fakülte	**Eğitim**		**Edebiyat**		**Mühendislik**		**İİB**	
	Sc	**Sıra**	**Sc**	**Sıra**	**Sc**	**Sıra**	**Sc**	**Sıra**
Nüfus artışı ve mekan sorunu	0,64	6	0,55	5	0,68	8	0,63	6
Enerji sorunu	0,77	7	0,6	7	0,24	4	0,55	5
Tarımsal Sorunlar	0,78	8	0,76	8	0,59	7	1,12	8
Çölleşme	0,58	5	0,49	4	0,39	6	0,54	4
Çoraklaşma ve Kuraklık	0,39	3	0,32	2	0,24	3	0,43	3
Ormansızlaşma	0,49	4	0,57	6	0,3	5	0,69	7
Canlı türlerinin yok olması	0,23	2	0,37	3	0,13	2	0,22	2
Küresel ısınma	0,00	1	0,00	1	0,00	1	0,00	1

Tablo 6'da uygulama yapılan her bir fakültedeki öğrencilerin ekolojik sorunları önem düzeyine göre sıralama sonuçları bir arada verilmiştir. Tablo 6'da görüldüğü gibi Eğitim Fakültesi, Edebiyat fakültesi, Mühendislik Fakültesi ve İktisadi ve İdari Bilimler Fakültesi öğrencilerine göre ekolojik sorunların en önemlisi küresel ısınmadır. Ayrıca tabloda, sekiz ekolojik sorunun önem sırasının fakülteler arasında farklılık gösterdiği görülmektedir.

Uygulamalara ait ölçek değerlerinin iç tutarlılıklarına ilişkin bulgular

Tablo 5 ve Tablo 6'da verilen ve uygulama yapılan gruplara ait ölçekleme sonuçlarının ölçek değerlerinin iç tutarlılıklarını bulmak için ki-kare analizleri yapılmıştır. Analizleri yapılan her bir ölçeklemeye ait ortalama hata ve $\chi 2$ istatistik değerleri Tablo 7'de verilmektedir.

Tablo 7: Ölçek değerlerinin iç tutarlılıklarına ait bilgiler

Gruplar	**N**	**Hata**	**Ortalama Hata**	**sd**	**X^2**	**Relative X^2**
Genel	482	0,85	0,03	21	93,9	4,5
Eğitim fakültesi	166	0,86	0,03	21	36,1	1,7
Edebiyat fakültesi	177	1,02	0,04	21	43,9	2,1
Mühendislik fakültesi	74	1,04	0,04	21	18	0,9
İİB Fakültesi	57	1,28	0,05	21	23,7	1,1

Ki-kare tablo değeri $\chi 2_{(0,05;21)}$= 32,671

Tablo 7 incelendiğinde ölçeklemelere ait ortalama hata değerlerinin 0'a çok yakın olduğu gözlenmektedir. Bir başka ifadeyle gözlenen ortalama puanlar ile kestirilen ortalama puanlar arasındaki farklılık sıfıra yakındır, yani kestirilen ile gözlenen puanlar birbirine benzerdir. Mühendislik Fakültesi ve İBBF için yapılan ölçeklemelere ait $\chi 2$ değerlerinin $\chi 2$ tablo değerinden küçük olması, ölçek değerlerinin iç tutarlığa sahip olduğunu göstermektedir.

Genel, Eğitim Fakültesi ve Edebiyat Fakültesi için yapılan ölçeklemelerde ise $\chi 2$ değeri anlamlı bulunmuştur. Bu durumun nedeni olarak örneklem büyüklüklerindeki farklılık olduğu düşünülmektedir. $\chi 2$ istatistiğinin formülündeki N, formülde çarpım durumunda olduğu için örneklem sayısının artması ortalama hata düşük olsa bile $\chi 2$'nin anlamlı çıkmasına neden olabilmektedir. Dolayısıyla, sonuçları, $\chi 2$'yi daha az örneklem büyüklüğüne bağımlı hale getiren bir yöntem olan ve ki-kare'nin serbestlik derecesine bölümünden elde edilen Göreli Ki-Kare İndeks (Relative Chi Square Index) istatistiğine göre yorumlamanın daha doğru olacağı düşünülmektedir.

Göreli Ki-Kare İndeks değerinin bazı araştırmacılara göre 3'ün altında bazı araştırmacılara göre ise 5'in altında olması model-veri uyumu için yeterli sayılmaktadır. Buna göre Tablo 7'deki sonuçlar değerlendirildiğinde, uygulama sonuçlarına ait Relative $\chi 2$ değerlerinin 5'in altında hatta Genel uygulama dışındaki uygulamalarda 3'ün altında olması, ölçek değerlerinin iç tutarlığa sahip olduğunu göstermektedir.

Sonuç

Üniversite öğrencilerinin görüşlerine dayanarak insanlığın belli başlı ekolojik sorunlarının önem sırasının belirlemek amacıyla oluşturulan Ekolojik Sorunlar Anketi'nin Thurstone'un karşılaştırmalı yargılar kanununun V.Hal denklemi ile ölçeklenmesi sonucu öğrencilerin sekiz ekolojik sorun içinde en önemli buldukları sorunun küresel ısınma olduğu gözlenmektedir. Bu durumun nedeni, televizyon gibi kitle iletişim araçlarında, günümüz ekolojik sorunlarından en fazla küresel ısınmadan bahsedilmesi gösterilebilir. Genel olarak üniversite öğrencilerine göre diğer yedi ekolojik sorunun önem sırası şöyledir; canlı türlerinin yok olması, çoraklaşma ve kuraklık, ormansızlaşma, çölleşme, enerji sorunu, nüfus artışı ve mekan sorunu ve tarımsal sorunlar.

Eğitim Fakültesi, Edebiyat Fakültesi, Mühendislik Fakültesi ve İktisadi ve İdari Bilimler Fakültesi'nde okuyan öğrencilere ekolojik sorunların sıralaması incelendiğinde bu dört fakülte için de en önemli sorun küresel ısınmadır. Eğitim Fakültesi öğrencilerine göre diğer yedi ekolojik sorunun önem sırası şöyledir; canlı türlerinin yok olması, çoraklaşma ve kuraklık, ormansızlaşma, çölleşme, nüfus artışı ve mekan sorunu, enerji sorunu ve tarımsal sorunlar. Edebiyat Fakültesi öğrencilerine göre diğer yedi ekolojik sorunun önem sırası şöyledir; çoraklaşma ve kuraklık, canlı türlerinin yok olması, çölleşme, nüfus artışı ve mekan sorunu, ormansızlaşma, enerji sorunu ve tarımsal sorunlar. Mühendislik Fakültesi öğrencilerine göre diğer yedi ekolojik sorunun önem sırası şöyledir; canlı türlerinin yok olması, çoraklaşma ve kuraklık, enerji sorunu, ormansızlaşma, çölleşme, tarımsal sorunlar ve nüfus artışı ve mekan sorunu. İktisadi ve İdari Bilimler Fakültesi öğrencilerine

göre diğer yedi ekolojik sorunun önem sırası şöyledir; canlı türlerinin yok olması, çoraklaşma ve kuraklık, çölleşme, enerji sorunu, nüfus artışı ve mekan sorunu, ormansızlaşma ve tarımsal sorunlar. Uygulama yapılan beş gruba ait ölçekleme sonuçlarının ölçek değerlerinin iç tutarlılıklarının olduğu yani yapılan ölçekleme işlemlerinin güvenilir olduğunu göreli ki-kare değerlerine dayalı olarak söylemek mümkündür.

Alan yazın incelendiğinde bu çalışmanın sonuçlarına benzer şekilde Yoloğlu ve Halisdemir'in (2020) üniversite öğrencileri ile yürüttüğü çalışmasında öğrencilerin en önemli gördüğü üç çevre sorunu sırasıyla küresel ısınma, su kaynaklarının kirlenmesi, kimyasal atıklar olduğu raporlanmıştır. Bu çalışmanın sonucundan farklı olarak, Yılmaz, Samsunlu ve Peker'in (2017) üniversite öğrencileri ile yaptığı çalışmalarının sonucunda; öğrencilerin çoğunluğunun çevre sorunu kavramına ''Toplumsal yaşama verilen zarar'' anlamı yükledikleri, yaşadıkları çevrede çevre sorunu olarak en çok 'Çöpler'i gördükleri belirtilmiştir. Kırılmaz'ın (2018) ortaokul öğrencileri ile yürüttüğü çalışmasında öğrenciler rahatsızlık duydukları çevre sorunlarını ilk üç olarak çöp sorunu, hava kirliliği ve su kirliliğini olarak sıralamışlardır. Aynı şekilde Yalçınkaya (2013) tarafından yapılan başka bir çalışmada ise 8.sınıf öğrencilerine göre en önemli çevre sorunları olarak su kirliliği, hava kirliliği, gürültü kirliliği ormanların yok olması, çöp sorunu, doğal afetler, trafik kazaları, gecekondu sorunu ve görüntü kirliliği olarak belirlenmiştir.

Bu araştırmanın bulguları, üniversite öğrencilerinin ekolojik sorunlara ilişkin algılarının, toplumsal bilinçlenme süreçlerinin ve medya etkilerinin bir yansıması olduğunu göstermektedir.

Öğrencilerin en önemli çevre sorunu olarak küresel ısınmayı seçmeleri, bu konuda medya aracılığıyla yayılan bilgilendirme kampanyalarının ve güncel haberlerin etkisini ortaya koymaktadır. Küresel ısınma, özellikle televizyon gibi kitle iletişim araçlarında sıklıkla gündeme geldiği için, bu konuda duyarlılığın yüksek olması anlaşılabilir bir durumdur. Ayrıca, öğrencilerin çevresel sorunları sıralarken farklı fakültelere göre değişen öncelikler, bireylerin eğitimsel ve mesleki arka planlarının, çevre algılarını nasıl şekillendirdiğini göstermektedir. Örneğin, mühendislik fakültesi öğrencilerinin enerji sorunu konusunda daha fazla duyarlılık gösteriyor olmaları, bu konuda aldıkları teknik eğitimin ve mesleki odaklarının bir yansımasıdır.

Araştırmanın sonuçları, ekolojik sorunlara yönelik farkındalığın artırılması gerektiğini ortaya koymaktadır. Eğitim kurumlarının, öğrencilere çevre bilincini aşılamak adına daha kapsamlı ve disiplinler arası bir yaklaşım benimsemeleri gerektiini düşüncesini ortaya çıkarmaktadır. Üniversite öğrencilerinin çevresel sorunlara ilişkin farkındalık düzeylerinin yüksek olmasına rağmen, bu farkındalığın toplumsal düzeyde daha geniş kitlelere yayılabilmesi için farklı stratejiler geliştirilmelidir. Çevresel sorunların daha etkili bir şekilde ele alınabilmesi için, üniversitelerde çevre eğitiminin yanı sıra, öğrencilere uygulamalı proje ve etkinlikler sunulması, ekolojik sorunlara yönelik çözüm odaklı düşünme becerilerinin geliştirilmesi önemlidir.

Sonuç olarak, bu çalışmanın bulguları, üniversite öğrencilerinin çevresel sorunlara olan duyarlılıklarının arttırılmasının ve bu sorunların çözümü için kolektif bir bilinç geliştirilmesinin gerekliliğini vurgulamaktadır. Öğrencilerin

çevresel sorunları daha derinlemesine anlamaları ve bu konuda aktif rol almaları, sürdürülebilir bir gelecek için kritik öneme sahiptir. Gelecekteki araştırmalarda, öğrenci gruplarının çevresel sorunlara yönelik davranışlarını değiştirmeye yönelik müdahalelerin etkinliği üzerine çalışmalar yapılabilir. Ayrıca, diğer yaş grupları ve toplum kesimleriyle yapılan karşılaştırmalı çalışmalar, çevre bilincinin toplum genelinde nasıl geliştirilebileceğine dair daha kapsamlı veriler sağlayabilir.

Kaynakça

Akyüz, E. (2015). Çevre Sorunlari ve İnsan Haklari Ilişkisi. *The Journal of Academic Social Science*, *15*(15), 427-436.

Arık, R. S., & Kutlu, Ö. (2013). Öğretmenlerin ölçme ve değerlendirme alanı yeterliklerinin yargıcı kararlarına dayalı ölçeklenmesi. Eğitim Bilimleri Araştırmaları Dergisi, 3(2), 163-196.

Anıl, D., Ertuna, L., & Uysal, İ. (2015). Türkiye'deki araştırma görevlilerinin mesleki sorunlarının ikili karşılaştırma yoluyla ölçeklenmesi. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6(2), 279- 292.

Anıl, D., & Güler, N. (2006). İkili karşılaştırma yöntemi ile ölçekleme çalışmasına bir örnek. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 30*, 36-39.

Baykal, H., & Baykal, T. (2008). Küreselleşen dünya'da çevre sorunları. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5*(9),1-17

Bramley, T., Bell, J. F., & Pollitt, A. (1998). Assessing changes in standards over time using Thurstone paired comparisons. Education Research and Perspectives, 25 (2), 1-24.

Güler, N. & 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.

Güler, N., İlhan, M., & Taşdelen Teker, G. (2018). İkili Karşılaştırmalarla Ölçekleme Yöntemi ile Rasch Analizinden Elde Edilen Ölçek Değerlerinin Karşılaştırılması. İnönü Üniversitesi Eğitim Fakültesi Dergisi, 19(1), 31-48.

Görmez, K. (2003). *Çevre sorunları ve Türkiye*. Ankara: Gazi Kitapevi.

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, *37*(2), 1-19.

Kan, A. (2008). Yargıcı kararlarına dayalı ölçekleme yöntemlerinin karşılaştırılması üzerine ampirik bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 35*, 186-194.

Kırılmazkaya, G. (2022). Ortaokul öğrencilerinin çevre ve çevre sorunlarına ilişkin görüşleri. *Adnan Menderes Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Dergisi, 13*(1), 1-11.

Nartgün, Z. (2006). Öğretmenlik meslek bilgisi derslerinin önem düzeyinin ikili karşılaştırmalarla ölçeklenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 6*(2), 161-176.

Ömür, S., & Erkuş, A. (2013). Dereceli puanlama anahtarıyla, genel izlenimle ve ikili karşılaştırmalar yöntemiyle yapılan değerlendirmelerin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *28*(28-2), 308-320.

Polat, B., & Göksel, H.G. (2014). Öğretmen adaylarının sosyal aktivite tercihlerinin ikili karşılaştırmalı ölçekleme yöntemiyle belirlenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 5*(1), 88- 100.

Schultz, P. W. (2002). *Knowledge, education, and household recycling: Examining the knowledge-deficit model of behavior change*. In T. C. Kasser & A. S. Blom (Eds.), *The psychology of environmental issues* (pp. 141-165). Psychology Press.

Torgerson, W.S. (1958). Theory and methods of scaling. Newyork: John Wiley & Sons Inc.

Turgut, M. F., & Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM Yayınları.

Türkmen, L. (2010). Ekolojik Konu ve Sorunlar. Orçun Bozkurt (Ed.), *Çevre eğitimi* içinde (s. 153-178). Ankara: Pegem Akademi yayınlar

UNEP (2020). *Global Environment Outlook (GEO-6): Healthy planet, healthy people*. United Nations Environment Programme.

UNFPA (2024). *World population dashboard*. (18/10/2024) tarihinde https://www.unfpa.org/data/world-population-dashboardadresinden ulaşılmıştır).

Yalçınkaya, E. (2013). İlköğretim 8. sınıf öğrencilerine göre çevre sorunları: Nitel bir çalışma. *Marmara Coğrafya Dergisi, 27*, 416-439.

Yılmaz, O., Samsunlu, Ü., & Peker, R. (2017). Üniversite öğrencilerinin çevre sorunlarına ilişkin görüşleri: Isparta İli Örneği. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 43*, 283-301.

Yoloğlu, Ö. Ü. A. C., & Halisdemir, b. (2021). Üniversite öğrencilerinin çevre bilinci ve çevresel tutumları üzerine ampirik bir çalışma: mersin üniversitesi örneği. *The journal of academic social science, 103*, 91-107.

BÖLÜM VI

Bilgisayar Ortamında Bireye Uyarlanan Test Uygulamalarında Kullanıcı Deneyimine Etki Eden Faktörler

Alper TOSUN[1]
Alper SİNAN[2]

Giriş

Teknolojinin gelişmesi ve yaygınlaşmasıyla birlikte bilgisayarlar da eğitim sürecine dahil olmuş ve dolayısıyla ölçme ve değerlendirme amacıyla kullanılan testler de bilgisayar ortamında yapılmaya başlanmıştır (Linden & Glas, 2002). Dijitalleşmiş bir çerçeve içinde geliştirilmiş olan BOBUT uygulamaları geleneksel kağıt kalem test türlerine kıyasla daha esnek ve birey odaklı ölçüm sağlar. Geleneksel kağıt kalem testleri ile BOBUT uygulamaları arasındaki farklar test yapısının bireye uyarlanabilir olması, soru

[1] Akdeniz Üniversitesi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, Antalya/TÜRKİYE, Orcid: 0000-0001-9715-5209 alpertosun.003@gmail.com
[2] Doç.Dr., Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, Antalya/TÜRKİYE, Orcid: 0000-0001-6632-5500, alpsin25@gmail.com

seçimi, test süresi ve sonuç hassasiyeti gibi çeşitli noktalarda farklılık göstermektedir. Geleneksel kağıt kalem testlerinde bireyler aynı soruları aynı sırayla çözmek durumundadır. Bu durum aynı testin yüksek yetenek düzeyindeki bireylere çok kolay gelirken, düşük yetenek düzeyindeki bireylere çok zor gelmesine sebep olabilir. Bu nedenle farklı yetenek düzeyindeki bireylerin değerlendirilmesinde tutarsızlıklar olabilmektedir(Lord, 2012).

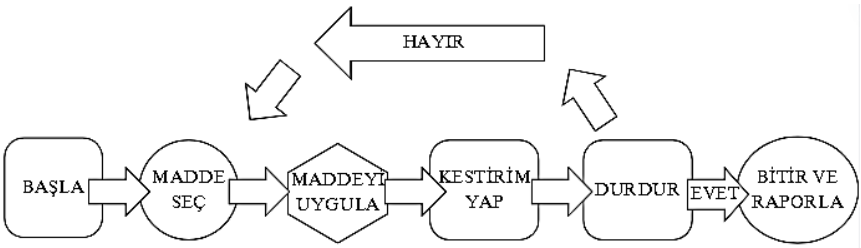
BOBUT uygulamalarında ise test bireyin verdiği cevaplara göre anlık uyarlandığı için bireyler kendi yetenek düzeyinin çok üzerinde ya da çok altında sorularla karşılaşmazlar. Böylece bireylerin yetenek düzeyine uyarlanmış bir test süreci oluşturularak bireyselleştirilmiş ve doğruluğu yüksek sonuçlar elde edilebilir. Bu bakımdan BOBUT uygulamaları kullanıcı deneyimine olumlu katkı vermektedir(Weiss, 1981). Ayrıca kağıt kalem testlerinde soruların önceden düzenlenmiş olmasından kaynaklı olarak madde güçlükleri 0,50 civarında yani orta güçlükte oluşturulmaktadır. Ancak BOBUT uygulamaları için madde güçlükleri bireyin sorulara verdiği cevaplara göre azalış ya da artış gösterebilir(Thorndike & ark. 1991). Geleneksel kağıt kalem testlerinin yapısından kaynaklı olarak test süresi genellikle uzundur ve bireyler kendi yetenek seviyesine uygun olmayan sorular ile karşılaştığı için gerçek performanslarını yansıtamayabilir. Ancak BOBUT uygulamasının amacı daha az sayıda soru ile bireyin yetenek seviyesini kestirmektir. Bu durum geleneksel kağıt kalem testlerine göre BOBUT uygulamalarının daha kısa sürede tamamlanmasına ve bireyin yetenek düzeyine uygun soruların sorulmasıyla bireylerin yetenek düzeylerini daha iyi yansıtmaktadır(Reckase, 1997; Hambleton & Swaminathan, 1985). Ayrıca geleneksel kağıt kalem uygulamaları sonrasında değerlendirme süreci belli bir zaman alırken BOBUT

uygulamalarında testi alan bireyler uygulama sonrasında sonuçlarını alabilmektedir.

Bilgisayar Ortamında Bireye Uyarlanmış Testler(BOBUT) bireylerin yetenek düzeyine uygun soruların bireylere sorulması ve bu sayede bireyler ile ilgili maksimum performansın ölçülmesini amaçlayan testlerdir. BOBUT uygulamalarında yüksek yetenek düzeyine sahip bireyler daha zor sorular ile karşılaşırken, düşük yetenek düzeyine sahip bireyler daha kolay sorular ile karşılaşırlar. Süreç bireyin ilk soruyu cevaplaması ile başlayarak bireyin cevaplarına göre testin bireye uyarlanmış bir şekilde gelmesi ile devam eder. Yani birey eğer ilk soruya doğru cevap verdiyse daha zor bir soru ile yanlış cevap verdiyse daha kolay bir soru ile karşılaşır(Wainer & vd. 2000). BOBUT birey ile ilgili maksimum bilginin elde edilmesini bireye özelleştirilmiş olan test formları ile elde etmektedir(Bartram & Hambleton, 2006). BOBUT uygulamalarında bireylerin yetenek düzeyine göre bireyselleşmiş testler alabilmesine imkan sağlanması geniş bir madde havuzu gerekliliğini de ortaya koymaktadır. BOBUT için madde havuzu çalışması planlanan konuları kapsayan ve farklı güçlük düzeylerine sahip sorulardan oluşturulmalıdır(Weiss, 1981). Geniş madde havuzunun olması gerekliliği BOBUT uygulamalarının en büyük sınırlılığıdır. BOBUT uygulamaları bireyin cevapları doğrultusunda yetenek düzeyini farklı istatistiksel yöntemleri kullanarak kestiren ve test sonucunda bireyin yeteneğini yüksek güvenirlikle ortaya koymaya çalışan uygulamalardır(Frederiksen & vd. 1993). BOBUT bireylerin yetenek kestirimlerini daha hassas bir şekilde ölçerken bireye gereksiz soruların sorulmaması açısından test uzunluğu ve süresi bakımından kullanışlılık sağlar. Sonuç olarak, değerlendirme süreci bireye göre uyarlanır ve en uygun sonuçları verir.

Yöntem

BOBUT uygulamaları bilgisayar ortamında hazırlanmış olup doğası gereği tüm bireyler için bireyselleştirmiş bir sınav süreci oluşturmaktadır. BOBUT uygulaması öncelikle bireye sorulacak ilk soru ile başlayan, öğrencinin yanıtı doğrultusunda yetenek kestirimi yapılan ve sonraki soru seçimi ile devam eden iteratif bir süreçtir. Test sonlandırma kuralına göre gerekli kriterler sağlandığı takdirde test sonlandırılır ve değerlendirme uygulanır(Van Der Linden & Glas, 2010; Kingsbury & Weiss, 1983; Chang & Ying, 1996). BOBUT uygulamalarında sürecin gerçekleşmesi şekil 1'deki gibidir.



Şekil 1. *BOBUT Akış Şeması(Davey, 2011).*

Wainer(2000) BOBUT uygulamalarında sürecin nasıl gerçekleştirileceği ile ilgili olarak üç temel soru üzerinden durmaktadır. Bunlardan ilki testi başlatmak için madde seçiminin nasıl yapılacağı; ikincisi birey maddeyi cevapladıktan sonra yetenek kestirimin nasıl yapılacağı ve sonraki sorunun nasıl seçileceği ve son olarak da testin nasıl sonlandırılacağıdır. Araştırmacıları BOBUT uygulamalarını tasarlarken bu üç soru çerçevesinde çeşitli seçimler yaparak BOBUT uygulamalarını tasarlamaktadırlar. BOBUT uygulamalarının tasarlanması sürecinde üzerinde durulması gereken bir diğer husus ise madde tepki kuramının kullanımıdır.

BOBUT uygulamalarında madde tepki kuramı

Eğitimde ve psikolojide testlerin oluşturulması sürecinde ve testler aracılığıyla elde edilen puanların yorumlanmasında kullanılan modeller test geliştirme uzmanlarına ve test kullanıcılarına uzun bir süredir hizmet etmektedir. Testlerin geliştirilmesinde kullanılan modeller incelendiğinde alan yazında uzun süredir var olan model klasik test kuramıdır. Klasik test kuramı ile oluşturulan modeller içerisinde barındırdığı çeşitli yaklaşımlarla birlikte test geliştiricilere uzun süre hizmet etmiştir(Hambleton,2013). Ancak klasik test kuramının bazı sınırlılıkları farklı test geliştirme modellerinin ortaya çıkmasına sebep olmuştur. Test modellerinin gelişiminde temel amaç bireyin yeteneğinin geçerli ve güvenilir bir şekilde ölçülmesidir. Klasik test kuramı yeteneğin ölçümü ile ilgili bulunan bazı sorunların çözümü adına test geliştiricileri madde tepki kuramını geliştirmiştir. Madde tepki kuramının klasik test kuramı ile farklılıkları incelendiğinde(Koğar, 2021);

Güvenirliğin arttırılması için KTK' da yapılan uygulamalardan birisi de test duyarlılığının arttırılmasıdır. Testin duyarlılığının arttırılması test ile daha hassas ölçümler yapılabilmesine olanak sağlamaktadır. KTK madde sayısı ile güvenirliğin doğrusal bir ilişki içerisinde olduğunu savunmaktadır. Yani madde sayısı arttıkça güvenirlikte artmaktadır ve test daha uygun olmaktadır(Alkan & Sinan, 2020). Bu durum test için soru doyum noktasına ulaşana kadar devam eder. Ancak MTK için bireyin yetenek düzeyi ile ilgili en çok bilgiyi veren birkaç madde ile maksimum güvenirlik sağlanabilmektedir. MTK algoritması gereği doğru madde ile doğru bireyi eşleştirme yapabildiği için birey için en yüksek bilgi sağlayan madde seçilmesi mümkündür. Yani testin güvenirliği yetenek düzeyine bağlı olarak değişkenlik

gösterebilmektedir(Baker, 2001; Hambleton, 1991). Bu sayede güvenirlik daha az madde seçimi ile sağlanabilmektedir.

Bazı geniş ölçekli sınavlarda güvenirliği sağlamak ve kopya riskinin önüne geçebilmek adına farklı test formları geliştirilmektedir(Örn ALES, CAT). Farklı test formlarından elde edilen puanların karşılaştırılması gerekmektedir. KTK farklı test formlarından elde edilen puanların karşılaştırılabilmesi için testlerin paralel olması varsayımını gerektirir. Ancak uygulamalar arasında testlerin paralel olmasını sağlamak oldukça güçtür. MTK madde odaklı yaklaşımı ve farklı testlerde bulunan maddelerin aynı ölçek düzeyinde doğrudan karşılaştırılabilmesi açısından doğrudan karşılaştırma yapabilmektedir. MTK doğrusal olmayan yöntemler sayesinde testlerin paralelliği koşulunu gerektirmez(Kolen, 2004; Baker & Kim, 2004; Lord, 2012).

Bireyin grup içerisindeki sıralaması ve başarısını belirlemek için KTK'da dağılımın normallik göstermesi ve puanların standart skorlara dönüştürülerek bireyin grup içerisindeki yerinin belirlenmesi gerekmektedir. Bu uygulama kriter referanslı ölçümler için hiçbir anlama gelmezken norm referanslı ölçümlerde kullanılabilir. Ancak MTK kişi ve maddeyi aynı ölçek düzeyinde değerlendirdiği için yetenek düzeylerini karşılaştırmak mümkündür. Normallik varsayımı sağlanmasa dahi ve test puanların standart skorlara dönüştürülmesine gerek olmaksızın yetenek düzeylerine göre bireyin sıralaması ve başarısı belirlenebilir(Ayala, 2013; Kleine, 2005; Cohen & Swerdlik & Phillips, 1996).

Ölçmenin standart hatası KTK için tüm yetenek seviyelerinde eşit kabul edilmektedir. Ancak MTK, her yetenek seviyeleri için özel bir güvenirlik sunmaktadır. Bu durum bireyselleştirilmiş testler için

MTK ‘yı daha etkin ve tercih edilir kılar. KTK’da güvenirlik madde sayısına bağlıyken MTK’da daha az sayıda madde ile güvenirlik sağlanabilmektedir. Bu durum bireylerin daha az madde ile performanslarını yansıtmasına imkan sağlayarak MTK’yı daha kullanışlı yapar(Baker, 2001; Hambleton, 1991).

Klasik test kuramı uygulamalarında madde parametreleri gruptan bağımsız kestirilemez ve kestirimler testi alan grubun özelliklerine bağımlıdır. Buna karşılık MTK, madde parametrelerini gruptan bağımsız doğru bir şekilde kestirebilmekte ve farklı örneklemlerde dahi yanlılığa sebep olmamaktadır. MTK’nın bu avantajı sayesinde geçerliliği ve güvenirliği kanıtlanmış geniş madde havuzları oluşturulabilir olması özellikle BOBUT gibi uygulamalarda bireyin yetenek seviyesine uygun soruların sorulması ile test süresini kısaltıcı ve kullanıcı deneyimine iyileştirici bir etki oluşturmaktadır(Embretson & Raise, 2000; Camilli, 1994; Baker, 2001; Hambleton, 1991).

Eşit aralık düzeyinde verilerin elde edilmesinin KTK için zor olmasından dolayı, düşük yetenek düzeyine sahip bir bireyin puanının 1 puan arttırması ile yüksek yetenek düzeyine sahip bireyin puanını 1 puan arttırması aynı anlama gelmemektedir. Bu durum özellikle deneysel desenlerde etkinin gözlemlenmesini zorlaştırmaktadır. MTK için ise eşit aralık düzeyi ölçeğin elde edilmesi daha kolay olduğu için puan artışlarının ne anlama geldiğini ortaya koyabilmek çok daha kolaydır. Kullanıcı deneyimi açısından MTK’nın sunduğu bu avantaj bireylerin sonuçlara olan güvenini arttırır ve test sürecini daha anlamlı hale getirir(Embretson & Raise, 2000; Crocker & Algina, 1986).

KTK ile MTK arasındaki çeşitli farklılıkların yanı sıra iki model de testi alan bireylerin verdiği yanıtlara odaklanmaktadır. Ayrıca iki model için de test ölçülmesi hedeflenen özelliği başka özelliklere karışmadan ölçebilmeli ve bireylerin bir maddeye verdiği cevabın diğer cevaplarını etkilemeyecek şekilde kurgulanması gerekmektedir(Çıkrıkçı & Demirtaşlı 1998).

Madde tepki kuramı bir ölçme aracı ile ölçülen yetenek ile bir madde arasındaki ilişkinin incelenmesi adına geliştirilmiş olan bir kuramdır. Maddeye doğru yanıt verme ve yetenek arasında ortaya çıkan *lojisti*k bir regresyondur. Madde tepki kuramı lojistik bir özellik gösteren madde karakteristik eğrisine(MKE) sahiptir. MKE madde tepki kuramının anlaşılabilmesinde en önemli gerekliliktir. MKE lojistik temeli gereği "S" şeklinde bir eğridir. (Sünbül, 2020; Crocker & Algina, 2006b)

Madde tepki kuramının kullanılabilmesi için iki temel varsayım bulunmaktadır. Bunlar yerel bağımsızlık ve tek boyutluluktur. Yerel bağımsızlık, bireylerin maddelere verdikleri yanıtların istatistiksel olarak birbirlerinden bağımsız olmasıdır (Sünbül, 2020). Tek boyutluluk, geliştirilmek istenen testin tek bir faktörü ölçüyor olması gerekliliğidir. (Hambleton vd., 1991).

Madde tepki kuramı modelleri maddelerin puanlanma şekillerine ve testin amacına göre farklılık gösterebilmektedir. İkili(dichotomous) ve çoklu(polytomous) puanlanan maddeler farklı modelleri ile puanlanabilmektedir. İkili puanlanan maddeler için madde tepki kuramı modelleri a, b, c ve d olmak üzere dört parametreye sahiptir. "a" parametresi madde ayırt edicilik parametresidir ve maddenin bilen ve bilmeyeni ne kadar ayırt ettiğini göstermektedir. "b" parametresi madde güçlük parametresidir ve

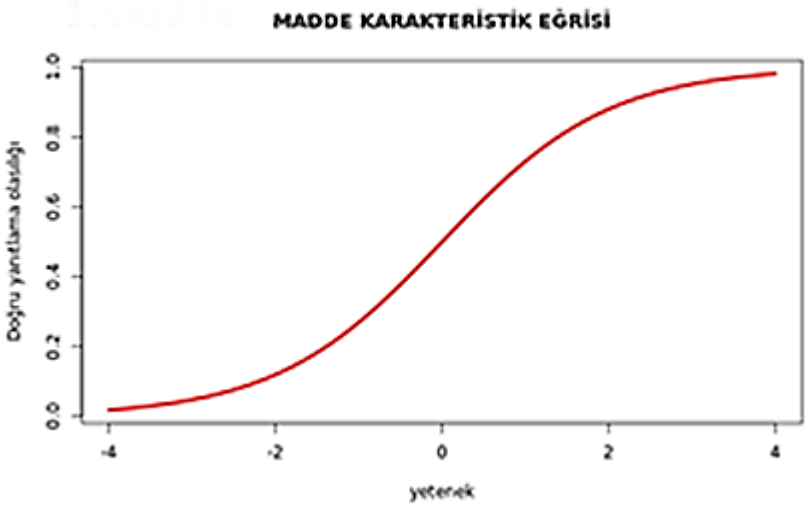
madde karakteristik eğrisinin konumunu göstermektedir. “c” parametresi şans parametresidir ve düşük asimptot olarak ifade edilmektedir yani düşük yetenek düzeyine sahip bireyin maddeyi tahmin ile doğru cevaplayabilmesi anlamına gelmektedir. “d” parametresi üst asimptot parametresidir yani yüksek yetenek düzeyine sahip bireyin maddeyi dikkatsizlik ile yanlış cevaplandırmasıdır (Baker & ark., 2016). İkili(dichotomous) ve çoklu(polytomous) puanlanan maddeler farklı modelleri ile puanlanabilmektedir. 1PL model için madde güçlükleri(b) farklılık gösterirken madde ayırt edicilikleri(a) sabit tutulmaktadır. 2PL model için madde güçlükleri(b) ve madde ayırt edicilikleri(a) kestirilirken modelde şans parametresi(c) modelde bulunmamaktadır. 3PL model için madde güçlükleri(b), madde ayırt edicilikleri(a) ve düşük asimptot olarak adlandırılan şans parametreleri(c) kestirilmektedir(Baker & Kim, 2017).4PL model madde güçlükleri(b), madde ayırt edicilikleri(a), şans parametresi(c) ve dikkatsizlik parametresini(d) içermektedir. (Veerkamp & Berger 1997).

1PL model için madde güçlükleri(b) farklılık gösterirken madde ayırt edicilikleri(a) sabit tutulmaktadır. Şans parametresi(c) ise modele dahil edilmemektedir. Madde ayırt edicilikleri(a) 1’e sabitlenirse model “Rasch Model” olarak isimlendirilmektedir. 1PL model ve rasch model için kullanılan olasılık formülü;

$$P(\theta) = \frac{e^{\theta-b}}{1+e^{-1(\theta-b)}}$$

İncelendiğinde “b” madde güçlüğünü, “θ” yeteneği, “e” matematiksel sabittir ve değeri 2.718 olarak alınmaktadır. Model için sadece madde güçlük(b) parametresi kullanıldığı için farklı maddeler için madde karakteristik eğrileri kesişmemektedir. 1PLM

için madde parametreleri kestirilmiş olan dört farklı maddeye ait MKE'ler incelendiğinde;



***Şekil 2**: 1PLM İçin MKE -1*

Kaynak: Veerkamp ve Berger 1997

Şekil 2'de görülmekte olan madde karakteristik eğrisi 1PL modele aittir. Modelde sadece madde güçlük parametresi bulunmaktadır. Şekil için madde güçlüğü 0.0 düzeyindedir.

Şekil 3*: 1PLM İçin MKE -2*

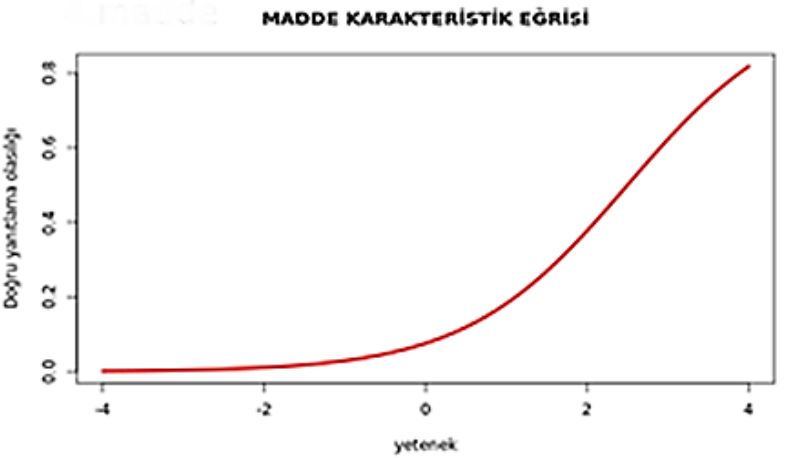
Kaynak: Veerkamp ve Berger 1997

Şekil 3'de görülmekte olan madde karakteristik eğrisi 1PL modele aittir. Modelde sadece madde güçlük parametresi bulunmaktadır. Şekil için madde güçlüğü -1.5 düzeyindedir. Şekil 2'de gösterilen maddeye göre daha kolay bir soruyu temsil etmektedir.

Şekil 4*: 1PLM İçin MKE -3*

Kaynak: Veerkamp ve Berger 1997

Şekil 4'de görülmekte olan madde karakteristik eğrisi 1PL modele aittir. Modelde sadece madde güçlük parametresi bulunmaktadır. Şekil için madde güçlüğü 1.0 düzeyindedir. Şekil 2 ve şekil 3' de gösterilen maddelere göre daha zor bir maddeyi temsil etmektedir.



Şekil 5*: 1PLM İçin MKE -4*

Kaynak: Veerkamp ve Berger 1997

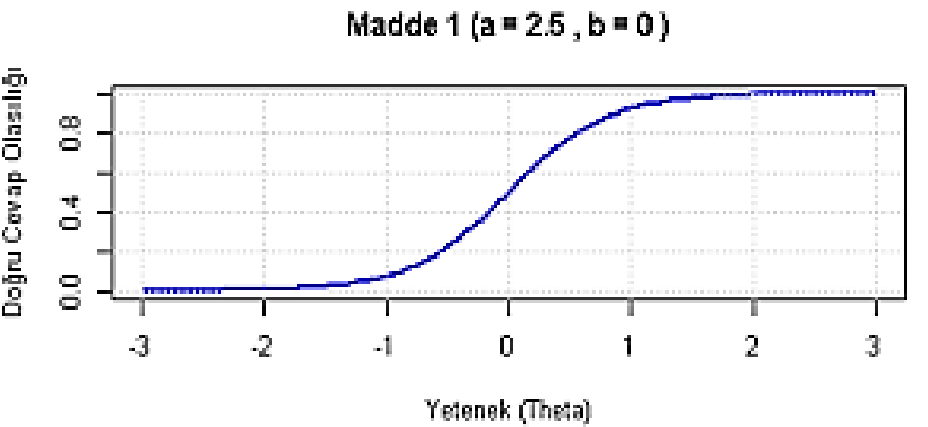
Şekil 5'de görülmekte olan madde karakteristik eğrisi 1PL modele aittir. Modelde sadece madde güçlük parametresi bulunmaktadır. Şekil için madde güçlüğü 2.0 düzeyindedir. Şekil 2, şekil 3 ve şekil 4'de gösterilen maddelere göre daha zor bir maddeyi temsil etmektedir.

MKE'ler lojistik temeli gereği "S" şeklinde oluşma eğilimindedir. Farklı güçlüklere sahip MKE'ler incelendiğinde eğrinin 0.5 doğru cevaplanma olasılığına geldiği noktalar maddelerin güçlük parametreleri vermektedir. Buna göre 1.madde 0 güçlük parametresine, 2.madde -1.5 güçlük parametresine, 3.madde 1.0 güçlük parametresine ve 4.madde 2.0 güçlük parametresine sahiptir. Dört madde içerisinden 2.madde en kolay madde iken, 4.madde en zor maddeyi temsil etmektedir.

2PL model için madde güçlükleri(b) ve madde ayırt edicilikleri(a) kestirilirken modelde şans parametresi(c) modelde bulunmamaktadır. 2PL model için olasılık formülü;

$$P(\theta) = \frac{e^{a(\theta-b)}}{1+e^{-a(\theta-b)}}$$

İncelendiğinde 1PL model için kullanılan olasılık fonksiyonuna sadece "a" parametresi yani madde ayırt edicilik parametresinin eklenmesi ile oluşmaktadır. 2PL model için madde karakteristik eğrileri "a" parametresinin modele dahil edilmesinden dolayı kesişebilmektedir. 2PLM için madde parametreleri kestirilmiş olan dört farklı maddeye ait MKE'ler incelendiğinde;

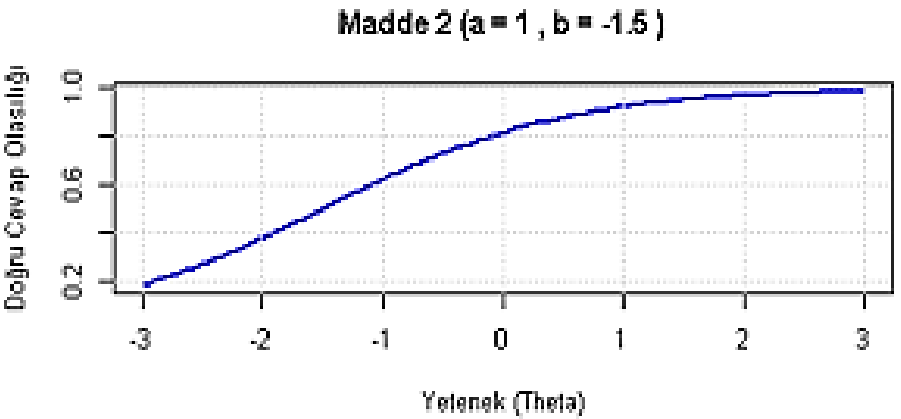


***Şekil 6**: 2PLM İçin MKE-1*

Kaynak: Veerkamp ve Berger 1997

Şekil 6'da görülmekte olan madde karakteristik eğrisi 2PL modele aittir. Modelde madde güçlük parametresi ve madde ayırt

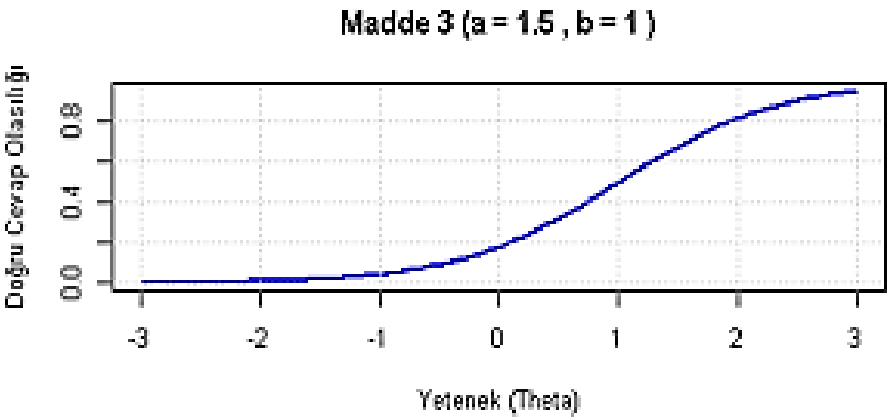
edicilik parametreleri bulunmaktadır. Şekil için madde güçlüğü 0.0 ve madde ayırt ediciliği 2.5 düzeyindedir.



***Şekil** 7: 2PLM İçin MKE-2*

Kaynak: Veerkamp ve Berger 1997

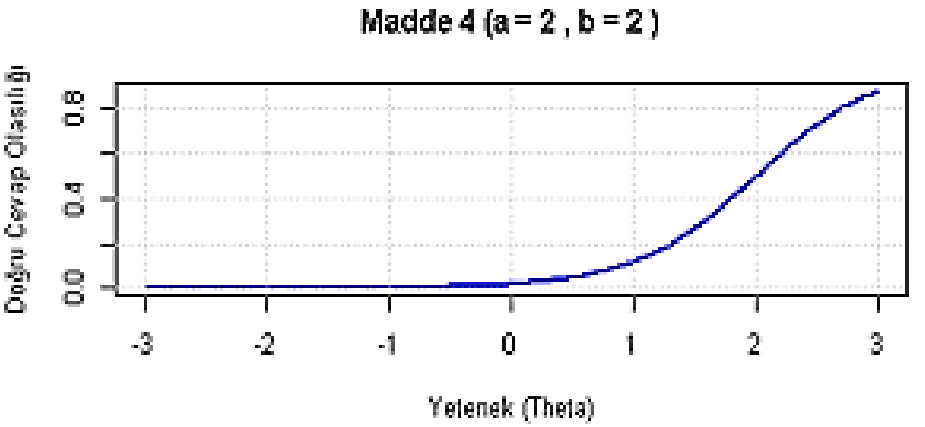
Şekil 7'de görülmekte olan madde karakteristik eğrisi 2PL modele aittir. Modelde madde güçlük parametresi ve madde ayırt edicilik parametreleri bulunmaktadır. Şekil için madde güçlüğü -1.5 ve madde ayırt ediciliği 1.0 düzeyindedir. Şekil 6'da gösterilen maddeye göre madde ayırt ediciliği daha düşük düzeydedir. Ayrıca daha kolay bir soruyu temsil etmektedir.



Şekil 8*: 2PLM İçin MKE-3*

Kaynak: Veerkamp ve Berger 1997

Şekil 8'de görülmekte olan madde karakteristik eğrisi 2PL modele aittir. Modelde madde güçlük parametresi ve madde ayırt edicilik parametreleri bulunmaktadır. Şekil için madde güçlüğü 1.0 ve madde ayırt ediciliği 1.5 düzeyindedir. Şekil 6'da gösterilen maddeye göre madde ayırt ediciliği daha düşük düzeydedir. Ayrıca daha zor bir soruyu temsil etmektedir. Şekil 7'de gösterilen maddeye göre madde ayırt ediciliği daha yüksek düzeydedir. Ayrıca daha zor bir maddeyi temsil etmektedir.



Şekil 9*: 2PLM İçin MKE-4*

Kaynak: Veerkamp ve Berger 1997

Şekil 9'da görülmekte olan madde karakteristik eğrisi 2PL modele aittir. Modelde madde güçlük parametresi ve madde ayırt edicilik parametreleri bulunmaktadır. Şekil için madde güçlüğü 2.0 ve madde ayırt ediciliği 2.0 düzeyindedir. Şekil 6'da gösterilen maddeye göre madde ayırt ediciliği daha düşük düzeydedir. Ayrıca daha zor bir soruyu temsil etmektedir. Şekil 7'de gösterilen maddeye göre madde ayırt ediciliği daha yüksek düzeydedir. Ayrıca daha zor bir maddeyi temsil etmektedir. Şekil 8'de görülen maddeye göre madde ayırt ediciliği daha yüksek düzeydedir. Ayrıca daha zor bir maddeyi temsil etmektedir.

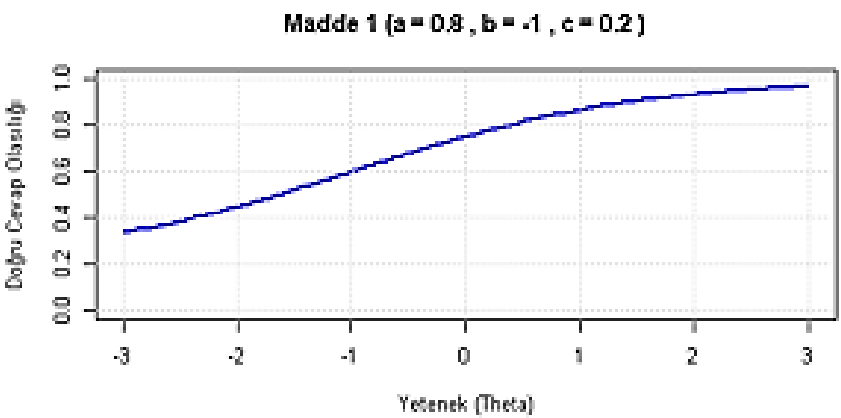
MKE'lerin çizilmesinde madde güçlük parametresi ve madde ayırt edicilik parametreleri kullanılmıştır. "a" parametresi madde ayırt ediciliklerini temsil etmektedir ve "a" parametresindeki artış MKE'nin eğiminin artması ile doğru orantılıdır. Buna göre en yüksek ayırt edicilik parametresine sahip olan madde 1 için eğimin daha keskin bir artış trendinde olduğu gözlenirken, en düşük ayırt

edicilik parametresine sahip olan madde 2 için eğimin artış trendinin keskinliği daha düşüktür.

3PL model için madde güçlükleri(b), madde ayırt edicilikleri(a) ve düşük asimptot olarak adlandırılan şans parametreleri(c) kestirilmektedir 3PL model için olasılık formülü;

$$P(\theta) = c + (1 - c)\frac{e^{a(\theta-b)}}{1+e^{-a(\theta-b)}}$$

2PL modele şans parametresinin(c) eklenmesi ile oluşmaktadır. "c" parametresinin modele dahil edilmesi ile birlikte madde karakteristik eğrileri sıfır düzeyinden farklı düzeylerde başlayabilmektedir(Baker ve Kim, 2017). 3PLM için madde parametreleri kestirilmiş olan dört farklı maddeye ait MKE'ler incelendiğinde;

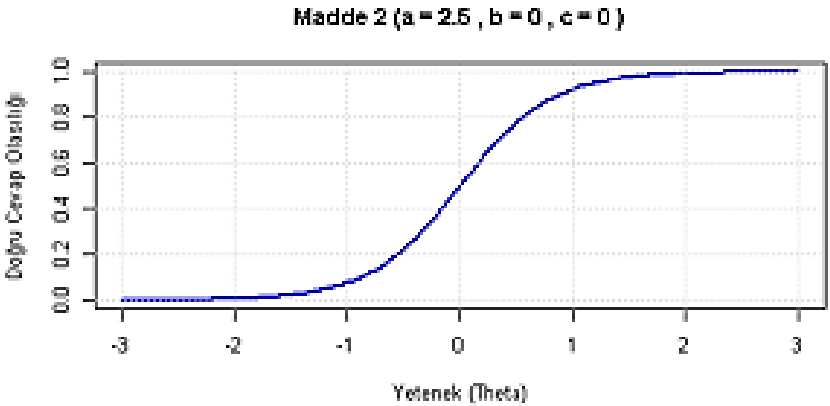


***Şekil 10**: 3PLM İçin MKE-1*

Kaynak: Veerkamp ve Berger 1997

Şekil 10'da görülmekte olan madde karakteristik eğrisi 3PL modele aittir. Modelde madde güçlük parametresi, madde ayırt

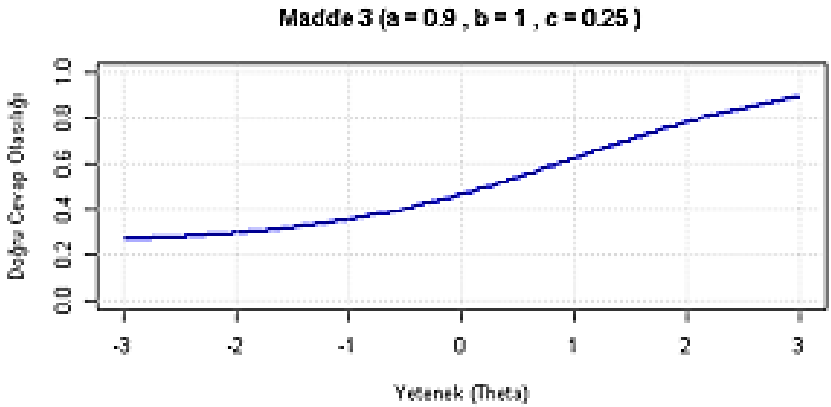
edicilik parametresi ve şans parametresi bulunmaktadır. Şekil için madde güçlüğü -1.0, madde ayırt ediciliği 0.8 ve şans parametresi 0.2 düzeyindedir.



***Şekil 11**: 3PLM İçin MKE-2*

Kaynak: Veerkamp ve Berger 1997

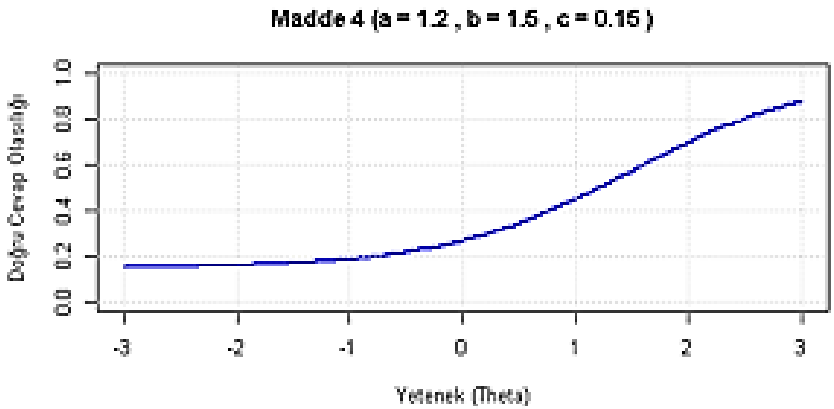
Şekil 11'de görülmekte olan madde karakteristik eğrisi 3PL modele aittir. Modelde madde güçlük parametresi, madde ayırt edicilik parametresi ve şans parametresi bulunmaktadır. Şekil için madde güçlüğü 0.0, madde ayırt ediciliği 2.5 ve şans parametresi 0.0 düzeyindedir. Şekil 10'da gösterilen maddeye göre madde ayırt ediciliği daha yüksektir. Ayrıca daha zor bir soruyu temsil etmektedir. Madde için şans ile doğru cevaplanama olasılığı şekil 10'da temsil edilen maddeye göre daha düşüktür.



***Şekil 12**: 3PLM İçin MKE-3*

Kaynak: Veerkamp ve Berger 1997

Şekil 12'de görülmekte olan madde karakteristik eğrisi 3PL modele aittir. Modelde madde güçlük parametresi, madde ayırt edicilik parametresi ve şans parametresi bulunmaktadır. Şekil için madde güçlüğü 1.0, madde ayırt ediciliği 0.9 ve şans parametresi 0.25 düzeyindedir. Şekil 10'da gösterilen maddeye göre madde ayırt ediciliği daha yüksektir. Ayrıca daha zor bir soruyu temsil etmektedir. Madde için şans ile doğru cevaplanama olasılığı şekil 10'da temsil edilen maddeye göre daha yüksektir. Şekil 11'de temsil edilen maddeye göre madde ayırt ediciliği daha düşük düzeydedir. Ayrıca madde daha zor bir soruyu temsil etmektedir ve şans parametresi daha yüksektir.



***Şekil 13**: 3PLM İçin MKE-4*

Kaynak: Veerkamp ve Berger 1997

Şekil 13'de görülmekte olan madde karakteristik eğrisi 3PL modele aittir. Modelde madde güçlük parametresi, madde ayırt edicilik parametresi ve şans parametresi bulunmaktadır. Şekil için madde güçlüğü 1.5, madde ayırt ediciliği 1.2 ve şans parametresi 0.15 düzeyindedir. Şekil 10'da gösterilen maddeye göre madde ayırt ediciliği daha yüksektir. Ayrıca daha zor bir soruyu temsil etmektedir. Madde için şans ile doğru cevaplanama olasılığı şekil 10'da temsil edilen maddeye göre daha düşüktür. Şekil 11'de temsil edilen maddeye göre madde ayırt ediciliği daha düşük düzeydedir. Ayrıca madde daha zor bir soruyu temsil etmektedir ve şans parametresi daha yüksektir. Şekil 12'de gösterilen maddeye göre madde ayırt ediciliği daha yüksek düzeydedir. Ayrıca daha zor bir soruyu temsil etmektedir ve şans parametresi daha düşük düzeydedir.

MKE'lerin çizilmesinde madde güçlüğü, madde ayırt ediciliği ve şans parametreleri kullanılmıştır. 3PLM ile modele dahil

edilen “c” parametresi şans parametresini temsil etmektedir. “c” parametresi düşük yetenek düzeyine sahip bireylerin maddeyi şans tercihi ile doğru cevaplandırmasıdır ve MKE’nin başlangıç noktasını etkilemektedir. Buna göre 0.0 “c” parametresine sahip 2.madde için MKE 0.0 doğru cevaplanma olasılığından başlarken; 0.25 “c” parametresine sahip madde 3 için MKE daha yüksek bir doğru cevaplanma olasılığından başlamaktadır.

4PL model ise araştırmalarda kullanılan yeni bir yaklaşımdır. 4PL model madde güçlükleri(b), madde ayırt edicilikleri(a), şans parametresi(c) ve üst asimptot parametresini(d) içermektedir. 4PL model için olasılık fonksiyonu;

$$P(\theta) = c + (d - c)\frac{e^{a(\theta-b)}}{1+e^{-a(\theta-b)}}$$

3PL modele “d” parametresinin eklenmesi ile olasılık fonksiyonu oluşmaktadır. “d” parametresi yüksek yetenek düzeyine sahip bireylerin maddeyi doğru cevaplandırmalarında ulaşabilecekleri maksimum değeri belirtmektedir(Veerkamp & Berger 1997).

İlk madde seçimi

BOBUT uygulamalarında madde tepki kuramı modeli seçiminden sonra test geliştiricisinin karar vermesi gereken sonraki aşama bireylerin sınavda çözecekleri ilk sorunun belirlenmesidir. İlk sorunun seçimi ölçüm sürecine yüksek etkiye sahip olmasa da testi alan bireylere psikolojik etkileri olduğu göz ardı edilmemelidir. Emraston ve Reise(2000) çalışmalarında ilk soru konusundaki fikirlerini şu şekilde açıklamaktadır; Eğer ilk soru çok kolay olursa testi alan birey rehavete kapılarak dikkatsiz yanıtlar verebilmektedir. Öte yandan ilk soru bireye çok zor gelecek şekilde seçilirse bireyin

motivasyonu düşebilir ve başarısızlık hissine kapılabilir. Bu tarz durumların oluşmaması için bireye sorulacak olan ilk sorunun seçimi testin güvenirliği açısından son derece önemlidir. Testin ilk maddesinin seçimi için uzmanlar çeşitli yöntemlere başvurmaktadır. Testi alan grup ile ilgili önsel bilgiye sahip olunması durumunda ilk sorunun grubun yetenek düzeyine göre belirlenmesi yaygın kullanılan bir yöntemdir(Thissen & Wainer, 2001). Test geliştirme uzmanlarının kullandığı bir diğer yöntem ise asıl BOBUT uygulaması öncesi bireylere pilot madde seti uygulaması yaparak bireylerin yetenek seviyeleri doğrultusunda ilk madde seçiminin yapılmasıdır(Van Der Linden & Glas, 2010). Bazı durumlarda BOBUT uygulamasını yapan test geliştirici grup hakkında önsel bilgiye sahip olmayabilir ya da pilot uygulama yapabilme imkanı bulunmayabilir. Bu tarz durumlarda test geliştirme uzmanları orta güçlüğe sahip tek bir soru veya orta güçlükte olduğu bilinen bir grup sorudan rastgele bir sorunun seçimini kullanabilmektedir(Eggen, 1999).

Yetenek kestirimi

İlk soruya bireyin cevap vermesinin ardından bireyin cevabına göre yetenek kestirimi yapılmaya başlanmaktadır. Yetenek kestirim sürecinde test geliştirme uzmanları yaygın olarak En Çok Olabilirlik Tahmini(Maximum Likelihood Estimation-MLE) ve Bayesian yöntemlerden faydalanmaktadır. Ancak test geliştirme uzmanları farklı durumlarda ihtiyaçlarına yönelik olarak farklı yetenek kestirim yöntemlerinden de faydalanmaktadır.

En çok olabilirlik yöntemi(maximum likelihood estimation-MLE)

Testi alan bireylerin yetenek düzeyi tahminlerini belirlemek için olabilirlik fonksiyonunu maksimum yapan θ tahmin değerini arar. Bu yöntem bireyin cevaplarına göre en olası yetenek düzeyini tahmin etmektedir. MLE Newton-Rabson, Fisher Skorlama ve beklenti maksimizasyonu gibi iteratif yöntemler ile çözülmektedir. MLE için olabilirlik fonksiyonu;

$$L(\theta) = \prod_{i=1}^{n} P_i(k_i|\theta)$$

Eşitlikte θ; yetenek düzeyini, $P_i(k_i|\theta)$; Testi alan bireyin θ yetenek düzeyinde i. maddeye doğru cevap verme olasılığını, n; madde sayısını temsil etmektedir. Eşitlik optimizasyon algoritmaları kullanılarak çözüme kavuşturulmaktadır(Lord, 2012). MLE yöntemi tarafsız bir kestirim yöntemi olmasından dolayı yaygın olarak kullanılmaktadır. Ancak mükemmel cevap örüntüsünün olduğu durumlarda yani testi alan bireyin tüm soruları doğru veya tüm soruları yanlış cevapladığı durumlarda kestirim yapamamaktadır. Ayrıca madde havuzunun dar olması durumunda kestirim doğruluğu düşebilmektedir(Baker, 1998).

Bayesian yöntemler

Bayesian yöntemler MLE'den farklı olarak önsel(prior) ve sonsal(posterior) dağılımları kullanarak yetenek tahmininde bulunmaktadır. Testi alan bireylerin yetenek düzeyi hakkındaki varsayımlara dayanarak tahmin süreci başlatılır ve bireyin verdiği cevaplara göre θ tahmini sürekli güncellenerek devam etmektedir. Mükemmel cevap örüntülerinde yani bir katılımcının tüm sorulara

doğru yanıt verdiği ya da yanlış yanıt verdiği durumlarda kestirim yapabiliyor olması MLE'ye göre Bayesian yöntemleri kullanışlı hale getirmektedir.

Maximum a priori(MAP)

Kestirim testi alan bireyin testi almadan önceki önsel bilgisine dayalı olarak maksimum tahmin değerine ulaşmaya çalışmaktadır. Başlangıçta aritmetik ortalaması 0 ve standart sapması 1 olan ve normal dağılımdan gelen önsel bilgi dağılımına göre en olası tahmin değeri bulunur. Daha sonra bireyin maddelere verdiği doğru veya yanlış yanıtlara göre oluşan olabilirlik fonksiyonu kullanarak tahmin süreci devam eder(Bock & Aitkin, 1981).

Expected a posteriori(EAP)

EAP kestirimi MAP kestiriminden farklı olarak tüm sonsal dağılıma göre yetenek kestirimi yapmaktadır. MAP kestirimi tek bir maksimum değeri hedeflerken, EAP kestirimi sonsal dağılımın aritmetik ortalamasına göre kestirim yapmaktadır. EAP kestirimi bu süreçte;

$$\theta_{EAP} = \int_{-\infty}^{\infty} \theta . P(\theta|k) d\theta$$

İle çözülmektedir. Eşitlikte θ_{EAP} sonsal dağılımın ortalamasını, $P(\theta|k)$ sonsal dağılımı, $d\theta$ entegrasyon işlemini ifade etmektedir(Ho, 2010). EAP test sürecinde yanıt dengesizliği olan durumlarda daha dayanıklıdır ve az sayıda madde yanıtlandığı durumlarda daha kestirim yapabilmesinden dolayı test geliştirme uzmanları tarafında yaygın bir şekilde kullanılmaktadır(Hambleton & Swaminathan, 2013).

Madde seçme yöntemleri

Test geliştirme uzmanının araştırmanın amacına yönelik olarak belirlediği yetenek kestirim yöntemi sonrasında testi alan bireylerin doğru veya yanlış cevaplar verdiği durumda sorulacak bir sonraki sorunun belirlenmesi BOBUT uygulamalarında önemli noktalardan birisidir. Test geliştirme uzmanları beklentilerine göre farklı madde seçme yöntemleri kullanmaktadır. BOBUT uygulamalarında yaygın olarak Maksimum Fisher Bilgi(MFB), Kullback-Leibler(KL) Bilgisi, Aralık Bilgisi Ölçütü(ABÖ), En Uygun b Değeri, Bayesian Yöntemler gibi çeşitli yöntemler kullanmaktadırlar.

Maksimum fisher bilgi(MFB)

Maksimum Fisher Bilgi ölçütü testi alan bireyin çözdüğü her bir sorudan sonra bireyin yetenek düzeyi ile ilgili en yüksek bilgiyi verecek maddenin seçilmesi temeline dayanmaktadır. Ancak uygulamanın başında birey ile ilgili önsel bilgi mevcut değilse seçilecek sorunun yanlılık oluşturma ihtimali olabilir, madde seçim algoritması yüksek ayırt ediciliğe sahip maddeleri seçmeye meyillidir(Han, 2009). MFB eşitliği;

$$I_i(\theta) = \frac{a_i^2 . e^{-Da_i(\theta - b_i)}}{(1 + e^{-Da_i(\theta - b_i)})^2}$$

Şeklindedir. Formülde $I_i(\theta)$ yetenek düzeyine göre fisher bilgi fonksiyonunu, $a_i$madde ayırt edicilik parametresini, $b_i$madde güçlük parametresini, Dölçekleme sabitini, $e^{-Da_i(\theta - b_i)}$ekspansiyel terimi, $\left(1 + e^{-Da_i(\theta - b_i)}\right)^2$ normalizasyon faktörünü ifade etmektedir(Birnbaum, 1968; Veerkamp & Berger 1997).

Bayesian yöntemler

Bayesian yöntemler bireyin her bir cevap verdiği sorudan sonra yetenek seviyesini güncellenerek yetenek tahminini en doğru şekilde yapacak soruyu seçmeyi amaçlamaktadır. Temelde süreç bireyin sorulara verdiği cevaplar ve önsel dağılımın olabilirlik fonksiyonun çarpımı ile güncellenir ve böylece her bir yeni madde için daha güvenilir tahmin yapılır. Güncellenen sonsal dağılım temel alınarak birey hakkında en fazla bilgi verecek madde seçimi yapılarak devam edilir. Bu özelliğinden dolayı test süreci boyunca doğru kestirim yapabilme imkanı vermektedir ve ayrıca kısıtlı test uzunluğunda da güvenilir tahminler yapabilmektedir. Her bir sorudan sonra sonsal dağılımın güncellenmesi süreci bayesian yöntemleri büyük test modüllerinde ve çok kişinin katıldığı sınavlarda daha kullanışsız hale getirmektedir(Van Der Linden & Pashley, 2009).

Aralık bilgi ölçütü(ABÖ)

Veerkamp ve Berger(1997) tarafından ortaya atılmış olan ABÖ, MFB'ne alternatif olarak geliştirilmiştir. Her bir madde için bilgi fonksiyonunu belirli bir güven aralığında ortalayarak daha dengeli bir madde seçimi yapmaya imkan sağlamaktadır. ABÖ için bireyin yetenek seviyesi etrafında bir güven aralığı belirlenir ve bu aralık boyunca sağlanan toplam bilgiye göre madde seçimi yapılır. ABÖ eşitliği

$$I_i = \int_{\hat{\theta}_L}^{\hat{\theta}_R} I(\theta)d\theta$$

$\hat{\theta}_R$ ve $\hat{\theta}_L$tahmin edilen yetenek seviyesinin alt ve üst sınırını gösterir. $I(\theta)$ θ seviyesindeki bilgi fonksiyonudur. Kestirilen yetenek seviyesinin güven aralığında ele alınması, standart hata sorunu azaltır ve test süreci başında daha dengeli madde seçimi yapılmasına imkan tanır. Ancak güven aralığı geniş olduğu için bilgi miktarı düşük olabilmektedir ve bu durumda seçilen maddenin bireyin yetenek kestirimi için bilgi sağlayıcılığını düşürebilmektedir(Chang & Ying, 1999).

Test sonlandırma kuralı

BOBUT uygulamalarında test sürecinin kontrolünü sağlamak ve test süresini ayarlamak adına kullanılan test sonlandırma kuralları ikiye ayrılmaktadır. Bunlar sabit test uzunluğu kuralı ve değişken test uzunluğu yöntemleridir(Embretson & Reise, 2000).

Sabit test uzunluğu kuralı

Sabit test uzunluğu kuralına göre hazırlanmış bir BOBUT uygulamasında kaç soru sorulduğunda uygulamanın biteceği önceden belirlenmektedir(Lord, 2012)

Değişken test uzunluğu yöntemleri

Değişken test uzunluğu yöntemleri BOBUT uygulamalarının daha adaptif bir şekilde uygulanmasına imkan sağlayan yöntemlerdir. Bu amaçla kullanılan değişken test uzunluğu yöntemlerinden Standart Hata Temelli Sonlandırma ve Kesme Puanı Temelli Sonlandırma literatürde en yaygın kullanılan yöntemlerdir

Standart hata temelli test sonlandırma

Bu yöntem için BOBUT uygulaması öncesi bireyin yetenek düzeyi için belirlenen standart hata değerinin altına düşüldüğünde

test sonlandırılır. Algoritmanın adaptif niteliğinden dolayı test sürecinde yetenek kestirimlerinde hızla sonuca ulaşılabilmektedir(Weiss, 1982).

Kesme puanı temelli sonlandırma

Bireyin performansına göre belirli bir kesme puanına ulaşıldığında uygulama sonlandırılır. Örneğin birey matematik sınavı için yeterliliği belirten kesme puanına ulaştığında başarılı kabul edilerek test süreci sonlandırılır(Reckase & Reckase, 2009).

3.sonuç

BOBUT uygulamalarında kullanıcı deneyimi

BOBUT uygulamalarının doğası gereği bireye uyarlanması durumu bireylerin test stresini azaltıcı ve test doğruluğunu arttırıcı bir etkiye sahiptir. Bu açıdan klasik kağıt-kalem testlerine göre BOBUT uygulamaları kullanıcı deneyimini olumlu şekilde etkilemektedir(Lu & Cong, 2016). Eggen(1999), BOBUT uygulamalarında bireylere uygulanacak olan ilk soru seçiminde madde güçlüğünün doğru belirlenmesinin, testi alan bireylerin sınav deneyimine etkisinden bahsetmektedir. BOBUT uygulamalarındaki adaptif algoritma ve doğru ilk soru seçimi testi alan bireylerin motivasyonuna olumlu katkı vermektedir. BOBUT uygulamaları geleneksel kağıt-kalem testlerine göre daha doğru sonuçlar vermektedir. Ancak sonuçların psikolojik etkilerle optimize edilmesi gerekmektedir. Gölge test yaklaşımı(shadow testing) gibi yaklaşımlarla bireyde stres yaratmadan en yüksek bilgi doğruluğu sağlanabilir. Ayrıca test uzunluğunun optimize edilmesi kullanıcı deneyimi ve test etkililiği üzerinde olumlu etkiler sağlamaktadır(Kimura, 2017). Test uzunluğunun kısa tutulması BOBUT uygulamasına katılan bireylerde motivasyonun artmasına

ve özellikle zaman kısıtlaması olan testlerde stres düzeyinin azalmasına imkan vermektedir. Öte yandan daha uzun testler kullanıcı deneyiminde olumsuz etkilere sebep olduğu gibi, bireylerde dikkatsizliğe de ortam hazırlamaktadır. Ayrıca anında geri bildirim mekanizmalarının testlere entegrasyonu sürecinde kısa testlerin kullanımı anında geri bildirimin etkisini arttırmaktadır.(Thissen vd., 2007)

Geribildirim, testi alan bireylerin maddelerin çözümünde strateji geliştirmesinde, bireylerin motivasyonunun arttırılmasında ve daha az hata ile öğrenimin sağlanmasında biçimlendirici bir etkiye sahiptir(Bangert-Drowns & ark., 1991). BOBUT uygulamalarında testi alan bireylerin test esnasında daha rahat geribildirim almasına olanak tanıdığı için bireyler daha bilinçli ve planlı hareket edebilmektedir. Bireylerin sorulara verdiği yanıtların doğru veya yanlış olması durumu bireylerin çözüm stratejilerini optimize etmesi ve hangi konu veya becerilere odaklanmaları gerektiğini kavramlarına destek olur. Oluşturulacak geribildirimlerin pozitif nitelikte olması bireylerin motivasyonu ve yetkinliğine olumlu katkı vermektedir(Vollmeyer & Rheinberg, 2005). Ayrıca geribildirimlerin bireyselleştirilmesi öğrenme etkilerini arttıracağı gibi test sürecini de olumlu bir şekilde etkiler. Ancak bireyselleştirilmiş geribildirimlerin oluşturulması adına testi alacak bireylerle ilgili önsel bilgilerin bulundurulması gerekmektedir(Bimba vd. 2017). Bireylere verilecek olan geribildirimin detaylı bir şekilde oluşturulması özellikle daha yüksek öğrenmelerin gerektiği alanlarda daha etkili olduğu bulunmuştur(Van der Kleij & ark., 2013; Wisniewski & ark., 2020). Hattie ve Timperley(2007) çalışmalarında, oluşturulacak

geribildirimlerin pozitif ve anında geribildirimin özellikle motive edici etkilerinden bahsetmektedir.

BOBUT uygulamalarında testi alan bireylere sürenin gösterilmesi, bireylerin test sürecini planlamalarında ve kullanıcı stresinin azaltılmasında olumlu katkılar vermektedir. Ayrıca süre göstergelerinin stratejik kullanımları bireylerin testin sonuna yaklaştıkları algısının pekişmesine imkan sağlamaktadır(Rossi & ark., 2014; Mead & Drasgow, 1993). Leeson(2006) çalışmasında bilgisayarlı testlerde zaman gösteriminin bireylerde sınav baskısını azalttığını ve performansı arttırdığını öne sürmektedir.

Ayrıca BOBUT uygulamalarının oluşturulması sürecinde algoritmanın kullanıcı dostu bir ara yüz ile tasarlanması kullanıcı deneyimine olumlu katkı vermektedir(Thompson & Weiss, 2011). Bireylerin katılacağı BOBUT uygulamasında tasarlanacak olan ara yüzün aşırı sadelikten yana olması durumu kullanıcı deneyimine olumsuz etkilerde bulunabilmektedir. Ara yüzün oluşturulmasında bireylerin kullanabileceği yön okları(navigatörler) ve zaman göstergeleri gibi unsurlar kullanıcı deneyimine olumlu katkı vereceği düşünülen ara yüz parçalarıdır(Leeson, 2006).

BOBUT uygulamalarında kullanıcı deneyimine etki eden bir diğer değişken bireyin testi aldığı ekipmanların teknik performansıdır. Bireylerin BOBUT uygulamasına katıldığı yazılımın sorunsuz bir şekilde çalışıyor olması, bireylerin testi tamamlama süresini ve oranını arttırmaktadır. Öte yandan sorunlu çalışan bir yazılımın kullanıcı deneyimi üzerine olumsuz etkileri olmaktadır. Bireylerin BOBUT uygulamasına katıldıkları cihazın türü de test uygulaması üzerinde son derece önemlidir. Örneğin uygulamaya mobil bir cihaz ile katılan bireyin testin navigatörleri

konusunda sıkıntı çekebilmesi son derece olasıdır. Ayrıca bireylerin testi aldığı süreçte aktif çalışan bir teknik destek biriminin olması kullanıcı deneyimine olumlu katkı vermektedir(Spenassato & ark., 2015).

BOBUT uygulamasının yapılacağı ortamın da kullanıcı deneyimine olumlu ve olumsuz çeşitli katkıları olabilmektedir. Test uygulamasını yapacak olan bireyin ortamda gürültü ve ışıklandırma gibi faktörleri doğru ayarlamasının kullanıcı deneyimine olumlu katkı sunmaktadır(Kimura, 2017). Ayrıca bireylerin testi aldığı ortam (ev veya laboratuvar) da kullanıcı deneyimini etkilemektedir. Bireylerin daha rahat ve dikkat dağıtıcı unsurlardan arındırılmış bir ortamda testi alması bireylerin kullanıcı deneyimini arttırmaktadır(Leeson, 2006; Keskin & Günay, 2021). Ayrıca test ortamının engelli bireyleri için ergonomik ve teknik standartlar açısından düzenlenmesi kullanıcı deneyimini etkilemektedir(Harniss & ark., 2007).

Sonuç

Günümüz dünyasında bilgisayar temelli test yaklaşımlarının yaygınlaşması ile birlikte bilgisayar algoritmaları kullanılarak bireylerin test deneyimlerinin optimize edilmesi son derece kritik bir hal almıştır. Bilgisayar uygulamalı testlerin çıkış sürecinde özellikle bazı gruplara(iyi derecede bilgisayar kullanma becerisine sahip bireyler) çeşitli avantaj sağladığı gözlense de günümüzde bu fark hızla kapanmaktadır. Bilgisayar temelli test uygulamalarında son zamanlarda uluslararası çeşitli sınavlarda da kullanılan BOBUT uygulamalarında, kağıt-kalem testlerine göre bireyler çok daha az sayıda madde ile geçerli ve güvenilir bir şekilde ölçülebilmektedir. Ancak BOBUT uygulamalarının yapısı gereği bireylerin alacakları

testlerde kullanıcı deneyiminin optimize edilmesi BOBUT uygulamaları açısından son derece önemlidir. BOBUT uygulamalarında kullanıcı dostu ara yüz tasarımı yapılması, bireye verilen geribildirimin açıklayıcı ve katkı bir şekilde dönüt verebiliyor olması, Süre ve navigatörlerin doğru bir şekilde konumlandırılması bireyin aldığı test ile ilgili kullanıcı deneyimini etkileyen çeşitli değişkenlerdir. BOBUT uygulaması tasarlayan test geliştirme uzmanları değişkenleri göz önünde bulundurarak uygulamanın kullanıcı deneyimi açısından son derece önemlidir.

Büyük veri kavramı, son yıllarda dünyada popüler hale gelmiş ve birçok alanda uygulamaları yapılmaya başlanmıştır (Yılmaz & Alkan, 2024). Bilgisayar uygulamalarının hayatımızda daha çok yer kaplaması birlikte gelişen BOBUT uygulamaları da daha geniş ölçekli sınavlarda da kullanılmaya başlamıştır. Thompson(2017) çalışmasında Avusturalya'da uygulanmakta olan NAPLAN sınavını örneği ile BOBUT uygulamalarında büyük veri analitiğinin eğitim üzerindeki etkilerini ele almaktadır. Ayrıca alan uzmanları büyük veri analizine yönelik derin öğrenme tabanlı adaptif yöntemleri çalışarak, BOBUT uygulamalarında büyük veri analizinde yapay zeka ve derin öğrenme tekniklerini bir arada kullanmaya başlamıştır(Juuso, 2018). Yapay zeka algoritmaları geçmiş verilerin detaylı analiz edebiliyor olduğundan gelecekteki olaylar hakkında daha doğru tahminler yapılabilmesine imkan sunmaktadır(Kölük & Alkan, 2024). Ayrıca BOBUT uygulamalarının tasarlanması sürecinde yapay zeka destekli oluşturulan madde havuzunda otomatik madde üretimi, ara yüzün bireyselleştirilmesi ve otomatik geri bildirim uygulamalarının kullanılması kullanıcı deneyiminin optimize edilebilmesi adına katkılar sunmaktadır. Bütün bu çalışmalar göz önüne alındığında BOBUT uygulamaları daha

dikkatli ve çok yönden ele alınarak geliştirilmesi gereken çalışmalardır. Bu çalışmaların eğitim alanına katkıları ilerleyen yıllarda daha çok görülecektir

Kaynakça

Alkan, B.B., Sinan A., (2020). Veri Toplama Yöntemlerinde Geçerlik ve Güvenirlik. Şefika Ş. Erçetin (Ed.), Araştırma Teknikleri(187-197). Ankara: Nobel Akademik Yayıncılık

Baker, F. B. (2001). The basics of item response theory (Cilt 2). USA: ERIC Clearinghouse on Assessment and Evaluation

Baker, F.B. & Ho-Kim, S.H. (2017). The basics of Item response theory using R. Springer International Publishing.

Baker, F. B. & Kim S. H. (2004). Item response theory: Parameter estimation techniques.

Baker, N., İlhan, M., & Güler, N. (2016). Madde tepki kuramının temelleri. Pegem Akademi.

Baker, F. B. (1998). Item Response Theory: Parameter Estimation Techniques. *Applied Psychological Measurement*, *22*(1), 89-91.

Bangert-Drowns, R., Kulik, C., Kulik, J., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, 61, 213 - 238. https://doi.org/10.3102/00346543061002213.

Bartram, D., ve Hambleton, R. K. (2006). Computer-based testing and the Internet. Issues and advantes. England: John Wiley and Sons.

Bimba, A., Idris, N., Al-Hunaiyyan, A., Mahmud, R., & Shuib, N. (2017). Adaptive feedback in computer-based learning environments: a review. *Adaptive Behavior*, 25, 217 - 234. https://doi.org/10.1177/1059712317727590.

Birnbaum, A. (1968). Some latent trait models. *Statistical theories of mental test scores*.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443-459.

Camilli, G. (1994). *Methods for identifying biased test items*. Sage

Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, *20*(3), 213-229.

Chang, H. H., & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied psychological measurement*, *23*(3), 211-222.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.

Crocker, L., & Algina, J. (2006b). Introduction to item response theory. İçinde Introduction to Classical and Modern Test Theory (s. 339-375). Mason, Ohio Cengage Learning.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Çıkrıkçı-Demirtaşlı N. (1998). Test Geliştirmede Yeni Yaklaşımlar: Örtük Özellikler Kuramı - TemelÖzellikleri, Varsayımları, Model ve Sınırlılıkları. Ankara Üniversitesi Eğitim Fakültesi Dergisi, 2(28), 161-173.

Davey, T. (2011). A guide to computer adaptive testing systems. Council of Chief State School Officers.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Eggen, T. J. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, *23*(3), 249-261.

Embredson Susan, E., & Reise Steven, P. (2000). Item response theory for psychologists. *Mahway, NJ: Erlbaum*.

Frederiksen, N., Mislevy, R. J., & Bejar, I. I. (Eds.). (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hambleton, R. K. & Swaminathan, H. (1985). Item response theory: principles and applications. Boston: Kluwer Nijhoff Publishing

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. London: Sage Publications.

Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing. *Reston (VA): Graduate Management Admission Council*.

Harniss, M., Amtmann, D., Cook, D., & Johnson, K. (2007). Considerations for Developing Interfaces for Collecting Patient-Reported Outcomes That Allow the Inclusion of Individuals With

Disabilities. *Medical Care*, 45, S48-S54. https://doi.org/10.1097/01.mlr.0000250822.41093.ca.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 112 - 81. https://doi.org/10.3102/003465430298487.

Ho, T. (2010). A Comparison of Item Selection Procedures Using Different Ability Estimation Methods in Computerized Adaptive Testing Based on Generalized Partial Credit Model. Doctoral Dissertation, Graduate School of the University of Texas at Austin

Keskin, B., & Gunay, M. (2021). A Survey On Computerized Adaptive Testing. *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1-6. https://doi.org/10.1109/ASYU52992.2021.9598952.

Kimura, T. (2017). The impacts of computer adaptive testing from a variety of perspectives. *Journal of Educational Evaluation for Health Professions*, 14. https://doi.org/10.3352/jeehp.2017.14.12.

Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In *New horizons in testing* (pp. 257-283). Academic Press.

Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage publications.

Koğar, H. (2021). R ile geçerlik ve güvenirlik analizleri. *PEGEM Akademi*.

Kolen, M. J. (2004). Test equating, scaling, and linking.

Kölük, U. K., Alkan, N. (2024). Yapay Zeka Destekli İşletme Yönetimi: Uygulamalar, Zorluklar Ve Gelecek Yönelimler, İktısadi Ve İdari Bilimlerde Uluslararası Çalışma Ve Değerlendirmeler, Serüven Yayınevi, 157

Leeson, H. (2006). The Mode Effect: A Literature Review of Human and Technological Issues in Computerized Testing. *International Journal of Testing*, 6, 1 - 24. https://doi.org/10.1207/s15327574ijt0601_1.

Linden, W. J and Glas, G. A. W., 2002. Computerized Adaptive Testing: theory and practice. USA: Kluwer Academic Publishers.

Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.

Lu, P., & Cong, X. (2016). The Research on Computerized Adaptive Testing. Journal of Physics: Conference Series, 710. https://doi.org/10.1088/1742-6596/710/1/012029.

Juuso, E. (2018). Smart Adaptive Big Data Analysis with Advanced Deep Learning. *Open Engineering*, 8, 403 - 416. https://doi.org/10.1515/eng-2018-0043.

Mead, A., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis.. *Psychological Bulletin*, 114, 449-458. https://doi.org/10.1037/0033-2909.114.3.449.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25-36.

Reckase, M. D., & Reckase, M. D. (2009). Historical background for multidimensional item response theory (MIRT). *Multidimensional item response theory*, 57-77.

Rossi, D., Ferreira, W., Mansur, W., & Calenzani, A. (2014). A review of automatic time-stepping strategies on numerical time integration for structural dynamics analysis. *Engineering Structures*, 80, 118-136. https://doi.org/10.1016/J.ENGSTRUCT.2014.08.016.

Spenassato, D., Bornia, A., & Tezza, R. (2015). Computerized Adaptive Testing: A Review of Research and Technical Characteristics. *IEEE Latin America Transactions*, 13, 3890-3898. https://doi.org/10.1109/TLA.2015.7404924.

Sünbül, Ö. (2020). Madde tepki kuramı (MTK). İçinde Psikolojide Ölçme ve Ölçek Geliştirme II: Ölçme Araçlarının Psikometrik Nitelikleri ve Ölçme Kuramları (pp. 122-141). Ankara: Pegem Akademi.

Thissen, D., & Wainer, H. (2001). An overview of test scoring. *Test scoring*, 13-32.

Thissen, D., Reeve, B., Bjorner, J., & Chang, C. (2007). Methodological issues for building item banks and computerized adaptive scales. *Quality of Life Research*, 16, 109-119. https://doi.org/10.1007/s11136-007-9169-5.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education*. Macmillan Publishing Co, Inc.

Thompson, N., & Weiss, D. (2011). A Framework for the Development of Computerized Adaptive Tests.. *Practical Assessment, Research and Evaluation*, 16, 1-9. https://doi.org/10.7275/WQZT-9427.

Thompson, G. (2017). Computer adaptive testing, big data and algorithmic approaches to education. *British Journal of Sociology of Education*, 38, 827 - 840. https://doi.org/10.1080/01425692.2016.1158640.

Van Der Kleij, F., Feskens, R., & Eggen, T. (2013). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes. *Review of Educational Research*, 85, 475 - 511. https://doi.org/10.3102/0034654314564881.

Van der Linden, W. J., & Glas, C. A. (Eds.). (2010). *Elements of adaptive testing* (Vol. 10, pp. 978-0). New York: Springer.

Van der Linden, W. J., & Pashley, P. J. (2009). Item selection and ability estimation in adaptive testing. In *Elements of adaptive testing* (pp. 3-30). New York, NY: Springer New York.

Veerkamp, W. J., & Berger, M. P. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, *22*(2), 203-226.

Vollmeyer, R., & Rheinberg, F. (2005). A Surprising Effect of Feedback on Learning.. *Learning and Instruction*, 15, 589-602. https://doi.org/10.1016/J.LEARNINSTRUC.2005.08.001.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge.

Weiss, D. J. (1981). Computerized Adaptive Ability Testing, April 1981.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied psychological measurement*, *6*(4), 473-492.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10. https://doi.org/10.3389/fpsyg.2019.03087.

Yılmaz, D. E., Alkan, N. (2024) Öğrenen Örgütlerde Büyük Veriden Öğrenme Yoluyla Rekabet Avantajı Elde Etmek, İktisadi Ve İdari Bilimlerde Uluslararası Çalışmalar, Serüven Yayınevi, 55

BÖLÜM VII

Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Test Sonlandırma Kuralları

Dilek GÜR ALP[1]
Alper SİNAN[2]

Giriş

Geleneksel testler bilgi, beceri ve yeteneklerin ölçülmesinde dünyada yaygın olarak kullanılmaktadır. Klasik Tepki Kuramı veya Madde Tepki Kuramı temel alınarak hazırlanmış geleneksel testler diğer bir ifade ile doğrusal testler tüm bireylerin sabit sayıda aynı maddeleri aldığı bir test türüdür. Katılımcılara testte bulunan tüm maddeler kolay veya zor olmasına bakılmaksızın kağıt kalem ortamında sunulur. Yüksek yetenek düzeyine sahip katılımcının kolay soruları, düşük yetenek düzeyine sahip katılımcının ise zor soruları cevaplaması gerekir. Yüksek yetenekli bir adaya kolay

[1] Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Bölümü, Antalya/Türkiye, Orcid: 0000-0002-2262-9059, dilekgur.dilek@gmail.com
[2] Doç. Dr., Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Bölümü, Antalya/Türkiye, Orcid: 0000-0001-6632-5500, asinan@akdeniz.edu.tr

maddeler vermek, doğru cevapların adayın yeteneği hakkında nispeten az bilgi sağlaması nedeniyle zaman kaybıdır. Ayrıca, kişi kendisini zorlamayan maddelerle sıkılabilir ve dikkatsizce cevap verebilir. Bu durum ölçme hatasının yükselmesine neden olur. Benzer şekilde, düşük yetenekli bir adaya zor maddeler vermek de verimsizdir, çünkü yanlış cevaplar o kişinin yeteneği hakkında fazla bilgi vermez. Dahası, düşük yetenekli adaylar genellikle maddeleri çok zor bulabilir ve bu durum onları hayal kırıklığına uğratarak rastgele yanıt vermelerine yol açabilir, bu da test sürecine ek hata katabilir (Sands & Waters & McBride, 1997). Klasik Test Kuramı iç tutarlılık güvenirliğinin yüksek olmasına odaklanır. Klasik Test Kuramında iç tutarlılık güvenirliğinin yüksek olabilmesi için test maddelerinin orta güçlükte (p=0.50) olması ve toplam puanla yüksek korelasyon göstermesi yani testin madde ayırt ediciliği yüksek olan maddeleri içermesi gerekmektedir. Böyle bir testin güvenirliği yüksek, ölçme hatası minimum seviyesinde olur. Böylece, ortalama yetenek düzeyindeki bireyler için ölçme hatası daha düşük sonuçlar elde edilir. Orta güçlükteki maddelerden oluşan bir test, yetenek seviyesi ortalama veya ortalamaya yakın olan bireylerin yeteneklerini oldukça iyi ölçmektedir (Kezer, 2020). Fakat yüksek ve düşük yetenek/gizil özellik gösteren bireylerin yetenek dağılımında üst ve alt uçlarda yer almasına neden olur. Uçlara gittikçe ölçme hatası artmakta, geçerlik ve güvenirlik düşmektedir. Dolasıyla uçlarda yer alan bireylerin yetenek kestirimlerine büyük miktarda hata karışmaktadır. Bu durum ölçmenin doğruluğunun sağlanamamasına neden olur (Hambeleton & Swaminathan, 1985; Kezer, 2020; Weiss & Şahin, 2024).

Klasik Test Kuramının getirdiği bazı sınırlılıklar ve dezavantajların giderilmesi amacıyla Madde Tepki Kuramı ortaya

atılmıştır. Günümüzde Madde Tepki Kuramı temelli kağıt kalem tabanlı testler ve bilgisayar temelli uygulamalar yapılmaktadır. Uyarlanmış testler ilk kağıt kalem tabanlı uygulanmak üzere geliştirilmiş ve daha sonra bilgisayar teknolojisindeki gelişmeler ile birlikte Bilgisayar Ortamında Bireye Uyarlanmış Testler-BOBUT (Computerized Adaptive Testing-CAT) halini almıştır.

BOBUT, önceden sabitlenmiş maddelerin ve sıralamanın olduğu geleneksel bir testin bilgisayarda uygulanması değil, yapay zeka ve makine öğrenimi ilkelerini kullanarak her bir birey için bir test tasarlamaktadır. Günümüzde yetenek, başarı, yeterlilik, kişilik, tercihler, tutumlar ve diğer psikolojik, eğitimsel ve insan kaynakları değişkenlerinin ölçüldüğü testler, bir bilgisayar tarafından uygulanmak üzere yeniden tasarlamaktır. Test uygulaması sırasında her cevabın ardından hemen puanlama yapılır. Testi alan birey maddelere cevap verdikçe algoritma yeni test maddelerini seçer ve yanıtları hemen puanlar. Algoritma testi alan bireyin sürekli yetenek kestirimini yapar ve maddelerin seçimini bireye göre uyarlayarak ilerler. Sonuç olarak testi alan her birey benzersiz bir test alma potansiyeline sahiptir (Weiss & Şahin, 2024).

Bilgisayar ortamında bireye uyarlanmış testlerin altında yatan istatistiksel model Madde Tepki Kuramıdır. Madde Tepki Kuramı olmadan uyarlanabilir testlerin mümkün olmayacağını düşünen araştırmacılar mevcuttur. Her birey farklı zorluk seviyelerine sahip maddeler almasına rağmen MTK'nın farklı bireylerin yetenek tahminlerinin karşılaştırılabileceği bir çerçeve sunması bilgisayar destekli uyarlanabilir testler için MTK'nın en cazip özelliklerinden biridir (Hambleton, 1991; Sireci, 2003).

Günümüzde bilgisayar ortamında bireye uyarlanmış testlerin temel ilkeleri Alfred Binet tarafından 1905 yılında ortaya konmuştur. İlk bireye uyarlanmış test Alfred Binet ve Theodore Simon tarafından geliştirilen Binet-Simon Zeka Testidir. Bu test geleneksel eğitim sisteminden fayda sağlayamayacak, zeka düzeyi yeterli olmayan öğrencileri belirlemek amacıyla geliştirilmiştir Kolay bir soruya cevap veremeyen bir çocuğun zor bir soruyu da cevaplayamayacağını bilen Binet, test maddelerini zorluk derecesine göre artan bir sırada 30 alt testten oluşacak şekilde düzenlemiştir. Binet-Simon zeka testinin uygulanmasında öncelikle testin hangi seviyeden başlatılacağının belirlenmesi için çocuğun seviyesinin bilinmesi gerekmektedir. Eğer testi uygulayan kişi çocuğun testteki olası performansına dair bir şey biliyorsa testin başlangıç seviyesini bu bilgiye göre belirleyebilir. Başlangıç seviyesi belirlendiğinde o seviyedeki ilk madde uygulanır ve hemen puanlanır. Bu şekilde seçilen seviyedeki tüm maddeler uygulanır. Daha sonra bir sonraki uygulanacak daha düşük ya da daha yüksek yaş seviyesine ait madde grupları belirlenir. Eğer çocuk başlangıç seviyesinde uygulanan maddelere doğru cevap verdiyse o seviyedeki tüm maddelere doğru cevap vereceği kabul edilir ve daha zor olan bir sonraki seviyenin maddeleri uygulanır. Bu şekilde çocuğun tüm maddeleri yanlış cevaplayacağı seviyeye gelene kadar test uygulanır ve o noktada çocuğun tavan yaşı tespit edilmiş olur. Eğer çocuk başlangıç seviyesindeki maddelere yanlış cevap vermiş olsaydı bir alt seviyeye ait maddeler uygulanmaya devam edecekti. Daha sonra tavan yaş tespit edilene kadar artan zorlukta maddeler uygulanmaya devam edecekti. Farklı bireylere farklı sayıda maddelerin uygulandığı Binet-Simon testinde farklı sonlandırma kuralı mevcuttur (Sireci,

2003; Van der Linden & Glas, 2010; Kezer, 2020; Şenel, 2023; Weiss & Şahin, 2024;).

Bir uyarlanabilir testin geleneksel testlerden ayıran beş özelliği bulunmaktadır:

1. Uyarlanabilir bir test psikometrik nitelikleri bilinen test maddelerinden oluşan bir madde havuzuna dayanmaktadır. Binet-Simon testi gibi bireysel farklılıkları ölçmek için uyarlanabilir test kullanıldığında madde havuzu ölçülecek özelliğin geniş bir yelpazesini kapsayan maddelerden oluşur.
2. Test, birey hakkındaki mevcut bilgileri kullanarak katılımcı için madde havuzundan bir başlangıç noktası (başlangıç maddesi) seçebilir. Bu şekilde tüm adayların aynı madde veya madde setiyle başlaması gerekmeyebilir.
3. Önceden tanımlanmış bir ölçekte maddeler bireye uygulandıkları anda puanlanır. Farklı adaylara verilen farklı maddelerden test puanı elde edilir.
4. Bireyin önceki maddelere verdiği cevapların puanlamasına dayanan bir kural ile (testin öncesinde belirlenmiş) katılımcıya uygulanacak sonraki maddeler seçilir.
5. Önceden belirlenmiş bir test bitirme kuralında bitirme kriterine ulaşıldığında testi alan birey için test sonlandırılır. Her bireye sabit bir madde sayısı uygulanmak zorunda değildir.

Son dört özelliğin bir sonucu olarak, uyarlanabilir bir test bireyselleştirilmiş bir testtir. Bireyler mutlaka aynı maddelerle teste başlamayabilir, her birey farklı madde seti alabilir ve farklı sayıda maddelere cevap verebilir. Uyarlanabilir bir test dinamiktir. Test

uygulanma aşamasında bireyde ölçülmek istenen özellik seviyesine göre testin zorluğu ayarlanır (Weiss & Şahin, 2024). MTK tabanlı uyarlanabilir testlere kıyasla Binet'in uyarlanabilir testi daha basit olsa da bir uyarlanabilir testte bulunması gereken özellikler mevcuttur. Yapılan revizyonlar sonucu ismi Stanford Binet Zeka Testi olarak anılan test günümüzde hala kullanılmaktadır.

Erken dönem uyarlanabilir test yaklaşımlarında madde seçimi MTK temelli olmayan yöntemler kullanılarak oluşturulmuştur. Bu dönemde sıklıkla kullanılan yöntemler; *iki aşamalı uygulama, kendinden seçimli uygulama, piramitsel çok aşamalı uygulama, değişimlemeli uygulama, tabakalı uygulama ve çok aşamalı uygulamadır.* Erken dönem bireye uyarlanmış testler kağıt kalem tabanlı uygulanması ve kullanılan yöntemlerde hesaplamaların karmaşıklığı sebebiyle uzun vadede yaygınlaşmamıştır. Fakat 1970lerde bilgisayar teknolojisindeki gelişmeler sonucu bireye uyarlanmış testlerle ilgili araştırmalarda ve uygulamalarda artış yaşanmıştır. Kağıt kalem tabanlı uygulamada yaşanan zorluklar sonucu testlerin bilgisayar ortamına aktarılması ile ilk BOBUT çalışmaları yapılmıştır. Başlarda KTK tabanlı yapılan BOBUT uygulamaları MTK'nın bireyselleştirilmiş testler için daha elverişli psikometrik veriler sunması nedeniyle uygulanmaya başlamıştır. Bu alanda öncü olan Amerika Birleşik Devletler Savunma Bakanlığı Silahlı Kuvvetler Mesleki Yetenek Testi (ASVAB) ilk BOBUT versiyonu 1980'lerin ortalarında faaliyete geçmiştir (Van der Linden & Glas, 2010). Daha sonra uluslararası Lisansüstü Kabul Sınavı (GRE) gibi geniş ölçekli sınavlarda uygulanmaya devam etmiştir.

Madde Tepki Kuramı

MTK bireyin test alması durumunda göstermesi beklenen özellikleri yetenek veya gizil özellik olarak tanımlar. Daha sonra bu özellikler için yetenek olarak adlandırdığı puanları tahmin eder ve bu puanları kullanarak madde ve test performanslarını tahmin etmeye çalışır (Lord & Novick, 2008). Bu özellikler doğrudan ölçülemediği için gizil özellikler ya da yetenek olarak adlandırılır. Bir madde tepki modeli, bireyin gözlemlenebilir test performansı ile test performansının altında yatan varsayılan gözlemlenemez özellikler ya da yetenekler arasındaki ilişkiyi belirler. Maddeye doğru yanıt verme olasılığı, madde ve yetenek parametreleri arasındaki ilişkiyi tanımlayan fonksiyona *madde karakteristik fonksiyonu*, fonksiyonun oluşturduğu eğriye ise *madde karakteristik eğrisi (MKE)* denir. Madde karakteristik fonksiyonunda yeterlilik düzeyi (θ) arttıkça, maddeyi doğru cevaplama olasılığı $P(\theta)$ da artmaktadır. Madde karakteristik eğrileri S şeklinde olan matematiksel formu nedeniyle mevcutta olan birçok model uygulanabilir. MTK'da kullanılan bu modeller doğru ya da yanlış olarak gösterilmezken, belirli bir modelin herhangi bir test verisi setiyle uygunluğu uyum testi yapılarak tespit edilir (Hambleton & Swaminathan, 1985).

MTK modellerine ilişkin çeşitli sınıflamalar yapılmaktadır. Testi oluşturan maddelerin tek veya baskın olan gizil özelliği ölçmesi durumunda kullanılan modeller *tek boyutlu modeller*, birden fazla gizil özelliğin ölçülmesi durumunda kullanılan modeller *çok boyutlu modeller* olarak adlandırılmaktadır. Puanlanma yöntemine göre sınıflanan modeller ise *ikili puanlanan* ya da *çoklu puanlanan* modeller olarak ayrılmaktadır.

1. Tek Boyutlu MTK Modelleri

Tek boyutlu MTK modelleri ikili puanlanan modeller ve çoklu puanlanan modeller olarak ayrılmaktadır. İkili puanlanan modeller MTK'nın en temel modelleri olup daha sonra geliştirilen modellere zemin oluşturmuştur (De Ayala, 2009).

1.1.İkili Puanlanan Modeller

İkili puanlanan MTK modelleri, yetenek veya başarı testlerinde belirli özellik seviyelerine sahip bireylerin test maddelerini ne sıklıkta doğru yanıtladığını (doğru cevapların 1 ile puanlandığı) açıklamak için tasarlanmıştır. Aynı şekilde *katılıyorum/katılmıyorum* şeklinde ifadeler içeren kişilik, tutum gibi psikolojik testlerde en yüksek yapı 1 ile puanlanır. Örneğin depresyon ölçeğinde *katılıyorum* ifadesi yüksek depresyonu tarif ediyorsa 1 ile puanlanır. Olumsuz maddelerde ise *katılmıyorum* ifadesi 1 ile puanlanmalıdır.

1.1.1. 1 Parametreli Lojistik Model:

1PLM'de madde parametrelerinden sadece madde güçlük parametresi (b_i) yer alır. Testteki tüm maddelerin ayırt edicilik parametreleri (a_i) aynı değeri almaktadır. Şans parametresi ise sıfıra eşittir ($c_i = 0$). Dikkatsizlik parametresi (d_i) ise modelde eklenmemiştir. Yanıtlayıcının performansını etkileyen tek parametre madde güçlük parametresidir.

1.1.2. 2 Parametreli Lojistik Model

2 Parametreli Lojistik Modelde madde güçlük parametresine ek olarak madde ayırt edicilik parametresi de yer almaktadır. Bu

modelde maddelere ait ayırt edicilik parametresi farklı değerler almaktadır. Şans başarısı yine sıfırdır. Dikkatsizlik parametresi de modelde bulunmamaktadır.

1.1.3. 3 Parametreli Lojistik Model

3 Parametreli Lojistik Modelde madde güçlük parametresi, madde ayırt edicilik parametresi ve şans parametresi yer almaktadır. Dikkatsizlik parametresi ise modelde bulunmamaktadır.

1.2.Çoklu Puanlanan Modeller

Psikolojik değişkenler her zaman iki kategorili sonuçlar veren test maddeleriyle ölçülemez. Kişilik, tercih ve tutum gibi değişkenler genellikle daha fazla kategoriye ihtiyaç duyulan ölçeklerle ölçülür. Bunlar bazen tepki sıralaması olurken bazen eşit aralıklı tepkiler olabilmektedir (Kezer, 2020). Çoklu puanlanan veriler için sık kullanılan beş farklı model önerilmiştir: *Dereceleme Ölçeği Modeli, Sınıflamalı Tepki Modeli, Kısmi Puan Modeli, Genelleştirilmiş Kısmi Puan Modeli, Aşamalı Tepki Modeli*

1.2.1. Dereceleme Ölçeği Modeli

Özellikle eğitim ve psikoloji ile ilgili ölçmelerde Likert tipi dereceleme ölçeği olarak bilinen sıralama düzeyindeki ölçekler sıklıkla kullanılmaktadır. Örneğin "kesinlikle katılmıyorum, katılmıyorum, kararsızım, katılıyorum ve kesinlikle katılıyorum" gibi sıralı bir dizi yanıt seçeneğinden birini seçmesini gerektiren çeşitli türde derecelendirme ölçekleriyle ölçülür.

1.2.2. Sınıflamalı Tepki Modeli

Sınıflama ölçeği düzeyindeki verilerde uygulanabilmektedir. Çoktan seçmeli maddelerde uygulanmaktadır. Sınıflama düzeyinde

olduğu için çoktan seçmeli maddelerde çeldiricilerin modele dahil edilmesi mümkündür. Madde konum parametresi ve kesişim parametresi bulunmaktadır. Kesişim parametresi aynı zamanda k kategorisinin çekicilik yoğunluğu olarak da adlandırılmaktadır.

1.2.3. Kısmi Puan Modeli

Kısmi puan modeli 1PLM'nin uzantısıdır. Daha çok kısmi puanlama yapılan kişilik tutum gibi ölçekler ve yetenek testleri için uygundur. Veriler sıralama ölçeği düzeyinde olmalıdır. Kısmi puan modelinde, belirli bir kategoride yanıt verme olasılığı, madde için adım zorluğu parametresiyle ilişkilidir. Adım zorluğu veya yanıt eşiği, kategoriler arasında geçiş yapmanın zorluğu ile ilişkilidir.

1.2.4. Genelleştirilmiş Kısmi Puan Modeli

Kısmi Puan Modelinin genelleştirilmiş halidir. Bir maddeye doğru cevap verme olasılığı tahmin edilirken eğim parametresi (madde ayırt edicilik) serbest olarak tahmin edilir.

1.2.5. Aşamalı Tepki Modeli

Aşamalı Tepki Modelinde her bir kategoriyi veya daha yükseğini işaretleme olasılığı modellenmiştir. Kümülatif bir yapısı vardır ve sıralama düzeyindeki ölçekler için uygundur. Tutum, kaygı, ilgi ve motivasyon gibi psikolojik özelliklerin ölçülmesinde kullanılan testler için daha uygundur. Maddelerin dereceli ve aynı sayıda kategoriye sahip olması gerekmektedir.

2. Çok Boyutlu MTK Modelleri

Yaygın olarak kullanılan Tek boyutlu MTK modelleri her zaman testler için uygun değildir. Psikolojide kişilik envanterleri, mesleki envanterler, tutum ve ilgi gibi birçok psikolojik özellik çok

boyutlu yapılar sergilemektedir. Aynı zamanda yetenek testlerinde de birden fazla boyutu ölçmeyi amaçlayacak şekilde hazırlanabilir. Örneğin; matematik beceri ile okuduğunu anlama becerisini aynı anda ölçmeyi amaçlayan maddelerin yazıldığı bir test hazırlanabilir. Böyle bir durumda Çok boyutlu MTK Modellerinin kullanılması daha uygun olacaktır (Wang & Chen, 2004; Seo & Weiss, 2015).

Reckase'e (2009) göre Çok Boyutlu MTK modelleri telafi edici ve telafisel olmayan modeller olarak ikiye ayrılmaktadır. Telafi edici modeller; *çok boyutluluk için genişletilmiş rasch model, çok boyutluluk için genişletilmiş iki parametreli lojistik model, çok boyutluluk için genişletilmiş üç parametreli lojistik model, çok boyutluluk için genişletilmiş normal ogive model* olmak üzere dört modelden oluşmaktadır.

Embretson ve Reise (2000) Çok Boyutlu MTK modellerini *doğrulayıcı* ve *açımlayıcı modeller* olmak üzere iki kategoride incelemiştir. Açımlayıcı modeller tam bilgi madde faktör analizi olarak da bilinmektedir.

Wang ve Chen'e (2004) göre Çok Boyutlu MTK modelleri madde düzeyinde boyutluluk *madde içi çok boyutluluk ve maddeler arası çok boyutluluk* olmak üzere ikiye ayrılmaktadır. Bir test, aynı gizil özelliği ölçmeyen tek boyutlu maddeler içeriyorsa maddeler arası çok boyutludur. Test bataryaları (alt testler veya alt ölçekler içeren testler) buna bir örnektir. Bir test, çok boyutlu maddeler içeriyorsa madde içi çok boyutludur. Örneğin, bir kompozisyon yazma görevi, hem içerik anlayışını hem de dil becerilerini değerlendirmek için kullanılabilir veya bir matematik becerisini ölçmesi için sorulan madde hem matematik bilgisini hem okuma becerisini ölçebilir.

Madde Tepki Kuramı Temelli BOBUT Geliştirme

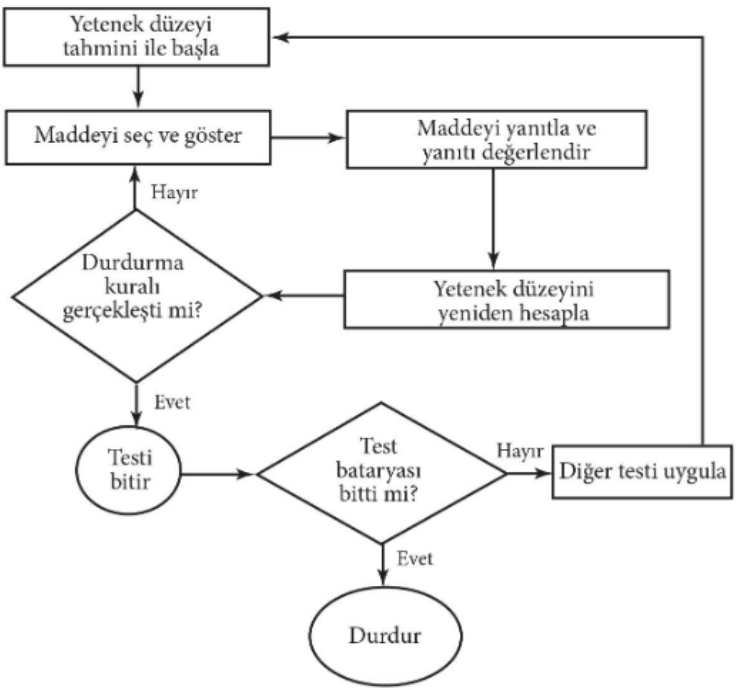
Çoğu Bilgisayar Ortamında Bireye Uyarlanmış Test-BOBUT, Madde Tepki Kuramı temelinde oluşturulmuştur. MTK, test geliştirme, madde analizi ve bireylerin yeteneklerinin kestirimi için birçok avantaj sunan güçlü psikometrik temellere sahiptir. MTK'nın BOBUT ile ilgili en önemli avantajı, maddeleri ve bireyleri aynı ölçeğe yerleştirmesidir. Bu şekilde bireylerin en uygun maddeler ile doğrudan eşleşebilmesi kolaylaşmaktadır. Bireyler kendi düzeylerine uygun maddeler yanıtlayacağından geleneksel kağıt kalem testlerinde yanıtladıkları madde sayısından daha az madde yanıtlayacaklardır. BOBUT'lar klasik test kuramı ile tasarlanabilse de bu avantaj nedeniyle BOBUT'ların büyük çoğunluğu MTK'ya dayanmaktadır. Bu nedenle, BOBUT'u anlamak için önce MTK hakkında bilgi sahibi olmak gerekmektedir (Thompson & Weiss, 2011).

Bir gerçek zamanlı BOBUT uygulamasının beş bileşeni şu şekildedir (Thompson & Weiss, 2011).

1. Kalibre edilmiş bir madde havuzu
2. Teste başlama kuralı
3. Madde seçim yöntemi
4. Yetenek kestirim yöntemi
5. Sonlandırma kuralı

Bileşenlerden yola çıkarak bir BOBUT uygulamasındaki algoritma, kalibre edilmiş madde havuzundan bireye sunulan ilk madde ile başlamakta ve sonrasında bireyin kestirilen yetenek düzeyi temel alınarak bireyin yetenek düzeyine en uygun maddenin

yöneltilmesi ile devam etmekte ve önceden belirlenmiş sonlandırma kuralına göre test sonlandırılmaktadır. Söz konusu BOBUT algoritmasının şeması Şekil 1'de verilmiştir.



Şekil 1. Bilgisayar ortamında bireye uyarlanmış test-BOBUT Algoritması

Kaynak: Wainer ve Diğerleri, 2000

Weiss ve Kingsbury'a göre (1984) BOBUT uygulamasının ilk basamağı madde havuzuna uygun madde tepki modelinin belirlenmesidir. Söz konusu bu bileşenler detaylı olarak açıklanmaya çalışılacaktır.

1.Madde Havuzunun Oluşturulması

Bir BOBUT uygulamasında en önemli bölüm madde havuzudur. Madde havuzunun yapısı BOBUT uygulamasının bireyleri ne kadar iyi veya kötü ölçebileceğini, tüm bireyleri belirli bir kesinlik seviyesinde ölçüp ölçemeyeceğini, kaç maddeye ihtiyaç duyulacağını ve testin uygulanma süresini belirlemektedir. Bu nedenle farklı güçlük düzeyinde, yüksek bilgi veren maddelerden oluşan geniş bir havuza ihtiyaç duyulmaktadır. Söz konusu madde havuzlarının 200 madde ve üzerinde olması (Reckase, 2009) veya en az 100 maddeden oluşmasının yeterli olabileceğini (Embretson & Reise, 2013) gösteren çalışmalar mevcuttur. Her ne kadar madde sayısı bir kriter olsa da madde güçlüğünün dengeli dağılmaması, maddelerin kalibrasyonunun yeterli büyüklükte örneklemden elde edilmemesi, maddelerin bireyler hakkında yeterli bilgi vermemesi gibi durumlarda tek başına anlam ifade etmemektedir. Ayrıca madde sayısının az olduğu BOBUT uygulamasında maddelerin teşhir olması sebebiyle test güvenliği sıkıntıya düşebilir.

2.Teste Başlama Kuralı

Bu basamakta BOBUT'un nasıl başlayacağına karar verilir. Her bireye madde sunulmadan önce başlangıç yetenek θ tahmini için birkaç seçenek mevcuttur. En sık tercih edilen yöntem hedef grubun yetenek düzeyinin ortalama (θ=0) düzey kabul edilip madde havuzundan en çok bilgi veren madde seçilmesidir. Bu yöntemin basit ve anlaşılır olmasına karşın testi alan her bireyin aynı başlangıç θ tahmini ile başlamasının bir dezavantajı söz konusudur. BOBUT algoritması, bireye en uygun maddeyi θ tahminine göre seçeceğinden dolayı o yetenek düzeyine karşılık gelen maddelerin çok fazla kullanılmasına neden olur. Bu durum da maddelerin teşhir

edilmesine neden olabilir. Bunun önüne geçmek için yetenek tahmininin -0.5 ile +0.5 aralığında rastgele seçilmesi veya rastgele madde seçimi yöntemi önerilmektedir (Thompson & Weiss, 2011).

3.Madde Seçim Yöntemi

BOBUT uygulamasının temel amacından biri bireyin yeteneğine en uygun maddelerin belirlenmesi ve uygulanmasıdır. Madde seçim algoritması ile bireylere uygulanan maddelerin yarısını doğru ve yarısını yanlış cevapladığı bir test formu oluşturulur. Alan yazında BOBUT uygulamasının tek boyutlu veya çok boyutlu olmasına göre kullanılan madde seçim yöntemleri değişmektedir. Çok boyutlu BOBUT uygulamasında daha az yöntem mevcuttur. Fakat her ikisinde de madde seçimi *Geleneksel* ve *Bayesçi* yöntemlere dayalıdır.

Tek boyutlu BOBUT uygulamaları için geleneksel yöntemlerden en çok kullanılan yöntemler *Fisher En Yüksek Bilgi, En Yüksek Olabilirlik Ağırlıklandırılmış Bilgi ve Kullback-Leibler* yöntemleridir. Yeteneğin sonsal dağılımını kullanan Bayes yöntemlerinden en çok kullanılanlar ise *Maksimum Sonsal Ağırlıklandırılmış Bilgi, Maksimum Beklenen Bilgi, Minimum Beklenen Sonsal Varyans ve Maksimum Beklenen Sonsal Ağırlıklandırılmış Bilgidir*. Bu yöntemlerin haricinde *Alanyazında A Tabakalama, En Uygun B Değeri, Aşamalı En Yüksek Bilgi* gibi başka madde seçim yöntemleri de mevcuttur. Çok boyutlu BOBUT uygulamaları için ise *A-rule, D-rule, T-rule, W-rule, E-rule,* ve bu yöntemlerin ağırlıklandırılmış hali olan *APrule, WPrule, DPrule, TPrule, EPrule* yöntemleri bulunmaktadır. Ayrıca yakın zamanda alanyazına girmiş *filtrelenmiş-A, filtrelenmiş-D, hafif kısıtlanmış-A, hafif kısıtlanmış-D, çok kısıtlanmış-A, çok kısıtlanmış-D* gibi madde

seçme yöntemleri bulunmaktadır. En çok kullanılan yöntemlerden bazıları açıklanmaya çalışılacaktır.

Fisher En Yüksek Bilgi: Bireyin her maddeyi yanıtlaması sonunda anlık kestirilen yetenek düzeyi için madde havuzundan en yüksek bilgiyi veren madde seçilir.

En Yüksek Olabilirlik Ağırlıklandırılmış Bilgi: Yine maksimum bilgiyi kullanmayı amaçlamaktadır. Maddelerin seçiminde olabilirlik fonksiyonu ile ağırlıklandırılmış en yüksek bilgiyi kullanmaktadır.

Maksimum Sonsal Ağırlıklandırılmış Bilgi: Sonsal dağılım ile ağırlıklandırılmış bilgi fonksiyonu kullanılmaktadır. Bir yetenek düzeyi için farklı ağırlıklandırılmalar yapmaktadır (Ho, 2010).

Maksimum Beklenen Bilgi: Ağırlıklandırılmış sonsal dağılım ile bilgi fonksiyonunu kullanmaktadır. Bireyin madde havuzundaki maddelere vereceği yanıtların olasılık dağılımı hesaplanır ve olasılık dağılımında hangi maddenin en yüksek bilgi vermesi bekleniyor ise o madde seçilir.

Minimum Beklenen Sonsal Varyans: Maksimum beklenen bilgi yönteminin küçük gruplarda uygulanmış halidir. Buradaki farklılık sonsal bilgi fonksiyonu yerine sonsal varyans fonksiyonu kullanılmaktadır (Ho, 2010).

Maksimum Beklenen Sonsal Ağırlıklandırılmış Bilgi: En Yüksek Olabilirlik Ağırlıklandırılmış Bilgi yöntemine benzer bir yöntemdir. Fisher bilgi fonksiyonu ve olabilirlik fonksiyonu yerine sonsal dağılımlı bilgi fonksiyonu kullanılmaktadır.

4.Yetenek Kestirim Yöntemi

BOBUT algoritmasında geleneksel testlerden farklı olarak yetenek kestirimi yapılmaktadır. Geleneksel testlerde testin sonunda yapılan tek bir puanlama ile bireyin yetenek kestirimi yapılırken BOBUT uygulamasında bireye yöneltilen her maddeden sonra anlık yetenek kestirimi yapılmaktadır. Bu yetenek kestirimi ilk yöneltilen maddeden başlayarak yapılır ve bu kestirimlere göre bir sonraki madde seçimi ya da testin sonlandırılması gerçekleştirilir.

Yetenek kestirimi yapılırken *Maksimum Olabilirlik* ve *Bayesçi* yöntemler olmak üzere iki farklı matematiksel yaklaşım vardır.

Maksimum Olabilirlik; yetenek kestiriminde en çok kullanılan yöntemlerden birisidir. Birey hakkında en fazla bilgi veren maddeyi seçmeye dayanmaktadır. Olabilirlik fonksiyonunun en yüksek değerine ulaştığı noktada yetenek kestirimi yapmaktadır. Maksimum olabilirlik yönteminde en önemli sınırlılık birey maddelere tümüyle doğru veya tümüyle yanlış cevap verdiğinde yetenek tahmini yapamamaktadır. Yöntemin uygulanabilmesi için en az bir farklı cevabın verilmesi gerekmektedir (Hambleton & Swaminathan, 1985; De Ayala, 2009; DeMars; 2010). Maksimum olabilirliğin getirdiği bu sınırlılıkların giderilmesi adına Bayesçi yöntemler önerilmiştir. *Maksimum Sonsal Kestirim (En Yüksek Posteriori)* ve Beklenen Sonsal Kestirim (*Posteriori İçin Beklenen Değer)* sık kullanılan yöntemlerdir.

Maksimum Sonsal Kestirim; bireyin önsel dağılımlarını kullanan bir yöntemdir. Maksimum Olabilirlik yöntemine göre daha az maddede daha kararlı sonuçlar üretebilmektedir (Hambleton & Swaminathan, 1985).

Beklenen Sonsal Kestirim; bu yöntem, bireyin yetenek seviyesini tahmin ederken, verilen yanıtlara dayalı olasılıkları ve önceki bilgilere dayalı bir öncül dağılımı dikkate alarak bir tahmin yapmaktadır. Çoklu ve ikili puanlanan veriler için kolaylıkla kullanılabilir.

5.Test Sonlandırma Kuralı

BOBUT, testi sonlandırmak için durdurma kurallarına ihtiyaç duymaktadır. Bireylerin yetenek düzeylerinin doğru belirlenebilmesi için testin ne zaman sonlanacağı önem arz etmektedir. Bu nedenle ilk zamanlar sabit uzunluklu durdurma kuralı ortaya atılmıştır. Daha sonra sabit uzunluğun kullanıldığı testlerde yaşanan aksaklıklardan dolayı zaman içerisinde değişken uzunluklu durdurma kuralları türetilmiştir. Günümüzde hala yeni test sonlandırma kuralları türetilmekte ve diğer sonlandırma kuralları ile karşılaştırmaların yapıldığı çalışmalar mevcuttur.

5.1.Sabit Uzunluk Kuralı

Sabit uzunluk durdurma kuralı altında, uyarlamalı bir test, önceden belirlenmiş bir madde sayısı uygulandığında sonlandırılır. Buna göre, tüm sınavı alanlara, testin sonunda elde edilen ölçüm hassasiyeti derecesinden bağımsız olarak aynı sayıda madde uygulanır. En temel avantajı bu kuralın basitliğidir. Ancak sabit uzunluktaki testler, özellikle madde havuzunun boyutu sınırlı olduğunda, yetenekleri madde havuzunun ortalama zorluk seviyesinden oldukça farklı olan bireyler için daha az doğru sonuçlar üretebilir (Finkelman, Nering, & Roussos, 2009) veya orta yetenek düzeyindeki bireylere gereksiz maddeler sunabilir. Yani bireyin yetenek düzeyinde daha az bilgiye sahip maddelerin gereksiz yere uygulanması BOBUT'un etkililiğini sınırlayabilir. Ayrıca

bireyselleştirilmiş testlerin doğası gereği her bireyin cevaplayacağı madde sayısı farklılaşacağından dolayı standart hatası yüksek yetenek kestirimleri elde edilebilir. Bu nedenle sabit uzunluk yönteminin tek başına kullanımı yerine diğer yöntemler ile birlikte kullanılması daha yaygındır (Şenel, 2023).

Çok boyutlu BOBUT uygulamasında bireylerin her alt boyut için aynı veya farklı sayıda madde alması için sabit durdurma kuralı belirlenebilir.

Sabit uzunluk kuralında eşit hassasiyette ölçüm sağlanamaması, yani farklı standart hataların bulunması, daha büyük standart hatalar ile ölçülen adaylar için "adil olmayan" test puanlarına yol açar. Bu nedenle, sabit uzunluklu BOBUT uygulamaları önerilmemektedir. Değişken sonlandırma kuralları, bireysel ölçüm hassasiyetini artırmak ve test sonuçlarının adil olmasını sağlamak için daha uygun bir yaklaşımdır (Weiss & Şahin, 2024).

Sabit uzunluk ve değişken uzunluklu durdurma kuralları sıklıkla birleştirilir. Böylece bir değişken uzunluklu sonlandırma ölçütü, katılımcıya maksimum sayıda madde uygulanana kadar durdurma kriteri olarak kullanılır. Maksimum test uzunluğunun yanı sıra bir değişken uzunluklu durdurma kuralı kullanmak, madde havuzu dağılımı ile katılımcının yetenek seviyesi arasında bir uyumsuzluk olduğunda faydalıdır. Çünkü katılımcıya, değişken uzunluklu durdurma ölçütünü karşılayacak hiçbir madde kalmadığı zaman madde havuzundaki tüm maddeler uygulanabilir. Bu nedenle BOBUT uygulamasının verimli olmasını, katılımcı yükünü ve madde maruziyetini azaltmasını sağlamak amacıyla, genellikle belirli bir madde sayısından sonra madde uygulamasını durdurmak

için ikincil bir sonlandırma ölçütü olarak sabit uzunluklu durdurma kuralı kullanılmaktadır (Stafford, Runyon, Casabianca, & Dodd, 2019).

5.1.Değişken Uzunluk Kuralı

Değişken uzunluklu durdurma kuralında, bazı sınavı alan bireylere diğerlerinden daha fazla madde verilmesini gerektirse bile tüm sınavı alanlar için belirli bir ölçüm hassasiyetine ulaşmayı hedeflemektedir. Sabit uzunluklu durdurma kuralının aksine testi alan her birey farklı sayıda madde ile karşılaşmaktadır. Değişken uzunluklu yöntemler kullanmak ölçmenin etkililiğini artırırken, bireylerin yeteneklerine daha uygun maddeler ile karşılaşırken testin süresinin de kısalmasını sağlamaktadır. En çok kullanılan yöntemlerin başında *Standart Hata Kuralı*, *En Düşük Bilgi Kuralı* ve *θ Değişimi Kuralı* gelmektedir.

5.1.1.Standart Hata Kuralı

Durdurma kuralları içerisinde sık kullanılan bir yöntemdir. Bireylerin her maddeye verdiği yanıt sonucunda yetenek kestirimi yapılırken, bu yetenek kestirime ilişkin standart hata değeri de hesaplanmaktadır. Yetenek kestirim sonucu elde edilen standart hata değeri test başlamadan önce belirlenen kabul edilebilir standart hata değerinden daha düşük ise test sonlandırılmaktadır. Bu değerin altına inene kadar bireye maddeler uygulanmaya devam eder. Maddeler bireyin yeteneğine göre uygulanacağı için her bireyin test uzunluğu farklılaşacaktır. Maksimum olabilirlik puanlaması yapıldığında bir θ'nın Standart Hatası (SE) aşağıda verilen denklem ile hesaplanmaktadır (Babcock & Weiss, 2009).

$$SE(\hat{\theta}) = \sqrt{\frac{1}{\frac{-\partial^2 \log L}{\partial \theta^2}}}$$

Sonlandırma kuralı olarak standart hata değeri genellikle 0.32 veya 0.20 olarak belirlenir (Van der Linden & Glas, 2010). Yetenek kestirimlerine ait standart hatalar arasındaki fark 0.01'den küçük olduğunda da test sonlandırılabilir (Kezer & Koç, 2014). Ancak gerçek madde havuzlarının nadiren eşit hassasiyette ölçüm sağlayabildiği göz önüne alındığında, bu kural genellikle bir ikincil sonlandırma kriteri ile birleştirilmelidir. İkincil bir sonlandırma kuralı uygulanmadıkça BOBUT, gözlemlenen standart hata üzerinde çok az etkisi olan veya hiç etkisi olmayan maddeleri bireye sunmaya devam eder. Basit bir ikincil sonlandırma kuralı olarak testte kullanılacak maksimum madde sayısı belirlenebilir. Örneğin, maksimum teste uygulanacak madde sayısı rastgele 25 olarak belirlendiğinde standart hata durdurma kuralı 25 madde sonra sağlanamadığında test sonlandırılır (Weiss & Şahin, 2024).

5.1.2.En Düşük Bilgi Kuralı

Madde seçiminde en yüksek bilgi yöntemlerinin kullanıldığı zaman uygulanabilecek bir kuraldır. Her maddeden yapılan yetenek kestirimi sonucu o yetenek düzeyine uygun en yüksek bilgi veren madde bireye sunulur. Bu şekilde madde havuzunda, kestirilen yetenek düzeyine uygun bilgi verecek madde kalmamışsa test sonlandırılır (Maurelli & Weiss, 1981). Düşük bilgi sağlayan maddelerin gereksiz yere uygulanmasını önler. Ayrıca ölçüm hassasiyeti standart hata kuralına kıyasla daha düşük olabilir (Wang, Weiss & Shang, 2019).

Sabit uzunluklu durdurma kuralının ikincil bir kural olarak kullanıldığı çalışmalar mevcuttur. Dodd, Koch ve De Ayala (1989), Aşamalı Tepki Modeline dayalı simüle edilmiş uyarlamalı testlerde minimum standart hata ve en düşük bilgi durdurma kuralı yaklaşımlarını karşılaştırmışlardır. Standart hata durdurma kuralının bilinen ve tahmin edilen yetenekler arasında daha yüksek korelasyonlar sağladığını ve en az bilgi kuralını geride bıraktığını belirtmişlerdir. Ayrıca, Dodd ve arkadaşları (1993), kısmi kredili tepki modeliyle puanlanan uyarlamalı testler için standart hata durdurma kuralının en az bilgi durdurma kuralına göre daha kısa testlerle daha yüksek ölçüm hassasiyeti sağladığını bulmuşlardır.

5.1.3. Theta (θ) Değişimi Kuralı

Babcock ve Weiss (2012), bir madde havuzunun belirli özellik aralıklarında yeterli iyi maddeler sunamadığı durumlarda kullanışlı θ değişim kuralı veya mutlak theta değişimi kuralını önermiştir. Bu kural, testin theta değerindeki değişimin belirli bir eşik seviyesine ulaşmadığında sonlandırılmasını öngörür.

BOBUT uygulamasında her maddeden sonra kestirilen yetenek kestirim değeri değişmektedir. Bireyin yeteneğine doğru yaklaştıkça yetenek kestirimleri arasındaki fark gittikçe azalacağı için önceden belirlenen kritik yetenek değerinin altına düştüğünde test sonlandırılır (Weiss & Kingsbury, 1984).

Hem tek boyutlu 2PLM model ve hem de çok boyutlu derecelendirilmiş yanıt modeli kullanılarak gerçekleştirilen bir çalışmada θ değişim kuralının tek başına durdurma kuralı olarak kullanımının kararsız sonuçlar doğurduğu ve testin erken sonlandırılmasına neden olabilecek bir dengesizlik yaratabileceği tespit edilmiştir. Yalnızca destekleyici yani ikincil bir kural olarak

kullanılmasının doğru olacağı belirtilmiştir (Wang, Weiss & Shang, 2019).

Babcock ve Weiss (2012), θ değişim kuralının klinik ölçümlerde etkili bir durdurma kuralı olduğunu belirtmiştir.

θ değişim kuralı için kesme değerleri daha keyfi bir şekilde belirlenmektedir. Babcock ve Weiss (2012), kesme değeri olarak 0.02 ve 0.05 kullanmıştır. Bu, genellikle gizli özelliğin $\pm\ 3\theta$ aralığında yer aldığı varsayımına göre toplam ölçeğin sırasıyla %0.3'ü ve %1'ine karşılık gelmektedir.

5.1.4.Tahmini Standart Hata Azaltımı

Choi, Grady & Dodd (2011) tarafından durdurma kurallarıyla ilişkili verimlilik ve ölçüm hassasiyeti problemlerini ele almayı amaçlayan yeni bir durdurma kuralı önerilmektedir. Tahmini Standart Hata Azaltımı olarak adlandırılan bu yeni durdurma kuralı, ek maddelerin uygulanmasının getireceği ölçüm hassasiyetindeki tahmini değişikliği dikkate alarak ölçüm hassasiyeti ve test verimliliği arasında bir denge sağlamaya çalışmaktadır.

Bu yaklaşım, öngörülen posterior dağılımı ve sonraki madde uygulamalarının getireceği beklenen hassasiyet artışını değerlendirerek gerçekleştirilir. Bu yöntemle, ek bir madde uygulanmasının getireceği tahmini ölçüm hassasiyeti kazancı, önceden belirlenmiş bir kriterin altında olduğunda uyarlamalı test sonlandırılır. Eğer bu kazanç önceden belirlenmiş kriterin üzerinde olursa test devam ettirilir.

Standart hata azaltımı durdurma kuralı, theta'nın öngörülen posterior varyansına dayanmaktadır. Bu aynı zamanda minimum beklenen posterior varyans madde seçimi kriterinin de hedef

fonksiyonudur. Eşdeğer şekilde maksimum beklenen posterior hassasiyet kriteri şu şekilde ifade edilebilir:

$$i_k = \arg\min_j \left\{ \sum_{r=1}^{m_j} p_j(r|u_{i1}, u_{i2}, \dots, u_{ik-1})\, Var(\theta|u_{i1}, u_{i2}, \dots, u_{ik-1}, U_j = r) : j \in R_k \right\}$$

$p_j(r|u_{i1}, u_{i2}, \dots, u_{ik-1})$, madde j için r yanıtının verilme olasılığını ifade eden sonsal öngörü dağılımını temsil eder. Bu, önceki yanıt geçmişi göz önünde bulundurularak tahmin edilen, m_j kategorisine sahip bir maddeye verilen yanıtın olasılığıdır.

$Var(\theta|u_{i1}, u_{i2}, \dots, u_{ik-1}, U_j = r)$, madde j için r yanıtı tahmin edildiğinde elde edilen sonsal varyansı ifade eder.

Bu hedef fonksiyon, maksimum bilgi seçimi kriterinin Bayesian bir karşılığıdır. Bayesian yaklaşımı, madde seçiminde ek bilgiyi en üst düzeye çıkarmayı amaçlayan klasik yöntemlerin ötesine geçerek, yanıt geçmişi ve olasılık dağılımını dikkate alır. Bu sayede hem ölçüm hassasiyeti hem de verimlilik artırılır.

Minimum standart hata durdurma kuralı ile birlikte, standart hata azaltımı kuralı iki ek eşik parametresi tanımlar. Bu parametreler hyper ve hypo olarak adlandırılmaktadır.

Hyper parametresi, standart hatadaki minimum beklenen azaltmayı belirler. Eğer bu kriter sağlanırsa, minimum standart hata durdurma kuralı tarafından verilen "durdur" sinyali geçersiz kılınır ve uyarlamalı test devam eder.

Hypo parametresi; yine standart hatadaki minimum beklenen azaltmayı belirler. Eğer bu kriter sağlanmazsa, minimum standart hata durdurma kuralı tarafından verilen "devam et" sinyali geçersiz kılınır ve uyarlamalı test sonlandırılır.

Örneğin, hyper ve hypo parametrelerinin sırasıyla 0,03 ve 0,01 olarak belirlendiğini varsayalım. Mevcut theta tahminine ilişkin standart hata değeri hedef durdurma eşiğinin altına düştüğünde, minimum standart hata durdurma kuralı "durdur" sinyali çağırır. Ardından, standart hata azaltımı kuralı, madde havuzunda kalan herhangi bir maddenin standart hatası 0,03 veya daha fazla azaltmasının beklenip beklenmediğini değerlendirir. Eğer bu soruya olumlu bir cevap verilirse, "durdur" sinyali geçersiz kılınır ve test devam eder. Benzer şekilde, minimum standart hata eşiği sağlanmadığında "devam et" sinyali çağrıldığında, standart hata azaltımı kuralı havuzda kalan herhangi bir maddenin standart hatasını en az 0,01 azaltmasının beklenip beklenmediğini değerlendirir. Bu soruya olumsuz bir cevap verilirse, "devam et" sinyali geçersiz kılınır ve test sonlandırılır.

5.1.5. Fisher Bilgi Matrisi Determinant Kriteri (Determinant Kuralı (D-Kuralı))

Bundan önceki test sonlandırma kuralları tek boyutlu BOBUT uygulamalarında rahatlıkla uygulanmaktadır. Fakat çok boyutlu BOBUT uygulamasında bütün boyutların doğruluğunun aynı anda dikkate alınması gerekmesi nedeniyle şimdiye kadar bahsedilen durdurma kuralları doğrudan çok boyutlu BOBUT'a uygulanamamaktadır. Dolayısıyla çok boyutlu BOBUT için sonlandırma kuralları arayışına gidilmiştir. Wang, Chang & Boughton (2013) çok boyutlu BOBUT için sonlandırma kuralı

oluşturmak için önce bir yetenek vektörünün tahmin doğruluğunu nicelendiren bir indeks seçilmesi gerektiğini ve böyle indeksler arasında genelleştirilmiş varyans, toplam varyans veya entropi yer aldığını belirtmişlerdir. Bu doğrultuda çok boyutlu BOBUT için dört farklı durdurma kuralı önermişlerdir. Bunlardan biri olan D-Kuralı Fisher bilgi matrisinin determinantını kullanarak, testin yeterli bir doğrulukla tamamlanıp tamamlanmadığını belirler. Test, Fisher bilgi matrisinin determinantının önceden belirlenen bir eşik değerine ulaştığında sona erer.

Wang, Chang & Boughton (2013) kesme değerini belirlemek için ampirik bir yöntem kullanmayı önermektedir. Bu, her bir testin doğruluk ve hassasiyet gereksinimlerine uygun olarak, geçmiş veriler veya deneysel analizlerle kesme değerlerinin en uygun şekilde belirlenmesi anlamına gelmektedir.

5.1.6.Genişletilmiş Fisher Bilgi Matrisinin En Küçük Özdeğer Kriteri (Özdeğer Kuralı (E-Kuralı))

E-kuralı çok boyutlu BOBUT uygulamasında durdurma kuralı olarak geliştirilmiştir. Wang, Chang & Boughton (2013) Chang ve Martinsek'in (1992) lojistik regresyon modeli için yaptığı ardışık analiz araştırmasından ilham alınarak geliştirildiğini belirtmiştir. Bu makalelerinde, bilinmeyen bir parametrenin ardışık tahmininin, istenen güven düzeyine sahip olarak gerçek parametrenin etrafında kalmasını sağlamak amacıyla bir "sabit boyutlu güven bölgesi" kavramını geliştirmişlerdir. Sonuçları, burada iki parametreli çok boyutlu madde tepki kuramı modelinin çoklu lojistik regresyon modelinin özel bir durumu olarak görülebileceği için kolayca uyarlanmaktadır.

E-kuralı, Fisher bilgi matrisinin en küçük özdeğerini izler. Fisher bilgi matrisi, bir testin doğruluğunun ne kadar güvenilir olduğunu gösteren bir ölçüdür ve özdeğerlerinin en küçüğü, testin doğruluğunun yeterli olup olmadığını belirlemede kullanılır. Bu kural, parametre tahmininin güvenilirliğini sağlamak ve testin daha fazla madde sunmasına gerek olup olmadığını kontrol etmek için tasarlanmıştır.

5.1.7.Örnek Varyans-Kovaryans Matrisinin İzi Kuralı (İz Kuralı (T-Kuralı))

Wang, Chang & Boughton (2013) tarafından ortaya atılan çok boyutlu BOBUT için uygulanabilecek bir diğer durdurma kuralı T-Kuralıdır. D ve E kuralı güven elipsoidi hacminin veya maksimum ekseninin minimize edilmesine odaklanmıştır. T kuralında ise araştırmacılar örnek varyansı başka bir şekilde tanımlayarak toplam varyansı sunmuşlardır. Toplam varyans, örnek varyans-kovaryans matrisinin izinden hesaplanır ve asimptotik olarak ters Fisher bilgi matrisinin izine eşdeğer olarak hesaplanmaktadır. Yani T-kuralı, örnek varyans-kovaryans matrisinin izini izler ve bu sayede testin doğruluğu ve güvenilirliği üzerinde bir kontrol sağlar. Bu kural, testin doğru bir tahmin yapabilmesi için gerekli olan toplam varyansı izler ve belirli bir eşiğe ulaşıldığında testin sonlandırılmasını sağlar.

T-Kuralı bilgi matrisinin tersini almayı içerdiğinden, eğer bilgi matrisi tekil ise sayısal zorluklarla karşılaşılabilir. Bu nedenle, bu kuralın hem simülasyonlarda hem de pratikte uygulanırken dikkatli olunması gerektiği belirtilmektedir.

5.1.8. İki Ardışık Sonsal Dağılım Arasındaki Kullback–Leibler (KL) Mesafesi (Kullback–Leibler Sapma Kuralı (K-kuralı))

Wang, Chang & Boughton (2013) tarafından ortaya atılan çok boyutlu BOBUT için uygulanabilecek bir diğer durdurma kuralı K-Kuralıdır. Bir testte her yeni madde uygulandıktan sonra, bireyin yeteneği hakkında elde edilen bilgiye ne kadar değişiklik getirildiğini ölçen bir yöntemdir. K-Kuralı, önceden belirlenen bir önsel dağılımından, her yeni madde ile elde edilen sonsal dağılımına geçişin bilgi kazancını ölçer. Diğer bir ifade ile K-Kuralı, iki ardışık dağılım arasındaki farklılıkları belirleyerek, her yeni maddeye verilen cevabın modelin doğruluğunu ne kadar artırdığına dair bir gösterge sağlar.

KL mesafesi küçükse, yeni bilgi çok az değişim yaratıyordur. Büyükse, yeni bilgi modelin tahminini önemli ölçüde değiştirmiştir şeklinde yorumlanır. Her yeni maddenin bilgi kazancını izleyerek testin sonlandırılmasında kullanılır. Eğer KL mesafesi belirli bir eşik değeri geçerse, test sonlandırılabilir.

Daha bilgilendirici bir önsel bilgi kullanıldığında, K-kuralı daha kısa test uzunlukları üretme eğilimindedir.

K-kuralı, en düşük bilgi kuralına oldukça benzerdir. Çünkü her ikisi de sıradaki maddelerin bilgisini değerlendirir ve sınav katılımcısının özellik düzeyine çok az bilgi katkı sağlayan maddelerin uygulanmasını engeller.

K-kuralının çekici bir özelliği, tek boyutlu uyarlanabilir testlerden çok boyutlu uyarlanabilir testlere hızlıca genellenebilmesidir. Aynı zamanda KL mesafesini yeni bir yanıtın sonsal tahmin dağılımı ile birleştiren K-kuralı, standart hata azaltımı kuralı ile aynı özelliği paylaşır (Wang, Chang & Boughton, 2013).

Sonuç

Test geliştiriciler, sonlandırma kuralının seçiminin genellikle testin amacı, madde havuzu özellikleri ve operasyonel kısıtlamalara bağlı olduğunu belirtmişlerdir. Örneğin, birçok durumda adaylar arasında eşit derecede hassas puanların elde edilmesi önceliklidir. Bu, test puanlarına dayalı olarak yapılan kararların ve yorumların tüm adaylar için eşit derecede güvenilir olmasını sağlar. Ancak diğer durumlarda, değişken uzunluklu testlerde zaman zaman karşılaşılan uzun test süreleri hem adaylar hem de test yöneticileri için aşırı yük olarak değerlendirilebilir. Bu operasyonel zorlukları hafifletmek için değişken uzunluklu testler, uygulanan maksimum madde sayısına veya her aday için izin verilen maksimum test süresine üst sınır getirilerek uygulanmıştır. Bazı durumlarda ise, test geliştiriciler, test süresindeki değişkenliği ve buna bağlı olarak ortaya çıkabilecek yorgunluğu standartlaştırmak amacıyla sabit uzunluklu testleri tercih etmişlerdir (Boyd, Dodd & Choi, 2010; Segall, 2005).

Bir testin ne zaman sonlandırılacağına karar verirken, ilk olarak sabit bir madde sayısının tüm adaylara uygulanmasının mı yoksa farklı adayların farklı sayılarda madde yanıtlamasına olanak tanıyan bir değişken sonlandırma kuralının mı kullanılacağı belirlenmelidir. Bu iki sonlandırma kuralı, ölçümlerin farklı kalitelerde olmasına yol açacaktır (Weiss & Şahin, 2024). Farklı sonlandırma kurallarının bir arada kullanılması da önerilen bir yaklaşımdır (Babcock & Weiss, 2009).

Farklı durdurma kuralları çeşitli simülasyon ve gerçek veri çalışmalarında değerlendirilmiş olsa da, bu kurallar arasındaki ilişkiler tam olarak net değildir. Bu nedenle, hangi durumlarda hangi kuralın kullanılacağına dair net bir kılavuz eksikliği bulunmaktadır.

Bu doğrultuda alanda hala çeşitli çalışmalar yapılarak kurallar karşılaştırılmakta güçlü ve eksik yönleri tartışılmaktadır. Bu doğrultuda eksiklikleri giderebilecek doğru ve hassas ölçümler sağlayacak test sonlandırma kurallarının halen geliştirilmesine ihtiyaç bulunmaktadır.

Kaynakça

Babcock, B., & Weiss, D. J. (2009). Termination criteria in computerized adaptive tests: Variable-length CATs are not biased. In Proceedings of the 2009 GMAC conference on computerized adaptive testing (Vol. 14).

Babcock, B., & Weiss, D. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? Journal of Computerized Adaptive Testing. https://doi.org/10.7333/1212-0101001

Boyd, A. M., Dodd, B. G., & Choi, S. W. (2010). Polytomous models in computerized adaptive testing. Handbook of polytomous item response theory models, 229-255.

Choi, S. W., Grady, M. W., & Dodd, B. G. (2011). A new stopping rule for computerized adaptive testing. Educational and Psychological Measurement, 71(1), 37-53.

De Ayala, R. J. (2013). *The theory and practice of item response theory.* Guilford Publications.

DeMars, C. (2010). *Item response theory*. Oxford University Press.

Dodd, B. G., Koch, W. R., & De Ayala, R. J. (1989). Operational characteristics of adaptive testing procedures using the graded response model. *Applied Psychological Measurement, 13*(2), 129-143.

Dodd, B. G., Koch, W. R., & De Ayala, R. J. (1993). Computerized adaptive testing using the partial credit model: Effects

of item pool characteristics and different stopping rules. *Educational And Psychological Measurement, 53*(1), 61-77.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory.* Psychology Press.

Hambleton, R. K. (1991). Fundamentals of item response theory. Sage Puplications.

Hambleton, R. K., & Swaminathan, H. (2013). Item response theory: principles and applications. Springer Science & Business Media.

Ho, T. H. (2010). A comparison of item selection procedures using different ability estimation methods in computerized adaptive testing based on the generalized partial credit model. The University of Texas at Austin.

Kezer, F. (2020). *Yeni başlayanlar için bilgisayar ortamında bireye uyarlanmış test yöntemi*. Pegem Akademi.

Kezer, F., & Koç, N. (2014). Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması. *Eğitim Bilimleri Araştırmaları Dergisi, 4*(1), 145-174.

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores.* IAP.

Maurelli, V. A., & Weiss, D. J. (1981). *Factors ınfluencing the psychometric characteristics of an adaptive testing strategy for test batteries.*

Reckase, M., D. (2009). *Multidimensional item response theory: statistics for social and behavioral sciences.* New York, NY: Springer.

Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: from inquiry to operation.* American Psychological Association.

Segall, D. O. (2005). Computerized adaptive testing. *Encyclopedia of Social Measurement, 1*(429-438), 4.

Sireci, S. G. (2003). C*omputerized adaptive testing: an ıntroduction.*

Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and Psychological Measurement, 75*(6), 954-978.

Şenel, S. (2023). *Bilgisayar ortamında bireye uyarlanmış testler.* Pegem Akademi.

Stafford, R. E., Runyon, C. R., Casabianca, J. M., & Dodd, B. G. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods, 51*, 1305-1320.

Thompson, N. A. & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation, 16*(1), 1.

Van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika, 63*, 201-216.

Van der Linden, W. J., & Glas, C. A. (2010). *Elements of adaptive testing.* New York: Springer.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized adaptive testing: A primer.* Routledge.

Wang, C., Chang, H. H., & Boughton, K. A. (2013). Deriving stopping rules for multidimensional computerized adaptive testing. *Applied Psychological Measurement, 37*(2), 99-122.

Wang, W. C., & Chen, P. H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement, 28*(5), 295-316.

Wang, C., Weiss, D. J., & Shang, Z. (2019). Variable-length stopping rules for multidimensional computerized adaptive testing. *Psychometrika, 84*, 749-771.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of educational measurement, 21*(4), 361-375.

Weiss, D. J., & Sahin, A. (2024). *Computerized adaptive testing: from concept to ımplementation.* Guilford Publications.

BIDGE