



Eğitim Araştırmalarında Likert Ölçeği ve Likert-Tipi Soruların Kullanımı ve Analizi*

İbrahim TURAN¹ Ümit ŞİMŞEK² Hasan ASLAN³

Öz: Rensis Likert tarafından 1932 yılında ortaya atıldığından bu yana geçen sürede Likert ölçeği ve Likert-tipi sorular sosyal bilimler, siyaset bilimi, psikoloji, pazarlama ve eğitim gibi pek çok alanda en çok kullanılan tutum ve eğilim ölçüm tekniği haline gelmiştir. Ancak “Likert Ölçeği” ve “Likert-tipi soru” kavramlarının birbirleriyle karıştırılması bu ölçeklerden elde edilen verilerin yanlış analizine ve yorumlanmasına yol açmaktadır. Bir başka tartışma konusu da Likert ölçeği veya Likert-tipi sorular ile elde edilen verilerin analizinde parametrik testlerin mi yoksa parametrik olmayan testlerin mi kullanılacağı hususudur. Bu araştırmanın amacı literatürde “Likert Ölçeği” ve “Likert-tipi Soru” kavramları ve bu tür ölçeklerde kullanılacak analiz türleri üzerinde yapılan tartışmaları incelemek, eğitim araştırmalarında bu kavram ve yöntemlerden hangisinin tercih edildiğini tespit etmek ve son olarak bu tip ölçeklerden elde edilen verilerin istatistiksel analizinde hangi testlerin uygulanması gerektiğini örnek analizler ile ortaya koymaktır. Araştırma sonucunda konuyla ilgili kavramların birbirleriyle karıştırıldığı, Likert-tipi sorularda çoğunlukla parametrik olmayan, Likert ölçeklerinde hangi testin daha tutarlı ve güvenilir olduğu konusunda bir görüş birliği olmamasına rağmen genellikle parametrik testlerin kullanıldığı tespit edilmiştir.

Anahtar Kelimeler: Eğitim araştırması; likert ölçeği; likert-tipi soru; parametrik testler; parametrik olmayan testler.

* Bu çalışma 3-5 Ekim 2013 tarihleri arasında Trabzon’da düzenlenen VI. Sosyal Bilimler Eğitimi Kongresinde sunulan bildirinin gözden geçirilmiş ve genişletilmiş halidir.

¹ İstanbul Üniversitesi, Hasan Ali Yücel Eğitim Fakültesi, İlköğretim Bölümü, ibrahim.turan@istanbul.edu.tr

² Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi, İlköğretim Bölümü, simsekum@atauni.edu.tr

³ Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi, Ortaöğretim Bölümü, hasanaslan_25@hotmail.com

The Use and Analysis of Likert Scales and Likert-Type Items in Educational Research

Abstract: Since its introduction in 1932 by Rensis Likert, Likert scales and Likert-type items have become the most used attitude and tendency measurement technique in many areas like, social sciences, political sciences, psychology, marketing and education. However, confusion in "Likert Scale" and "Likert-type item" concepts leads to misinterpretation and false analysis of the data obtained from these scales. Another topic of discussion is whether parametric or non-parametric tests should be used in analyzing the Likert scale and Likert-type items. The purpose of this research, is to clarify concepts of "Likert Scale" and "Likert-type item", to determine which concept and method is more preferred in educational research, and to examine which test should be used in the statistical analysis of data obtained from this type of scales. The result of the study show that relevant concepts are mixed with each other, Likert-type items are mostly analyzed with non-parametric tests, and generally parametric tests were used in analyzing Likert scales despite the lack of a consensus on consistent and reliable method.

Keywords: Educational research; likert scale; likert-type item; parametric tests; non-parametric tests.

Giriř

Eđitim arařtırmalarında sıklıkla llmeye alıřılan deđiřkenlerden biri de tutumdur. Tutum, "belirli nesne, durum, kurum, kavram ya da diđer insanlara karřı đrenilmiř, olumlu ya da olumsuz tepkide bulunma eđilimi" (Tezbařaran, 2008, s.1) olarak tanımlanmıřtır. Birey ve grupların tutum, eđilim ve grřlerini lmek iin bugne kadar Bogardus tarafından geliřtirilen "Toplumsal Uzaklık leđi," L. L. Thurstone'un "Eřit Grnml Aralıklar" leđi, L. Guttman'ın "Yıđıřımlı lekleme" tekniđi ve Rensis Likert'in "Dereceleme Toplamlarıyla lekleme" modeli gibi farklı lekler kullanılmıřtır (Tezbařaran, 2008, s.5). Ancak bunlar arasında en yaygın kullanıma sahip olanı (Judd, Eliot ve Kidder, 1991, Akt. Tezbařaran, 2008) Rensis Likert (1932) tarafından Thurstone leđinin basitleřtirilmiř bir versiyonu olarak geliřtirilen Likert leđidir (Cramer ve Howitt, 2004: 89). Uygulaması, kodlaması ve lmesi gayet kolay olduđu iin (Spector, 1992) sosyal bilimler, siyaset bilimi, psikoloji, pazarlama ve eđitim gibi pek ok alanda sıklıkla bařvurulan teknik (Edmondson, 2005) haline gelmiřtir.

Likert-tipi sorular araştırılan konu hakkında tutum veya görüş içeren bir ifade ve bu ifadeye katılım düzeyini belirten seçenekler içerir. Likert-tipi sorularda katılım düzeyini belirlemek amacıyla iki aşırı uç arasında yer alan birden çok seçenek sunulur. Bu seçenekler “en yüksekte en düşüğe” veya “en iyiden en kötüye” doğru dereceli bir şekilde sıralanır. Analiz aşamasında bu seçenekler derecelerine göre birer sayısal değer atanarak kodlanır ve böylece nitel veri nicel veriye dönüştürülerek analiz edilir.

Likert “Bir Tutum Ölçüm Tekniği” isimli çalışmasında “Sosyal tutum ölçülebilir mi?” ve “İki kişinin tutumu birbirinden ayırt edilebilir mi?” sorularına cevap aramıştır (1932: 8). 3 seçenekli ifadeler ve çoktan seçmeli sorulara da yer verdiği bu çalışmasında kendi deyimiyile “Kesinlikle Onaylıyorum İfadeleri” olarak adlandırdığı 5’li Likert-tipi soruların diğer soru tipleri karşısında etkililiğini araştırmıştır.

Tablo 1. Likert’in Tezinde Kullandığı 5 Seçenekli Ölçüm Sistemi.

Kesinlikle Onaylıyorum	Onaylıyorum	Kararsızım	Onaylamıyorum	Kesinlikle Onaylamıyorum
------------------------	-------------	------------	---------------	--------------------------

Orijinali 5 seçenekten oluşan (Tablo 1) Likert-tipi sorular günümüzde 3’ten 7’ye kadar farklı seçenek sayısı ve farklı etiketleme sistemleri ile kullanılmaktadır. Şekil-1’de Likert-tipi soruların farklı kullanımlara dair örnekler sunulmuştur.

Şekil 1. Farklı Seçenek Sayısı ve Etiketleme Sistemi ile Hazırlanmış Likert-tipi Soru Örnekleri.

Örnek-1: 3’lü Likert-tipi soru						
Katılıyorum	Kararsızım	Katılmıyorum				
Örnek-2: 4’lü Likert-tipi soru						
Tamamen Destekliyorum	Destekliyorum	Desteklemiyorum	Hiç Desteklemiyorum			
Örnek-3: 5’li Likert-tipi soru						
Çok Seviyorum	Seviyorum	Ne Seviyorum Ne Sevmiyorum	Sevmiyorum	Hiç Sevmiyorum		
Örnek-4: 6’lı Likert-tipi soru						
Çok iyi	İyi	Biraz İyi	Biraz Kötü	Kötü	Çok Kötü	
Örnek-5: 7’li Likert-tipi soru						
Tamamen Uygun	Uygun	Biraz Uygun	Tarafsızım	Biraz Uygunsuz	Uygunsuz	Kesinlikle Uygunsuz

Yukarıda bahsedilen çalışmasında Likert katılımcıların uluslararasılık konusundaki tutumlarını ölçmek için 9 adet 5’li Likert-tipi soru, Afro-Amerikan hakları konusundaki

tutumlarını lmek iin 6 adet 5’li Likert-tipi soru ve smrgecilik konusundaki tutumlarını lmek iin 8 adet 5’li Likert-tipi soru kullanmıřtır (1932:15-20). Katılımcıların herhangi bir konu (boyut) hakkındaki tutumlarını ortaya koymak iin soruları tek tek ele almamıř o boyutu lmek iin kullandıđı tm soruları bir arada analiz etmiřtir.

Son derece yaygın bir kullanıma sahip olmakla birlikte Likert leđi ve Likert-tipi soruların dođru kullanımı ve analizi konusunda uzun sren bir kafa karıřıklıđı ve anlařmazlık sz konusudur (Carifio ve Perla, 2008: 1150). Bazı arařtırmacılar Likert verilerin *t* test gibi parametrik testler kullanılarak analiz edilebileceđini diđerleri ise bu tip verilerin analizinde parametrik olmayan testlerin kullanılmasının daha dođru olduđunu savunmaktadır.

Bu arařtırmanın amacı literatrde yer alan farklı grřlere yer vererek “Likert leđi” ve “Likert-tipi Soru” kavramlarına aıklık getirmek, eđitim arařtırmalarında bu kavram ve yntemlerden hangisinin tercih edildiđini tespit etmek ve bu tip leklerden elde edilen verilerin istatistiksel analizinde hangi testlerin uygulanması gerektiđini rnek analizler ile ortaya koymaktır.

“Likert-tipi Soru” ve “Likert leđi” Kavramlarının Ayrımı

Literatrde bu tip soru ve lekleri tanımlarken “Likert leđi”, “Likert tipi lek”, “Likert tipi soru”, “5’li Likert leđi”, “5’li Likert-tipi soru”, “Likert tipi anket” gibi ok farklı kavramların ve ođu zaman birbirinin yerine kullanıldıđını grmekteyiz. Bu arařtırma kapsamında YK Tez Merkezinde tarama yapılarak Eđitim Bilimleri alanında Likert-tipi soru veya Likert leđi kullanılarak yapılmıř tezler arasından ulařılabilir rneklem yolu ile seilmiř 64 adet Yksek Lisans ve Doktora tezi incelenmiřtir. İncelenen tezlerde ođunlukla “Likert tipi lek” ve “Likert tipi anket” ifadelerinin kullanıldıđı grlmřtr. Bu konudaki kavram karmařası incelenen tezlerde de gzlemlenmiř ve arařtırmacıların ođu zaman “Likert-tipi lek” ile “Likert tipi anket” kavramlarını veya “Likert-tipi soru” ile “Likert-tipi lek” kavramlarını bir arada ve birbirlerinin yerine kullandıkları saptanmıřtır.

Akademik yazımda bu kavramlar dikkatli seilmeli ve kesinlikle birbirinin yerine kullanılmamalıdır. nk kavramların birbirinin yerine kullanılması sadece yazım hatası deđil, daha byk istatistiksel hatalara yol aan bir durumdur. Yukarıda sayılan farklı kavramlar arasında tercih edilmesi gereken alan yazımda tanımlanmıř ve aıklanmıř olan

“Likert-tipi soru” veya “Likert ölçeği” kavramlarıdır. Bu iki kavram istatistiksel analizde doğru test seçimi için önemli olduğundan burada açıklanmaya çalışılacaktır.

Likert Tipi Soru

Likert’in kendi çalışmasında kullandığı soru tipine benzeyen ancak seçenek sayısı veya kullanılan kelimeler farklı olan tek soruları ifade eder. Likert tipi soruların kullanıldığı araştırmalarda her ne kadar birden çok soru kullanılsa da araştırmacının bu soruların ortalama değerlerini kullanarak genel bir çıkarımda bulunma amacı yoktur. Sorular birbirinden bağımsız olarak tek tek ele alınır. Tablo-2’de her biri başka bir değişkeni ölçmeye yönelik hazırlanmış dolayısıyla birbirinden bağımsız değerlendirilmesi gereken Likert-tipi sorulara örnek verilmiştir.

Tablo 2. Likert-tipi Sorular

	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
İstanbul Üniversitesini seçtiğim için memnunum					
Ailem üniversite tercihimde bana çok yardımcı oldu					
Danışman hocalarımız bize yeterince yardımcı olmaktadır					

Likert-tipi Sorular Konusunda Literatürde Yer Alan Tartışmalar

Seçenek Sayısı

Likert-tipi sorulara anketlerinde yer veren araştırmacılar Likert’in yaptığı gibi seçenek sayısı olarak genelde 5’li ölçek kullanmaktadırlar. Ancak literatürde 3’ten 18’e kadar farklı seçenek sayılarının kullanıldığı ve en uygun seçenek sayısının tartışma konusu olduğu görülmektedir (Preston ve Colman, 2000: 2). Tartışmanın kaynağı kullanılacak seçenek sayısının ölçeğin geçerlik ve güvenilirliği üzerine olası etkisidir. Jacoby ve Matell (1971: 499) seçenek sayısının araştırma sonucu üzerinde anlamlı bir etkisinin olmadığını bu nedenle 3 seçenekli bir ölçeğin yeterli olacağını ifade etmektedir. Diğer taraftan 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ve 101 seçenektan oluşan ölçekleri geçerlik, güvenilirlik ve kullanıcı tercihleri açısından karşılaştıran Preston ve Colman (2000) bunun aksini iddia etmektedir. Preston ve Colman (2000: 12) araştırmalarında 2, 3 ve 4 seçenektan oluşan ölçeklerin en düşük, 5 ve 6 seçenektan oluşan ölçeklerin orta, 7-10 seçenektan oluşan ölçeklerin ise en yüksek geçerlik, güvenilirlik ve

kullanıcı tercihi değerlerine sahip olduklarını, 7-10 seçenekten oluşan ölçekler ile 101 puanlı ölçek arasında ise anlamlı bir farklılık bulunmadığını ortaya koymuşlardır. Seçenek sayısının belirlemede katılımcı profili de önemli bir unsurdur. Küçük yaştaki katılımcılar için daha az seçenek sayısı önerilmektedir (Adelson ve McCoachk, 2010: 797).

Seçenek Etiketleri

Likert-tipi soruların avantajlarından biri de seçeneklerin etiketlenmesinde araştırmacılara sağladığı geniş serbestliktir (Bkz. Şekil-2). Bu etiketlerin anket veya ölçekte gösterilmesinde iki farklı uygulamanın var olduğunu görmekteyiz; a) Tüm seçeneklerin etiketlenmesi, b) Sadece uç değerlerin etiketlenerek ara değerlerin boş bırakılması. Weijters, Cabooter ve Schillewaert (2010: 244) bu iki sistem arasındaki farkı ortaya koydukları araştırmalarında tüm seçenekleri etiketlemenin algılama, fark edilme ve çekiciliği arttırdığını, yanlış işaretleme, boş bırakma ve aşırı uçlara yönelme eğilimini azalttığını ortaya koymuşlardır.

Tarafsızlık/Kararsızlık Seçeneği

Ölçek geliştiriciler Likert-tipi sorularda kullanılabilecek en uygun seçenek sayısının tek sayı mı yoksa çift sayı mı olduğu üzerinde tartışmaktadırlar (Adelson ve McCoachk, 2010: 797). Bu tartışmanın asıl amacı tarafsızlık veya kararsızlık ifade eden seçeneğine yer verilip verilmeyeceğidir. Bazı araştırmacılar tarafsızlık/kararsızlık seçeneğine yer verilmesinin ölçeğin ayırt ediciliğini ve dolayısıyla güvenilirliğini arttıracaklarını iddia ederken, bazı araştırmacılar ise bu seçeneğin kaldırılmasının katılımcıları daha çok düşünmeye sevk edip daha kesin cevaplar vermelerini sağlayacağını belirtmektedir (Garland, 1991: 70).

Burada tartışılan bir başka husus da araştırmacıların tarafsızlık seçeneğini kullanma amacıyla katılımcıların bu seçenekleri algılayış biçiminin her zaman örtüşmediği görüşüdür. Araştırmacılar bu seçeneği, katılımcılar soruda verilen (pozitif ve negatif) iki uç seçenek arasında, her iki seçeneğe eşit uzaklıkta kaldıklarında seçebilmeleri için kullanmaktadırlar. Yani araştırmacılar için bu seçenek çoğu zaman ortalama, ılımlı bir konumu belirtme amacı gütmektedir (Kulas ve Stachowski, 2009). Ancak bu seçenek katılımcılar için her zaman ortalama veya ılımlı bir anlam ifade etmemektedir (Kulas, Stachowski, & Haynes, 2008). Shaw ve Wright 'a (1967) göre katılımcılar üç durumda bu seçeneği işaretlemektedirler:

1. Konu hakkında her hangi bir tutum veya fikirleri olmadığı zaman,
2. Konu hakkındaki tutum veya fikirlerini tam olarak tanımlayamadıkları zaman,

3. Değerlendirme açısından dengede oldukları zaman.

Bunun dışında katılımcılar konuya ilgisiz veya kayıtsız olduklarında (Nowlis, Kahn ve Dhar, 2002), kişisel sorularda gerçek cevaplarını gizlemek istediklerinde (Tourangeau, Smith, & Rasinski, 1997), diğer cevapları kendilerine uygun bulmadıklarında, konu hakkında yeterli bilgiye sahip olmadıklarında veya kesin cevaplar vermekten kaçındıklarında (Stone, 2004) bu seçeneği işaretlemektedirler.

Tarafsızlık/Kararsızlık Seçeneğinde Kullanılacak İfade

Yukarıda katılımcıların sorulan sorular hakkında yeterli bilgiye sahip olmadıklarında ve kendilerine başka bir seçim hakkı verilmediğinde tarafsızlık/kararsızlık seçeneğini işaretleyerek soruyu geçıştirdiklerinden bahsedilmiştir. Bu durum veri toplama aracının geçerlik ve güvenilirliğini olumsuz etkilemektedir. Ryan ve Garland (1999, 109) yaptıkları üç uygulamada kullanılan anketlerde “bilmiyorum” seçeneğine yer verildiğinde katılımcıların ortalama % 20’sinin bu seçeneği işaretlediklerini ortaya koymuşlardır.

Likert-tipi Sorulara Getirilen Eleştiriler

Uygulaması, kodlaması ve ölçmesinin kolay olması, farklı sayıda seçenek kullanımına izin vermesi, seçeneklerin etiketlenmesinde araştırmacılara serbestlik tanınması ve katılımcılar için kolay anlaşılır olması (Tavakoli, 2012; Spector, 1992), tutum ve görüş araştırmalarında Likert-tipi soruların sıklıkla kullanılmasını sağlamıştır. Ancak Likert-tipi soruların eleştirildiği hususlar da söz konusudur. Literatürde Likert-tipi sorulara yöneltilen eleştiriler beş ana başlık altında toplanabilir:

1. *Eşit Aralık*: Seçenekler arasında eşit aralık söz konusu değildir (Tavakoli, 2012; Goldstein ve Hersen, 1984).
2. *Zorlanmış seçenek*: Birçok anket ve ölçek sorusunda olduğu gibi Likert tipi sorularda katılımcılardan verilen seçenekler arasından kendisine en uygun olanı seçmesini istenmesi çoğu zaman katılımcıları “kötünün en iyisini” seçmeye zorlamaktadır.
3. *Kabullenme eğilimi*: Katılımcıların sorunun içeriğine bakmadan olumlu seçeneği işaretleme eğilimidir.
4. *Merkeze yönelik eğilimi*: Yukarıda bahsedildiği üzere çeşitli nedenlere katılımcılar Likert tipi sorularda ortada yer alan seçeneği işaretleme eğilimindedir. Bu nedenle tarafsızlık/kararsızlık seçeneğine yer verilip verilmemesi tartışma konusudur. Yine bu

seçenekte kullanılan “Kararsızım”, “Fikrim Yok” gibi ifadeler araştırmacılar tarafından eleştirilmektedir.

5. *Aşırı uçlara yönelik eğilimi:* İnsanlar “katılıyorum – tamamen katılıyorum” veya “onaylamıyorum – kesinlikle onaylamıyorum” gibi iki dereceli bir seçenekte karşılaştıklarında en olumsuz veya en olumlu seçeneği seçme eğilimindedirler (Javaras ve Ripley, 2007; Baumgartner ve Steenkamp, 2001; Cronbach, 1950; Cronbach, 1946).

Likert Ölçeği

Likert ölçeği, birden çok Likert-tipi sorunun bir araya getirilerek kullanıldığı ölçekleri ifade eder. Tek bir araştırma problemini cevaplandırmak amacıyla iki veya daha fazla Likert tipi soru oluşturmak ve analiz aşamasında bu soruların ortalama (birleştirilmiş) değerlerini kullanmak şeklinde tanımlanmıştır (Clason ve Dormody, 1994). Bu ölçekte amaç tüm soruların birleştirilmiş değerlerinden insanların bu konular üzerindeki ortalama tutumlarını belirlemektir. Tablo-3’de sağlıklı beslenmeye yönelik tutumu ölçmek amacıyla birden çok Likert-tipi sorunun bir arada kullanılmasıyla oluşturulmuş Likert ölçeğine örnek verilmiştir.

Tablo 3. Likert Ölçeği Örneği

	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
Sağlıklı beslenmeye çalışırım.					
Fastfood her zaman ilk tercihimdir.					
Yemek pişirirken mümkün olduğunca az yağ kullanmaya çalışırım.					
Abur-cubur tarzı gıdaları sadece öğün aralarında tercih ederim					

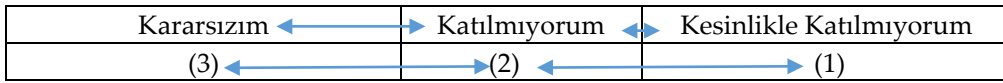
Likert ölçekleri toplamalı ölçeklerdir çünkü katılımcının bir konu hakkındaki genel görüşüne ulaşmak için tek tek soruları verdiği cevaplar toplanır. Bu nedenle Likert ölçeği aynı değişkeni ölçtüğü varsayılan Likert-tipi soruların rastgele bir araya getirilmesi ile oluşturulamaz. Öncelikle biçimsel ve dil açısından bütünlük sağlanmalı, sorularda olgusal ifadeler yer verilmemeli, ifadeler yanlış anlamalara yol açmayacak şekilde açık ve net olmalı, eşit miktarda olumlu ve olumsuz ifadeye yer verilmelidir (Tezbaşaran, 2008, 12-13).

İkinci olarak belirlenen soruların aynı değişkeni ölçtüğünden emin olmak için pilot çalışma üzerinde madde analizi yapmak gereklidir. Likert ölçeğinde kullanılan her bir sorunun ölçülmek istenilen tutum veya görüşü ölçme gücünü belirlemek için iki farklı analiz kullanılır: a) Korelasyonlara dayalı analiz, b) İç tutarlık ölçütüne (alt ve üst grup ortalamaları farkına) dayalı analiz (McIver ve Carmines 1982, Akt. Tezbaşaran, 2004, 78).

Likert-tipi Soruların İstatistiksel Analizi

Likert-tipi sorular 1'den başlayarak soruda kullanılan seçenek sayısına kodlanır. Bu kodlamada en olumsuz cevap en düşük (1) rakamı ile en olumlu cevap ise en yüksek rakam ile temsil edilir. 1'den başlayıp devam eden bu sıralı kodlamadan dolayı Likert tipi sorular sıralı (ordinal) veri olarak kabul edilir. Likert-tipi sorular aralı (interval) veri değildir. Çünkü yukarıda bahsedilen kodlama sisteminde kullanılan rakamlar arasında matematiksel olarak eşit aralık var iken bu rakamların temsil ettiği ifadeler arasında eşit bir aralık söz konusu değildir (Tavakoli, 2012: 326; Goldstein ve Hersen, 1984, 52). Şekil 2'de görüleceği üzere bu kodlama sisteminde kullanılan 3 ile 2 veya 2 ile 1 rakamları arasındaki aralık eşit iken bu soruları cevaplayan katılımcıların bu üç cevabı birbirine eşit uzaklıkta gördükleri söylenemez. Çoğunluk için "Katılmıyorum" ile "Kesinlikle Katılmıyorum" cevapları arasındaki fark "Kararsızım" ile "Katılmıyorum" cevapları arasındaki farktan daha azdır. Bu alanda yapılan ampirik araştırmalar da katılımcıların Likert tipi sorularda seçenekler arasındaki farkı eşit olarak algılamadıklarını ortaya koymaktadır (Hart, 1996).

Şekil 2. Likert Tipi Soruların Algılanışı.



Likert tipi sorular ordinal veri olarak kabul edildiğinden bu sorulardan elde edilen verilerin istatistiksel analizinde parametrik olmayan istatistiksel testlerin kullanılması gerekmektedir. Likert tipi sorulardan oluşan verilerin analizinde tanımlayıcı istatistik yani aritmetik ortalama yerine medyan ve mod, standart sapma yerine ranj, grafik yerine histogram kullanılması, fark ve korelasyon hesaplamalarında ise parametrik olmayan testlerin kullanılması daha doğrudur (Boone ve Boone, 2012).

Ancak literatür taramaları arařtırmacıların çođu zaman bu konuda hataya düřtüklerini ortaya koymaktadır. Clason ve Dormody (1994) Agricultural Education dergisinde yayınlanmış ve veri toplama aracı olarak Likert öleđi veya Likert tipi soruları kullanmış 188 makaleyi incelemiřlerdir. Bu inceleme sonucunda Likert tipi soruların analizinde farklı istatistiksel yöntemlerin kullanıldığını ortaya koymuřlardır. İncelenen 188 makaleden 95'inde Likert tipi sorular kullanılmış ve bunların % 54'ünde sadece tanımlayıcı istatistik (aritmetik ortalama, standart sapma, frekans vb.), % 13'ünde parametrik olmayan testler (ki-kare, Mann-Whitney U testi, Kruskal-Wallis testi vb.) ve 34'ünde parametrik testler (t-test, ANOVA vb.) kullanılarak istatistiksel analiz yapılmıştır (s.31).

Arařtırma kapsamında incelediđimiz 65 yüksek lisans ve doktora tezinden 4'ünde Likert-tipi soru kullanıldığını ve bunların 3'ünde verilerin analizi için parametrik olmayan testlerin kullanıldığını ancak birinde parametrik testlerin kullanıldığını tespit edilmiştir.

Likert Öleđinin İstatistiksel Analizi

Likert-tipi soruların analizinde parametrik olmayan testlerin kullanılması hususunda görüş birliđi varken Likert öleđinin analizinde kullanılacak test türü konusunda iki ayrı görüş söz öne sürülmektedir. Birinci görüşü destekleyen uzmanlara göre sıralı (ordinal) veri her zaman ordinal veridir ve hiçbir şekilde aralı (interval) veri haline getirilemez (Jamieson, 2004). Bu nedenle bazı arařtırmacılar ordinal veri analizinde parametrik testlerin kullanılmasını "istatistiksel analizin yedi ölümcül günahından biri" olarak tanımlamışlardır (Kuzon, Urbanchek, ve McCabe, 1996).

İkinci görüşe sahip uzmanlara göre ise Likert öleđinde yer alan sorulara verilen cevapların aritmetik ortalamasının alınması veriyi aralı (interval) veri haline getirir ve bu aritmetik ortalama üzerinde parametrik testler uygulanabilir (Carifio ve Perla, 2008). Yine bazı arařtırmacılara göre test türünün seçiminde veri türüne kıyasla örneklem büyüklüğü ve dağılımın türü daha önemli kıstaslardır (Knapp, 1990, Akt. Jamieson, 2004).

Bu durumda Likert öleđinin analizinde hem parametrik (t-test, ANOVA vb.), hem de parametrik olmayan (ki-kare, Mann-Whitney U testi, Kruskal-Wallis vb.) testler kullanılabileceđi iddia edilmektedir. İstatistik uzmanlarının arasındaki bu görüş ayrılıđı alan yazımda karşılık bulmuş ve arařtırmacılar Likert öleđinden elde ettikleri verilerin analizinde farklı testler kullanmışlardır.

Kaptein, Nass ve Markopoulos (2010, 2391), CHI (2009) Konferansında sunulan bildirileri incelenmiş ve bunların % 45'inde Likert tipi ölçeklerin kullanıldığını, Likert tipi ölçeği kullanan araştırmacıların büyük çoğunluğunun (% 80.6) veri analizinde parametrik testleri kullandığını ve sadece % 8.3'ünün parametrik olmayan testleri kullandığını tespit etmişlerdir.

Bu araştırma kapsamında incelediğimiz 65 yüksek lisans ve doktora tezinden 61'inde Likert tipi ölçek kullanılmıştır. Bu ölçeklerin % 90'ı (55) parametrik testler kullanılarak analiz edilmiştir. Geriye kalan 4 araştırmada (ölçekten elde edilen veriler parametrik testlerin ön koşullarını sağlamadığı için) parametrik olmayan testler kullanılmış, 2 araştırmada ise ölçek verileri hem parametrik hem de parametrik olmayan testler kullanılarak analiz edilmiştir.

Likert ölçeğinden elde edilen verilerin analizinde hangi yaklaşımın daha güvenli ve tutarlı sonuç verdiği farklı araştırmalarda test edilmiştir. Kaptein, Nass ve Markopoulos (2010) 7'li Likert ölçeğinden elde edilen verilerin analizinde parametrik testler ile parametrik olmayan testlerin güvenilirliğini incelemişlerdir. Araştırmacılar parametrik ANOVA testi ile bunun parametrik olmayan alternatifi Kruskal Wallis testini karşılaştırmışlar ve özellikle örneklem sayısının küçük olduğu durumlarda ($n < 50$) parametrik olmayan testin daha güvenilir sonuçlar verdiğini ortaya koymuşlardır (s.2393).

Nanna ve Sawilowsky (1998) 7'li Likert ölçeğinden elde edilen verilerin analizinde t-testi ve Mann-Whitney-Willcoxon testini karşılaştırmış ve parametrik olmayan Mann-Whitney-Willcoxon testinin daha güvenilir sonuçlar verdiğini ortaya koymuştur.

Glass, Peckham ve Sanders (1972:237) 5-7 seçenekli Likert ölçeğinden elde edilen verilerin analizinde parametrik ANOVA testinin kullanımının uygun olduğunu. Bu test kullanılarak aralı (interval) veri önkoşullarının ihlal edildiği durumlarda dahi (varyansların eşitliği önkoşulu hariç) son derece kararlı ve güvenilir sonuçlar elde edilebileceğini ortaya koymuşlardır.

Winter ve Dodou (2010) 5'li Likert ölçeğinden elde edilen verilerin analizinde t-testi ve Mann-Whitney-Willcoxon testini karşılaştırmış ve her iki testte de tip-1 hata yapma ihtimalinin %3'ün altında olduğu, her iki yaklaşımda da testin gücünün birbirine yakın seviyelerde olduğunu saptamışlar bu nedenle bunlardan herhangi birini seçmenin yanlış olmayacağı kanaatine varmışlardır.

Bu araştırmada Likert ölçeklerinden elde edilen verilerin analizinde parametrik ve parametrik olmayan testlerin ulaştığı sonuçları sınamak amacıyla ilk olarak parametrik bağımsız gruplar *t* testi ile bunun parametrik olmayan karşılığı Mann-Whitney U Testi karşılaştırılmıştır. Karşılaştırma amacıyla 1350 öğrenci üzerinde uygulanmış 5 adet 5’li Likert tipi sorudan oluşan tutum ölçeği verileri üzerinden SPSS-21 programı kullanılarak farklı büyüklüklerde örneklem alınmıştır. Bu örneklem üzerinde “Öğrencilerin tarih dersine karşı tutumları cinsiyete ve sınıfa göre değişir” tezleri test edilmiştir. Rastgele seçilen örneklem arasında parametrik *t* test ön koşullarını taşıyan (normal dağılıma sahip) 27 örneklem üzerinde karşılaştırma yapılmıştır. Tablo 5’de görüleceği her iki testte ulaşılan σ değerleri birbirine çok yakındır. Yine yapılan karşılaştırmaların 24’ünde (%89) *t* test ve Mann-Whitney U testleri aynı sonuca ulaşarak H_1 hipotezini ret edip H_0 hipotezini kabul etmiştir. Ancak yapılan karşılaştırmaların 3’ünde (%11) *t* test ve Mann-Whitney U testi farklı sonuçlara ulaşmışlardır.

Tablo 5. *t* Test ile Mann-Whitney U Testi Karşılaştırması.

n	<i>t</i> test	Mann-Whitney U	Fark (%)
	Sig. (2-tailed)	Sig. (2-tailed)	
24	,099*	,161*	62
24	,133	,204	71
25	,074*	,116*	42
26	,232	,314	82
26	,380	,602	222
35	,774	,778	4
35	,965	,935	30
38	,761	,843	82
42	,586	,565	21
42	,827	,795	32
51	,291	,265	26
51	,366	,396	30
51	,652	,712	60
51	,928	,985	57
51	,497	,361	136
51	,112*	,069*	43
61	,648	,370	278
61	,072	,080	8
62	,161	,157	4
67	,122	,190	68
68	,218	,308	90
78	,002	,003	1
99	,891	,717	174
116	,673	,652	21
140	,524	,445	79
145	,551	,504	47
197	,007	,010	3

* *t* test ve Mann-Whitney U testi farklı sonuçlara ulaşmıştır.

İkinci olarak Tablo 3'te gösterilen karşılaştırmanın bir benzeri parametrik tek yönlü varyans analizi (ANOVA) ile bunun parametrik olmayan karşılığı Kruskal-Wallis Testi arasında gerçekleştirilmiştir. 1350 kişilik çalışma evreni içerisinde SPSS-21 programı kullanılarak farklı büyüklüklerde örneklem seçilmiş ve bunlar arasında parametrik tek yönlü varyans analizi (ANOVA) testinin ön koşullarını karşılayan 15 örneklem üzerinde karşılaştırma yapılmıştır. Tablo 6'da görüleceği üzere her iki testte ulaşılan σ değerleri birbirine çok yakındır. Yapılan karşılaştırmaların büyük çoğunluğunda (14) ANOVA ve Kruskal-Wallis testleri aynı sonuca ulaşarak H1 hipotezini ret edip H0 hipotezini kabul etmiştir. 15 karşılaştırmadan sadece birinde kullanılan testler farklı sonuçlara ulaşmıştır.

Tablo 6. ANOVA ile Kruskal-Wallis Testi karşılaştırması

n	ANOVA	Kruskal-Wallis	Fark (%)
27	,214	,204	10
27	,150	,139	11
41	,168	,179	11
41	,640	,529	111
55	,003	,007	4
65	,792	,741	51
82	,642	,592	50
84	,198	,211	13
85	,729	,709	20
140	,687	,709	22
195	,308	,211	97
200	,622	,578	44
200	,093	,154	61
200	,023*	,105*	82
250	,318	,222	96

* ANOVA ve Kruskal-Wallis testleri farklı sonuca ulaşmıştır.

Son olarak Glass, Peckham ve Sanders'in (1972:237) "Likert ölçeğinden elde edilen verilerin analizinde parametrik ANOVA testinin veri önkoşullarının ihlal edildiği durumlarda dahi (varyansların eşitliği önkoşulu hariç) son derece kararlı ve güvenilir sonuç verdiği" tezini sınanmıştır. Tek yönlü varyans analizi (ANOVA) testinin normal dağılım ve varyansların eşitliği ön koşulunu yerine getirmeyen iki örneklem verisi üzerinde ANOVA ve Kruskal-Wallis testlerinin ne tür bir sonuca ulaşacağı sınanmıştır. Tablo 7'de "Öğrencilerin teknolojiye karşı tutumları öğrenme stillerine göre değişmektedir" hipotezi parametrik ANOVA ve parametrik olmayan alternatifi Kruskal-Wallis testi ile test edilmiştir. Hipotezi test etmek için yanlış test kullanıldığında (dağılım normal olmadığından burada ANOVA

kullanılmamalıdır) dağılımda aslında var olan istatistiksel farkın var olmadığı sonucuna varılacak dolayısıyla tip II (β) hataya düşülecektir.

Tablo 7. Normal Dağılım Ön Koşulunun Karşılanmadığı Durumda ANOVA ile Kruskal-Wallis Testi Karşılaştırması

ANOVA					
Attitude					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,577	4	,394	1,201	,309
Within Groups	368,577	1123	,328		
Total	370,154	1127			

Test Statistics ^{a,b}	
	Attitude
Chi-Square	11,020
df	4
Asymp. Sig.	,026

a. Kruskal Wallis Test

b. Grouping Variable: Q4LS

Tablo 8’de ise “Çalışanların işyeri memnuniyetleri eğitim seviyesine göre değişmektedir” hipotezi parametrik ANOVA ve parametrik olmayan alternatifi Kruskal-Wallis testi ile test edilmiştir. Hipotezi test etmek için yanlış test kullanıldığında (grupların varyansları eşit olmadığından burada ANOVA kullanılmamalıdır) dağılımda aslında olmayan istatistiksel farkın var olmadığı sonucuna varılacak dolayısıyla tip I (α) hataya düşülecektir.

Tablo 8. Varyansların Eşitliği Ön Koşulunun Karşılanmadığı Durumda ANOVA ile Kruskal-Wallis Testi Karşılaştırması

ANOVA					
Memnuniyet					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5,497	3	1,832	3,239	,023
Within Groups	110,889	196	,566		
Total	116,385	199			

Test Statistics ^{a,b}	
	Memnuniyet
Chi-Square	6,133
df	3
Asymp. Sig.	,105

a. Kruskal Wallis Test

b. Grouping Variable:
Eğitimdurumu

Sonuç, Tartışma ve Öneriler

Likert-tipi soru veya Likert ölçeği geliştirmesi, uygulaması ve değerlendirmesi kolay olduğu için eğitim araştırmalarında tutum, eğilim ve görüş ölçmek amacıyla sıklıkla kullanılmaktadır. Ancak özellikle “değerlendirme kolaylığı” konusu tartışma konusudur. Üst düzey istatistiksel analizler gerektirmediği için değerlendirmesi kolaydır fakat bu tür verilerin analizinde hangi test türünün kullanılacağı konusunda görüş birliği olmaması araştırmacıların işini zorlaştırmaktadır. Bu aşamada karşılaşılan sorunları iki başlık altında toparlayabiliriz:

- a) Kavram karmaşasından kaynaklanan sorunlar: Likert-tipi sorular (3'lü, 5'li, 6'lı vs.) kullanılarak oluşturulan bir anket ile Likert ölçeği aynı değildir. Yapılan incelemelerde yanlış analiz kullanıldığı için eleştirilen çalışmalar büyük oranda araştırmacıların Likert-tipi soruları Likert ölçeği gibi algıladığı araştırmalardır. Bu tür çalışmaların doğrudan hatalı olarak değerlendirilmesinin nedeni birbirinden bağımsız Likert-tipi sorularla oluşturulan bir ankette her sorunun ayrı ayrı analiz edilmesi ve bu analizde parametrik olmayan (mod, medyan, ranj, Mann-Whitney U, Kruskal-Wallis vb.) testlerin kullanılması gerektiği hususunda görüş birliği olmasıdır.
- b) Veri türü üzerinde yapılan tartışmalardan kaynaklanan sorunlar: Likert ölçeklerinden elde edilen verinin sıralı veya aralı veri olarak tanımlanmasından kaynaklanan bu sorunun mutlak çözümü mümkün görünmediğinden bu tip ölçeklerden elde edilen verilerin analizinde kullanılacak yöntem tartışma konusu olmaya devam edecektir.

Veri türü üzerinde yapılan tartışmalar nedeniyle Likert ölçeklerinin analizinde hangi test türünün kullanılacağı konusunda üç ayrı görüş ortaya çıkmaktadır. Birinci görüşe göre bu tip verilerin analizinde sadece parametrik olmayan testler kullanılabilir, ikinci görüşe göre hem parametrik hem de parametrik olmayan testler kullanılabilir ve üçüncü görüşe göre daha güvenli sonuçlar verdiği için sadece parametrik testler kullanılmalıdır.

Bu çalışma sonucunda araştırmacıların daha ziyade üçüncü görüşü benimsedikleri dolayısıyla Likert ölçeği verilerini genellikle parametrik testler kullanarak analiz ettikleri saptanmıştır. Her iki test türünün de birbirine yakın sonuçlara ulaşmasına rağmen

arařtırmacıların parametrik testleri daha ok tercih etmesinde bu testlerin daha gvenli sonular ortaya koyduđu inancının yanı sıra parametrik testlerin tanınırlıđı ve poplerliđi de etkilidir (Kaptein, Nass ve Markopoulos, 2010: 2391). Burada sıklıkla hatırlatılan husus eđer parametrik testler tercih edilecekse kullanılan verinin parametrik testlerin nkořullarına sahip olduđundan (yani rneklem sayısının 50'nin stnde, dađılımın normal ve varyansların eřit olduđundan) emin olunmasıdır (Boone ve Boone, 2012).

Kaynaka

- Adelson, J.L., ve McCoach, D.B. (2010). Measuring the mathematical attitudes of elementary students: the effects of a 4-point or 5-point likert-type scale. *Educational and Psychological Measurement*, 70, 796-807, DOI: 10.1177/0013164410366694.
- Baumgartner, H., & Steenkamp, J. B. E. M. (2001, May). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, ss. 143-156.
- Boone, H.N. ve Boone, D.A. (2012). Analyzing Likert data. *Journal of Extension*, April 2012, V.50, N.2.
- Carifio, J., ve Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, ss. 1150-1152.
- Clason, D. L. ve Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, 35(4), ss. 31- 35.
- Cramer, D. & Howitt, D.L. (2004). *The SAGE dictionary of statistics: A practical resource for students in the social sciences*. Londra: SAGE.
- Cronbach, L.J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, ss. 475-494.
- Cronbach, L.J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, ss. 3-31.
- Edmondson, D. R. (2005). Likert scales: A history. In L. C. Neilson (Ed.), *Proceedings of the 12th conference on historical analysis and research in marketing (CHARM)*(pp. 127-133). Eriřim: <http://faculty.quinnipiac.edu/charm>
- Garland, R. (1991). The mid-point on a Likert rating scale: Is it desirable? *Marketing Bulletin*, 2, ss. 66-70.
- Goldstein, G., & Hersen, M. (1984). *Handbook of psychological assessment*. New York: Pergamon.

- Hart, M.C. (1996). Improving the discrimination of SERVQUAL by using magnitude scaling. G. K. Kanji (Ed.), *Total Quality Management in Action*. London: Chapman and Hall.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 2004; 38: 1217–1218.
- Javaras, K.N. ve Ripley, B.D. (2007). An "unfolding" latent variable model for likert attitude data: drawing inferences adjusted for response style. *Journal of the American Statistical Association*, Vol. 102, No. 478 (Jun., 2007), ss. 454-463. Erişim: <http://www.jstor.org/stable/27639876> .
- Kaptein, M., Nass, C., ve Markopoulos, P. (2010). Powerful and consistent analysis of likert-type rating scales. CHI 2010 Proceedings, ss.2391-2394.
- Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, 22, ss.251–260.
- Kulas, J. T., Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43, ss.489–493.
- Kuzon, W.M., Urbanchek, M.G. ve McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 1996; 37:265–272.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, Vol. 22, ss. 5-55.
- Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods*, 3, ss.55–67.
- Nowlis, S. M., Kahn, B. E., & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research*, 29, ss.319–334.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, ss. 1-15.
- Ryan, C. ve Garland, R. (1999). The use of a specific non-response option on Likert-type scales. *Tourism Management*, 20, ss.107-113.
- Schutz, H. G., & Rucker, M. H. (1975). A comparison of variable configurations across scale lengths: an empirical study. *Educational and Psychological Measurement*, 35, ss.319-324.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.
- Stone, M. H. (2004). Substantive scale construction. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (201–225). Maple Grove, MN: JAM.

- Tavakoli, H. (2012). A dictionary of research methodology and statistics in applied linguistics. Tahran: Rahnama.
- Tezbařaran, A.A. (2008). Likert tipi lek hazırlama kılavuzu (e-kitap). Eriřim: [http://www.academia.edu/1288035/Likert Tipi lek Hazırlama Kılavuzu](http://www.academia.edu/1288035/Likert_Tipi_lek_Hazırlama_Kılavuzu)
- Tezbařaran, A.A. (2004). Likert tipi leklere madde semede geleneksel madde analizi tekniklerinin karřılařtırılması. *Trk Psikoloji Dergisi*, 19 (54), 77-87.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying contexteffects in attitude measurement. *Psychological Bulletin*, 103, ss. 299–314.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, ss.236–247.
- Winter, J.C.F. & Dodou, D. (2010). Five-point likert items: T test vs Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation*, Vol 15, No 11, ss. 1-16.