

Ölçek Geliştirme Konulu Tezlerin Ölçek Geliştirme Standartlarına Göre İncelenmesi*

Analysis of Theses on Scale Development According to Scale Development Standards

Esra EMİNOĞLU ÖZMERCAN ** 
Buket KEMER *** 

Öz

Bu araştırmanın amacı eğitim bilimleri ve psikoloji alanında geliştirilmiş ölçeklerin ölçek geliştirme süreçlerinin AERA, APA ve NCME tarafından yayınlanan ölçek geliştirme standartlarına göre uygunluğunun incelenmesidir. 2010 ile 2021 yılları arasında Türkiye’de eğitim bilimleri ve psikoloji alanında yayınlanan 50 yüksek lisans ve doktora tezi araştırmanın örneklemini oluşturmaktadır. Tezlerin incelenmesinde AERA, APA ve NCME tarafından yayınlanan standartlara göre hazırlanmış dereceleme ölçeği kullanılmıştır. Dereceleme ölçeği “Ölçek Yönergesi Standartları”, “Madde Geliştirme ve Gözden Geçirme Standartları” ve “Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları” olmak üzere üç ana başlıktan ve 27 maddeden oluşmaktadır. Çalışmada içerik analizi yöntemi kullanılmıştır. Elde edilen veriler doğrultusunda incelenen tezlerin büyük çoğunluğunda ölçülmek istenen yapı veya alan, araştırma evreni, ölçeğin kullanım amacı ve geliştirme gerekçeleri, maddelerin geliştirilme ve gözden geçirilme aşaması, ön deneme uygulamasının yapıldığı, madde havuzundan nasıl madde seçildiği, uzman görüşleri, örneklem özellikleri, yönergelerin yeterli açıklıkta

- * Makalenin ilk hali 21 – 23 Eylül 2022 tarihlerinde düzenlenen 8. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi’nde sunulmuştur.
- ** Doç. Dr., İstanbul Üniversitesi, Açık ve Uzaktan Eğitim Fakültesi, İstanbul/Türkiye, esemcan@gmail.com, ORCID ID: 0000-0003-4105-9837
- *** Öğr. Gör., İstanbul Üniversitesi, Açık ve Uzaktan Eğitim Fakültesi, İstanbul/Türkiye, buketkemer94@gmail.com, ORCID: 0000-0002-9170-012X

Atıf için: Eminoglu Özmercan, E. ve Kemer, B. (2024). Ölçek Geliştirme Konulu Tezlerin Ölçek Geliştirme Standartlarına Göre İncelenmesi. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 60(60), 190-206. DOI: 10.15285/maruaeabd.1486846

sunulduğu, güvenilirlik ve geçerlik kanıtlarının verildiği, ölçeğin ne amaçla kullanılacağı uygun bir şekilde verildiği belirlenmiştir. Ölçek geliştirme sürecinde kullanılan kuram ve ölçeğin içeriğinin yönergede tanımlanan alanı ne ölçüde temsil ettiği hakkında neredeyse hiç bir bilgiye verilmediği belirlenmiştir. Buna göre ölçek geliştirme çalışmalarında ölçme ve değerlendirme uzmanlarından görüş alınması ve yönergelerin ayrıntılı bir şekilde verilmesi önerilebilir.

Anahtar Kelimeler: Ölçek, ölçek geliştirme, AERA, APA ve NCME ölçek geliştirme standartları, içerik analizi, eğitim ve psikoloji.

Abstract

The purpose of this research is to analyse the scales developed in the field of educational science and psychology according to the standards for educational and psychological testing of AERA, APA and NCME. Between 2010 and 2021, 50 theses constitute the sample of the research. A rating scale prepared according to the standards published by AERA, APA and NCME was used to analyse theses. The rating scale consists of 27 items and three main headings. Content analysis method was used in the study. According to findings, it was determined that in the majority of theses the structure or domain to be measured, the examinee population, the purpose of the scale and the reasons for its development, the procedures used to develop and review the items, information that items tried out, the procedures used to select items from the item pool, expert review, the information about the sample's representativeness of the population, the directions for test administration, reliability and validity properties, and the purpose of the test were defined adequately. It has been determined that no information is given about documenting the model used for evaluating the psychometric properties of items and extent to which the content domain of a test represents the domain. According to this result, it may be recommended to get opinions from measurement and evaluation specialists in scale development studies and to give instructions detailly.

Keywords: Scale, scale development, AERA, APA and NCME testing standards, content analysis, education and psychology.

Giriş

Ölçme; insanlar, nesnelere, olaylar ve süreçler hakkında bilgi edinilmesine olanak sağlayan bilimin temel bir faaliyetidir (DeVellis, 2017). Ölçme, nesnelere ya da olaylara belli araç ve kurallar yardımıyla sayısal değer atamayı içermektedir (Stevens, 1946). Eğitim ve psikolojide kullanılan değişkenler; ilgi, tutum, akademik başarı ya da zekâ gibi özellikler olduğu için bu değişkenlere çoğu zaman sayısal değer verilememektedir (Baykul, 2021; Cohen ve Swerdlik, 2018). Ölçme araçları, doğrudan ve kolayca gözlemlenemeyen teorik değişkenlerin seviyelerini ortaya koyan maddeler bütünüdür. Bu nedenle teorik açıdan var olan fakat doğrudan değerlendirilemeyen olgular ölçülmek istendiği zaman ölçek geliştirme çalışmaları yapılmaktadır (DeVellis, 2017). Ölçeklerin farklı şekillerde sınıflandırıldığı göz önüne alındığında, birçok farklı ölçekleme yöntemi olduğu söylenebilir. Tek bir ölçme yöntemi ya da en iyi ölçek türü yoktur.

Ölçek geliştirme çalışmalarında; “Yeni bir ölçeğe ihtiyaç var mı?”, “Ölçek neyi ölçmeyi amaçlıyor?”, “Bu ölçek kimler için geliştirilecek?”, “Nasıl uygulanacak?” ve “Uygulama süresi önemli mi?” gibi sorulara yanıt aranmaktadır. Buna bağlı olarak öncelikle ölçülmek istenen özelliğin neden ölçüleceğine ve nasıl ölçüleceğine karar verilmelidir. Kişinin ölçülmek istenilen hangi özelliği daha kritikse o özelliğe ilişkin maddeler ölçekte yer almalıdır. Bu işlem yeni bir ölçek geliştirme çalışması olabileceği gibi yabancı bir dil ve kültürden uyarlama çalışması da olabilir (Cohen ve Swerdlik, 2018).

Ölçek geliştirme sürecinde takip edilmesi gereken adımları açıklayan (Crocker ve Algina, 2008; Cohen ve Swerdlik, 2018; DeVellis, 2017; Erkuş, 2012; Ifelunni 1997'den aktaran Eze vd., 2017; Tezbaşaran, 2008) pek çok çalışma bulunmaktadır. Cohen ve Swerdlik (2018) ölçek geliştirmenin, ölçek oluşturma sürecini içerdiğini; ölçek oluşturma ise ölçek kavramsallaştırma, ölçek yapısı, ölçek deneme, madde analizi ve ölçek revizyonu şeklinde beş basamaktan oluştuğunu belirtmiştir. Bu aşamalardan ilki olan ölçeğin kavramsallaştırılması ile ölçek yapısını oluşturma başlamaktadır. Ölçeğin yapısı; maddelerin yazımı, düzeltilmiş maddelerin yazımı, maddelerin puanlanması ve ölçeğin yapısının tasarlanması ile oluşturulur. Ön deneme formu geliştirildikten sonra nihai formun uygulanacağı koşullara benzer koşullarda ölçek denemesi yapılır. Sonrasında istatistiksel prosedürler uygulanarak maddelerin iyi bir madde olup olmadığına karar verilip, düzeltilmesine ya da ölçekten çıkarılmasına karar verilir. En son aşamada ise ölçek madde analizlerine göre revize edilerek test geliştirme amacına uygun hale getirilir. Testin revize edilmiş hali başka yeni bir örnekleme uygulanır. Crocker ve Algina'ya (2008) göre izlenmesi gereken ölçek geliştirme adımları; ölçek puanlarının kullanılacağı temel amaçları belirleme, yapıyı temsil eden davranışları belirleme, madde havuzu oluşturma, maddeleri gözden geçirme ve düzenleme, ön deneme ve gerekliyse revize etme, uygulama yapma, maddelerin analizlerini yapma ve uygun olmayanları çıkarma, formun son hali için geçerlik ve güvenilirlik analizlerini yapma ve ölçeği uygulama, puanlama ve puanlarını yorumlamaya yönelik yönerge oluşturma şeklindedir. Tezbaşaran (2008) ölçek geliştirme sürecini; ölçülecek tutumun (özelliğin) tanımlanması, deneme ölçeğinin düzenlenmesi ve deneme uygulaması, deneme ölçeğinden elde edilen verilerin analizi şeklinde üç ana başlıkta açıklamaktadır. Ölçülecek tutumun tanımlanması tutum kapsamının belirlenmesi ve denemelik tutum ifadelerinin belirlenmesinden oluşmaktadır. Deneme ölçeğinin düzenlenmesi ve deneme uygulaması olan ikinci basamak ölçekte kullanılacak materyallerin hazırlanması, yönerge ve cevaplama düzeninin oluşturulması, maddelerin ölçek içindeki düzeni, ön inceleme ve deneme şeklindeki adımlardan meydana gelmektedir. Deneme ölçeğinden elde edilen verilerin analizi olan son basamakta ise maddelere verilen cevapların puanlanması, ölçekten alınan ham puanların hesaplanması, ham puan dağılımının özellikleri, madde puanları dağılımının özellikleri ve madde analizi adımlarından oluşmaktadır. DeVellis (2017) ölçek geliştirme sürecini; ölçülmesi istenen yapıya karar verilmesi, madde havuzu oluşturulması, ölçme biçiminin belirlenmesi, uzmanlar tarafından madde havuzunun gözden geçirilmesi, geçerlik maddelerinin ölçeğe dahil etme durumunun değerlendirilmesi, maddelerin ölçeğin geliştirildiği örnekleme uygulanması, maddelerin değerlendirilmesi ve ölçek uzunluğunu en uygun şekilde getirilmesi şeklinde sekiz basamakta değerlendirmiştir. Erkuş (2012) ölçek geliştirme sürecinde ölçülmesi hedeflenen değişken için kavramsal ya da teorik genel hatlarının çizilmesi ve tanımlanması, kavramın tanımının davranışsal göstergelerinin bulunması, değişkenin doğasına ya da amaca bağlı olarak ölçme geliştirme tekniğine karar verilmesi, uyarıcıların ve uygun tepki kategorilerinin üretilmesi, ölçülmesi hedeflenen yapıya uygun madde türünün belirlenmesi, madde yazımı öncesi ön hazırlık yapılması, maddeleri gözden geçirilmesi ve düzenlenmesi, ölçeğin yönerge ve açıklamasının yazılması, ölçeğin biçimsel yapısının incelenmesi, ön deneme yapıp sonrasında tekrar inceleme – düzeltme ile deneme uygulamasına hazırlık, deneme uygulaması, madde ve test analizlerinin yapılması ve tekrar deneme uygulamalarına başvurma şeklinde basamaklara ayırmıştır. Ifelunni ölçek geliştirme sürecini; ölçeği planlama, ölçeği hazırlama, ön deneme yapma ve ölçeği değerlendirme şeklinde dört basamakta

açıklamıştır (Ifelunni 1997'den aktaran Eze vd., 2017). Bu araştırmanın kapsamı ölçek geliştirme ile ilgili çalışmalar olduğu için uyarlamayla ilgili bilgilere yer verilmemiştir.

Psikolojik özellikler ve durumların varlığı kabul edildiğinde, ölçülen ve sayısal değer verilen belirli özellik ve durumlar dikkatli bir şekilde tanımlanmalıdır. Ölçek geliştiriciler ve araştırmacılar aynı olguya birçok farklı şekilde bakıp tanımlarlar. Ölçülecek özellik, durum ya da diğer yapılar tanımlandıktan sonra madde içerikleri belirlenmelidir (Cohen ve Swerdlik, 2018). Bunlara ek olarak Amerikan Eğitim Araştırmaları Derneği (American Educational Research Association [AERA]), Amerikan Psikologlar Derneği (American Psychological Association [APA]) ve Eğitimde Ölçme ve Değerlendirme Ulusal Konseyi (National Council on Measurement in Education [NCME]) tarafından yayınlanan “Eğitim ve Psikolojide Ölçme Standartları (Standards for Educational and Psychological Testing) (2014)” eğitimde ve psikolojide ölçme sürecinde yol gösterme amacı taşımaktadır. AERA, APA ve NCME'nin (2014) hazırlamış olduğu Eğitim ve Psikolojide Ölçme Standartları'nda ölçek geliştirme süreci; “Ölçek Yönergesi Standartları”, “Madde Geliştirme ve Gözden Geçirme Standartları” ve “Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları” ve “Ölçeği Gözden Geçirme” olmak üzere dört ana başlık altında ele alınmıştır.

İlgili alanyazın incelendiğinde çok sayıda ölçek geliştirme çalışmasının olduğu görülmektedir. Aktaş (2023) ölçek geliştirme çalışmalarının sayısında 2004 yılından sonra bir artış gözlemlendiğini belirtmektedir. Ayrıca yapılan bibliyometrik makale incelendiğine 2010 yılı sonrasındaki ölçek geliştirme çalışmalarının daha da arttığı göze çarpmaktadır. Bunun yanı sıra geliştirilen ölçeklerin, ölçek geliştirme aşamalarını inceleyen çalışmalar da yer almaktadır (Barış Pekmezci ve Ayan, 2020; Çelik, 2020; Çüm ve Koç, 2013; Ergene, 2020; Erkuş, 2016; Gül ve Sözbilir, 2015; Güler ve Ayan, 2020; Hinkin, 1995; İlic ve Haseski 2019; Kaya Uyanık, Güler, Taşdelen Teker ve Demir, 2017; Mor Dirlik, 2014; Mor Dirlik ve Koç, 2017; Olgun ve Alath, 2021; Slavec ve Drnovsek, 2012; Soycan ve Babacan, 2019; Şahin ve Boztunç Öztürk, 2018; Tatlı Harmanlı ve Güngör Aytar, 2021; Tosun ve Taşkesenligil, 2015; Worthington ve Whittaker, 2006). Var olan çalışmalar incelendiğinde ölçek geliştirme süreci açısından pek çok sorun ve eksiklerin olduğu göze çarpmaktadır. Geliştirilme sürecinde ölçek geliştirme adımlarını tam anlamıyla takip etmeyen ölçekler veri elde etmede kullanılmakta ve bu verilere dayalı olarak bireyler hakkında karar verilebilmektedir. Bu gibi durumlarda bireylerin yaşamlarını etkileyen durumlarla ilgili yanlış kararlar verilebilmekte ve bireyleri hayatları boyunca olumsuz etkileyen durumlara neden olabilmektedir. Ayrıca, ölçekler standartlara uygun hazırlanmadığında öncelikle araştırmacılar açısından emek ve zaman kaybına sebep olmakta, uygun olmayan ölçekler kullanılmakta bu nedenle de alanyazında çokça hatalı ve yanlış bilgiler yer almaktadır. Ölçek geliştirme sürecinde ölçek geliştirme standartlarına uyulmasıyla elde edilen sonuçların geçerli ve güvenilir olması sağlanabilir. Böylece olası ölçme hataları azaltılabilir. İlgili ölçek geliştirme aşamalarını inceleyen çalışmalara göre, birçok ölçeğin, ölçek geliştirme standartlarına uygun hazırlanmadığı görülmektedir. Bunun yanı sıra ilgili çalışmaların neredeyse tamamı makalelerden oluşmaktadır. Tezlerle ilgili geniş kapsamlı bir araştırmanın olmadığı görüldüğünden bu çalışmanın yapılması önem arz etmektedir.

Bu araştırmanın amacı 2010-2021 yılları arasında Türkiye'de eğitim bilimleri ve psikoloji alanında hazırlanan ve Yükseköğretim Kurulu Başkanlığı Tez Merkezinde erişime açık yüksek

lisans ve doktora tezlerinin AERA, APA ve NCME tarafından yayınlanan Eğitimde ve Psikolojide Ölçme Standartları'na göre ölçek geliştirme standartlarının bir dereceleme ölçeği oluşturularak incelenmesidir. Araştırma kapsamında aşağıdaki sorulara yanıt aranmıştır:

İlgili tezlerde incelenen ölçekler;

1. Ölçek Yönergesi Standartları'na uygun bir şekilde hazırlanmış mıdır?
2. Madde Geliştirme ve Gözden Geçirme Standartları'na uygun bir şekilde hazırlanmış mıdır?
3. Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları'na uygun bir şekilde hazırlanmış mıdır?

Yöntem

Bu çalışma tarama modellerinden genel tarama modelindedir. Tarama modelinde geçmişte ya da halen varolan bir durumu varolduğu şekliyle betimlemeyi amaçlayan araştırma yaklaşımlarıdır. Genel tarama modeli ise çok sayıda elemandan oluşan bir evren hakkında genel bir kaniya varma amacıyla evren ya da belli bir örneklem üzerinde yapılan tarama çalışmalarıdır (Karasar, 2012). Araştırmanın evreni belirlenirken Yükseköğretim Kurulu Başkanlığı Tez Merkezi'nin gelişmiş arama kısmında aranacak kelime yerine "Ölçeği geliştir" ve "Ölçek geliştir", izin durumu "İzinli", dili "Türkçe", Yıl 2010-2021 arasında olacak şekilde anahtar kelimeler girilerek tezler taranmıştır. Daha sonrasında çıkan tezlerin konusu "Eğitim ve Öğretim" ve/veya "Psikoloji" alanlarında olacak şekilde filtreleme yapılmıştır. Bunun sonucunda 199 teze ulaşılmıştır. Araştırma evrenini 171'i eğitim öğretim alanından 28'i ise psikoloji alanından olmak üzere toplam 199 tez oluşturmaktadır.

Tablo 1. Araştırma Evreni ve Örneklemi Özellikleri

Değişken	Evren Özellikleri		Örneklem Özellikleri		
	f	%	f	%	
Program türü	Tezli Yüksek Lisans	159	80	38	86
	Doktora	40	20	6	14
Tez konusu	Eğitim ve Öğretim	171	86	36	82
	Psikoloji	28	14	8	18
Tez Yılı	2010	5	3	1	2
	2011	8	4	2	5
	2012	10	5	1	2
	2013	8	4	1	2
	2014	14	7	3	7
	2015	10	5	2	5
	2016	12	6	2	5
	2017	8	4	3	7
	2018	28	14	7	16
	2019	48	24	10	23
	2020	26	13	7	16
2021	22	11	5	11	

Örneklem belirleme yöntemlerinden seçkisiz örnekleme, evreni oluşturan öğelerin tümünün örneklem belirlemede aynı şansa sahip olmasıdır. Basit seçkisiz örneklem oluşturmada çalışılan konunun kapsamı içindeki her bir öğenin, nesnenin, üyenin veya paydaşın örneklem olma şansı eşittir (Korkmaz, 2020). Evrenden seçkisiz örnekleme yöntemlerinden basit seçkisiz örnekleme kullanılarak 50 tez seçilmiştir.

Veri toplama araçları

Veri toplama aracı olarak AERA, APA ve NCME'nin yayınladığı “Eğitim ve Psikolojide Ölçme Standartları (2014)” araştırmacılar tarafından Türkçeye çevrilmiş ve uzman görüşleri (bir ölçme ve değerlendirme uzmanı, bir program geliştirme uzmanı) doğrultusunda bir dereceleme ölçeği hazırlanmıştır. Standartlar belli zaman aralıklarıyla güncellendiği için araştırma kapsamında en son yayınlanan “Eğitim ve Psikolojide Ölçme Standartları (2014)” kullanılmıştır. Dereceleme ölçeği “Ölçek Yönergesi Standartları”, “Madde Geliştirme ve Gözden Geçirme Standartları” ve “Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları” üç ana başlık altında toplanmış 27 standarttan oluşmaktadır. Diğer bir başlık olan “Ölçeği Gözden Geçirme” bölümü bitmiş ve hâlihazırda kullanılan bir ölçeğin yeniden gözden geçirilmesi adımlarını içerdiğinden veri toplama aracına bu bölüm dâhil edilmemiştir. “Ölçek yönergesi standartları” beş maddeden, “Madde geliştirme ve gözden geçirme standartları” 15 maddeden ve “Ölçek uygulama, puanlama prosedürü ve materyallerin standartları” ise yedi maddeden oluşmaktadır. İncelenen tezler ilgili her bir standart için “Hiç uygun değil (0)”, “Kısmen uygun (1)” ve “Tamamen uygun (2)” olarak derecelendirilmiştir. Belirlenen standartlar hakkında tezde bilgi verilmemesi durumunda ise “Bilgi verilmemiştir (3)” olarak kodlanmıştır. Dereceleme ölçeği Ek-1’de verilmiştir.

Verilerin analizi

Ölçek geliştirme ile ilgili tezler içerik analizi yöntemi kullanılarak incelenmiştir. İçerik analizi elde edilen verileri açıklayan kavram ve ilişkiler belirleme amacındadır. Bu amaç doğrultusunda benzer veriler belli tema ve kavram çerçevesinde birleştirildikten sonra düzenlenir ve yorumlanır (Yıldırım ve Şimşek, 2013). İçerik analizine başlamadan önce, araştırmacılar arasındaki tutarlılığı sağlamak için 50 tez içinden tesadüfi olarak seçilen altı tez her iki araştırmacı tarafından ayrı ayrı incelenmiştir. Elde edilen veriler ile Miles ve Huberman (1994) tarafından önerilen tutarlılık anlamında güvenilirlik kestirilmiştir. Elde edilen değerler 0,70 ve üstünde olması puanlayıcılar arası uyumun göstergesi olarak kabul edilmektedir (Tavşancıl ve Aslan, 2001).

$$\text{Güvenirlilik} = \frac{\text{Uzlaşma Sayısı}}{\text{Uzlaşma Sayısı} + \text{Uzlaşmama Sayısı}}$$

Araştırmacılar arası tutarlılık katsayısı 0,84 olarak hesaplanmıştır. Bu sonuç yapılan değerlendirmenin güvenilir olduğuna ilişkin kanıt sağlamaktadır. Bununla birlikte, Krippendorff Alfa katsayısına da bakılmıştır. Krippendorff Alfa katsayısı (α) iki veya daha fazla puanlayıcı arası uyumu belirleme amacıyla kullanılmaktadır. Hesaplaması yapılırken gözlenen uyumsuzluk (D_0)

değeri beklenen uyumsuzluk (D_c) değerine bölünmektedir. Bulunan değer in 1'den çıkarılmasıyla Krippendorff Alfa katsayısı (α) elde edilir.

$$\alpha = 1 - \frac{D_o}{D_e}$$

Krippendorff Alfa istatistiğinde α 'nın 0 olması puanlayıcılar arasında tam uyumsuzluk, 1 olması ise puanlayıcılar arasında mükemmel uyumu göstermektedir. Uyumun gücü 0,67'den küçükse “zayıf”; 0,67 ile 0,80 arasında ise “orta” ve 0,80'den daha fazla ise “yüksek” şeklinde yorumlanır (Krippendorff, 1995). Araştırmacılar arası Krippendorff Alfa katsayısı 0,85 olarak hesaplanmış olup puanlayıcılar arası uyumun gücü yüksek şeklinde yorumlanabilir. Ayrıca, puanlayıcılar arası güvenilirliğin belirlenmesinde bir diğer yöntem olan Cohen (1960) tarafından önerilen Kappa istatistiği (κ) kullanılmıştır. Nominal puanlama yapan iki puanlayıcı arasındaki uyumu belirlemek amacıyla geliştirilen Kappa istatistiğini aldığı değerler - 1 ile + 1 arasında değişmektedir. Kappa istatistiği (κ) formülü aşağıda verilmiştir:

$$\kappa = \frac{\bar{p} - \bar{p}_e}{1 - \bar{p}_e}$$

Kappa istatistiği 0'dan düşük olduğunda “zayıf”; 0 ile 0,20 arasında ise “önemsiz”; 0,21 ile 0,40 arasında ise “düşük”; 0,41 ile 0,60 arasında ise “orta”; 0,61 ile 0,80 arasında ise “önemli” ve 0,81 ve üstünde ise “çok yüksek” şeklinde yorumlanmaktadır (Landis ve Koch, 1977). Araştırmacılar arası Kappa istatistiği 0,76 olarak hesaplanmış olup puanlayıcılar arası uyum gücünün önemli olduğu şeklinde yorumlanabilir. Bu sonuçlar ışığında kodlamalar incelenmiş ve farklı kodlamalar üzerinde tartışılıp ortak bir karara varılmıştır.

Geriyeye kalan 44 tez araştırmacılar tarafından eşit bir şekilde paylaşılmış ve bireysel olarak incelenmiştir. Tezler incelendikten sonra veriler Microsoft Office programlarından Excel'e aktarılmış, tek bir dosyada birleştirilmiştir. Hatalı kodlamaların olup olmadığı incelenmiş, gerekli düzeltmeler yapıldıktan sonra dereceleme ölçeğinde yer alan başlıklar altında elde edilen verilerin frekans ve yüzde analizleri yapılmış ve tablolaştırılmıştır.

Bulgular

Tezlerin ilgili standartlara uygunluk düzeylerine ilişkin bulgular “Ölçek Yönergesi Standartları”, “Madde Geliştirme ve Gözden Geçirme Standartları” ve “Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları” şeklinde üç ana başlık altında verilmiştir. Her tez çalışması için standartların frekans ve yüzdeleri “Hiç uygun değil (0)”, “Kısmen uygun (1)”, “Tamamen uygun (2)” ve “Bilgi verilmemiş (3)” şeklinde tablolaştırılmıştır.

Ölçek yönergesi standartlarına ilişkin bulgular

Tezler dereceleme ölçeğinde yer alan beş adet ölçek yönergesi standardına göre incelenmiş ve buna ilişkin bulgular Tablo 2'de verilmiştir.

Tablo 2. Ölçek Yönergesi Standartlarına İlişkin Frekans ve Yüzde Dağılımları

Ölçek Yönergesi Standartları	Hiç uygun değil		Kısmen uygun		Tamamen uygun		Bilgi verilmemiş	
	f	%	f	%	f	%	f	%
Ölçülmek istenen yapı veya alan tanımlanmıştır.	0	0	9	20	35	80	0	0
Araştırmanın evreni belirtilmiştir.	0	0	13	30	31	70	0	0
Ölçeğin kullanım amacı ve geliştirme gerekçeleri belirtilmiştir.	0	0	18	41	26	59	0	0
Ölçeğin kapsamı, ölçek uzunluğu, madde formatları, maddelerin ve ölçeğin psikometrik özellikleri ile maddelerin ve bölümlerin sıralanışı belirtilmiştir.	4	9	14	32	20	45	6	14
Ölçeği cevaplamaya ayrılan süre, ölçeği alanlar için uygulama adımları, ölçekte uygulanacak prosedürler (izin verilen değişiklikler dâhil), varsa kullanılacak materyaller, puanlama ve raporlama prosedürleri belirtilmiştir.	5	11	16	36	10	23	13	30

Tablo 2 incelendiğinde ölçülmek istenen yapı veya alan, incelenen tezlerin %80'inde tamamen, %20'sinde ise kısmen uygun olarak tanımlandığı görülmektedir. Araştırma evreninin %70'inde tamamen, %30'unda ise kısmen uygun olduğu belirtilmiştir. Ölçeğin kullanım amacı ve geliştirme gerekçelerinin %59'unda tamamen, %41'inde ise kısmen uygun olduğu bulunmuştur. Yönergede ölçeğin kapsamı, ölçek uzunluğu, madde formatları, maddelerin ve ölçeğin psikometrik özellikleri ile maddelerin ve bölümlerin sıralanışının %45'inde tamamen, %32'sinde kısmen uygun olduğu belirtilmiştir. Bu standart tezlerin %9'unda hiç uygun değilken %14'ünde bu standarda ilişkin herhangi bir bilgi verilmemiştir. Yönergede ölçeği cevaplamaya ayrılan süre, ölçeği alanlar için uygulama adımları, ölçekte uygulanacak prosedürler (izin verilen değişiklikler dâhil), varsa kullanılacak materyaller, puanlama ve raporlama prosedürlerinin %23'ünde tamamen, %36'sında kısmen uygun olduğu belirlenmiştir. Bununla ilgili %30'unda herhangi bir bilgi verilmediği görülmektedir.

Madde geliştirme ve gözden geçirme standartlarına ilişkin bulgular

Tezler dereceleme ölçeğinde yer alan 15 adet madde geliştirme ve gözden geçirme standardına göre incelenmiş ve buna ilişkin bulgular Tablo 3'te verilmiştir.

Tablo 3. Madde Geliştirme ve Gözden Geçirme Standartlarına İlişkin Frekans ve Yüzde Dağılımları

Madde Geliştirme ve Gözden Geçirme Standartları	Hiç uygun değil		Kısmen uygun		Tamamen uygun		Bilgi verilmemiş	
	f	%	f	%	f	%	f	%
Maddelerin geliştirilme ve gözden geçirilme aşaması raporlanmıştır.	4	9	21	48	19	43	0	0
Ön deneme uygulamasının yapıldığı raporlanmıştır.	2	5	2	5	31	70	9	20
Ön deneme ile ilgili bilgiler verilmiştir.	1	2	11	25	22	50	10	23
Madde havuzundan nasıl madde seçildiği raporlanmıştır.	1	2	18	41	23	52	2	5
Ölçek geliştirme süreci ampirik analizleri veya madde ve puanlama kriterlerinin incelenmesine yönelik uzman görüşünün alınmasını içermiştir.	1	2	11	25	24	55	8	18
Uzman görüşleri alış süreci ve sonuçları raporlanmıştır.	8	18	18	41	15	34	3	7

Uzmanların nitelikleri, ilgili alandaki deneyimleri ve demografik özellikleri raporlanmıştır.	9	20	20	45	11	25	4	9
Madde seçiminde ölçü alanların örneklem özellikleri raporlanmıştır.	4	9	18	41	18	41	4	9
Örneklemin, ölçü hedef evrenini olabildiğince temsil edici özellikte olduğu belirtilmiştir.	7	16	19	43	15	34	3	7
Maddelerin psikometrik özelliklerini değerlendirmede kullanılan kuram raporlanmıştır.	2	5	0	0	1	2	41	93
Maddenin psikometrik özelliklerini belirlemede kullanılan örneklemin yeterli büyüklük ve çeşitlilikte olduğu raporlanmıştır.	1	2	15	34	22	50	6	14
Maddeler değerlendirilirken büyük örneklem gruplarında madde gücü, madde ayırt ediciliği veya madde işlev farklılığı gibi veriler raporlanmıştır.	2	5	12	27	7	16	23	52
Maddeler veya ölçek deneysel ilişkiler temelinde seçildiğinde ölçek ile ilgili çapraz geçerlik yöntemleri uygulanmıştır.	2	5	10	23	29	66	3	7
Farklı çalışmaların ne ölçüde tutarlı sonuçlar gösterdiği raporlanmıştır.	0	0	6	14	9	20	29	66
Ölçü içeriğinin yönergede tanımlanan alanı ne ölçüde temsil ettiği raporlanmıştır.	1	2	9	20	10	23	24	55

Tablo 3 incelendiğinde maddelerin geliştirilme ve gözden geçirilme aşaması incelenen tezlerin %43'ünde tamamen, %48'inde kısmen uygun olduğu raporlanmıştır. Ölçeklerde ön deneme uygulamasının yapıldığı %70'inde tamamen uygun olarak verildiği belirtilmiştir. Bu standart ile ilgili tezlerin %20'sinde herhangi bir bilgi verilmemiştir. Ön deneme ile ilgili tezlerin %50'sinde tamamen, %25'inde kısmen uygun bilgi verilmiş olup %23'ünde ise herhangi bir bilgi verilmemiştir. Madde havuzundan nasıl madde seçildiği %52'sinde tamamen %41'inde kısmen uygun verildiği belirlenmiştir. Ölçek geliştirme sürecinin ampirik analizleri veya madde ve puanlama kriterlerini incelemeye yönelik uzman görüşünün alınması %55'inde tamamen uygun, %25'inde kısmen uygunken, %18'inde herhangi bir bilgi verilmemiştir. Uzman görüşleri alışı süreci ve sonuçlarının %34'ünde tamamen, %41'inde kısmen uygun verildiği tespit edilmiştir. Uzmanların nitelikleri, ilgili alandaki deneyimleri ve demografik özellikleri %25'inde tamamen, %45'inde kısmen uygun verilirken %20'sinde hiç uygun raporlanmadığı belirlenmiştir. Madde seçiminde ölçü alanların örneklem özellikleri %41'inde tamamen, %41'inde ise kısmen uygun raporlanmıştır. Örneklemin, ölçü hedef evrenini olabildiğince temsil edici özellikte olduğu %34'ünde tamamen, %43'ünde kısmen uygun olduğu tespit edilmiştir. Maddelerin psikometrik özelliklerini değerlendirmede kullanılan kuramın raporlanması %2'sinde tamamen uygundur. Bu standart %5'inde hiç uygun raporlanmamışken, %93'ünde ise herhangi bir bilgi verilmemiştir. Maddenin psikometrik özelliklerini belirlemede kullanılan örneklemin yeterli büyüklük ve çeşitlilikte olduğu %50'sinde tamamen, %34'ünde kısmen uygun olduğu raporlanmıştır. Maddeler değerlendirilirken büyük örneklem gruplarında madde gücü, madde ayırt ediciliği veya madde işlev farklılığı gibi veriler %27'sinde kısmen uygun olarak verilmiştir. Bu standart ile ilgili %52'sinde herhangi bir bilgi verilmemiştir. Maddeler veya ölçek deneysel ilişkiler temelinde seçildiğinde ölçek ile ilgili çapraz geçerlik yöntemleri %66'sında tamamen uygun bir şekilde uygulanmıştır. Farklı çalışmaların ne ölçüde tutarlı sonuçlar gösterdiği %20'sinde tamamen uygun raporlanırken tezlerin %66'sında bu

konuda herhangi bir bilgi verilmemiştir. Ölçeğin içeriğinin yönergede tanımlanan alanı ne ölçüde temsil ettiği %23'ünde tamamen, %20'sinde kısmen uygun olduğu bulunmuştur. Bu standart ile ilgili %55'inde bu konuda herhangi bir bilgi verilmemiştir.

Ölçek uygulama, puanlama prosedürü ve materyallerin standartlarına ilişkin bulgular

Tezler dereceleme ölçeğinde yer alan yedi adet ölçek yönergesi standardına göre incelenmiş ve buna ilişkin bulgular Tablo 4'te verilmiştir.

Tablo 4. Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartlarına İlişkin Frekans ve Yüzde Dağılımları

Ölçek Uygulama, Puanlama Prosedürü ve Materyal Standartları	Hiç uygun değil		Kısmen uygun		Tamamen uygun		Bilgi verilmemiş	
	f	%	f	%	f	%	f	%
Ölçeğin uygulanmasında test uygulayıcılarına ilişkin yönergeler yeterli açıklıkta sunulmuştur.	4	9	19	43	12	27	9	20
Güvenirlik kanıtları yeterli açıklıkta sunulmuştur.	3	7	12	27	29	66	0	0
Geçerlik kanıtları yeterli açıklıkta sunulmuştur.	3	7	18	41	23	52	0	0
Ölçeğin uygulanmasında testi alanlara yönelik yönerge ayrıntılı bir şekilde ifade edilmiştir.	9	20	18	41	12	27	5	11
Ölçeğin ne amaçla kullanılacağı açıkça belirtilmiştir.	1	2	11	25	32	73	0	0
Puanlama süreci ayrıntılı bir şekilde sunulmuştur.	5	11	14	32	7	16	18	41
Puanları yorumlamak için kullanılan prosedürler raporlanmıştır.	4	9	4	9	4	9	32	73

Tablo 4 incelendiğinde ölçeğin uygulanmasında test uygulayıcılarına ilişkin yönergelerin yeterli açıklıkta sunulması %27'sinde tamamen, %43'ünde kısmen uygundur. Bu standart ile ilgili %20'sinde ise herhangi bir bilgi verilmemiştir. Ölçeğin güvenilirlik kanıtlarının yeterli açıklıkta sunulması %66'sında tamamen, %27'sinde kısmen uygun olduğu belirlenmiştir. Ölçeğin geçerlik kanıtlarının yeterli açıklıkta sunulması %52'sinde tamamen, %41'inde kısmen uygun olduğu bulunmuştur. Ölçeğin uygulanmasında testi alanlara yönelik yönergenin ayrıntılı bir şekilde ifade edilmesi tezlerin %27'sinde tamamen %41'inde kısmen uygun olduğu saptanmıştır. Bu standart %20'sinde ise hiç uygun değildir. Ölçeğin ne amaçla kullanılacağı (araştırma veya operasyonel) açıkça belirtilmesi %73'ünde tamamen uygun, %25'inde kısmen uygun olduğu görülmüştür. Ölçeğin puanlama sürecinin ayrıntılı bir şekilde sunulması %16'sında tamamen %32'sinde kısmen uygun olduğu belirtilmiştir. Bu standart ile ilgili %41'inde ise herhangi bir bilgi verilmemiştir. Ölçek puanlarını yorumlamak için kullanılan prosedürlerin raporlanması %9'unda tamamen %9'unda kısmen uygun olarak verilmiştir. Bu standart %9'unda ise hiç uygun değilken %73'ünde ise herhangi bir bilgi verilmemiştir.

Sonuç ve Tartışma

Araştırmanın sonuç ve tartışma kısmı "Ölçek Yönergesi Standartları", "Madde Geliştirme ve Gözden Geçirme Standartları" ile "Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartları" şeklinde üç ana başlık altında sunulmuştur.

Ölçek yönergesi standartlarına ilişkin sonuç ve tartışma

İncelenen tezlerin %80'inde ölçülmek istenen yapı veya alanın tanımlandığı belirlenmiştir. Elde edilen bu sonuç alanyazındaki diğer bazı çalışmalarla da benzerlik göstermektedir (Çüm ve Koç, 2013; Olgun ve Alatl, 2021; Şahin ve Boztunç Öztürk, 2018). Araştırmacıların ölçek geliştirirken ölçülmek istenilen alanı önceden tanımlamaları bu adımın önemini anladıklarını gösteriyor nitelikte olduğu söylenebilir.

Araştırma evreni, tezlerin %70'inde tamamen, %30'unda da kısmen uygun olarak verildiği belirlenmiştir. Bu bulgu Mor Dirlik ve Koç'un (2017) yapmış oldukları çalışmada, inceledikleri iki ölçeğin bu standardı tamamen karşıladığı bulgusu ile benzerlik göstermiştir.

Ölçeğin kullanım amacı ve geliştirme gerekçeleri tezlerin neredeyse tamamında uygun olarak verildiği sonucuna ulaşılmıştır. Bu bulgular alanyazında ölçeklerin geliştirme sürecini inceleyen diğer çalışmaları destekler niteliktedir (Barış Pekmezci ve Ayan, 2020; Çüm ve Koç, 2013; Ergene, 2020; Mor Dirlik, 2014; Olgun ve Alatl, 2021; Şahin ve Boztunç Öztürk, 2018). Mor Dirlik ve Koç'un (2017) çalışmasında ise incelenen iki ölçeğin bu standardı kısmen karşıladığı belirtilmiştir. Ölçülmek istenen özelliklere ya da kullanım amacına uygun olmayan ölçeklerin kullanılması ölçülen özellik ve alan açısından sorun yaratabileceği için bu adımın mutlaka detaylı bir şekilde verilmesi önemlidir.

Yönergede ölçeğin kapsamı, ölçek uzunluğu, madde formatları, maddelerin ve ölçeğin psikometrik özellikleri ile maddelerin ve bölümlerin sıralanışının %45'inde tamamen, %32'sinde kısmen uygun olduğu belirlenmiştir. Bu standart %9'unda hiç uygun değilken %14'ünde bu konu ile ilgili herhangi bir bilginin vermediği görülmektedir. Bir ölçekte kaç madde olduğu, neyi ölçtüğü, nasıl ölçtüğü, maddelerin karışık sırada verilmeyişi ve maddelerin özelliklerinin bilinmesinin o ölçeğin kullanımı açısından kolaylık sağlayacağı söylenebilir.

Ölçeği cevaplama süresi, ölçeği alanlar için uygulama adımları, puanlama ve raporlama gibi prosedürler tezlerin %23'ünde tamamen uygun, %36'sında kısmen uygun bir şekilde verildiği, %30'unda ise herhangi bir bilgi vermediği belirlenmiştir. Çüm ve Koç'un (2013) bulgularına göre çalışmaların %86'sında ölçek geliştirme adımlarından ölçeğin açıklama ve yönergesinin yazılması ile ilgili bilgilerin rapor edilmemesi bu araştırmanın bulguları ile kısmen örtüşmektedir. Bir ölçeğin puanlamasının nasıl yapılacağı ve çıkan sonuçların nasıl raporlanacağını bilmesi ölçeği alan kişilerin özelliklerinin doğru bir şekilde değerlendirileceğini gösterdiğinden bu adımların belirtilmesi oldukça önemlidir.

Madde Geliştirme ve Gözden Geçirme Standartlarına İlişkin Sonuç ve Tartışma

Maddelerin geliştirilme ve gözden geçirilme aşaması incelenen tezlerin %43'ünde tamamen, %48'inde kısmen uygun olarak verilmiştir. APA'nın 1999 yılında yayınladığı "Eğitim ve Psikolojide Ölçme Standartları" kitabında yer alan standartları kullanarak çalışma yapan Mor Dirlik (2014) bu standardın; incelediği beş tezin ikisinde tamamen uygun olduğunu, üçünde kısmen uygun olduğunu belirtmiştir. Bu bulgu araştırma bulgusu ile benzer niteliktedir. Alanyazın incelendiğinde maddelerin oluşturulma süreci ile ilgili benzer bulguları olan, araştırma bulgularını destekleyen çalışmalar yer almaktadır (Ergene, 2020; Olgun ve Alatl, 2021; Şahin ve Boztunç Öztürk, 2018).

İncelenen tezlerin %70'inde ön deneme uygulamasının yapıldığı belirlenmiştir. Bu bulgu Olgun ve Alatlının (2021) çalışması ile neredeyse tamamen (%72) paralellik göstermektedir. Ergene (2020) yapmış olduğu çalışmada ölçek geliştirme makalelerinin %46'sında ön deneme yapıldığını, %54'ünde ise yapılmadığını belirtmiştir. Neredeyse yarı yarıya olan bu durum dikkat çekicidir. Bununla beraber alanyazında ön denemenin yapılmadığı bulgularına rastlanmıştır. Şahin ve Boztunç Öztürk (2018) (%75) ile Çüm ve Koç'un (2013) bulgusu (%28) araştırma bulgularıyla farklılık göstermektedir. Ayrıca bu standart ile ilgili incelenen tezlerin %20'sinde herhangi bir bilgi verilmemiştir ve bu bulgu Çüm ve Koç'un (2013) (%72) bulgusu ile benzerlik göstermektedir. Slavec ve Drnovsek (2012) ise girişimcilik alanında yapılan ölçek geliştirme makalelerini incelediklerinde çalışmaların sadece %10'unda ön deneme yapıldığını raporlamıştır. Bunun yanında incelenen tezlerde ön deneme süreci ile ilgili bilgilerin %50'sinde tamamen uygun, %25'inde kısmen uygun bir şekilde verildiği raporlanmıştır. Bu sonuç alanyazındaki Çüm ve Koç (2013) ile benzerlik göstermektedir. Tezlerin %23'ünde ise herhangi bir bilgi verilmemiştir. Ön deneme uygulamasının yapılmadığı ya da ön deneme uygulaması ile ilgili bilgi verilmemesi durumu çok önemli olan bu aşama için kabul edilebilir bir durum değildir. Ön deneme uygulamasının ne kadar önemli olduğu araştırmacılar tarafından dikkate alınmalıdır.

Madde havuzundan nasıl madde seçildiği incelenen tezlerin büyük çoğunluğunda (%52 tamamen uygun, %41 kısmen uygun) raporlanmıştır. Mor Dirlik ve Koç (2017) inceledikleri ölçeklerin ikisinin bu standardı kısmen karşıladığını ifade etmiştir.

Ölçek geliştirme sürecinin ampirik analizleri veya madde ve puanlama kriterlerini incelemeye yönelik uzman görüşünün alınması incelenen tezlerin %55'inde tamamen uygun, %25'inde ise kısmen uygun olduğu belirlenmiştir. Ölçek geliştirme sürecinde uzman görüşüne başvurulmasıyla ilgili Ergene (2020) %96, Çüm ve Koç (2013) %90, Şahin ve Boztunç Öztürk (2018) %99 ile Tosun ve Taşkesenligil'in (2015) %90 oranında uygun bulunan bulguları ile benzerlik göstermektedir. Bu çalışmalarda neredeyse tamamen uzman görüşüne başvurulduğu belirtildiğinden bu bulgu araştırma bulgusunu destekler niteliktedir. Bu araştırma kapsamında incelenen tezlerin %18'inde uzman görüşü ile ilgili herhangi bir bilgi yer almadığı belirlenmiştir. Ölçek geliştirme süreci uzmanlık gerektiren bir alan olduğu için uzman görüşlerine yer verilmesi önemlidir.

Uzman görüşleri alış süreci ve sonuçlarının tezlerin %34'ünde tamamen uygun, %41'inde ise kısmen uygun olduğu belirlenmiştir. İncelenen tezlerin %25'inde uzmanların nitelikleri, ilgili alandaki deneyimleri ve demografik özellikleri detaylı raporlanırken %45'inde de kısmen raporlandığı belirlenmiştir. Benzer bulgular alanyazındaki diğer çalışmalarda da yer almaktadır (Ergene, 2020; Mor Dirlik, 2014; Mor Dirlik ve Koç, 2017; Olgun ve Alatl, 2021; Tosun ve Taşkesenligil, 2015). Barış Pekmezci ve Ayan (2020) yaptıkları çalışmada ölçek geliştirirken uzman görüşü alınması sürecinde dil uzmanı, ölçme-değerlendirme uzmanı ve konu alanı uzmanını birlikte ele alarak değerlendirmiştir. İnceledikleri ulusal araştırmalarda tüm uzmanların bir arada olduğu bir çalışmaya rastlamadıklarını, inceledikleri araştırmaların büyük çoğunluğunda uzmanların bazılarını yer verildiğini, bazı çalışmalarda da hiç bir bilgiye yer verilmediğini ifade etmişlerdir. İnceledikleri uluslararası çalışmalarda ise çok az bir çalışmada bütün uzmanların yer aldığını, çalışmaların yarısında uzmanlara kısmen yer verildiğini ve üçte birinde de hiçbir bilginin yer almadığını belirtmişlerdir. Güler ve Ayan'ın (2020) çalışmasında incelenen çalışmaların üçte birinde

bütün uzmanlık alanlarının yer aldığı, yarısına yakınında kısmen yer aldığı, çok az bir kısmında ise hiçbir bilginin yer almadığı belirtilmiştir. Ölçek geliştirme sürecinde konu alanı uzmanı, dil uzmanı, ölçme ve değerlendirme uzmanı ve diğer uzmanların bir arada olması ve bu uzmanların uzmanlık alanlarından yararlanılması geçerli ve güvenilir bir ölçek elde etmek için gereklidir.

Madde seçiminde ölçeği alanların örneklem özellikleri incelenen tezlerin %41'inde tamamen, %41'inde ise kısmen uygun olarak verildiği belirlenmiştir. Örneklem, ölçeğin hedef evrenini olabildiğince temsil edici özellikte olduğu %34'ünde tamamen, %43'ünde kısmen uygun olduğu sonucuna ulaşılmıştır. Barış Pekmezci ve Ayan (2020) çalışmasında incelenen çalışmaların büyük çoğunluğunun uygun olduğunu belirtmiştir ve bu bulguyla benzerlik göstermektedir.

Maddelerin psikometrik özelliklerini değerlendirmede kullanılan kuram %2'sinde tamamen uygun verildiği, %5'inde hiç uygun raporlanmadığı ve %93'ünde ise herhangi bir bilgi verilmediği sonucuna ulaşılmıştır. Her kuramda farklı adımlar izlenebileceği için ölçeklerin hangi kurama göre geliştirildiği verilmelidir.

Maddenin psikometrik özelliklerini belirlemede kullanılan örneklem yeterli büyüklük ve çeşitlilikte olduğu incelenen tezlerin yarısında tamamen uygun, %34'ünde kısmen uygun olduğu tespit edilmiştir. Barış Pekmezci ve Ayan (2020) inceledikleri çalışmaların büyük çoğunluğunda örneklem yeterli büyüklükte olduğunu belirtmiştir. Olgun ve Alatlının (2021), deneme uygulamasının yapıldığı grupta madde başına düşen katılımcı oranının 20 ve üstünde olduğunu belirledikleri bulgu ile de benzerlik göstermektedir. Gül ve Sözbilir'in (2015) çalışmalarındaki örneklem büyüklüğünün yaklaşık yarısının 301-500 kişi arası, geriye kalanların yarısının ise 300 ve altındaki kişilerle gerçekleştirildiği; Şahin ve Boztunç Öztürk (2018) ölçek geliştirme çalışmalarının yarısından fazlasının örneklem büyüklüğünün 300 kişi ve üzerinde olması bulgusu ile paraleldir. İlic ve Haseski (2019) örneklem sayısı olarak 501 ve üzerini kullanan çalışmaların oldukça az (%6) olduğunu, 0-100 örneklem büyüklüğünü kullanan araştırmaların büyük çoğunlukta (%52) olduğunu belirttiği çalışma ile araştırma bulguları farklılaşmaktadır. Nunnally (1978) deneme uygulaması yapılacak grubun 300 kişilik olmasını yeterli görürken; Comrey ve Lee (2013) ise 1000 kişinin mükemmel, 500 kişinin çok iyi, 300 kişinin iyi, 200 kişinin orta ve 100 kişinin zayıf örneklem büyüklüğü olduğunu belirtmektedir.

Maddeler değerlendirilirken büyük örneklem gruplarında madde güçlüğü, madde ayırt ediciliği veya madde işlev farklılığı gibi veriler incelenen tezlerin %16'sında tamamen %27'sinde ise kısmen raporlanmıştır. Tezlerin %52'sinde ise herhangi bir bilgi verilmediğine rastlanmıştır. Alanyazın incelendiğinde Olgun ve Alatlın (2021) yaptıkları çalışmada madde analizinin yapılma oranının %78; Şahin ve Boztunç Öztürk (2018) %72,5; Çüm ve Koç (2013) %69 olduğunu Gül ve Sözbilir (2015) ise inceledikleri makalelerin yarısından fazlasında madde analizinin yapıldığını belirtmiştir. Bu bulgular araştırma bulguları ile farklılık göstermektedir. Madde analizleri ile ilgili ölçek geliştirme çalışmalarında herhangi bir bilgi verilmemesi ya da madde analizlerinin uygun bir şekilde yapılmaması ölçek geliştirme çalışmalarında önemli bir eksikliğin olduğunu düşündürmektedir.

Maddeler veya ölçek deneysel ilişkiler temelinde seçildiğinde ölçek ile ilgili çapraz geçerlik yöntemlerinin tezlerin %66'sında tamamen uygun olduğu belirlenmiştir. Farklı çalışmaların

ne ölçüde tutarlı sonuçlar gösterdiği incelenen çalışmaların %66'sında belirtilmediği sonucuna ulaşılmıştır. Alanyazında ise buna benzer bir inceleme yapılmadığı görülmüştür.

Ölçeğin içeriğinin yönergede tanımlanan alanı ne ölçüde temsil ettiği tezlerin %23'ünde tamamen uygun, %20'sinde kısmen uygun olduğu belirlenmiştir. Bununla ilgili tezlerin %55'inde ise herhangi bir bilgi verilmediği belirlenmiştir. Ölçek geliştirme çalışmalarında kuramsal bilgiye hakim olmayan kişilerin hazırladıkları ölçeklerin yapılarında doğru tanımlama yapılamayacağı ve uygun maddelerin yazılamayacağı göz önünde bulundurulmalıdır.

Ölçek Uygulama, Puanlama Prosedürü ve Materyallerin Standartlarına İlişkin Sonuç ve Tartışma

Ölçek uygulamalarında test uygulayıcılarına yönelik yönergeler %27'sinde tamamen yeterli açıklıkta sunulurken %43'ünde kısmen uygun verildiği belirlenmiştir. Araştırma bulgusu Mor Dirlik ve Koç'un (2017) inceledikleri iki ölçek için de tamamen karşılıyor bulgusu ile Mor Dirlik'in (2014) incelenen beş tezin üçü ile tamamen, ikisi ile de kısmen uygun olduğu bulgularıyla farklılık göstermektedir. Çüm ve Koç'un (2013) çalışmasında ise ölçeğin yönergesinin yazılması standardına %14'ünde uyulduğu belirtilmiştir. Yönergelerin test alanları ve test uygulayıcıları doğru bir şekilde yönlendirmesi ve ölçeklerin uygun bir şekilde uygulanması için ayrıntılı yazılması önemlidir.

İncelenen çalışmaların %66'sında ölçeğin güvenilirlik kanıtlarının yeterli açıklıkta raporlandığı sonucuna ulaşılmıştır. Alanyazında da incelenen çalışmaların büyük bir çoğunluğunda güvenilirlik çalışmalarına yer verilmesi bu bulguları desteklemektedir (Barış Pekmezci ve Ayan, 2020; Çüm ve Koç, 2013; Gül ve Sözbilir, 2015; Güler ve Ayan, 2020; Hinkin, 1995; Şahin ve Boztunç Öztürk, 2018). Yapılan çalışmalardaki güvenilirlik analizi oranı Hinkin'de (1995) %74; Güler ve Ayan (2020) %92; Çüm ve Koç (2013), Gül ve Sözbilir (2015), Olgun ve Alatl (2021) ile Şahin ve Boztunç Öztürk'te (2018) %100 olarak belirlenmiştir. Barış Pekmezci ve Ayan (2020) ise çalışmaların yarısından fazlasında bu analizin yapıldığını belirtmiştir.

Tezlerin %52'sinde geçerlik kanıtlarının yeterli açıklıkta sunulduğu, %41'inde ise kısmen yeterli olduğu belirlenmiştir. Bu bulgu Güler ve Ayan'ın (2020) yapmış oldukları çalışmada geçerlik kanıtının %54'ünde tamamen uygun, %24'ünde ise kısmen uygun olması ile benzerlik göstermektedir. Ayrıca, alanyazında incelenen diğer çalışmaların büyük bir çoğunluğunda Çüm ve Koç (2013) %69; Barış Pekmezci ve Ayan (2020) %96; Şahin ve Boztunç Öztürk (2018) %97; Gül ve Sözbilir (2015), Olgun ve Alatl (2021) %100 oranında uygun olduğu belirtilmiştir. Ölçek geliştiren araştırmacıların büyük çoğunluğunun ölçekle ilgili geçerlik ve güvenilirlik belirleme yöntemlerini kullanma konusunda bilgi sahibi olduğu söylenebilir.

Ölçeğin uygulanmasında testi alanlara yönelik yönergenin tezlerin %27'sinde tamamen, %41'inde ise kısmen ayrıntılı bir şekilde ifade edildiği belirlenmiştir. Araştırma bulgusu Mor Dirlik'in (2014) yapmış olduğu çalışmadaki beş tezin ikisinde tamamen, ikisinde de kısmen uygun olduğu bulgusu ile benzerlik göstermektedir. Mor Dirlik ve Koç'un (2017) çalışmasındaki incelenen iki ölçeğin ikisi için de tamamen karşılıyor bulgusu ile farklılık göstermektedir.

Ölçeğin ne amaçla kullanılacağı tezlerin %73'ünde tamamen uygun olarak verilmiştir. Ölçeğin puanlama süreci tezlerin %32'sinde kısmen ayrıntılı bir şekilde verilirken %41'inde herhangi bir bilgi

verilmediği bulgusuna ulaşılmıştır. Bu bulgu; Kaya Uyanık vd.'nin (2017) yapmış olduğu çalışmadaki ölçeklerin puanlama bilgisinin yaklaşık %44'ünde verilmediği bulgusuyla benzerlik göstermektedir. Ölçek geliştirenler tarafından ölçeğin puanlaması ile ilgili bilgilerin verilmesi daha sonra bu ölçeği kullanacak araştırmacılar tarafından hatalı puanlama yapılmasını ve kişiler hakkında eksik veya yanlış karar verilmesini önleyecektir.

Öneriler

Ölçeklere ilişkin hazırlanan yönergelerde ölçülmek istenen yapı, ölçeğin kapsamı, ölçek uzunluğu, madde formatları, ölçeğin psikometrik özellikleri ile ölçekteki maddelerin ve ilgili bölümlerin nasıl sıralandığı daha ayrıntılı şekilde verilebilir. Direkt olarak uygulama sürecini içeren ölçeğin ne kadar sürede cevaplandırılacağı, ölçeğin nasıl puanlanacağı vb. prosedürler açıkça belirtilebilir. Hem ölçek uygulayıcıları hem de ölçeğin uygulandığı grup açısından kullanılan yönerge ile ölçeğin puanlama süreci ayrıntılı bir şekilde ifade edilebilir.

Ölçeklerde yer alacak maddelerin geliştirilmesi ve bu maddelerin yeniden gözden geçirilmesi gibi aşamaları ile asıl uygulamadan önce yapılması gereken ön deneme çalışmaları ölçek geliştirme araştırmalarında yer alması ve ölçeğin asıl formunda kullanılacak maddelerin özenli bir şekilde seçilmesi önerilmektedir. Ayrıca bu maddeler seçilirken özellikle dil, ölçme değerlendirme ve alan uzmanları tarafından görüş alınmalıdır. Bu uzmanların nitelik ve deneyimleri ile görüş alınan uzman sayısı çalışmalarda belirtilebilir. Maddelerin psikometrik özellikleri verilirken her bir kuramda farklı adımlar izlenebileceği için hangi kurama göre bu özelliklerin değerlendirildiği çalışmalarda yer alabilir.

Ayrıca bu araştırma farklı yıl, konu ve alanlar kullanılarak; farklı ölçütler veya standartlar belirlenerek diğer araştırmacılar tarafından tekrarlanabilir. Buna ek olarak lisansüstü eğitim sürecinde ölçek geliştirme ile ilgili bir dersin zorunlu ders olarak okutulmalı, ölçek geliştirme ile ilgili tezler yazılırken ölçme ve değerlendirme uzmanlarından görüş alınmalıdır.

Kaynaklar

- Aktaş, M. (2023). *Ölçek geliştirme üzerine bir bibliyometrik analiz*. Batı Anadolu Eğitim Bilimleri Dergisi, 14(2), 1273-1292. <https://doi.org/10.51460/baebd.1275572>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. [Eğitim ve psikolojide ölçme standartları]. American Educational Research Association.
- Barış Pekmezci, F. & Ayan, C. (2020). Confusion of scale development: Investigation of self-efficacy scales. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 48, 130-151. <https://doi.org/10.9779/pauefd.529986>
- Baykul, Y. (2021). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Pegem Akademi.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001.316.44600200104>
- Cohen, R. J. & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement*. (9th ed.) McGraw-Hill Education.
- Comrey, A. L., & Lee, E. B. (2013). *A first course in factor analysis*. (2. Baskı). New York: Psychology.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston, Inc.

- Çelik, D. & Yüksel, G. (2020). Müzik Eğitimi Kapsamında Yapılan Ölçek Geliştirme Çalışmalarının Çok Yönlü İncelenmesi. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 9(5) , 4059-4087. <https://doi.org/10.15869/itobiad.793488>
- Çüm, S. & Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Journal of Educational Sciences & Practices*, 12(24), 115-135. http://ebuline.com/pdfs/24Sayi/EBU24_3.pdf sayfasından erişilmiştir.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.
- Ergene, Ö. (2020). Matematik eğitimi alanında ölçek geliştirme ve ölçek uyarlama makaleleri: Betimsel içerik analizi. *Yaşadıkça Eğitim*, 34(2), 360-383. <https://doi.org/10.33308/26674.874.2020342207>
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi.
- Erkuş, A. (2016). Ölçek geliştirme ve uyarlama çalışmalarındaki sorunlar ile yazım ve değerlendirilmesi. Ö. Demirel & S. Dinçer (Ed.), *Eğitim bilimlerinde yenilik ve nitelik arayışı* içinde (s. 1212 – 1224). Ankara: Pegem Akademi.
- Eze, P. O., Victor-Aigbodion, V., Ossai, O. V., Ugwoezunu, A. U., Egenti, N. T., & Eseadi, C. (2017) Procedures for the construction and development of psychological test. *European Journal of Scientific Research*, 145(2), 146 – 151. https://www.researchgate.net/publication/319330667_Procedures_for_the_Construction_and_Development_of_Psychological_Test_published_by_European_Journal_of_Scientific_Research sayfasından erişilmiştir.
- Gül, Ş. & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178), 85-102. <https://doi.org/10.15390/EB.2015.4070>
- Güler, G. & Ayan, C. (2020). Review of Attitude Scales Developed in Turkey Between 2002-2018 Regarding the Scale Development Process. *Journal of Faculty of Educational Sciences*, 53(3), 839-863. <https://doi.org/10.30964/auebfd.658488>
- Hinkin, T. R. (1995). A review of scale development practices in the study in organizations. *Journal of Management*, 21(5), 967-988. <https://doi.org/10.1177/014920.639502100509>
- İlic, U., & Haseski, H. İ. (2019). *Bilgi işlemsel düşünmeyi ölçmeye yönelik geliştirilen veri toplama araçlarının incelenmesi*. I. Uluslararası Bilim Eğitim Sanat ve Teknoloji Sempozyumu’nda sunulmuş bildiri, Dokuz Eylül Üniversitesi, İzmir. https://www.researchgate.net/profile/Ali-Ilhan3/publication/337114093_9_Sinif_Cografya_Dersi_Ogretim_Programi_Kazanimlarinin_Yenilenmis-Bloom_Taksonomisine_Bilissel_Alan_Gore_Degerlendirilmesi/links/5dfc710192851c83648b16ee/9-Sinif-Cografya-Dersi-Ogretim-Programi-Kazanimlarinin-Yenilenmis-Bloom-Taksonomisine-Bilissel-Alan-Goere-Degerlendirilmesi.pdf sayfasından erişilmiştir.
- Karasar, N. (2010). *Bilimsel araştırma yöntemi-kavramlar, ilkeler, teknikler*. (23. Baskı). Ankara: Nobel.
- Kaya Uyanık, G., Güler, N., Taşdelen Teker, G. & Demir, S. (2017). Türkiye’de eğitim alanında yayımlanan ölçek geliştirme çalışmalarının uygunluğunun çok yüzeyli Rasch modeli ile incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 8(2), 183-199. <https://doi.org/10.21031/epod.291367>
- Korkmaz, İ. (2020). Nicel araştırmalarda evren, örneklem, örnekleme teknikleri. B. Oral, & A. Çoban içinde, *Kuramdan uygulamaya eğitimde bilimsel araştırma yöntemleri* içinde (s. 147-159). Ankara: Pegem Akademi.
- Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, 25, 47-76. <https://doi.org/10.2307/271061>
- Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. <https://doi.org/10.2307/2529310>

- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Mor Dirlik, E. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 5(2), 62-78. <https://doi.org/10.21031/epod.63138>
- Mor Dirlik, E. & Koç, N. (2017). The Analysis of the Psychological Tests Using In Educational Institutions According To the Testing Standards. *Journal of Measurement and Evaluation in Education and Psychology*, 8(4), 453-468. <https://doi.org/10.21031/epod.365141>
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Olgun, G. & Alatlı, B. (2021). Türkiye’de ergenlere yönelik ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Türk Eğitim Bilimleri Dergisi*, 19(1), 568-592. <https://doi.org/10.37217/tebd.849954>
- Slavec, A. & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review*, 14(1), 39-62. <https://doi.org/10.15458/2335-4216.1203>
- Soycan, M., & Babacan, E. (2019). Müziksel İşitme Okuma ve Yazma İle İlgili Geliştirilmiş Ölçme Araçlarının İncelenmesi: İçerik Analizi Çalışması. *Elektronik Sosyal Bilimler Dergisi*, 18(69), 343-353. <https://doi.org/10.17755/esosder.409006>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680. <https://doi.org/10.1126/science.103.2684.677>
- Şahin, M.G. & Boztunç Öztürk, N. (2018). Eğitim alanında ölçek geliştirme süreci: Bir içerik analizi çalışması. *Kastamonu Eğitim Dergisi*, 26(1), 191-199. <https://doi.org/10.24106/kefdergi.375863>
- Tatlı Harmancı, S. & Güngör Aytar, A. F. (2021). Duygu düzenleme ile ilgili Türkiye’de geliştirilen veya uyarlanan ölçek çalışmalarının incelenmesi. *Pearson Journal of Social Sciences – Humanities*, 6(13), 170-184. <https://doi.org/10.46872/pj.291>
- Tavşancıl, E. ve Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri*. Epsilon Yayınevi, İstanbul.
- Tezbaşaran, A. (2008). *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları, 27-48.
- Tosun, C. & Taşkesenligil, Y. (2015). The instruments used in science education in Turkey: a descriptive content analysis. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 364 – 383. <https://doi.org/10.17240/aibuefd.2015.15.2-500.016.1329>
- Worthington, R. L. & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838. <https://doi.org/10.1177/001.100.0006288127>
- Yıldırım, A. & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.