



Ölçek Seçiminde Arayış: Doğru Ölçeği Bulma Yolculuğu

Meltem ACAR GÜVENDİR¹, Abdullah Faruk KILIÇ², Yeşim ÖZER ÖZKAN³

Özet

Bu çalışmanın amacı, geliştirilmiş ya da uyarlanmış bir ölçek kullanacak araştırmacılara rehberlik edecek kontrol listesi geliştirmektir. Bu amaçla araştırmacılar tarafından alanyazına dayalı olarak ölçek geliştirme ve uyarlama adımları hazırlanmış ve tüm adımların ağırlıkları tanımlanmıştır. Hazırlanmış olan adımlar ve ağırlıkları eğitimde ölçme ve değerlendirme alanında doktora derecesine sahip olan 10 akademisyene gönderilmiştir. Akademisyenlerden hem adımlara yönelik görüşleri hem de takdir ettikleri ağırlıklar istenmiştir. Gelen dönütlere göre düzenlemeler yapılarak ölçek geliştirme ve uyarlama adımlarının ağırlıkları tanımlanmış, adımların ağırlıklarına göre elde edilecek toplam puan göz önüne alınarak bir ölçüt belirlenmiştir. Hâlihazırda geliştirilmiş/uyarlanmış bir ölçeği seçecek araştırmacılara önerilen ölçüt temel alınarak bu ölçekleri seçip seçmeme konusunda bir referans sunulmuştur. Bu yönüyle çalışmanın geliştirilmiş ya da uyarlanmış ölçeklerin seçiminde rehber oluşturması açısından önemli katkıları olacağı düşünülmektedir.

Makale Bilgileri

Araştırma

Makalesi

Gönderim Tarihi

14/06/2024

Kabul Tarihi

17/04/2025

Anahtar

Kelimeler

Ölçek
geliştirme,
Ölçek uyarlama,
Ölçek seçimi,
Geçerlik,
Güvenirlilik,
Kontrol listesi

¹ Trakya Üniversitesi, 0000-0002-3847-0724, meltemacar@trakya.edu.tr

² Trakya Üniversitesi, 0000-0003-3129-1763, afarukkilic@trakya.edu.tr

³ Gaziantep Üniversitesi, 0000-0002-7712-658X, yesimozkan@gantep.edu.tr

Atıf:

Acar Güvendir, M., Kılıç, A. F. ve Özer Özkan, Y. (2025). Ölçek seçiminde arayış: Doğru ölçeği bulma yolculuğu. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi [PAÜEFD]*, 64, 338-373. <https://doi.org/10.9779/pauefd.1501330>

Giriş

Eğitim süreçlerinin temelinde insan yer alır ve bu süreçlerin başarılı bir şekilde değerlendirilmesi, genellikle soyut kavramlar üzerinden gerçekleştirilir. Öğrencilerin öğrenme becerileri, motivasyon seviyeleri, problem çözme yetenekleri, eleştirel düşünme kapasiteleri gibi önemli özellikler doğrudan gözlemlenemez. Bu noktada, ölçekler önemli bir araç olarak devreye girer. Ölçekler, bu soyut özellikleri yapılandırılmış ve nicel ifadelerle ölçülebilir ve anlaşılır hale getirir. Böylece eğitim süreçlerinin etkinliği, öğrenci performansı ve öğretmen etkililiği gibi kritik unsurlar objektif bir şekilde analiz edilebilir. Ölçekler, eğitim ve sosyal bilimler araştırmalarında sıklıkla kullanılan popüler psikometrik araçlardan biridir (Aybek ve Toraman, 2022; Bistop ve Herron, 2015; Chen ve Liu, 2020; Joshi ve diğerleri, 2015; Tanujaya ve diğerleri, 2022). Türkiye’de eğitim alanında yapılmış çalışmaların tematik ve metodolojik olarak incelendiği araştırmalarda (Alper ve Gülbahar, 2009; Doğan ve Tok, 2018; Erdem, 2011; İşçi, 2013; Karadağ, 2009; Ozan ve Köse, 2014; Seçer ve diğerleri, 2014; Şenyurt ve Özer Özkan, 2017; Tosuntaş ve diğerleri, 2019) en çok kullanılan veri toplama aracının ölçek olduğu belirtilmektedir.

Son yıllarda, yayın baskısının da etkisiyle, araştırmacılar tarafından yürütülen ölçek geliştirme ve uyarlama çalışmalarında belirgin bir artış olduğu gözlemlenmektedir. Ancak bu durum, söz konusu çalışmalarda pek çok hata ve eksikliğin ortaya çıkmasına neden olmaktadır (Erkuş, 2007). Bu sorunların hazırlanan makalelerde giderek daha fazla görülmesi, araştırmacıların dikkatini çekmiş ve bu eksiklikleri tespit etmek amacıyla çeşitli çalışmalar yapılmıştır (Acar Güvendir ve Özer Özkan, 2015; Boztunç Öztürk ve diğerleri, 2014; Carpenter, 2018; Chen ve Liu, 2020; Çüm ve Koç, 2013; Morgado ve diğerleri, 2017; Tavsancıl ve diğerleri, 2014). Bu tür çalışmaların varlığı bile, ölçek geliştirme süreçlerinde ciddi sorunlar yaşandığının önemli bir göstergesidir.

Belirlenen sorunlar arasında, ölçek geliştirme süreçlerinde işlevsel tanımların yapılmaması, ön deneme uygulamalarının gerçekleştirilmemesi veya raporlanmaması, madde analizleriyle ilgili yetersiz bilgi sunulması gibi eksiklikler öne çıkmaktadır. Ayrıca, açımlayıcı faktör analizi (AFA) yaygın olarak kullanılırken doğrulayıcı faktör analizine (DFA) daha az yer verilmesi de önemli bir eksiklik olarak dikkat çekmektedir (Acar Güvendir ve Özer Özkan, 2015; Çüm ve Koç, 2013). Bu sorunlar, ölçek geliştirme çalışmalarında standartların ve yönergelerin eksikliğine işaret etmektedir. Bu nedenle, süreçle ilgili yönergelerin açık, net ve standart hale getirilmesi büyük önem taşımaktadır. Böylece kullanılan ölçeklerin daha güvenilir sonuçlar vermesi sağlanabilir.

Ölçek geliştirme basamakları dikkate alınarak gerçekleştirilecek çalışmalar sonucunda elde edilecek ölçeklerin yeterli olacağı

söylenbilir. Ölçek geliştirme basamakları ile ilgili farklı araştırmacılar tarafından farklı öneriler yapılırsa da temelde benzerdir. Buna göre ölçek geliştirme basamakları ölçülen özelliğin tanımlanması ile başlamaktadır. Daha sonra bu tanım doğrultusunda maddeler yazılmaktadır. Maddelere yönelik uzman görüşleri alındıktan sonra ön deneme ve deneme uygulamaları yapılmaktadır. Elde edilen verilerle geçerlik ve güvenirlik analizleri yapılmakta, daha sonra ölçeğin standartlaştırılması gerçekleştirilmektedir (Morgado ve diğerleri, 2017). Bu çalışmanın amacı ölçek geliştirme süreçlerini tanıtmak olmadığından kısaca ölçek geliştirme süreçlerinden bahsedilmiştir. Ölçek geliştirmeyle ilgili daha ayrıntılı açıklamalara ilgilenen araştırmacılar ulaşabilir (Acar Güvendir ve Özer Özkan, 2015; Erkuş, 2019; 2022; Kılıç, 2023).

Ölçülen Özelliğin Tanımlanması

Bu süreç ilk olarak genellikle alanyazın taraması ile başlamaktadır (Kalkbrenner, 2021). Ölçülen özellik tanımlanırken, yapının benzer ve farklı olan yapılardan ayrımı ortaya konmalıdır (Furr, 2011). Diğer bir deyişle tanımlama yapılırken hem kapsayıcı hem de diğer değişkenlerden ayrılan yönlerine vurgu yapılmalıdır (Erkuş, 2019). Bu aşamada alanyazın taramasının yanında uzman görüşlerine de başvurulabilir. Yapı tanımlandıktan sonra bu yapıyı ölçecek maddeler yazılmalıdır.

Madde Yazım Süreci

Yeterli sayıda ve ilgili teorik yapıyı bütünüyle kapsadığı düşünülen düzeyde madde yazımı ölçek geliştirme sürecinin önemli bir aşamasıdır. Bu aşamanın tamamlanmasıyla araştırmacı madde havuzunu elde etmiş olur. Havuzdaki maddelerin ölçmeyi amaçlanan yapıyı ölçüp ölçmeyeceğini, diğer bir ifade ile amaca hizmet etme derecesini ("geçerliliğini") belirlemeye yönelik işlem adımlarını yapmak gerekmektedir. Bu aşama, ölçeğin kapsam geçerliğine ilişkin kanıtların elde edilmesini içermektedir. Kapsam geçerliği kanıtları alanyazında yaygın olarak uzman görüşüne dayalı olarak elde edilmektedir. Uzman görüşü bir anlamda madde yazarı dışında başka bir uzman tarafından da sürecin okunmasıdır. Bu süreç, hedef gruba uygulama öncesi maddelerin gözden geçirilerek düzeltilmesi olarak adlandırılabilir. Bu aşamada ölçek geliştirmenin farklı alandan uzmanların iş bölümünü gerektiren; ölçme değerlendirme, dil ve alan olmak üzere üç alandan uzmanların desteğiyle yürütülmesi gereken bir ekip çalışması olarak tanımlanabilir. Uzmanlar, her maddeyi ayrı ayrı değerlendirir ve önerilerde bulunur. Bu geri bildirimler, ölçeğin içerik geçerliği üzerinde önemli bir rol oynar ve ölçeğin güvenirliği ve geçerliği açısından kritik bir aşama olarak kabul edilir.

Geçerlik ve Güvenirlik Analizleri

Geçerlik ve güvenirlik analizleri, öncelikle deneme uygulamasından elde edilen verilerle AFA yapılarak ölçeğin yapısının ortaya çıkarılmasını, ardından bu yapının doğrulanması amacıyla yeni bir veri seti toplanarak DFA'nın gerçekleştirilmesini ve güvenirlik için iç tutarlık (örneğin, Cronbach alfa) ile test-tekrar test yöntemlerinin uygulanmasını kapsayan bir süreçtir. Bu noktada örneklemin kaç kişiden oluşturulması gerektiği araştırmacılar tarafından cevaplanması gereken bir önemli bir sorudur.

Örneklem Büyüklüğü

Ölçek geliştirme sürecinde araştırmacıların en çok sorguladığı konulardan biri de ölçeğin kaç kişi üzerinde geliştirilmesi gerektiğidir. Öncelikle, ölçek geliştirme sürecinin yalnızca maddelerin yazılıp bireylere uygulanarak, geçerlik ve güvenirlik süreçlerinin izlendiği bir süreç olmadığı, baştan sona her adımının titizlikle yapılması gerektiği unutulmamalıdır. Bu yüzden, geliştirilmek istenen ölçeğin kaç kişi üzerinde geliştirileceği, ölçülmeye çalışılan yapıya, elde edilen verinin yapısına ve kullanılacak olan istatistiksel tekniklere göre farklılaşacaktır.

Ölçek geliştirme sürecinde deneme uygulaması yapılmadan önce ön deneme uygulamasının yapılması gerekir. Bunun amacı, hazırlamış olduğunuz maddelerin ölçmeye çalıştığınız yapıyı ölçüp ölçmediğini belirlemek, maddelerin açık, net ve anlaşılır olup olmadığını kontrol etmek, ölçeğin ortalama cevaplama süresine karar vermektir. Deneme uygulamasının esas deneme uygulamasında yer alacak hedef kitleden özellikle uç grupları (ilgili yapının/değişkenin ranjı) temsil eden 10-15 kişi üzerinde yapılması yeterli olacaktır (Erkuş, 2019). Bu süreç mutlaka birebir ve mümkünse yüz yüze yapılmalıdır; çünkü araştırmacı sorular sormalı, katılımcıların davranışlarını gözlemeli ve notlar almalıdır. Ölçeğin, psikometrik özelliklerinin belirlenmesinin amaçlandığı esas deneme uygulamasında kaç kişi olması gerektiği konusu incelendiğinde bu sayı için farklı ölçütler öne sürülmüştür. Bryman ve Cramer (2001), Everitt (1975), Hair vd. (1998), Ho (2006), MacCallum vd. (1999), Steven (1996), Sönmez ve Alacapınar (2016) ve Şencan (2005) madde sayısına göre örneklem büyüklüğüne karar verilmesi gerektiğini belirtmişlerdir. Onlara göre örneklem için ölçekte bulunan madde sayısının 5, 10 veya 20 katı kadar sayıya ulaşmak gereklidir. Ancak güncel alanyazında madde sayısının temel alınarak bu sayının katları şeklinde verilen örneklem büyüklüğünün doğru bir ölçüt olmayacağı belirtilmektedir (Erkuş, 2019; Goretzko ve diğerleri, 2021).

Arrindell ve Van der Ende (1985), Kline (1994) ve DeVellis (2017) örneklem büyüklüğünün belirlenmesinde faktör sayısının önemine dikkat çekmişlerdir. Onlara göre bu süreçte özellikle dikkat edilmesi gereken nokta, faktör sayısının açık ve faktör yüklerinin büyük olmasına dikkat

edilmesidir. MacCallum vd. (1999), Guttman (1954) ve Kaiser (1970) ise örneklem büyüklüğü arttıkça örneklem hatası da azalacağından, ortak varyanslar yüksek (0.60 ve üstü) olduğunda örneklem büyüklüğünün göreceli olarak küçük olabileceğini; ancak ortak varyanslar büyük olmadığında daha büyük örneklem gereksinim olacağını dile getirmişlerdir.

Örneklem seçiminde Cattell (1978), Comrey ve Lee (1992), Gorsuch (1974), Kline (1994), Nunnally (1978), Rouquette ve Falissard (2011) Tabacknick ve Fidell (2019), her biri farklı sayıda (200, 300, 500) örneklem önerisinde bulunarak kesin bir sayı önerisi getirmişlerdir. DeVellis (2017), Henson ve Roberts (2006) ve Osborne ve Costello (2004) ise ölçek geliştirme sürecinde kullanılacak faktör analizleri için büyük örneklemelerin her durumda daha iyi sonuçlar elde etmede önemli olduğuna vurgu yapmışlardır. Ancak büyük örneklemelerin aksine Sapnas ve Zeller (2002), De Winter vd. (2009), Mundfrom vd. (2005), Preacher ve MacCallum (2002) ve Steven (1996) ise belirli durumlarda küçük örneklemelerin de yeterli olabileceğini savunmuşlardır. Erkuş (2019) ile Şeker ve Gençdoğan'a (2006) göre örneklemin büyük ya da küçük olmasından ziyade geliştirilen ölçekle ölçülmeye çalışılan özelliğin seçilen örneklemle temsil edilmesi daha önemlidir.

Ölçek geliştirme sürecinde faktör yapısını ortaya koymak amacıyla yapılan faktör analizi sürecini temel alarak bu aşamada örneklem büyüklüğünü belirlemek için tek bir ölçütün yeterli olmayacağını, birden fazla ölçütün göz önünde bulundurulması gerektiğini belirten araştırmacılar da bulunmaktadır (Çokluk ve diğerleri, 2014; Erkuş, 2019; Guadagnoli ve Velicer, 1988).

Faktör yapısı ortaya konulan bir ölçeğin farklı bir grup üzerinde bu faktör yapısının doğrulanması amacıyla yapılan DFA için seçilecek örneklem büyüklüğünün kaç kişi olması gerektiğine ilişkin görüşler (Anderson ve Gerbing, 1984; Boomsma, 1982; 1985; Cohen ve Cohen, 1983; De Winter ve diğerleri, 2009; Goodwin, 1999; Guadagnoli ve Velicer, 1988; Helm ve diğerleri, 2017; Hu ve Bentler, 1999; Jackson, 2001; MacCallum ve diğerleri, 1999; Mundfrom ve diğerleri, 2005; Muthén ve Muthén, 2002; Nevitt ve Hancock, 2004; Osborne ve Costello, 2004; Rouquette ve Falissard, 2011; Shevlin ve Miles, 1998; Steven, 1996; Uyumaz ve Sırgancı, 2020; Yu ve Muthén, 2002) belirtilmekle birlikte, bu görüşler kendi içinde de farklılık göstermektedir.

Ölçek geliştirme sürecinde seçilecek olan örneklemin ne kadar olması gerektiği ile ilgili olarak verilen önerilerin ne kadar farklı olduğu görülmektedir. Madde başına düşen kişi sayısı önemlidir; ancak asıl önemli olan seçilen kişilerin ölçülen psikolojik yapıyı temsil ediciliğidir. Ayrıca ölçülecek özelliğin tek ya da çok boyutlu, çok faktörlü olmasına bağlı olarak örneklem büyüklüğü değişeceği için sadece kesin bir sayı önermek de geçersiz kalabilir (Erkuş, 2019). Ayrıca elde edilen faktör

yüklerinin ve ortak varyansların yeterli olacağını önceden bilmek olası olmadığı için başka önemli değişkenleri göz önüne almadan kesin bir sayıya bağlı kalmak yanıltıcı olabilir. Ancak yine de veri toplama aşamasına geçen birçok araştırmacı örneklemin büyüklüğünün ne kadar olması gerektiğiyle ilgili sorulara sahip olacaktır. Bu noktada bir sayı önerisinde bulunulurken faktör analizinin kullanılacağı, bu analizin varsayımları açısından veri setinin inceleneceği ve güncel alanyazındaki öneriler dikkate alınarak en az 400 kişilik örnekleme ulaşılmamasının çalışmanın sonuca ulaşması açısından daha iyi olacağı söylenebilir (Goretzko ve diğerleri, 2021).

Örneklemin Özelliği

Bu süreçte seçilecek olan örneklem, ilgili özelliğin ranjını en iyi şekilde temsil edebilecek kişilerden oluşmalıdır. Örnekleme belirlerken, niteliğin yanı sıra varyansın da büyük önemi bulunmaktadır. Maddeler ve ölçek, bireyler arasındaki farklılıkları ayırt edebilmesi için geliştirilecek olan ölçeğin ölçülecek tüm özellikleri temsil edecek bireylere uygulanması gerekir (DeVellis, 2017; Erkuş, 2019). Sonuç itibarıyla seçilecek grup için homojenlik içinde heterojenliğin sağlanması yerinde olacaktır. Örneğin, ölçeğin geliştirileceği grubun öğretmenler olduğu bir durumda, gruptaki tüm bireylerin öğretmen olması homojenliği, öğretmenlerin tüm branşlardan seçilmesi de heterojenliği ifade etmektedir.

Ölçek geliştirme sürecinde örneklemin seçkisiz seçilmesi gönüllülük ilkesine ters düşeceğinden uygun olmaz. Amaçlı bir şekilde seçim yapılırsa ölçeğin yapısının ortaya çıkarılabileceği düşünülen bir gruba ulaşılmış olur.

Yapı Geçerliliği

Ölçek geliştirme veya uyarlama süreçlerinde faktör analitik yöntemlerden sıklıkla yararlanılmaktadır. Ölçülen yapının ortaya çıkarılmasında ya da ölçüldüğü düşünülen yapı ile eldeki verinin uyumlu olup olmadığının değerlendirilmesinde faktör analizine başvurulmaktadır (Nunnally, 1978). Faktör analizi açıklayıcı ve doğrulayıcı olarak iki şekilde ele alınabilir.

Açımlayıcı Faktör Analizi (AFA)

Ölçek geliştirme sürecinde AFA çok önemli bir yere sahip olup altta yatan gizil yapıyı ortaya çıkarmak için kullanılmaktadır (Fabrigar ve Wegener, 2012). AFA, değişkenler arasında neden korelasyon olduğunu açıklayan daha az sayıda (bir ya da birkaç tane) gizil değişken (yapı/faktör) elde etmek için kullanılan bir yöntemdir (Goretzko ve diğerleri, 2021). Ölçek geliştirme sürecinde AFA'nın basamakları doğru şekilde takip edilmeli ve her bir aşamada kullanılan yöntemler raporlanmış olmalıdır. AFA'da öncelikle veri seti analiz varsayımları açısından incelenmelidir. Daha sonra faktör çıkarma yöntemine karar

verilmelidir. Faktör sayısına farklı yöntemler kullanılarak karar verilmeli ve analiz gerçekleştirilmelidir. Analiz sonucunda elde edilen bulgular uygun şekilde raporlanmalıdır. AFA'nın işlem basamakları aşağıdaki gibi ele alınmıştır.

Veri Setinin Analiz Varsayımları Açısından İncelenmesi

AFA çok değişkenli istatistiksel bir yöntem olduğu için çok değişkenli analiz varsayımlarına sahiptir. Bu nedenle analiz edilecek veri seti; çok değişkenli normal dağılım, çok değişkenli uç değerler, çoklu doğrusal bağlantı, kayıp değerler ve değişkenler arasında doğrusal ilişki bulunması açısından incelenmelidir. Bunların yanında AFA için veri setinin faktör analizine uygunluğunun değerlendirilmesi amacıyla Bartlett'in küresellik testi (Bartlett, 1950) ve Kaiser-Meyer-Olkin (KMO) değeri de kullanılmaktadır.

Çok değişkenli normal dağılım kategorik verilerde genellikle sağlanamamaktadır. Bu nedenle çok değişkenli normal dağılım varsayımının ihlaline karşı daha güçlü kestirimler sunan temel eksenler (principle axis) faktör çıkarma yöntemi (Costello ve Osborne, 2005; Fabrigar ve diğerleri, 1999) veya minimum kalıntı (minimum residual) (Zygmunt ve Smith, 2014) faktör çıkarma yöntemlerinden biri kullanılabilir.

Çok değişkenli uç değerler açısından veri seti incelenmeli ve yanlış veri girişi olmuşsa düzeltilmelidir. Veri setinde uç değerler varsa analizler öncesinde veri setinden çıkarılabilir ya da hem uç değerlerle hem de uç değerler çıkarılarak analiz yapılabilir (Kılıç, 2022b). Çoklu doğrusal bağlantı açısından veri seti incelenmeli ve değişkenler arasında çoklu doğrusal bağlantı olup olmadığı raporlanmış olmalıdır. Çoklu doğrusal bağlantı AFA sonuçlarını etkiyebileceğinden bu varsayımın incelenerek sonuçlarının raporlanması önemlidir. Benzer şekilde kayıp değerler açısından veri seti incelenmeli ve kayıp değerlere ilişkin alınan önlemler raporlanmalıdır. Bu varsayımların yanında Bartlett'in küresellik testi ve KMO değeri de raporlanabilir. Bunların olması AFA açısından çok aşırı önemli olmasa da raporlanmasının daha uygun olduğu söylenebilir. AFA'nın ne kadar bir örnekleme gerçekleştirileceği de önemlidir. Örneklemin büyük olmasının yanında ölçülen özelliğin her düzeyinden bireyi de içermesi önemlidir. Örneklem büyüklüğüyle ilgili ayrıntılı bilgi daha önce verildiği için tekrar üzerinde durulmamıştır.

Açımlayıcı Faktör Analizi Süreci

Veri seti AFA varsayımları açısından incelenerek faktör çıkarma yöntemine karar verildikten sonra faktör sayısına karar verilmektedir. Faktör sayısına karar vermek için birden fazla yöntem kullanılmalı ve alanyazında yer alan Paralel analiz, MAP analizi ya da HULL yöntemi gibi objektif yöntemler tercih edilmelidir. Faktör sayısının belirlenmesiyle ilgili daha ayrıntılı bilgiye Kılıç (2022a) tarafından gerçekleştirilen

çalışmadan ulaşılabilir. AFA'da faktör sayısını belirledikten sonra eğer çok boyutlu bir yapı öngörülüyorsa bu durumda hangi döndürme yönteminin tercih edileceği de önemlidir. Alanyazında değişkenler arasında ilişki olmasının daha muhtemel olması ve ilişkisiz yapılar olsa dahi eğik döndürme yöntemlerinin ilişkisiz yapıya da izin vermesi nedeniyle eğik döndürme yöntemlerinin kullanılması önerilmektedir (Cooper, 2019; Costello ve Osborne, 2005; Fabrigar ve diğerleri, 1999; Osborne, 2015). Çok boyutlu yapılarda faktörler arası korelasyonların raporlanması da gelecekte bu ölçeği kullanmayı planlayan araştırmacılar için önemlidir. Faktörler arası korelasyonların yüksek olması ölçekten toplam bir puan alınabileceğini ifade ederken, faktörler arası korelasyonlar düşük ise bu durumda her bir faktör kendi başına kullanılabilir; ancak ölçeğin tamamından toplam bir puan elde edilemez. Bu nedenle ölçek geliştirme ya da uyarlama çalışmalarında faktörler arası korelasyon matrisi de raporlanmalıdır. Faktör analizi gerçekleştirilirken kullanılan korelasyon matrisi de tüm analiz sonuçlarını etkileyeceğinden önemlidir. Değişkenlerin yapısına göre uygun olan korelasyon matrisi tercih edilmelidir. Değişkenler iki kategoriden oluşuyorsa tetrakorik, 3-6 kategoriden oluşuyorsa polikorik yedi ve üzerinde kategoriye sahip ise bu durumda da pearson korelasyon matrisi kullanılabilir (Kılıç, 2022b). Ancak değişkenler kategorik ise öncelikle polikorik korelasyon matrisiyle analiz yapılmalı ancak kestirim yapılamadıysa (bazı durumlarda kestirim yapılamamaktadır) Pearson korelasyon matrisiyle analizlere devam edilebilir.

AFA gerçekleştirildikten sonra analiz sonuçlarının yorumlanmasına geçilir. Bu kısımda faktör yüklerinin en az olabileceği değer belirlenir, birden çok faktöre yüklenen maddeler için ne yapılacağına karar verilir ve açıklanan varyans oranları incelenir. Faktör yüklerinin en az kaç olması gerektiğiyle ilgili alanyazında farklı öneriler mevcuttur. Ancak genel olarak faktör yüklerinin en az 0.30 (Costello ve Osborne, 2005; Hinkin, 1995) ya da 0.32 (Tabachnick ve Fidell, 2019) olması önerilmektedir. Diğer taraftan bir maddenin birden fazla faktöre yüklenmesi de basit yapıların oluşmasını engellediği için istenmemektedir. Bununla ilgili olarak 0.40-0.30-0.20 kuralı (Howard, 2016) kullanılabilir. Buna göre maddelerin kendi faktörlerine en az 0.40, diğer faktörlere en fazla 0.30 ve birden fazla faktöre yüklenen maddelerin faktör yükleri arasında 0.20 fark olması durumunda maddeler ölçekte kalabilir aksi durumda ölçekten çıkarabilir. Faktör yüklerinin yanında faktör başına düşen madde sayısı da faktörün tanımlanabilmesi için önemlidir. Genellikle bir faktörde en az 3 madde olması gerektiği (Brown, 2015; Schönrock-Adema ve diğerleri, 2009) belirtilmektedir. Yapının kapsam geçerliğinin sağlanabilmesi ve güvenilirlik düşünüldüğünde bir faktörün ölçülebilmesi için 3'ten de fazla maddeye ihtiyaç duyulacağı söylenebilir. Son olarak AFA'da

açıklanan varyans oranı da ölçeğin gücü hakkında bilgi vermektedir. Tek boyutlu yapılarda açıklanan varyans oranının en az %30 (Büyüköztürk, 2020) olması gerektiği belirtilse de boyut sayısı arttıkça açıklanan varyans oranı da artacaktır. Ancak bu durumda örneğin beş boyutlu bir ölçeğin toplam varyansın %40'ını açıklamasıyla, tek boyutlu bir ölçeğin toplam varyansın %40'ını açıklaması ölçeklerin sahip olduğu gücünün farklı olduğunu ifade edebilir. Doğrudan boyut sayısına göre önerilen bir açıklanan varyans oranı olmasa da her bir boyutun açıklanan varyansı en az %10 artırması o boyutların önemli olduğuna işaret edebilir (Kılıç, 2022b). Bu nedenle beş boyutlu bir ölçeğin en az %50 civarında varyans açıklaması beklenebilir. Ölçekler seçilirken bu durum da göz önünde bulundurulmalıdır.

AFA sürecinde kullanılan yazılım da bir ölçüde seçilecek teknikleri, yöntemleri ve kararları etkilemektedir. Örneğin, polikorik korelasyon matrisiyle AFA gerçekleştirilecekse SPSS kullanılamamaktadır ya da boyut sayısına karar vermek için paralel analiz sonuçları incelenecekse yine SPSS'te bu analizler bulunmamaktadır. AFA gerçekleştirmek için JASP, JAMOVİ ya da Factor yazılımları kullanılabilir. R programlama dilindeki paketlerle de analizler gerçekleştirilebilir. Factor yazılımında optimal paralel analiz (Timmerman ve Lorenzo-Seva, 2011), HULL yöntemi (Lorenzo-Seva ve diğerleri, 2011), MAP testi (Velicer, 1976) bulunmaktadır. Ayrıca tetrakorik/polikorik korelasyon matrisiyle analize imkân vermektedir. Benzer şekilde R programlama dilinde bulunan psych (Revelle, 2024) paketinde faktör analizi hem Pearson hem polikorik korelasyon matrisiyle gerçekleştirilebilmektedir. Boyut sayısına karar vermek için EFA.MRFA (Navarro-Gonzalez ve Lorenzo-Seva, 2021), EFAtools (Steiner ve Grieder, 2020) ya da EFA.dimensions (O'Connor, 2022) gibi paketler kullanılabilir.

Doğrulayıcı Faktör Analizi (DFA)

Doğrulayıcı faktör analizi, yapısal eşitlik modellemesinin bir bölümünü oluşturmaktadır. Gözlenen değişkenlerin hangi faktöre ait olduğunun bilindiği durumda model-veri uyumunun incelenmesi için doğrulayıcı faktör analizi kullanılmaktadır (Raykov ve Marcoulides, 2006). Diğer bir deyişle ölçülen yapının ne olduğu, hangi maddelerin hangi faktörlerle ilişkili olduğu bilindiği için elde edilen veri seti düşünüldüğünde modele uyum gösterip göstermediği incelenmektedir. DFA gerçekleştirilirken bazı işlem adımları bulunmaktadır: Bunlar; i) betimleme, ii) tanımlama, iii) parametre kestirimi iv) modelin test edilmesi ve v) (gerekliyse) model modifikasyonudur (Byrne, 2016). Uyarlanmış bir ölçeğin seçiminde bu adımlardan; parametre kestirimi, modelin test edilmesi ve model modifikasyonu daha önemli olduğu için bu adımlar üzerinde durulmuştur. Ayrıca DFA varsayımları da AFA'nın varsayımlarıyla aynı olduğu için burada ayrıca yer verilmemiştir. Bu araştırma kapsamında

araştırmacılar tarafından geliştirilen kontrol listesinde varsayımlara ilişkin maddeler yer almaktadır.

DFA süreci

DFA'da model tanımlandıktan sonra parametre kestirimi yapılmaktadır. Parametre kestirimi için farklı kestirim yöntemleri mevcuttur. Değişkenler sürekli ve çok değişkenli normal dağılım varsayımı sağlanmışsa Maximum Likelihood (ML) kestirim yöntemi kullanılabilir. Ancak ölçeklerden elde edilen veriler çok büyük oranda ordinal olmaktadır. Bu nedenle Diagonally Weighted Least Squares (DWLS) ya da benzer şekilde kestirim yapan weighted Least Squares Mean-And-Variance Adjusted (WLSMV) kestirim yöntemi olarak kullanılabilir. Bu yöntem yerine Unweighted Least Squares Mean-And-Variance Adjusted (ULSMV) de tercih edilebilir. Değişkenlerin kategorik olarak ele alınıp kestirim yöntemi olarak da WLSM/ULSMV kullanılmasının daha uygun olacağı alanyazında kabul gören bir görüştür (Kılıç ve Doğan, 2021; Lei, 2009; Li, 2016a, 2016b; Liang ve Yang, 2014; Moshagen ve Musch, 2014; Oranje, 2003). AFA'ya benzer şekilde kategori sayısı iki ise tetrakorik kovaryans matrisi, 3-6 kategori için polikorik kovaryans matrisi ile analizler gerçekleştirilmelidir. Kategori sayısı 7 ve üzerinde olduğunda ise Pearson kovaryans matrisi kullanılabilir.

Kestirimler yapıldıktan sonra modelden elde edilen parametreler ile veriden elde edilen parametreler karşılaştırılarak model-veri uyumu test edilmektedir. Model-veri uyumunun değerlendirilmesinde kullanılabilecek indeksler; χ^2 , χ^2/sd , comparative fit index (CFI), Tucker-Lewis indeksi (TLI ya da non-normed fit index [NNFI]), root mean squared error approximation (RMSEA), standardized root mean square residual (SRMR), goodness of fit (GFI), adjusted goodness of fit (AGFI) gibidir. Bu indeksler haricinde flexible cutoffs (Niemand ve Mai, 2022) ya da dynamic cutoffs (McNeish ve Wolf, 2023) da kullanılabilir. DFA sonucunda model-veri uyumunun değerlendirilmesinde; i) χ^2 , serbestlik derecesi ve p değeri, ii) RMSEA değeri ve bu değer %90 güven aralığı, iii) CFI ve iv) SRMR indekslerinin raporlanması önerilmektedir (Kline, 2016). Bu nedenle bu çalışmadaki kontrol listesine bu indeksler alınmıştır. Bu indeksler için önerilen aralıklar bulunmaktadır. Örneğin, χ^2/sd 'nin 5'ten küçük olması (Anderson ve Gerbing, 1984), TLI ve CFI'nin 0.90'ın üzerinde olması (Hu ve Bentler, 1998, 1999), RMSEA'nın 0.10'dan küçük olması (Maccallum ve diğerleri, 1996) ve SRMR'nin ise 0.08'den küçük olması (Hu ve Bentler, 1999) önerilmektedir.

Model-veri uyumu incelendikten sonra istenilen uyum düzeyine ulaşamadığında model modifikasyonu yapılabilmektedir. Modifikasyonlar ile maddelerin hata varyansları ilişkilendirilmektedir. Ancak bu hata varyanslarının ilişkilendirilmesinin, ilişkilendirilen bu maddelerin farklı bir dış kaynaktan etkilendiğini (Green ve Hershberger,

2000) kabul etmek olduğu göz önünde bulundurulmalıdır. Bu nedenle modifikasyon sayısı modeli bozmayacak kadar olmalıdır. Yazılım tarafından önerilen tüm modifikasyon önerileri gerçekleştirildiğinde model-veri uyumu belli bir seviyeye kadar çıkabilir ancak bu durumda kurulan model ile modifikasyonun yapıldığı model ciddi şekilde farklı olacaktır.

Mevcut çalışma

Araştırmacılar ölçek geliştirme ya da uyarlama süreçlerini işe koştuktan sonra geliştirdikleri/uyarladıkları ölçekleri ilgili dergilerde yayınlamaktadır. Bunun sonucunda ilgilenilen değişkeni ölçmek isteyen diğer araştırmacılar da bu geliştirilmiş/uyarlanmış ölçekleri kullanarak çalışmalarını gerçekleştirmektedir. Ancak bu çalışmalar gerçekleştirilirken ölçekten elde edilen verilerin yapı geçerliğine yönelik analizler çoğunlukla yapılmamaktadır. Dahası analizler gerçekleştirildiğinde ölçeğin yayınlandığı çalışma ile farklılaşabilmektedir. Bu nedenle hangi ölçeğin kullanılacağına karar vermek araştırmacılar için önemli bir dönüm noktasını oluşturmaktadır.

Alanyazındaki benzer çalışmalar incelendiğinde bu çalışmalarda çoğunlukla işlem adımları dizisinin bulunduğu ve ölçek geliştirme/uyarlama çalışmalarının bu işlem adımları ölçüt alınarak incelendiği gözlenmektedir (Acar-Güvendir ve Özer-Özkan, 2015; Çüm ve Koç, 2013; Mor Dirlik, 2014). Bu ölçütlerin yayınlanması alanyazına katkı sunsa da ölçek seçecek araştırmacılara yol göstermekten ziyade daha çok ölçek geliştirecek araştırmacılara ışık tuttuğu söylenebilir. Ancak kullanacağı ölçeği nasıl seçeceği konusunda kafa karışıklığı yaşayan araştırmacılar için yol gösterici bir rehber ihtiyacı olabilir.

Geliştirilen/uyarlanan ölçeğin niteliğinin sayısal olarak ifade edilmesine imkân veren bir kontrol listesinin geliştirilmesiyle, ölçek seçecek araştırmacılara rehberlik edilebilir. Bu kontrol listesinde kullanılacak ağırlıkların eşit olmaması ile kontrol listesinden elde edilen puanlama sayesinde ölçeklere puan verilebilecek ve ölçekler birbiriyle nicel olarak kıyaslanabilecektir. Bu durum ölçülmesi istenilen özellikle ilgili birden fazla ölçme aracının olduğu durumda kullanışlı olacaktır. Bu nedenle bu çalışmanın amacı, ölçek seçiminde kullanılabilecek bir kontrol listesinin geliştirilmesidir. Bu amaç doğrultusunda şu alt problemlere yanıt aranmıştır:

- 1- Geliştirilmiş/uyarlanmış ölçeklerin seçiminde kullanılacak kontrol listesinin maddelerine yönelik uzman görüşleri nasıl dağılmaktadır?
- 2- Geliştirilmiş/Uyarlanmış ölçeklerin seçiminde kullanılacak kontrol listesi maddelerinin uzmanlara göre ağırlıkları nasıldır?

Ölçek seçme kontrol listesinin geliştirilmesi amacıyla gerçekleştirilen bu çalışma ile alanyazına beş açıdan fayda sağlanacağı

düşünülmektedir: i) araştırmacılara kullanacakları ölçeği seçme konusunda karar alırken dikkate alabilecekleri bir ölçme aracı sunulması, ii) ölçek geliştirme/uyarlama çalışmalarına hakemlik yapan araştırmacılar için kaynak oluşturmaları, iii) farklı ölçekler arasından en iyi sonuçları verebilecek ölçeğin seçilmesinin sağlanması, iv) ölçekten elde edilen veriler için yapı geçerliğine yönelik kanıtlar ortaya konulmadığında en azından ölçeğin ne kadar uygun sonuçlar verebileceğiyle ilgili bilgi sunması açısından alanyazına katkı sunacağı düşünülmektedir.

Yöntem

Araştırmanın Modeli

Bu çalışma bir kontrol listesi geliştirme çalışmasıdır. Bu nedenle düzeyine göre betimsel bir çalışma olduğu söylenebilir (Büyüköztürk ve diğerleri, 2020). Kontrol listesi de bir ölçme aracı olacağı için bir ölçüde ölçek geliştirme çalışmasının aşamalarına sahiptir.

Kontrol Listesi Geliştirme Süreci

Kontrol listesinin geliştirilmesi için öncelikle ölçek geliştirme ve uyarlama basamakları incelenmiştir. Daha sonra alanyazındaki diğer ölçek geliştirme süreçlerine odaklanan çalışmalar (Acar-Güvendir ve Özer-Özkan, 2015; Çüm ve Koç, 2013; Mor Dirlik, 2014) ve ölçek geliştirme/uyarlama çalışmalarında kullanılabilecek kaynaklar incelenmiştir (DeVellis, 2017; Erkuş, 2019; Kılıç, 2023b; Uysal, 2023). Buna göre kontrol listesinin maddeleri hazırlanmıştır. Bu doğrultuda kontrol listesi geliştirme süreci de ölçek geliştirme süreci olarak ele alındığında madde yazımının tümevarım yöntemiyle gerçekleştirildiği söylenebilir (Kılıç, 2023b). Kontrol listesinin maddeleri hazırlanırken ölçek geliştirme ve uyarlamayla ilgili alanyazın taraması yapmış olan çalışmalar (Boztunç Öztürk ve diğerleri, 2014; Kılıç ve diğerleri, 2023; Kılıç ve Koyuncu, 2017; Koyuncu ve Kılıç, 2019) incelenmiştir. Böylece denemelik kontrol listesi maddeleri ölçek geliştirme ve uyarlama süreçleri için ayrı ayrı oluşturulmuştur.

Kontrol listesinin maddeleri oluşturulduktan sonra her maddenin ağırlıklarının nasıl olması gerektiğiyle ilgili araştırmacılar tarafından beyin fırtınası ile önerilen ağırlıklar belirlenmiştir. Her bir maddenin ağırlığı belirlenirken ölçek geliştirme/uyarlama sürecine etkisi göz önüne alınmıştır. Örneğin, konu ile ilgili alanyazın taraması yapılması da önemli olmakla birlikte hâlihazırda geliştirilmiş bir ölçeği kullanacak araştırmacılar için elde ettikleri sonuçları daha az etkileyeceği söylenebilir. Ancak ölçeğin yapısı doğru yöntemlerle ortaya konmadığında bu ölçeği, kullanacak araştırmacılar için de geçersiz sonuçlar vereceğinden ölçeğin yapısının doğru yöntemlerle ortaya konması görece olarak daha önemli hale gelmektedir. Bu nedenle ağırlıklar bahsedildiği şekilde belirlenmeye çalışılmıştır.

Kontrol listesindeki maddeler ve ağırlıkları belirlendikten sonra eğitimde ölçme ve değerlendirme alanında en az doktora derecesine sahip 10 alan uzmanına kontrol listeleri gönderilmiştir. Burada hem kontrol listesindeki maddeler ile hem de bu maddelerin ağırlıklarıyla ilgili uzman görüşü alınmıştır.

Ağırlıkların Belirlenmesi

Hâlihazırda geliştirilmiş/uyarlanmış ölçekleri seçerken kullanılabilecek bir kontrol listesinin oluşturulmasında her maddenin aynı ağırlığa sahip olması seçilecek ölçeğin niteliksiz olmasına da neden olabilir. Şöyle ki ölçek geliştirme basamaklarından bir kısmını takip eden bir çalışmada AFA gerçekleştirilirken kullanılması gereken yöntem ve adımlar yanlış bir şekilde takip edilirse bu durumda yapı da yanlış tanımlanmış olacaktır. Çünkü AFA aynı zamanda ölçeğin işe vuruk tanımının yapılması anlamına gelmektedir (Erkuş, 2019; Kılıç ve Uysal, 2023). Bu açıdan bakıldığında, ölçeğin yapısının doğru bir şekilde ortaya konması diğer bazı aşamalardan daha önemli olabilir. Çünkü ölçeği kullanacak araştırmacılar ölçeğin yapısıyla daha çok ilgilenmektedir. Ölçekten toplam puan alınabiliyor mu, ölçek tek boyutlu mu, boyutlar için ayrı puanlar mı hesaplanmalı gibi problemler ölçeği kullanacak araştırmacıları daha çok ilgilendirebilmektedir. Bu noktadan hareketle, araştırmacılara kullanacakları ölçeği seçebilmelerine olanak tanımak için kontrol listesindeki her bir maddeye ağırlık verilmiş, toplam puan üzerinden ölçeğin seçilip seçilmeyeceğine karar verilmesi sağlanmıştır.

Ağırlıklarla ilgili olarak 10 uzmandan görüş alınmıştır. Buna göre araştırmacıların ve 10 uzmanın önerdiği ağırlıklar birlikte değerlendirilmiş ve ağırlıkların ortalaması alınarak her bir ölçütün ağırlıkları elde edilmiştir.

Uzman Görüşlerinin Alınması

Her bir madde için uzmanlardan hem maddenin uygunluğu hem de maddenin ağırlığı açısından görüş alınmıştır. Buna göre uzmanlardan maddeleri, ölçek seçme rehberine uygunluğu açısından 1 (hiç uygun değil) ile 5 (tamamen uygun) arasında değerlendirmeleri istenmiştir. Bu değerlendirmeler ve maddelere yazılan açıklamalar sonucunda ölçek seçme rehberi olarak geliştirilen kontrol listesine son hali verilmiştir (bknz Tablo 2 ve Tablo 3). Ayrıca uzmanlardan araştırmacıların önerdikleri ağırlıkların yanında kendi önerdikleri ağırlıklar da istenmiştir. Ağırlık önerisinde bulunmayan uzmanlar için araştırmacıların önermiş olduğu ağırlıkları uygun buldukları düşünülerek değerlendirme yapılmıştır. Ağırlık önerisinde bulunan araştırmacıların ise kendi ağırlıkları kullanılarak her bir maddenin ağırlığı elde edilmiştir. Böylece kontrol listesine göre ölçek seçecek araştırmacılar belirlenen kesme puanının üzerinde puana sahipse aday ölçeği araştırmada kullanılabilecek olarak değerlendirecektir. Belirlenen

kesme puanının altında puan alan aday ölçekler için alternatifler bakma yoluna gidilebilecektir.

Veri Analizi

Uzmanlar arası uyum, Fleiss'in Kappa katsayısı genellikle düşük değerler verdiği için (Kılıç ve Uysal, 2023) Gwet'in AC1 katsayısı ile değerlendirilmiştir (Wongpakaran ve diğerleri, 2013). Ayrıca her bir madde için uzmanların 1-5 arası puanlamalarına göre dağılımları da sunulmuştur. Gwet'in AC1 katsayısı hesaplanırken R programlama dilinde bulunan *irrCAC* paketi (Gwet, 2022) kullanılmıştır. Ayrıca ağırlık ortalamaları ve her bir maddeye uzmanların verdikleri puanların (1-5 arasında) dağılımları Microsoft Office yazılımı içinde bulunan Excel yazılımıyla elde edilmiştir.

Bulgular

Ölçek seçme rehberinin kontrol listesi olarak geliştirilmesi amacıyla gerçekleştirilen bu çalışmada öncelikle uzmanların maddelere verdikleri puanlara yönelik bulgulara yer verilmiştir. Daha sonra bu maddelerin ağırlıklarına yönelik bulgular sunulmuştur.

Geliştirilmiş/Uyarlanmış Ölçek Seçerken Kullanılacak Kontrol Listesi Maddelerinin Uzmanlara Göre Değerlendirilmesi

Geliştirilmiş veya uyarlanmış bir ölçeğin seçilmesinde kullanılması planlanan maddelerin bu amaç için uygunluğu uzmanlar tarafından değerlendirilmiştir. 10 uzmandan gelen dönütler doğrultusunda elde edilen sonuçlar Tablo 1'de sunulmuştur.

Tablo 1

Uzmanların Kontrol Listesi Maddeleri İçin Görüşlerinin Dağılımı (%)

Maddeler	Geliştirilmiş Ölçeği Seçerken Kullanılacak Maddeler					Uyarlanmış Ölçeği Seçerken Kullanılacak Maddeler				
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
1	0	0	10	10	80	0	0	20	10	70
2	0	0	20	10	70	0	0	20	20	60
3	0	0	10	0	90	0	0	0	10	90
4	0	0	0	20	80	0	0	0	10	90
5	0	0	20	0	80	0	0	0	0	100
6	10	0	0	30	60	10	0	0	30	60
7	0	0	0	20	80	0	0	0	40	60
8	0	0	30	0	70	0	0	0	10	90
9	0	0	30	10	60	0	0	10	10	80
10	0	0	30	10	60	0	10	0	20	70
11	10	0	10	0	80	10	0	0	20	70
12	0	0	10	0	90	0	0	20	20	60
13	0	0	0	10	90	0	0	0	20	80
14	0	0	0	20	80	0	10	10	10	70
15	0	0	0	10	90	0	10	0	0	90

16	0	0	10	10	80	0	0	10	20	70
17	0	0	0	10	90	10	0	0	20	70
18	0	10	20	10	60	0	10	30	0	60
19	0	0	20	0	80	0	0	20	10	70
20	0	0	10	0	90	0	0	10	10	80
21	0	0	20	0	80	0	0	10	10	80
22	0	0	20	0	80	0	0	20	0	80
23	0	0	10	10	80	0	0	20	0	80
24	0	0	10	10	80	0	0	10	10	80
25	0	0	10	20	70	10	0	20	10	60
26	0	0	30	20	50	0	10	10	10	70
27	0	0	10	20	70	0	0	10	0	90
28	0	10	10	10	70	0	0	10	10	80
29	0	0	10	0	90	0	10	10	20	60
30	0	0	10	10	80	0	0	30	0	70
31	0	10	10	10	70	0	0	20	10	70
32	0	0	10	10	80	0	0	20	10	70
33	0	0	10	10	80	0	0	10	20	70
34	0	0	10	10	80	0	0	10	20	70
35	0	0	10	10	80	0	0	20	20	60
36	0	0	20	0	80	0	0	30	20	50
37	0	0	10	10	80	0	0	10	10	80
38	10	0	20	20	50	0	0	10	10	80
39	0	10	10	10	70	0	0	10	0	90
40	0	0	10	10	80	0	0	10	20	70
41	0	0	10	10	80	0	10	10	10	70
42	0	10	20	10	60	0	10	0	30	60
43	0	10	0	10	80	0	20	30	0	50
44	10	10	0	10	70	0	10	0	0	90
45	0	10	0	20	70					
46	0	20	10	20	50					
47	0	0	10	0	90					

*1: Hiç uygun değil, 2: Uygun değil, 3: Kararsızım, 4: Uygun, 5: Tamamen Uygun

Tablo 1 incelendiğinde geliştirilmiş ölçeği seçerken kullanılacak maddelerle ilgili olarak uzmanların büyük kısmı maddelerin uygun ya da tamamen uygun olduğunu belirtmiştir. Maddelere verilen puanlardan 4 ve 5 olanlar dikkate alındığında en düşük %70 olduğu görülmektedir. Buna göre tüm maddeler için en az uzmanların %70'i (n=7) maddenin geliştirilmiş bir ölçeği seçmek için uygun olduğunu belirtmiştir. Bazı maddeler için bu oran %100'e kadar çıkmaktadır. Buna göre geliştirilmiş ölçeğin seçilmesinde kullanılacak kontrol listesi maddelerinin uzmanlara göre yeterli olduğu söylenebilir. Uzmanlar arasındaki uyumu değerlendirmek için hesaplanan Gwet'in AC1 katsayısı geliştirilen ölçeği değerlendiren kontrol listesi için 0.54 olarak elde edilmiştir. Buna göre uzmanlar arasındaki uyumun orta düzeyde olduğu söylenebilir (Wongpakaran ve diğerleri, 2013). Ayrıca maddelere 4 ve 5 puan veren uzmanların ortalaması incelendiğinde geliştirilen

ölçeği seçmek için kullanılacak kontrol listesi maddeleri için 8.55 olduğu gözlenmiştir. Buna göre maddelerin genel olarak uzmanlar tarafından uygun bulunduğu söylenebilir.

Uyarlanmış ölçek seçerken kullanılacak maddeler için de geliştirilmiş ölçek seçerken kullanılacak kontrol listesine benzer şekilde uzmanların büyük kısmı maddelerin uygun ya da tamamen uygun olduğunu belirtmiştir. Maddelere verilen puanlardan 4 ve 5 olanlar dikkate alındığında en düşük %50 olduğu görülmektedir (madde 43). Bazı maddeler için bu oran %100'e kadar çıkmaktadır. Buna göre uyarlanmış ölçeğin seçilmesinde kullanılacak kontrol listesi maddelerinin uzmanlara göre yeterli olduğu söylenebilir. Uzmanlar arasındaki uyumu değerlendirmek için hesaplanan Gwet'in AC1 katsayısı geliştirilen ölçeği değerlendiren kontrol listesi için 0.50 olarak elde edilmiştir. Buna göre uzmanlar arasındaki uyumun orta düzeyde olduğu söylenebilir (Wongpakaran ve diğerleri, 2013). Ayrıca maddelere 4 ve 5 puan veren uzmanların ortalaması incelendiğinde geliştirilen ölçeğin seçileceği kontrol listesi maddeleri için 8.55 olduğu gözlenmiştir. Buna göre maddelerin genel olarak uzmanlar tarafından uygun bulunduğu söylenebilir.

Geliştirilmiş/Uyarlanmış Ölçek Seçerken Kullanılacak Kontrol Listesi Maddelerinin Uzmanlara Göre Ağırlıklarının Belirlenmesi

Geliştirilmiş veya uyarlanmış bir ölçeğin seçilmesinde kullanılması planlanan maddelerin ağırlıklarının belirlenmesi için uzmanların önerileriyle bu araştırmayı gerçekleştiren araştırmacıların (bir kişi olarak ele alınmıştır) ağırlık ortalamaları elde edilmiştir. Bu doğrultuda elde edilen sonuçlar Tablo 2 ve 3'te sunulmuştur.

Tablo 2

Uzmanların Geliştirilmiş Ölçeği Seçerken Kullanılacak Kontrol Listesi İçin Ağırlık Öneri Ortalamaları

No	Madde	Ağırlık
1	Konu ile ilgili alanyazın taraması yapılmıştır.	1
2	Ölçülecek özellik tam olarak tanımlanmıştır.	1
3	Madde havuzu oluşturulma aşamasında alanyazından yararlanılmıştır.	1
4	Madde havuzu oluşturulma aşamasında hedef gruba kompozisyon yazdırılmış ya da grupla görüşmeler yapılmıştır.	1
5	Kapsam geçerliliği kanıtları sunulmuştur (kapsam geçerlik oranı-indeksi vb.).	1
6	Kapsam geçerliliği belirlenirken uzmanlar arası uyum raporlanmıştır (Fleiss'in kappası, Gwet'in AC1- Cohen'nin kappası gibi)	1
7	Madde havuzu oluşturulurken hedeflenen madde sayısının iki ya da üç katı madde yazılmıştır.	1
8	Madde havuzu oluşturulma aşamasında en az iki alan uzmanından yararlanılmıştır.	1

9	Hazırlanan maddelerle ilgili en az iki ölçme uzmanından görüş alınmıştır.	1
10	Hazırlanan maddelerle ilgili en az iki dil uzmanından görüş alınmıştır.	1
11	Madde havuzunda uzmanlardan gelen geribildirimler doğrultusunda gerekli düzeltmeler yapılmıştır.	1
12	Tepki kategorileri (beşli Likert derecelemesi gibi) ile madde ifadeleri uyumludur.	1
13	Her bir madde ile tek bir davranış ölçülmüştür.	1
14	Madde yazımında dilbilgisi ve yazım kurallarına uyulmuştur.	1
15	Maddelerde herkes tarafından farklı anlaşılabilirlik, muğlak ifadeler yer almamaktadır.	1
16	Hedef grubun (örneklem) dışındaki daha küçük bir gruba (15-20 kişi) anlaşılabilirlik açısından değerlendirilmek üzere maddelerin ön deneme uygulaması yapılmıştır.	2
17	Madde havuzunda ön uygulama sonrasında gerekli düzeltmeler yapılmıştır.	2
18	Hedef grup büyüklüğü 400 ve üstü olarak belirlenmiştir. (AFA için)	3
19	Hedef grup büyüklüğü belirlenirken maksimum çeşitliliğin olmasına dikkat edilmiştir.	3
20	Geçerlik çalışmalarında açımlayıcı faktör analizi (AFA) yapılmıştır.	20
21	AFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1
22	AFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1
23	AFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1
24	AFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1
25	AFA'da boyut sayısı belirlenirken objektif yöntemlerden (optimal paralel analiz, HULL yöntemi ve MAP testi) yararlanılmıştır.	6
26	AFA'da kestirim yöntemi olarak temel bileşenler haricindeki bir yöntem kullanılmıştır.	5
27	Ölçeğin AFA sonucunda açıklanan varyans oranı tek boyutlu ise %30'un üzerindedir.	5
28	Ölçeğin AFA sonucunda açıklanan varyans oranı çok boyutlu ise makul orandadır (her bir boyut için %10 ve üzerindedir).	5
29	Ölçeğin AFA sonucunda elde edilen faktör yükleri en az 0.30'un üzerindedir.	5
30	AFA sürecinde binişik (crossloading) maddelerin nasıl çıkarıldığı raporlanmıştır.	6
31	AFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson korelasyon matrisi kullanılmıştır.	5
32	Geçerlik çalışmalarında doğrulayıcı faktör analizi (DFA) yapılmıştır.	13

33	DFA, AFA uygulanan grubun dışında farklı bir örneklem üzerinde yapılmıştır.	5
34	DFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1
35	DFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır	1
36	DFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	2
37	DFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	2
38	DFA'da kestirim yöntemi olarak DWLS, WLSMV veya ULSMV kullanılmıştır.	6
39	DFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson kovaryans matrisi kullanılmıştır.	5
40	DFA sonucunda elde edilen χ^2 , serbestlik derecesi ve p değeri, RMSEA değeri ve bu değer %90 güven aralığı, CFI, TLI ve SRMR indeksleri raporlanmıştır.	5
41	DFA sonucunda elde edilen faktör yükleri 0.30 ve üzerindedir.	5
42	DFA sonucunda uygulanan modifikasyon sayısı 3'ün altındadır.	4
43	Madde analizleri yapılmıştır.	2
44	Ölçüt geçerliğine yönelik kanıtlar sunulmuştur.	2
45	İç tutarlılık güvenirlik belirleme yöntemlerinden en az iki tanesi (Cronbach'ın Alfa ve McDonald'ın Omegası gibi) kullanılmıştır.	5
46	Test tekrar test güvenirliğine yönelik kanıtlar sunulmuştur.	2
47	Ölçeğin uygulama yönergesi hazırlanmıştır.	1
Toplam		150

Tablo 2 incelendiğinde geliştirilmiş bir ölçeği seçmede kullanılabilecek 47 ölçüt görülmektedir. Buna göre araştırmacılar kullanmak istedikleri ölçekleri Tablo 2'de sunulan ölçütler çerçevesinde değerlendirebilir. Kontrol listeleri bir özelliğin varlığı-yokluğuyla ilgili bilgi vermektedir. Buna göre buradaki ölçütlerden ölçek geliştirme çalışmasında raporlananlar için 1, raporlanmayanlar için 0 olarak kodlama yapıldıktan sonra her bir ölçütün önerilen ağırlığı bu kodlarla çarpılarak ölçeğin kullanılıp kullanılmayacağına karar verilebilir. Tablo 2'de görülebildiği üzere 47 maddeden alınabilecek toplam puan 150'dir. Buna göre geliştirilen bir ölçeği seçecek araştırmacılar, buradaki kontrol listesinden elde edilecek puanın 75 ve üzerinde olması durumunda ölçeği kullanmaya karar verebilecekleri söylenebilir.

Örnek bir hesaplama için düşünecek olursak, bir ölçek geliştirme çalışmasında Tablo 2'de sunulan ölçütlerden 1 ile 25 arasındaki ölçütler sağlanmış, diğer ölçütler sağlanmamış olsun. Bu durumda 26 ile 47 arasındaki ölçütlerden 0 puan, 1 ile 25 arasındaki ölçütlerden $1 \times (\text{maddenin ağırlığı}) = 59$ puan elde edilmiş olacaktır. Bu durumda

59<75 olduğu için ölçeğin seçilmemesi yönünde karar verilebilir. Daha ayrıntılı bir sınıflama için şu puan aralıkları kullanılabilir: 0-74 seçilmez, 75-90 kabul edilebilir, 91-116 iyi, 117-150 çok iyi şeklinde sınıflandırılabilir. Buradaki sınıflamada ölçek geliştirme sürecinde AFA ve DFA'nın önemi dikkate alınmış ve bu kategorilerdeki puanlamalara göre sınıflama geliştirilmiştir. Tablo 3'te uyarlanmış ölçeklerin seçiminde kullanılabilecek ölçütlere ve bunların ağırlıklarına yer verilmiştir.

Tablo 3. Uzmanların Uyarlanmış Ölçeği Seçerken Kullanılacak Kontrol Listesi İçin Ağırlık Öneri Ortalamaları

No	Maddeler	Ağırlıklar
1	Konu ile ilgili alanyazın taraması yapılmıştır.	2
2	Ölçülecek özellik tam olarak tanımlanmıştır.	2
3	Ölçeğin uyarlamasının yapılmasına gereklilik duyulduğu ayrıntılı raporlanmıştır.	1
4	Ölçeğin orijinal kültüründeki özelliğin, uyarlandığı kültürdeki özelliklerle eşdeğer olduğu tartışılmıştır.	3
5	Ölçeğin uyarlanması için gerekli izin alınmıştır.	2
6	Ölçeğin uyarlanabilirliğine ilişkin alan uzmanlarından görüş alınmıştır.	1
7	Ölçeğin uyarlanabilirliğine ilişkin ölçme uzmanlarından görüş alınmıştır.	1
8	Ölçeğin orijinal dilden hedef dile çevirisi için her iki kültüre hâkim çevirmenlerden yararlanılmıştır.	2
9	Ölçeğin geri çevirisi için (hedef dilden orijinal dile çevirisi) her iki kültüre hâkim çevirmenlerden yararlanılmıştır.	2
10	Ölçeğin iki formu arasında anlamsal, deneyimsel, kavramsal ve deyimisel açıdan eşitliğin sağlanıp sağlanmadığı kontrol edilmiştir.	3
11	Ölçeğin dil eş değeri sağlanmıştır.	1
12	Tepki kategorileri (beşli Likert derecelemesi gibi) ile madde ifadeleri uyumludur.	1
13	Her bir maddede tek bir davranış ölçülmüştür.	1
14	Maddelerde dilbilgisi ve yazım kurallarına uyulmuştur.	1
15	Maddelerde herkes tarafından farklı anlaşılabilecek, muğlak ifadeler yer almamaktadır.	1
16	Hedef grubun (örneklem) dışındaki daha küçük bir gruba (15-20 kişi) anlaşılabilirlik açısından değerlendirilmek üzere maddelerin ön deneme uygulaması yapılmıştır.	2
17	Ön uygulama sonrasında gerekli düzeltmeler yapılmıştır.	2
18	Hedef grup büyüklüğü 400 ve üstü olarak belirlemiştir. (DFA için)	3
19	Hedef grup büyüklüğü belirlenirken maksimum çeşitliliğin olmasına dikkat edilmiştir.	3
20	Geçerlik çalışmalarına doğrulayıcı faktör analizi (DFA) ile başlanmıştır.	24
21	DFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1

22	DFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır	1
23	DFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1
24	DFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1
25	DFA'da kestirim yöntemi olarak DWLS, WLSMV veya ULSMV kullanılmıştır.	6
26	DFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson kovaryans matrisi kullanılmıştır.	5
27	DFA sonucunda elde edilen χ^2 , serbestlik derecesi ve p değeri, RMSEA değeri ve bu değerlerin %90 güven aralığı, CFI, TLI ve SRMR indeksleri raporlanmıştır.	7
28	DFA sonucunda elde edilen faktör yükleri 0.30 ve üzerindedir.	5
29	DFA sonucunda uygulanan modifikasyon sayısı 3'ün altındadır.	5
30	DFA sonucunda model uyumu sağlanamadıysa faktör yapısı AFA ile kontrol edilmiştir.	13
31	AFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1
32	AFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1
33	AFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1
34	AFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1
35	AFA'da boyut sayısı belirlenirken objektif yöntemlerden (optimal paralel analiz, HULL yöntemi ve MAP testi) yararlanılmıştır.	6
36	AFA'da kestirim yöntemi olarak temel bileşenler haricindeki bir yöntem kullanılmıştır.	5
37	Ölçeğin AFA sonucunda açıklanan varyans oranı tek boyutlu ise %30'un üzerindedir.	5
38	Ölçeğin AFA sonucunda açıklanan varyans oranı çok boyutlu ise makul orandadır (her bir boyut için %10 ve üzerindedir).	5
39	Ölçeğin AFA sonucunda elde edilen faktör yükleri en az 0.30'un üzerindedir.	5
40	AFA sürecinde binişik (crossloading) maddelerin nasıl çıkarıldığı raporlanmıştır.	4
41	AFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson korelasyon matrisi kullanılmıştır.	5
42	İç tutarlılık güvenirlik belirleme yöntemlerinden en az iki tanesi (Cronbach'ın Alfa ve McDonald'ın Omegası gibi) kullanılmıştır.	6

43	Test tekrar test güvenilirliğine yönelik kanıtlar sunulmuştur.	1
44	Ölçeğin uygulama yönergesi hazırlanmıştır.	1
	Toplam	150

Tablo 3 incelendiğinde uyarlanmış ölçekler seçilirken kullanılabilecek 44 ölçüt görülmektedir. Buna göre araştırmacılar kullanmak istedikleri Türk kültürüne uyarlanmış ölçekleri Tablo 3'te sunulan ölçütler çerçevesinde değerlendirebilir. Kontrol listeleri bir özelliğin varlığı-yokluğuyla ilgili bilgi vermektedir. Buna göre buradaki ölçütlerden ölçek uyarlama çalışmasında raporlananlar için 1, raporlanmayanlar için 0 olarak kodlama yapıldıktan sonra her bir ölçütün önerilen ağırlığı bu kodlarla çarpılarak ölçeğin kullanılıp kullanılmayacağına karar verilebilir. Tablo 3'te görülebildiği üzere 44 maddeden alınabilecek toplam puan 150'dir. Buna göre uyarlanan bir ölçeği seçecek araştırmacılar buradaki kontrol listesinden elde edilecek puanın 75 ve üzerinde olması durumunda ölçeği kullanmaya karar verebilecekleri söylenebilir.

Örnek bir hesaplama için düşünecek olursak, bir ölçek geliştirme çalışmasında Tablo 3'te sunulan ölçütlerden 1 ile 25 arasındakiler sağlanmış, diğer ölçütler sağlanmamış olsun. Bu durumda 26 ile 44 arasındaki ölçütlerden 0 puan, 1 ile 25 arasındakilerden $1 \times (\text{maddenin ağırlığı}) = 68$ puan elde edilmiş olacaktır. Bu durumda $68 < 75$ olduğu için ölçeğin seçilmemesi yönünde karar verilebilir. Daha ayrıntılı bir sınıflama için şu puan aralıkları da kullanılabilir: 0-74 seçilmez, 75-90 kabul edilebilir, 91-116 iyi, 117-150 çok iyi şeklinde sınıflandırılabilir. Buradaki sınıflamada ölçek uyarlama sürecinde özellikle DFA'nın önemi dikkate alınmış ve bu kategorilerdeki puanlamalara göre sınıflama geliştirilmiştir. Ölçütlerin en az %50'sinin sağlanmadığı durumda ölçeğin seçilmemesi kararı verilmesi amacıyla 75 puan kesme noktası olarak ele alınmıştır. Kesme noktası belirlenmesi genellikle standart belirleme süreçlerinin işe koşulmasıyla gerçekleştirilmektedir. Bu noktaya çalışmanın sınırlılığında ayrıca değinilmiştir.

Sonuç, Tartışma ve Öneriler

Araştırmacılar alanyazındaki geliştirilmiş/uyarlanmış ölçekleri kullanarak kendi araştırma problemlerine uygun şekilde analizler gerçekleştirmektedir. Ancak ölçek geliştirme/uyarlama çalışmalarını inceleme çalışmaları, alanyazında olsa da bu ölçeklerin nasıl seçileceğiyle ilgili somut bir öneri listesine rastlanmamıştır. Bu nedenle bu çalışmada araştırmacıların pratik olarak kullanabilecekleri bir kontrol listesi geliştirilmesi hedeflenmiştir. Bu hedef doğrultusunda hem geliştirilmiş hem de uyarlanmış ölçekleri seçerken kullanılabilecek iki kontrol listesi geliştirilmiştir.

Geliştirilen ölçekleri seçerken kullanılabilecek kontrol listesi ve ağırlıkları incelendiğinde listenin 47 maddeden oluştuğu ve ölçütlerin asıl ağırlığının AFA'da yoğunlaştığı söylenebilir. Bu durum araştırmacılar tarafından yanlış anlaşılmalıdır. Buradaki amaç, ölçek geliştirme süreçlerini AFA düzeyine indirmek değildir. Ancak hâlihazırda geliştirilmiş ölçekleri seçen araştırmacılar kendi çalışmalarında yapı geçerliğine yönelik kanıtları ortaya koyarken ölçeğin beklenen şekilde çalışmamasından dolayı sorun yaşayabilmektedir. Bazı araştırmacılar ise yapı geçerliğine yönelik kanıtları hiç çalışmalarına eklememekte, sadece ölçek geliştirme çalışmasındaki bulguları raporlamaktadır. Bu durum, elde edilen sonuçların geçerliğine gölge düşürebilmektedir. Bilindiği gibi güvenirlik ve geçerlik ölçme aracının değil, ölçme sonuçlarının bir özelliğidir (American Educational Research Association [AERA] ve diğerleri, 2014). Bu nedenle AFA süreçleri doğru şekilde gerçekleştirilen bir ölçek geliştirme çalışmasından elde edilen ölçek kullanılarak, yeni yapılacak çalışmada da büyük oranda ölçek geliştirme çalışmasına benzer sonuçlar elde edilebilecektir. Bu nedenle geliştirilmiş ölçeğin seçileceği kontrol listesinde AFA süreçlerine daha fazla ağırlık verilmiştir. Diğer taraftan bu önerilerin hâlihazırda geliştirilmiş ölçeklerle ilgili olduğu unutulmamalıdır. Eğer ilgilenilen özelliği ölçmek için geliştirilmiş ölçme aracı tekse, bu durumda araştırmacıya da çok fazla alternatif kalmamaktadır. Bu durumda yine mevcut çalışmada önerilen kontrol listesini kullanarak hâlihazırda var olan ölçeğin kullanılıp kullanılmamasına karar verilebilir. Kullanmama yönünde karar verilirse, yeni bir ölçek geliştirme yolu da tercih edilebilir. AFA yapılmayan/yapılamayan durumlarda (örneğin projektif kişilik testleri ya da faktör analizi yapmaya yetecek kadar verinin toplanamayacağı durumlarda) ölçüt geçerliği çalışmaları yapılmış olabilir. Mevcut çalışmada önerilen kontrol listesi ile bu ölçme araçlarının tercih edilmemesi yönünde karar alınabilir. Ancak bu durum ölçeğin niteliğinin düşük olduğunu yansıtmayabilir. Bazı durumlarda ölçüt geçerliği kullanılabilir ve daha anlamlı olabilir. Bu şekilde geliştirilen ölçeği kullanacak araştırmacıların ölçüt geçerliği için (ilgilenmeseler bile) benzer ölçeklerle ekstra veri toplaması gerektiği göz önünde bulundurulmalıdır. Bu çalışmada ele alınan kontrol listesi bir ölçeğe ulaşıp bunu kullanan daha sonra kendi örnekleminde bu ölçeğin çalışmadığını fark eden araştırmacıları hedeflemektedir. Çünkü projektif kişilik testleri gibi testleri uygulayabilmek için uzmanlık da gerekmektedir. Bu uzmanlığa sahip olmayan araştırmacılar bu testleri uygulayamamaktadır. Bu nedenle mevcut çalışmada, yayınlanan ölçekleri kullanarak veri toplayacak araştırmacı grubu hedeflenmektedir. Bu açıdan bakıldığında AFA'ya daha fazla ağırlık verilmesinin anlaşılabilir olduğu düşünülmektedir.

Uyarılama çalışmaları düşünüldüğünde orijinal kültürdeki yapının hedef kültürde de benzer olup olmadığı incelendiğinden (Kılıç, 2023a) DFA

süreçlerine daha fazla ağırlık verilmiştir. Benzer şekilde bu durum uyarılama çalışmalarını DFA düzeyine indirme amacı taşımamaktadır. Hâlihazırda uyarılama çalışması yapılmış ve yayınlanmış ölçeklerde, izlenmesi gereken süreçlerin doğru şekilde uygulanmadığı veya raporlanmadığı durumlar olabilir. Bu nedenle uyarılanmış ölçekleri seçen araştırmacılara bir fikir vermesi ve kendi çalışmalarında da tekrar edilebilir yapılar elde edebilmeleri amacıyla DFA süreçlerine daha fazla ağırlık verilmiştir. Çünkü bu süreçlerin işe koşulmadığı, sadece dilsel geçerlik, kapsam geçerliği ve ölçüt geçerliği çalışmaları yapıldığında uyarılan bu ölçeği kullanan araştırmacıların geçerlik için kullanabileceği tek kanıt (eğer ölçüt geçerliği için birkaç ölçek daha kullanmadılarsa) faktör analizi olacaktır. Uyarılan ölçeği kullanan araştırmacılar, çoğunlukla bu ölçekle birlikte başka bir ya da birkaç ölçek ile de veri toplamaktadır. Bu durumda ölçüt geçerliği için de veri toplandığında dört ya da beş ölçekten veri toplanması anlamına gelecektir. Bu durum, hem yorgunluk etkisi nedeniyle güvenirliliğin düşmesine neden olacak, hem de kullanışlı olmayacaktır. Bu gibi durumlar göz önüne alınarak faktör analizi yöntemlerine daha fazla ağırlık verilmiştir. Böylece uyarılan ölçeği kullanacak araştırmacıların da benzer sonuçları alma ihtimali artırılmaya çalışılmıştır. Buradaki ağırlıklara göre geçerlik kanıtlarının önem düzeyi hakkında yorum yapmak yanlış olacaktır. Çünkü geliştirme/uyarlama çalışmalarında imkanlar ölçüsünde tüm geçerlik türlerine yönelik kanıtların toplanması önerilmektedir (AERA ve diğerleri, 2014). Mevcut çalışmanın bakış açısı ise ölçek seçerken önerilen ölçütlere dikkat ederek araştırmacının kendi örnekleminde sürprizle karşılaşmasının önüne geçmektir.

Mevcut çalışmanın odak noktası, araştırmacılar ölçek seçerken onlara rehberlik etmek olsa da aynı zamanda ölçek geliştirme/uyarlama çalışmalarına hakemlik yapacak araştırmacılara da geliştirilen kontrol listesi yol gösterici olabilir. Bunun yanında ölçek geliştirme/uyarlama çalışması gerçekleştirmek isteyen araştırmacılar da burada sunulan kontrol listesine göre çalışmalarını planlayabilir. Bu durumda ölçütlerin ağırlıklarından ziyade ölçütlerin kendileriyle ilgilenmek daha doğru olacaktır. Mevcut çalışmanın, ölçeği kullanacak, geliştirecek, uyarlayacak ve çalışmalara hakemlik yapacak araştırmacılara yol gösterici olacağı söylenebilir. Gelecek çalışmalarda geliştirilmiş/uyarlanmış ölçeklerin burada önerilen kontrol listesinden aldıkları puanlar incelenebilir. Yeterli sayıda çalışma kontrol listesiyle incelenirse faktör analiziyle kontrol listesinin yapısı incelenebileceği gibi ölçütlerin ağırlıkları istatistiksel yöntemlerle de belirlenebilir.

Sınırlılıklar

Bu çalışmanın her çalışmada olduğu gibi birtakım sınırlılıkları bulunmaktadır. Çalışma, katılım gösteren 10 ölçme ve değerlendirme

uzmanının görüşleriyle sınırlıdır. Ayrıca, kontrol listesinde önerilen bilgiler güncel olmakla birlikte, gelecekte alanyazındaki gelişmelere bağlı olarak değişiklik gösterebilir. Örneğin “DFA’da kestirim yöntemi olarak WLSMV veya ULSMV” kullanılması önerisinde olduğu gibi. Gelecek çalışmalarda daha iyi yöntemler geliştirilmesi durumunda bu ölçüt, güncelliğini yitirme potansiyeline sahiptir. Bu durum çalışmanın diğer bir sınırlığı olarak ele alınabilir.

Çalışmanın bir diğer sınırlığı olarak düşünülebilecek durum standart belirleme çalışmasının yapılmamış olmasıdır. Ancak standart belirleme çalışmalarının amacı ölçülen özelliğe göre bir kesme puanı belirlemektir (Çetin, 2023). Bu konuda, ölçülen özelliğin asgari düzeyine sahip olup olunmadığının belirlenmesine yönelik bir çalışma yapılması gereklidir. Ancak ölçek geliştirme basamaklarından hangilerinin asgari olduğuyla ilgili uzman görüşü almak çok mantıklı olmayabilir. Çünkü ölçek geliştirmenin basamakları zaten bellidir ve ölçek geliştirme çalışması yapacak araştırmacıların da bu basamakları takip etmesi beklenmektedir. Uzmanlar da bu basamakların tümünün gerekli olduğu yönünde görüşler gelmesi muhtemeldir. Bu nedenle bu çalışmada standart belirleme yöntemleri kullanılarak bir standart belirlenmemiştir. Ölçek geliştirme basamaklarının tamamının doğru bir şekilde işe koşulması gerekse de alanyazında bu konuda problemler olduğu görüldüğünden geliştirilmiş ölçeklerin seçiminde araştırmacılara yol gösterici olması açısından bu çalışma gerçekleştirilmiştir. Önerilen kesme puanları da öneri mahiyetinde olup kesin sınırları ifade etmemektedir. AFA ya da DFA’nın doğru şekilde yapılmasına daha fazla ağırlık verilmesinin nedeni ise bu ölçekleri kullanacak araştırmacıların kendi örneklemelerinde de geçerli sonuçlar elde etme ihtimallerini artırma gayretidir. Bu noktadan hareketle kesme noktası olarak 75 puan belirlenmiştir.

Etik Kurul İzin Bilgisi: *Bu araştırma, Trakya Üniversitesi Sosyal ve Beşeri Bilimler Araştırmaları Etik Kurulu'nun 29.05.2024 tarihli toplantısında alınan 05/09 numaralı kararı ile alınan izinle yürütülmüştür.*

Yazar Çıkar Çatışması Bilgisi: *Yazarların beyan edeceği bir çıkar çatışması yoktur.*

Yazar Katkısı: *MAG: Çalışmanın tasarlanması, giriş kısmının yazılması ve verilerin toplanması, AFK: Çalışmasının tasarlanması, giriş kısmının yazılması, verilerin toplanması ve analizi, YÖÖ: Giriş ve tartışma kısmının yazılması.*

Teşekkür: *Araştırmada ölçek geliştirme ve uyarlama adımlarına yönelik görüşlerini aldığımız ve her bir adım için uygun gördükleri ağırlıkları vermesini istediğimiz uzmanlar; Dr. Başak Erdem Kara, Dr.*

Durmuş Özbaşı, Dr. Ergül Demir, Dr. Funda Nalbantoğlu Yılmaz, Dr. Gül Güler, Dr. Hatice Kumantaş Öztürk, Dr. İlhan Koyuncu, Dr. Levent Ertuna, Dr. Mustafa İlhan, Dr. Özge Altıntaş, Dr. Seher Yalçın ve Dr. Sema Sulak'a verdikleri önemli katkılarından dolayı teşekkür ederiz.

Kaynakça

- Acar Güvendir, M. ve Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Electronic Journal of Social Sciences*, 14(52), 23-33. <https://doi.org/10.17755/esosder.54872>
- Acar Güvendir, M. ve Özer Özkan, Y. (2022). *Tüm yönleriyle ölçek geliştirme süreci*. Pegem Akademi.
- Alper, A. ve Gülbahar, Y. (2009). Trends and issues in educational technologies: A Review of recent research in TOJET. *The Turkish Online Journal of Educational Technology- TOJET*, 8(2), 124-135.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173. <https://doi.org/10.1007/BF02294170>
- Arrindell, W. A., & Van der Ende, J. (1985). Cross-sample invariance of the structure of self-reported distress and difficulty in assertiveness: Experiences with the scale for interpersonal behaviour. *Advances in Behaviour Research and Therapy*, 7(4), 205-243. [https://doi.org/10.1016/0146-6402\(85\)90013-x](https://doi.org/10.1016/0146-6402(85)90013-x)
- Aybek, E. C. ve Toraman, C. (2022). How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory. *International Journal of Assessment Tools in Education*, 9(2), 534-547. <https://doi.org/10.21449/ijate.1132931>
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), 77-85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal Of Exercise Science*, 8(3), 297.
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In: Joreskog, K.G., Wold, H. (Eds.), *Systems under indirect observation: causality, structure, prediction* (pp. 149-173). Elsevier.
- Boztunç Öztürk, N., Eroğlu, M.G. ve Kelecioğlu, H. (2014, June 9-13). Eğitim bilimleri alanında yapılan ölçek uyarlama makalelerinin incelenmesi, IV.

Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Ankara.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford.

Bryman, A., & Cramer, D. (2001). *Quantitative data analysis with SPSS Release 10 for Windows: A guide for social scientists*. Routledge.
<https://doi.org/10.4324/9780203471548>

Büyüköztürk, Ş. (2020). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (28. baskı). Pegem Akademi.

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2020). *Bilimsel araştırma yöntemleri* (29. baskı). Pegem Akademi.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.
<https://doi.org/10.4324/9781315757421>

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25-44.
<https://doi.org/10.1080/19312458.2017.1396583>

Cattell, R. B. (2012). *The scientific use of factor analysis in behavioral and life sciences*, Springer Science & Business Media.

Chen, L. T., & Liu, L. (2020). Methods to analyze likert-type data in educational technology research. *Journal of Educational Technology Development and Exchange (JETDE)*, 13(2), 39-60. <https://doi.org/10.18785/jetde.1302.04>

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. L. NJ Erlbaum.

Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (2. baskı). Pegem Akademi.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (Second Edition). Lawrence Erlbaum Associate

Cooper, C. (2019). *Psychological testing: Theory and practice*. Routledge.
<https://doi.org/10.4324/9781315391342>

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 27-29.
<https://doi.org/10.7275/jyj1-4868>

Çetin, S. (2023). Standart belirleme. İçinde A. F. Kılıç (Ed.), *R programlama diliyle A'dan Z'ye ölçek geliştirme* (ss. 257-289). Nobel.

Çüm, S. ve Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.

- De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147-181. <https://doi.org/10.1080/00273170902794206>
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.
- Doğan, H. ve Tok, T. N. (2018). Türkiye'de eğitim bilimleri alanında yayınlanan makalelerin incelenmesi: Eğitim ve Bilim Dergisi örneği. *Current Research in Education*, 4(2), 94-109.
- Erdem, D. (2011). Türkiye'de 2005–2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Journal of Measurement and Evaluation in Education and Psychology*, 2(1), 140-147.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar, *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2019). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (4nd ed.). Pegem Akademi.
- Everitt, B. S. (1975). Multivariate analysis: the need for data and other problems. *British Journal of Psychiatry*, 126, 237-240. <https://doi.org/10.1192/bjp.126.3.237>
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University. <https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Sage.
- Goodwin, L. D. (1999). The role of factor analysis in the estimation of construct validity. *Measurement in Physical Education and Exercise Science*, 3(2), 85-100. https://doi.org/10.1207/s15327841mpee0302_2
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510-3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Gorsuch, R. L. (1974). *Factor analysis*. Saunders.
- Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(2), 251-270. https://doi.org/10.1207/S15328007SEM0702_6
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265–275. <https://doi.org/10.1037/0033-2909.103.2.265>

- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19, 149-161. <https://doi.org/10.1007/bf02289162>
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Pearson.
- Helm, J. L., Castro-Schilo, L., & Oravecz, Z. (2017). Bayesian versus maximum likelihood estimation of multitrait-multimethod confirmatory factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 17-30. <https://doi.org/10.1080/10705511.2016.1236261>
- Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393-416. <https://doi.org/10.1177/0013164405282485>
- Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988. [https://doi.org/10.1016/0149-2063\(95\)90050-0](https://doi.org/10.1016/0149-2063(95)90050-0)
- Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Chapman & Hall.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51-62. <https://doi.org/10.1080/10447318.2015.1087664>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 37-41. <https://doi.org/10.1080/10705519909540118>
- İşçi, S. (2013). *Türkiye’de eğitim yönetimi alanında yapılmış lisansüstü tezlerin tematik, metodolojik ve istatistiksel açıdan incelenmesi* [Yüksek lisans tezi, Eskişehir Osmangazi Üniversitesi]. Ulusal Tez Merkezi. Erişim No: 344302
- Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, 8, 205-223. https://doi.org/10.1207/s15328007sem0802_3
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403. <https://doi.org/10.9734/BJAST/2015/14975>
- Karadağ, E. (2009). *Türkiye’de eğitim bilimleri alanında yapılmış doktora tezlerinin tematik ve metodolojik açıdan incelenmesi; Bir durum çalışması*. [Doktora tezi, Marmara Üniversitesi]. Ulusal Tez Merkezi. Erişim No: 250896

- Kaiser, H.F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401-415.
<https://doi.org/10.1007/bf02291817>
- Kalkbrenner, M. T. (2021). Alpha, Omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, 1-12.
<https://doi.org/10.1080/21501378.2021.1940118>
- Kılıç, A. F. (2022a). Açıklayıcı faktör analizinde boyut sayısına karar verme: Yöntemlere kısa bir bakış. *Pamukkale University Journal of Social Sciences Institute*, 51 (Özel sayı 1), Ö305-Ö318.
<https://doi.org/10.30794/pausbed.1095936>
- Kılıç, A. F. (2022b). Ölçek geliştirme sürecinde açıklayıcı faktör analizi. İçinde M. Acar-Güvendir & Y. Özer-Özkan (Ed.), *Tüm yönleriyle ölçek geliştirme süreci* (ss. 69-129). Pegem Akademi.
- Kılıç, A. F. (2023a). Geçerlik ve kanıtları. *R programlama diliyle A'dan Z'ye ölçek uyarlama* (ss. 61-113) içinde. Nobel.
- Kılıç, A. F. (2023b). Ölçek geliştirme süreci. *R programlama diliyle A'dan Z'ye ölçek geliştirme* (33-59) içinde. Nobel.
- Kılıç, A. F. ve Doğan, N. (2021). Comparison of confirmatory factor analysis estimation methods on mixed-format data. *International Journal of Assessment Tools in Education*, 8(1), 21-37.
<https://doi.org/10.21449/ijate.782351>
- Kılıç, A. F. ve Koyuncu, İ. (2017). Ölçek uyarlama çalışmalarının yapı geçerliği açısından incelenmesi. *Küreselleşen dünyada eğitim* (ss. 1202-1205) içinde. Pegem Akademi.
- Kılıç, A. F., Koyuncu, İ. ve Uysal, İ. (2023). Scale development based on item response theory: A systematic review. *International Journal of Psychology and Educational Studies*, 10(1), 209-223.
<https://doi.org/10.52380/ijpes.2023.10.1.982>
- Kılıç, A. F. ve Uysal, İ. (2023). Ölçek puanlarının geçerliği. *R programlama diliyle A'dan Z'ye ölçek geliştirme* (95-150) içinde. Nobel.
- Kline, R. B. (2016). *Principle and practice of structural equation modeling* (4th ed.). The Guilford.
- Kline, P. (1994). *An easy guide to factor analysis*. Routledge.
- Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In *Applied quantitative analysis in the social sciences* (pp. 171- 207). Routledge.
- Koyuncu, İ. ve Kılıç, A. F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Eğitim ve Bilim*, 44(198), 361-388.
<https://doi.org/10.15390/EB.2019.7665>
- Lei, P. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, 43(3), 495-507.
<https://doi.org/10.1007/s11135-007-9133-z>
- Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares.

- Behavior Research Methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21(3), 369-387. <https://doi.org/10.1037/met0000093>
- Liang, X., & Yang, Y. (2014). An evaluation of WLSMV and Bayesian methods for confirmatory factor analysis with categorical indicators. *International Journal of Quantitative Research in Education*, 2(1), 17-38. <https://doi.org/10.1504/IJQRE.2014.060972>
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364. <https://doi.org/10.1080/00273171.2011.564527>
- Maccallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling of fit involving a particular measure of model. *Psychological Methods*, 13(2), 130-149. <https://doi.org/10.1037/1082-989X.1.2.130>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61-88. <https://doi.org/10.1037/met0000425>
- 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.
- Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(3), 1-20. <https://doi.org/10.1186/s41155-016-0057-1>
- Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, 10(2), 60-70. <https://doi.org/10.1027/1614-2241/a000068>
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5, 159-168. https://doi.org/10.1207/s15327574ijt0502_4
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599-620. https://doi.org/10.1207/s15328007sem0904_8
- Navarro-Gonzalez, D., & Lorenzo-Seva, U. (2021). *EFA.MRFA: Dimensionality assessment using minimum rank factor analysis*. <https://cran.r-project.org/package=EFA.MRFA>
- Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate*

- Behavioral Research*, 39(3), 439-478. https://doi.org/10.1207/s15327906mbr3903_3
- Niemand, T., & Mai, R. (2022). Flexible cutoffs for model fit evaluation in covariance-based structural models. *Journal of the Academy of Marketing Science*, 46, 1148-1172. <https://doi.org/10.1007/s11747-018-0602-9>
- Nunnally, J.C. (1978). *Psychometric theory*. McGraw-Hill.
- O'Connor, B. P. (2022). *EFA.dimensions: Exploratory factor analysis functions for assessing dimensionality*. <https://cran.r-project.org/package=EFA.dimensions>
- Oranje, A. (2003). *Comparison of estimation methods in factor analysis with categorized variables: Applications to NEAP data*.
- Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment Research & Evaluation*, 20(2), 1-7. <https://doi.org/10.7275/hb2g-m060>
- Osborne, J.W., & Costello, A.B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, 9(11), 1-9. <https://doi.org/10.7275/ktzq-jq66>
- Ozan, C. ve Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, 4(1), 116-136. <https://doi.org/10.19126/suje.76547>
- Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics*, 32, 153-161. <https://doi.org/10.1023/a:1015210025234>
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates.
- Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* (2.2.9) [Software]. Northwestern University. <https://cran.r-project.org/package=psych>
- Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research*, 20(4), 235-249. <https://doi.org/10.1002/mpr.352>
- Sapnas, K. G., & Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement*, 10, 135-154. <https://doi.org/10.1891/jnum.10.2.135.52552>
- Schönrock-Adema, J., Heijne-Penninga, M., van Hell, E. A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher*, 31(6), 226-232. <https://doi.org/10.1080/01421590802516756>
- Seçer, İ., Ay, İ., Ozan, C. ve Yılmaz, B. Y. (2014). Research Trends in the Field of Guidance and Psychological Counseling: A Content Analysis. *Turkish Psychological Counseling & Guidance Journal*, 5(41), 49-60.

- Shevlin, M. & Miles, J. N. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual differences*, 25(1), 85-90. [https://doi.org/10.1016/s0191-8869\(98\)00055-5](https://doi.org/10.1016/s0191-8869(98)00055-5)
- Sönmez, V. ve Alacapınar, G. (2016). *Sosyal bilimlerde ölçme aracı hazırlama*. Anı Yayıncılık.
- Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53), 2521. <https://doi.org/10.21105/joss.02521>
- Steven, J. (1996). *Applied multivariate statistics for social sciences* (third edition). Lawrence Erlbaum Associates, Inc.
- Şeker, H. ve Gençdoğan, B. (2006). *Psikolojide ve eğitimde ölçme aracı geliştirme*. Nobel.
- Şencan, H. (2005). Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik. Seçkin Yayıncılık.
- Şenyurt, S. ve Özer Özkan, Y. (2017). Eğitimde ölçme ve değerlendirme alanında yapılan yüksek lisans tezlerinin tematik ve metodolojik açıdan incelenmesi. *İlköğretim Online*, 16(2), 628-653.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. <https://doi.org/10.1037/a0023353>
- Tanujaya, B., Prahmana, R. C. I., & Mumu, J. (2022). Likert scale in social sciences research: Problems and difficulties. *FWU Journal of Social Sciences*, 16(4), 89-101.
- Tavşancıl, E., Güler, G. ve Ayan, C. (2014). 2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreci açısından incelenmesi. In IV. *Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı)* Hacettepe Üniversitesi, Ankara.
- Tosuntaş, Ş. B., Emirtekin, E. ve Süral, İ. (2019). Eğitim ve öğretim teknolojileri konusunda yapılan tezlerin incelenmesi 2013-2018. *Yükseköğretim ve Bilim Dergisi*, (2), 277-286.
- Uyumaz, G. ve Sırgancı, G. (2020). Doğrulayıcı faktör analizi için gerekli örneklem büyüklüğü kaç kişidir?: Bayes yaklaşımı ve maksimum olabilirlik kestirimi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 16(32), 5302-5340. <https://doi.org/10.26466/opus.826895>
- Uysal, İ. (2023). Ölçek uyarılama aşamaları. *R Programlama Diliyle A'dan Z'ye Ölçek Uyarılama* (ss. 1-9) içinde. Nobel.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327. <https://doi.org/10.1007/BF02293557>

- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61. <https://doi.org/10.1186/1471-2288-13-61>
- Yu, C.-Y., & Muthén, B. O. (2002, April 4). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. In *American Educational Research Association Annual Conference*. New Orleans, LA.
- Zygmunt, C., & Smith, M. R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology*, 10(1), 40-55. <https://doi.org/10.20982/tqmp.10.1.p040>

Ek A. Geliştirilmiş Ölçeği Seçerken Kullanılacak Kontrol Listesi

No	Madde	Ağırlık	Evet	Hayır
1	Konu ile ilgili alanyazın taraması yapılmıştır.	1		
2	Ölçülecek özellik tam olarak tanımlanmıştır.	1		
3	Madde havuzu oluşturulma aşamasında alanyazından yararlanılmıştır.	1		
4	Madde havuzu oluşturulma aşamasında hedef gruba kompozisyon yazdırılmış ya da grupla görüşmeler yapılmıştır.	1		
5	Kapsam geçerliliği kanıtları sunulmuştur (kapsam geçerlik oranı-indeksi vb.).	1		
6	Kapsam geçerliliği belirlenirken uzmanlar arası uyum raporlanmıştır (Fleiss'in kappası, Gwet'in AC1- Cohen'nin kappası gibi)	1		
7	Madde havuzu oluşturulurken hedeflenen madde sayısının iki ya da üç katı madde yazılmıştır.	1		
8	Madde havuzu oluşturulma aşamasında en az iki alan uzmanından yararlanılmıştır.	1		
9	Hazırlanan maddelerle ilgili en az iki ölçme uzmanından görüş alınmıştır.	1		
10	Hazırlanan maddelerle ilgili en az iki dil uzmanından görüş alınmıştır.	1		
11	Madde havuzunda uzmanlardan gelen geribildirimler doğrultusunda gerekli düzeltmeler yapılmıştır.	1		
12	Tepki kategorileri (beşli Likert derecelemesi gibi) ile madde ifadeleri uyumludur.	1		
13	Her bir madde ile tek bir davranış ölçülmüştür.	1		
14	Madde yazımında dilbilgisi ve yazım kurallarına uyulmuştur.	1		
15	Maddelerde herkes tarafından farklı anlaşılacak, muğlak ifadeler yer almamaktadır.	1		
16	Hedef grubun (örneklem) dışındaki daha küçük bir gruba (15-20 kişi) anlaşılabilirlik açısından değerlendirilmek üzere maddelerin ön deneme uygulaması yapılmıştır.	2		
17	Madde havuzunda ön uygulama sonrasında gerekli düzeltmeler yapılmıştır.	2		
18	Hedef grup büyüklüğü 400 ve üstü olarak belirlenmiştir. (AFA için)	3		
19	Hedef grup büyüklüğü belirlenirken maksimum çeşitliliğin olmasına dikkat edilmiştir.	3		

20	Geçerlik çalışmalarında açımlayıcı faktör analizi (AFA) yapılmıştır.	20		
21	AFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1		
22	AFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1		
23	AFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1		
24	AFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1		
25	AFA'da boyut sayısı belirlenirken objektif yöntemlerden (optimal paralel analiz, HULL yöntemi ve MAP testi) yararlanılmıştır.	6		
26	AFA'da kestirim yöntemi olarak temel bileşenler haricindeki bir yöntem kullanılmıştır.	5		
27	Ölçeğin AFA sonucunda açıklanan varyans oranı tek boyutlu ise %30'un üzerindedir.	5		
28	Ölçeğin AFA sonucunda açıklanan varyans oranı çok boyutlu ise makul orandadır (her bir boyut için %10 ve üzerindedir).	5		
29	Ölçeğin AFA sonucunda elde edilen faktör yükleri en az 0.30'un üzerindedir.	5		
30	AFA sürecinde binişik (crossloading) maddelerin nasıl çıkarıldığı raporlanmıştır.	6		
31	AFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson korelasyon matrisi kullanılmıştır.	5		
32	Geçerlik çalışmalarında doğrulayıcı faktör analizi (DFA) yapılmıştır.	13		
33	DFA, AFA uygulanan grubun dışında farklı bir örneklem üzerinde yapılmıştır.	5		
34	DFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1		
35	DFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1		
36	DFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	2		
37	DFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	2		
38	DFA'da kestirim yöntemi olarak DWLS, WLSMV veya ULSMV kullanılmıştır.	6		
39	DFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson kovaryans matrisi kullanılmıştır.	5		
40	DFA sonucunda elde edilen χ^2 , serbestlik derecesi ve p değeri, RMSEA değeri ve bu değer için %90 güven aralığı, CFI, TLI ve SRMR indeksleri raporlanmıştır.	5		
41	DFA sonucunda elde edilen faktör yükleri 0.30 ve üzerindedir.	5		
42	DFA sonucunda uygulanan modifikasyon sayısı 3'ün altındadır.	4		
43	Madde analizleri yapılmıştır.	2		
44	Ölçüt geçerliğine yönelik kanıtlar sunulmuştur.	2		
45	İç tutarlılık güvenirlik belirleme yöntemlerinden en az iki tanesi (Cronbach'ın Alfa ve McDonald'ın Omegası gibi) kullanılmıştır.	5		
46	Test tekrar test güvenirliğine yönelik kanıtlar sunulmuştur.	2		

47	Ölçeğin uygulama yönergesi hazırlanmıştır.	1		
Toplam		150		

*Evet işaretlenen satırlar 1 olarak ele alınıp ağırlıklar ile çarpılarak toplam puan elde edilebilir. Örneğin 47. Maddeye evet işaretlenmiş ise 1x1 = 1 puan, 46. Maddeye 1 işaretlenmiş ise 1x2 = 2 puan elde edilecektir. Her madde için elde edilen puanlar toplanarak toplam puan elde edilebilir.

Ek B: Uyarlanmış Ölçeği Seçerken Kullanılacak Kontrol Listesi

No	Madde	Ağırlık	Evet	Hayır
1	Konu ile ilgili alanyazın taraması yapılmıştır.	2		
2	Ölçülecek özellik tam olarak tanımlanmıştır.	2		
3	Ölçeğin uyarlanmasının yapılmasına gereklilik duyulduğu ayrıntılı raporlanmıştır.	1		
4	Ölçeğin orijinal kültüründeki özelliğin, uyarlandığı kültürdeki özellikle eşdeğer olduğu tartışılmıştır.	3		
5	Ölçeğin uyarlanması için gerekli izin alınmıştır.	2		
6	Ölçeğin uyarlanabilirliğine ilişkin alan uzmanlarından görüş alınmıştır.	1		
7	Ölçeğin uyarlanabilirliğine ilişkin ölçme uzmanlarından görüş alınmıştır.	1		
8	Ölçeğin orijinal dilden hedef dile çevirisi için her iki kültüre hâkim çevirmenlerden yararlanılmıştır.	2		
9	Ölçeğin geri çevirisi için (hedef dilden orijinal dile çevirisi) her iki kültüre hâkim çevirmenlerden yararlanılmıştır.	2		
10	Ölçeğin iki formu arasında anlamsal, deneyimsel, kavramsal ve deyimse açıdan eşitliğin sağlanıp sağlanmadığı kontrol edilmiştir.	3		
11	Ölçeğin dil eş değeri sağlanmıştır.	1		
12	Tepki kategorileri (beşli Likert derecelemesi gibi) ile madde ifadeleri uyumludur.	1		
13	Her bir maddede tek bir davranış ölçülmüştür.	1		
14	Maddelerde dilbilgisi ve yazım kurallarına uyulmuştur.	1		
15	Maddelerde herkes tarafından farklı anlaşılabilir, muğlak ifadeler yer almamaktadır.	1		
16	Hedef grubun (örneklem) dışındaki daha küçük bir gruba (15-20 kişi) anlaşılabilirlik açısından değerlendirilmek üzere maddelerin ön deneme uygulaması yapılmıştır.	2		
17	Ön uygulama sonrasında gerekli düzeltmeler yapılmıştır.	2		
18	Hedef grup büyüklüğü 400 ve üstü olarak belirlenmiştir. (DFA için)	3		
19	Hedef grup büyüklüğü belirlenirken maksimum çeşitliliğin olmasına dikkat edilmiştir.	3		
20	Geçerlik çalışmalarına doğrulayıcı faktör analizi (DFA) ile başlanmıştır.	24		
21	DFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1		
22	DFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1		
23	DFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1		
24	DFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1		
25	DFA'da kestirim yöntemi olarak DWLS, WLSMV veya ULMSV kullanılmıştır.	6		
26	DFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson kovaryans matrisi kullanılmıştır.	5		
27	DFA sonucunda elde edilen χ^2 , serbestlik derecesi ve p değeri, RMSEA değeri ve bu değerlerin %90 güven aralığı, CFI, TLI ve SRMR indeksleri raporlanmıştır.	7		
28	DFA sonucunda elde edilen faktör yükleri 0.30 ve üzerindedir.	5		
29	DFA sonucunda uygulanan modifikasyon sayısı 3'ün altındadır.	5		
30	DFA sonucunda model uyumu sağlanamadıysa faktör yapısı AFA ile kontrol edilmiştir.	13		

31	AFA'nın varsayımlarından kayıp verilerin kontrolü yapılmıştır.	1		
32	AFA'nın varsayımlarından çok değişkenli uç değer kontrolü yapılmıştır.	1		
33	AFA'nın varsayımlarından çok değişkenli normallik test edilmiştir.	1		
34	AFA'nın varsayımlarından çoklu doğrusal bağlantı test edilmiştir.	1		
35	AFA'da boyut sayısı belirlenirken objektif yöntemlerden (optimal paralel analiz, HULL yöntemi ve MAP testi) yararlanılmıştır.	6		
36	AFA'da kestirim yöntemi olarak temel bileşenler haricindeki bir yöntem kullanılmıştır.	5		
37	Ölçeğin AFA sonucunda açıklanan varyans oranı tek boyutlu ise %30'un üzerindedir.	5		
38	Ölçeğin AFA sonucunda açıklanan varyans oranı çok boyutlu ise makul orandadır (herbir boyut için %10 ve üzerindedir).	5		
39	Ölçeğin AFA sonucunda elde edilen faktör yükleri en az 0.30'un üzerindedir.	5		
40	AFA sürecinde binişik (crossloading) maddelerin nasıl çıkarıldığı raporlanmıştır.	4		
41	AFA sürecinde değişkenler iki kategoriden oluşuyorsa (1-0) tetrakorik, 3-6 kategoriden oluşuyorsa polikorik, 7 ve üzerinde kategoriden oluşuyorsa Pearson korelasyon matrisi kullanılmıştır.	5		
42	İç tutarlılık güvenirlik belirleme yöntemlerinden en az iki tanesi (Cronbach'ın Alfa ve McDonald'ın Omegası gibi) kullanılmıştır.	6		
43	Test tekrar test güvenirliğine yönelik kanıtlar sunulmuştur.	1		
44	Ölçeğin uygulama yönergesi hazırlanmıştır.	1		
Toplam		150		

*Evet işaretlenen satırlar 1 olarak ele alınıp ağırlıklar ile çarpılarak toplam puan elde edilebilir. Örneğin 44. Maddeye evet işaretlenmiş ise $1 \times 1 = 1$ puan, 42. Maddeye 1 işaretlenmiş ise $1 \times 6 = 6$ puan elde edilecektir. Her madde için elde edilen puanlar toplanarak toplam puan elde edilebilir.



Pursuit for Scale Selection: The Journey to Finding the Right Scale

Meltem ACAR GÜVENDİR¹, Abdullah Faruk KILIÇ², Yeşim ÖZER
ÖZKAN³

Abstract

The aim of this study is to develop a checklist to guide researchers in selecting scales that have already been developed or adapted. For this purpose, scale development and adaptation steps were prepared based on the relevant literature, and the weights of these steps were defined. The prepared steps and their respective weights were sent to 10 academics holding doctoral degrees in the field of measurement and evaluation in education. The academics were asked to provide feedback on the steps and assign weights they deemed appropriate. Based on the feedback received, adjustments were made, and the weights of the scale development and adaptation steps were finalized. A criterion was then established by considering the total score obtainable based on the weights of the steps. Using the proposed criterion, a reference point was provided to assist researchers in determining whether to select a given developed or adapted scale. In this respect, the study is expected to make significant contributions by serving as a guide for researchers in the selection of developed and adapted scales.

Article Details

Research Article

Received

14/06/2024

Accepted

17/04/2025

Key words

Scale
development,
Scale
adaptation,
Scale selection,
Validity,
Reliability,
Checklist

¹ Trakya University, 0000-0002-3847-0724, meltemacar@trakya.edu.tr

² Trakya University, 0000-0003-3129-1763, afarukkilic@trakya.edu.tr

³ Gaziantep University, 0000-0002-7712-658X, yesimozkan@gantep.edu.tr

Suggested Citation:

Acar Güvendir, M., Kılıç, A. F., & Özer Özkan, Y. (2025). Pursuit for scale selection: The journey to finding the right scale. *Pamukkale University Journal of Education [PUJE]*, 64, 338-373. <https://doi.org/10.9779/pauefd.1501330>

Introduction

At the core of educational processes lies the human element, and the successful evaluation of these processes often relies on abstract concepts. Crucial attributes such as students' learning skills, motivation levels, problem-solving abilities, and critical thinking capacities cannot be directly observed. At this point, scales serve as a vital tool. Scales structure these abstract characteristics and express them quantitatively, making them measurable and comprehensible. In doing so, they enable objective analysis of critical aspects such as the effectiveness of educational processes, student performance, and teacher efficacy. Scales are among the most frequently used psychometric tools in education and social science research (Aybek & Toraman, 2022; Bistop & Herron, 2015; Chen & Liu, 2020; Joshi et al., 2015; Tanujaya et al., 2022). Studies examining thematic and methodological aspects of educational research in Türkiye (Alper & Gülbahar, 2009; Doğan & Tok, 2018; Erdem, 2011; İşçi, 2013; Karadağ, 2009; Ozan & Köse, 2014; Seçer et al., 2014; Şenyurt & Özer Özkan, 2017; Tosuntaş et al., 2019) indicate that scales are the most commonly used data collection tools.

In recent years, influenced in part by the pressure to publish, there has been a noticeable increase in scale development and adaptation studies conducted by researchers. However, this trend has also led to the emergence of numerous errors and deficiencies in such studies (Erkuş, 2007). The growing prevalence of these issues in published articles has drawn the attention of researchers, prompting various studies aimed at identifying and addressing these shortcomings (Acar Güvendir & Özer Özkan, 2015; Boztunç Öztürk et al., 2014; Carpenter, 2018; Chen & Liu, 2020; Çüm & Koç, 2013; Morgado et al., 2017; Tavsancıl et al., 2014). The mere existence of such studies underscores the significant challenges encountered in scale development processes.

Among the identified issues, notable deficiencies include the lack of functional definitions in scale development processes, the absence or failure to report pilot testing, and the insufficient information provided regarding item analyses. Additionally, while exploratory factor analysis (EFA) is commonly used, the limited use of confirmatory factor analysis (CFA) stands out as a significant gap (Acar Güvendir & Özer Özkan, 2015; Çüm & Koç, 2013). These issues indicate a lack of standards and guidelines in scale development studies. Therefore, it is crucial to establish clear, concise, and standardized guidelines for the process. This would help ensure that the scales used yield more reliable results.

It can be claimed that the scales obtained from studies conducted by considering the steps of scale development will be sufficient. Although different researchers propose various suggestions regarding the steps of scale development, they are fundamentally similar. According to these steps, scale development begins with the definition of the

characteristic being measured. Following this, items are written in line with the definition. After expert opinions are sought on the items, pilot and trial applications are conducted. Validity and reliability analyses are performed with the obtained data, and then the scale is standardized (Morgado et al., 2017). Since the purpose of this study is not to introduce the scale development process, a brief overview of the steps has been provided. Researchers interested in more detailed explanations on scale development can refer to the relevant literature (Acar Güvendir & Özer Özkan, 2022; Erkuş, 2019; Kılıç, 2023).

Definition of the Measured Characteristic

This process typically begins with a literature review (Kalkbrenner, 2021). When defining the measured characteristic, it is essential to distinguish the construct from both similar and distinct constructs (Furr, 2011). In other words, the definition should highlight both the comprehensive nature of the construct and its differentiating aspects from other variables (Erkuş, 2019). At this stage, in addition to the literature review, expert opinions may also be consulted. After the construct is defined, items should be written to measure this construct.

Item Writing Process

Writing a sufficient number of items that comprehensively cover the relevant theoretical construct is a critical phase in the scale development process. Upon completing this stage, the researcher will have created an item pool. The next step involves determining whether the items in the pool truly measure the intended construct and, in other words, evaluating their “validity” in serving the intended purpose. This phase includes gathering evidence related to the scale’s content validity. Content validity evidence is typically obtained through expert reviews in the literature. Expert review, in this context, refers to having another specialist, besides the item writer, assess the process.

This process can be described as reviewing and revising the items before applying the scale to the target group. At this stage, scale development can be seen as a team effort requiring collaboration among experts from different fields: measurement and evaluation, language, and the specific subject area. Experts evaluate each item individually and provide suggestions. These feedbacks play a crucial role in the scale’s content validity and are considered a critical step for ensuring the scale’s reliability and validity.

Validity and Reliability Analyses

Validity and reliability analyses are processes that involve conducting EFA on data obtained from a pilot study to identify the structure of the scale, followed by CFA using a new dataset to validate the identified structure, and performing reliability analyses, such as internal consistency (e.g., Cronbach’s alpha) and test-retest methods. At this

point, a crucial question that researchers must address is the appropriate sample size for the analysis.

Sample Size

One of the most frequently questioned aspects of the scale development process is the number of participants on which the scale should be developed. First and foremost, it should be remembered that the scale development process is not merely about writing items and administering them to individuals while following validity and reliability procedures. Every step of the process must be carried out meticulously from start to finish. Therefore, the number of participants required for the development of the scale will vary depending on the construct being measured, the nature of the data collected, and the statistical techniques to be used.

The purpose of this is to determine whether the items you have developed measure the intended construct, check whether the items are clear, concise, and understandable, and decide on the average response time for the scale. The pilot test should be conducted on 10-15 individuals who represent the extreme groups (the range of the relevant construct/variable) from the target population of the main sample. This is typically sufficient (Erkuş, 2019). This process should be carried out face-to-face and, if possible, individually, as the researcher needs to ask questions, observe participants' behaviors, and take notes.

Regarding the number of participants required for the main sample, which aims to determine the psychometric properties of the scale, various criteria have been proposed. For example, Bryman and Cramer (2001), Everitt (1975), Hair et al. (1998), Ho (2006); MacCallum et al. (1999), Stevens (1996), Sönmez and Alacapınar (2016); Şencan (2005), and suggest that sample size should be determined based on the number of items in the scale. According to them, the sample size should be at least five, ten, or twenty times the number of items in the scale. However, current literature indicates that using multiples of the number of items as a criterion for sample size is not an accurate measure (Erkuş, 2019; Goretzko et al., 2021).

Arrindell and Van der Ende (1985), DeVellis (2017), and Kline (1994) have emphasized the importance of factor number in determining sample size. According to the scholars, a critical consideration in this process is ensuring that the number of factors is clear and that the factor loadings are large. MacCallum et al. (1999), Guttman (1954), and Kaiser (1970) argue that as the sample size increases, sampling error decreases. They suggest that when communalities are high (0.60 or above), the sample size may be relatively small; however, when communalities are low, a larger sample size is required.

In sample selection, Cattell (1978), Comrey and Lee (1992), Gorsuch (1974), Kline (1994), Nunnally (1978), Rouquette and Falissard (2011), Tabachnick and Fidell (2019) have each recommended different sample sizes (200, 300, 500), providing various suggestions but not a definitive number. DeVellis (2017), Henson and Roberts (2006), and Osborne and Costello (2004) emphasized that large sample sizes are crucial for obtaining better results in factor analyses used during the scale development process. However, in contrast to large sample sizes, Sapnas and Zeller (2002), De Winter et al. (2009), Mundfrom et al. (2005), Preacher and MacCallum (2002), and Stevens (1996) argued that in certain situations, smaller sample sizes could also be sufficient. According to Erkuş (2019) and Şeker and Gençdoğan (2006), the key factor is not whether the sample is large or small, but rather whether the selected sample adequately represents the construct being measured by the developed scale.

In the scale development process, some researchers argue that a single criterion is not sufficient to determine the sample size for factor analysis, which aims to reveal the factor structure. Instead, they suggest that multiple criteria should be considered (Çokluk et al., 2014; Erkuş, 2019; Guadagnoli & Velicer, 1988).

Regarding the sample size required for CFA, which is conducted to validate the factor structure of a scale on a different group, various opinions have been presented (Anderson & Gerbing, 1984; Boomsma, 1982; 1985; Cohen & Cohen, 1983; De Winter et al., 2009; Goodwin, 1999; Guadagnoli & Velicer, 1988; Helm et al., 2017; Hu & Bentler, 1999; Jackson, 2001; MacCallum et al., 1999; Mundfrom et al., 2005; Muthén & Muthén, 2002; Nevitt & Hancock, 2004; Osborne & Costello, 2004; Rouquette & Falissard, 2011; Shevlin & Miles, 1998; Stevens, 1996; Uyumaz & Sirgancı, 2020; Yu & Muthén, 2002). However, these opinions also vary within themselves.

The recommendations regarding the appropriate sample size for scale development processes vary significantly. While the number of participants per item is important, the key factor is the representativeness of the selected individuals concerning the psychological construct being measured. Additionally, since the sample size may differ depending on whether the construct being measured is unidimensional or multidimensional, suggesting a fixed number may be invalid (Erkuş, 2019). Furthermore, as it is not possible to know in advance whether the obtained factor loadings and communalities will be sufficient, relying solely on a fixed number without considering other critical variables can be misleading. Nevertheless, many researchers entering the data collection phase will still have questions about the appropriate sample size. In this context, it can be suggested that, taking into account the use of factor analysis, examining the dataset in terms

of the assumptions of this analysis, and considering the recommendations in the current literature, aiming for at least a sample size of 400 participants would be more beneficial for obtaining reliable results (Goretzko et al., 2021).

The Characteristics of the Sample

In this process, the sample selected should consist of individuals who can best represent the range of the relevant characteristic. When determining the sample, not only the quality but also the variance is of great importance. The scale and items should be applied to individuals who represent all the characteristics to be measured so that the scale can distinguish differences among individuals (DeVellis, 2017; Erkuş, 2019). Ultimately, it would be appropriate to achieve heterogeneity within homogeneity for the selected group. For example, in a situation where the group to be developed consists of teachers, all individuals being teachers represents homogeneity, while selecting teachers from all subject areas represents heterogeneity.

In the scale development process, a random selection of the sample would not be appropriate, as it contradicts the principle of voluntariness. If purposive selection is made, a group can be reached where the structure of the scale is believed to be revealed.

Construct Validity

Factor analytic methods are frequently used in scale development or adaptation processes. Factor analysis is employed to reveal the measured construct or to assess whether the data obtained aligns with the construct that is intended to be measured (Nunnally, 1978). Factor analysis can be approached in two ways: exploratory and confirmatory.

Exploratory Factor Analysis (EFA)

In the scale development process, EFA plays a crucial role and is used to uncover the underlying latent structure (Fabrigar & Wegener, 2012). EFA is a method used to identify a smaller number of latent variables (constructs/factors) that explain the correlations among observed variables (Goretzko et al., 2021). In the scale development process, the steps of EFA should be followed correctly, and the methods used at each stage should be reported. First, the data set should be examined for compliance with analysis assumptions. Then, a factor extraction method should be selected. The number of factors should be determined using different methods, and the analysis should be conducted. The findings obtained from the analysis should be reported appropriately. The steps of EFA are outlined as follows.

Examination of the Data Set in Terms of Analysis Assumptions

Since EFA is a multivariate statistical method, it has multivariate analysis assumptions. Therefore, the data set to be analyzed should be

examined for multivariate normality, multivariate outliers, multicollinearity, missing values, and linear relationships between variables. In addition to these, Bartlett's test of sphericity (Bartlett, 1950) and the Kaiser-Meyer-Olkin (KMO) value are also used to assess the suitability of the data set for factor analysis.

Multivariate normality is often not met in categorical data. Therefore, factor extraction methods that provide more robust estimates against violations of the multivariate normality assumption, such as the principal axis factor extraction method (Costello & Osborne, 2005; Fabrigar et al., 1999) or the minimum residual method (Zygmunt & Smith, 2014), can be used.

The dataset should be examined for multivariate outliers, and any erroneous data entries should be corrected. If outliers are present in the dataset, they can either be removed before analysis, or analyses can be conducted both with and without the outliers (Kılıç, 2022b). The dataset should also be examined for multicollinearity, and whether there is multicollinearity among the variables should be reported. Since multicollinearity can affect the results of the EFA, it is important to examine this assumption and report its findings. Similarly, the dataset should be examined for missing values, and any actions taken to handle missing values should be reported. In addition to these assumptions, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) value can also be reported. While these are not critically important for EFA, it is advisable to report them. The sample size for conducting EFA is also important. In addition to a large sample size, it is crucial that the sample includes individuals from all levels of the measured characteristic. Since detailed information on sample size was provided earlier, it will not be reiterated here.

Exploratory Factor Analysis Process

After examining the data set in terms of the assumptions of EFA and deciding on the factor extraction method, the next step is to determine the number of factors. Multiple methods should be used to decide on the number of factors, and objective methods such as Parallel Analysis, MAP analysis, or the HULL method, which are described in the literature, should be preferred. More detailed information on determining the number of factors can be found in the study conducted by Kılıç (2022a). After determining the number of factors in EFA, if a multidimensional structure is expected, it is important to choose a rotation method. The literature suggests that oblique rotation methods should be used because they allow for relationships between variables, and even if unrelated structures are present, oblique methods accommodate these as well (Cooper, 2019; Costello & Osborne, 2005; Fabrigar et al., 1999; Osborne, 2015). In multidimensional structures, reporting the correlations between factors is also important for future

researchers planning to use the scale. High correlations between factors imply that a total score can be obtained from the scale, whereas low correlations indicate that each factor can be used separately, but a total score cannot be derived from the whole scale. Therefore, in scale development or adaptation studies, the correlation matrix between factors should also be reported. The correlation matrix used during factor analysis is important as it will affect all analysis results. An appropriate correlation matrix should be selected based on the nature of the variables. If the variables are dichotomous, a tetrachoric correlation matrix should be used; if they have 3 to 6 categories, a polychoric correlation matrix should be used; and if the variables have 7 or more categories, a Pearson correlation matrix can be used (Kılıç, 2022b). However, if the variables are categorical, the analysis should initially be performed using the polychoric correlation matrix. If estimation cannot be performed (in some cases, estimation is not possible), the analysis can continue using the Pearson correlation matrix.

After conducting EFA, the results are interpreted. In this phase, the minimum value for factor loadings is determined, decisions are made regarding what to do with items that load on multiple factors, and the explained variance ratios are examined. There are different recommendations in the literature regarding the minimum factor loading. However, in general, it is suggested that factor loadings should be at least 0.30 (Costello & Osborne, 2005; Hinkin, 1995) or 0.32 (Tabachnick & Fidell, 2019). On the other hand, it is undesirable for an item to load on multiple factors, as this would prevent the formation of simple structures. To address this, the 0.40-0.30-0.20 rule (Howard, 2016) can be applied. According to this rule, items can remain in the scale if they load at least 0.40 on their primary factor, at most 0.30 on other factors, and if the difference in factor loadings for items that load on multiple factors is at least 0.20. Otherwise, these items should be removed from the scale. In addition to factor loadings, the number of items per factor is also important for defining the factor. Generally, it is stated that there should be at least 3 items in a factor (Brown, 2015; Schönrock-Adema et al., 2009). To ensure content validity and consider reliability, it can be said that more than 3 items are needed to measure a factor. Finally, the explained variance ratio in EFA also provides information about the scale's strength. It is suggested that the explained variance ratio should be at least 30% for unidimensional structures (Büyüköztürk, 2020). However, as the number of dimensions increases, the explained variance ratio will also increase. In this case, for example, the explanation of 40% of the total variance in a five-dimensional scale may indicate a different strength compared to a unidimensional scale explaining 40% of the total variance. Although there is no direct recommended explained variance ratio based on the

number of dimensions, an increase of at least 10% in the explained variance for each dimension may indicate the importance of those dimensions (Kılıç, 2022b). Therefore, a five-dimensional scale is expected to explain at least 50% of the variance. This should also be considered when selecting scales.

The software used in the EFA process can also influence the techniques, methods, and decisions to be selected. For example, if EFA is to be conducted using a polychoric correlation matrix, SPSS cannot be used, or if parallel analysis results are to be examined to determine the number of dimensions, these analyses are not available in SPSS. JASP, JAMOVl, or Factor software can be used to perform EFA. Analyses can also be conducted using R programming language packages. In Factor software, optimal parallel analysis (Timmerman & Lorenzo-Seva, 2011), the HULL method (Lorenzo-Seva et al., 2011), and the MAP test (Velicer, 1976) are available. Additionally, it allows analysis with a tetrachoric/polychoric correlation matrix. Similarly, the psych package in R programming (Revelle, 2024) can perform factor analysis using both Pearson and polychoric correlation matrices. For determining the number of dimensions, packages such as EFA.MRFA (Navarro-Gonzalez & Lorenzo-Seva, 2021), EFAtools (Steiner & Grieder, 2020), or EFA.dimensions (O'Connor, 2022) can be used.

Confirmatory Factor Analysis (CFA)

Confirmatory factor analysis is a part of structural equation modeling. CFA is used to examine model-data fit when the factors to which the observed variables belong are known (Raykov & Marcoulides, 2006). In other words, since the structure being measured and the items' relationships with specific factors are already known, the data set is examined to determine whether it fits the model. There are several steps involved in conducting CFA: these include i) description, ii) specification, iii) parameter estimation, iv) model testing, and v) (if necessary) model modification (Byrne, 2016). In the selection of an adapted scale, the steps of parameter estimation, model testing, and model modification are considered more important, so these steps are emphasized here. Additionally, since CFA assumptions are the same as those in EFA, they are not separately addressed here. In this study, the checklist developed by the researchers includes items related to the assumptions.

CFA Process

After the model is specified in CFA, parameter estimation is performed. There are different estimation methods available for parameter estimation. If the variables are continuous and the assumption of multivariate normal distribution is met, the Maximum Likelihood (ML) estimation method can be used. However, the data obtained from

scales are predominantly ordinal. Therefore, the Diagonally Weighted Least Squares (DWLS) or the Weighted Least Squares Mean-And-Variance Adjusted (WLSMV) estimation method, which performs similarly, can be used. Alternatively, the Unweighted Least Squares Mean-And-Variance Adjusted (ULSMV) method may also be preferred. It is widely accepted in the literature that considering the variables as categorical and using WLSM/ULSMV as the estimation method is more appropriate (Kılıç & Doğan, 2021; Lei, 2009; Li, 2016a, 2016b; Liang & Yang, 2014; Moshagen & Musch, 2014; Oranje, 2003). Similar to EFA, when the number of categories is two, the tetrachoric covariance matrix should be used, while for 3-6 categories, the polychoric covariance matrix should be used. For seven or more categories, the Pearson covariance matrix can be used.

After parameter estimation, the model parameters obtained from the model are compared with the parameters derived from the data to test model-data fit. Indices that can be used in evaluating model-data fit include χ^2 , χ^2/df , Comparative Fit Index (CFI), Tucker-Lewis Index (TLI, or Non-Normed Fit Index [NNFI]), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI). In addition to these indices, flexible cutoffs (Niemand & Mai, 2022) or dynamic cutoffs (McNeish & Wolf, 2023) can also be used. When evaluating model-data fit in CFA, it is recommended to report: i) χ^2 , degrees of freedom, and p-value, ii) RMSEA and its 90% confidence interval, iii) CFI, and iv) SRMR indices (Kline, 2016). Therefore, these indices have been included in the checklist for this study. There are recommended ranges for these indices. For example, it is suggested that χ^2/df should be less than 5 (Anderson & Gerbing, 1984), TLI and CFI should be above 0.90 (Hu & Bentler, 1998, 1999), RMSEA should be less than 0.10 (Maccallum et al., 1996), and SRMR should be less than 0.08 (Hu & Bentler, 1999).

After evaluating the model-data fit, if the desired level of fit is not achieved, model modification can be performed. Modifications involve correlating the error variances of items. However, it should be noted that correlating these error variances implies that the items are influenced by a different external source (Green & Hershberger, 2000). Therefore, the number of modifications should be limited to avoid overfitting the model. When all the modifications suggested by the software are implemented, the model-data fit may improve to a certain extent, but in this case, the model established and the modified model may be significantly different.

The Current Study

After researchers implement the process of scale development or adaptation, they publish the scales they have developed/adapted in

relevant journals. As a result, other researchers who wish to measure the variable of interest conduct their studies using these developed/adapted scales. However, during these studies, analyses related to the construct validity of the data obtained from the scale are often not conducted. Moreover, when analyses are performed, the results may differ from the study in which the scale was originally published. Therefore, deciding which scale to use represents an important turning point for researchers.

When reviewing similar studies in the literature, it is observed that these studies often contain a sequence of procedural steps, with scale development/ adaptation studies being examined based on these procedural criteria (Acar-Güvendir & Özer-Özkan, 2015; Çüm & Koç, 2013; Mor Dirlik, 2014). While the publication of these criteria contributes to the literature, it is more likely to illuminate researchers developing scales rather than guide those selecting a scale. However, researchers who are confused about how to choose the scale they will use may need a guiding reference.

By developing a checklist that enables the numerical expression of the quality of the developed/adapted scale, researchers selecting a scale can be guided. With this checklist, the weights used will not be equal, and by scoring the checklist, scales can be rated, and comparisons between scales can be made quantitatively. This would be particularly useful when multiple measurement tools exist for the feature to be measured. Therefore, the purpose of this study is to develop a checklist that can be used in scale selection. In this regard, the following sub-problems were addressed:

- 1- How are expert opinions distributed regarding the items of the checklist to be used in the selection of developed/adapted scales?
- 2- What are the weights of the items in the checklist for the selection of developed/adapted scales according to experts?

It is believed that this study, conducted with the aim of developing a scale selection checklist, will contribute to the literature in five ways: i) providing researchers with a measurement tool to consider when making decisions about which scale to use, ii) creating a resource for researchers who serve as reviewers for scale development/adaptation studies, iii) ensuring the selection of the scale that yields the best results among different scales, and iv) offering insight into how suitable a scale might be in providing appropriate results when construct validity evidence is not provided for the data obtained from the scale, thus contributing to the literature.

Method

Research Design

This study is a scale development checklist study. Therefore, it can be considered as a descriptive study in terms of its level (Büyüköztürk et al., 2020). Since the checklist is a measurement tool, it also involves certain stages of scale development.

Checklist Development Process

To develop the checklist, the stages of scale development and adaptation were first examined. Then, studies focused on other scale development processes in the literature (Acar-Güvendir & Özer-Özkan, 2015; Çüm & Koç, 2013; Mor Dirlik, 2014) and resources that could be used in scale development/adaptation studies were reviewed (DeVellis, 2017; Erkuş, 2019; Kılıç, 2023b; Uysal, 2023). Based on this, the items of the checklist were prepared. In this context, when the checklist development process is considered as part of the scale development process, it can be said that the item writing was carried out using the inductive method (Kılıç, 2023b). The studies that have conducted literature reviews on scale development and adaptation (Boztunç Öztürk et al., 2014; Kılıç et al., 2023; Kılıç & Koyuncu, 2017; Koyuncu & Kılıç, 2019) were examined. Thus, the trial checklist items were created separately for scale development and adaptation processes.

After the items of the checklist were created, the weights for each item were determined through brainstorming by the researchers. When determining the weight of each item, its impact on the scale development/adaptation process was taken into account. For instance, conducting a literature review on the subject is important, but it could be said that it would have less impact on researchers who will use an already developed scale. However, when the structure of the scale is not determined using correct methods, it can lead to invalid results for researchers using the scale. Therefore, determining the structure of the scale using proper methods becomes relatively more important. Hence, the weights were determined in the way described.

After the items and weights of the checklist were determined, the checklist was sent to 10 field experts with at least a doctoral degree in the field of measurement and evaluation. Experts' opinions were gathered regarding both the items in the checklist and their corresponding weights.

In this section, it can be stated who the study group of the research consists of.

Determining Weights

If each item in a checklist to select developed/adapted scales has the same weight, it could result in a scale selection that lacks quality.

Specifically, if methods and steps that should be followed in an EFA during a study that follows some stages of scale development are incorrectly applied, the structure of the scale will also be incorrectly defined. Because EFA is essentially a means of defining the scale's structure (Erkuş, 2019; Kılıç & Uysal, 2023), from this perspective, determining the structure of the scale correctly may be more important than some other stages. This is because researchers using the scale are more concerned with its structure, such as whether the scale provides a total score, whether it is unidimensional, and whether separate scores should be calculated for the factors. Based on this consideration, weights were assigned to each item in the checklist, allowing researchers to make decisions about whether the scale should be selected based on the total score.

Experts' opinions were gathered on the weights. The weights suggested by the researchers and the 10 experts were evaluated together, and the average of the weights was calculated to obtain the weights for each criterion.

Obtaining Expert Opinions

For each item, the experts were asked to provide feedback on both its suitability and its weight. The experts were asked to rate the items on a scale of 1 (not suitable at all) to 5 (completely suitable) with regard to their relevance to the scale selection guide. Based on these evaluations and the comments provided on the items, the final version of the checklist, developed as the scale selection guide, was produced (see Table 2 and Table 3). Additionally, experts were asked to provide their own suggested weights alongside those proposed by the researchers. For experts who did not provide weight suggestions, the weights proposed by the researchers were assumed to be appropriate and used in the evaluation. For those who did provide weight suggestions, their own proposed weights were used to determine the final weight for each item. In this way, researchers selecting a scale using the checklist will assess the candidate scale as usable in the study if it scores above the defined cutoff score. If a candidate scale scores below the cutoff, alternative options may be considered.

Data Analysis

Inter-rater agreement was assessed using Gwet's AC1 coefficient, as Fleiss' Kappa tends to yield low values (Kılıç & Uysal, 2023) (Wongpakaran et al., 2013). In addition, the distribution of expert ratings (1-5 scale) for each item was presented. Gwet's AC1 coefficient was calculated using the *irrCAC* package in the R programming language (Gwet, 2022). The average weights and the distribution of experts' ratings (1-5) for each item were calculated using Microsoft Excel software.

Findings

In this study, conducted with the aim of developing a scale selection guide as a checklist, the findings related to the ratings given by the experts to the items are presented first. Subsequently, the findings regarding the weights of these items are provided.

Evaluation of the Items in the Checklist for Selecting Developed/Adapted Scales According to Experts

The suitability of the items planned for use in selecting a developed or adapted scale was evaluated by experts. The results obtained based on the feedback from 10 experts are presented in Table 1.

Table 1

Distribution of Experts' Opinions on the Control List Items (%)

Items	Items to be Used When Selecting a Developed Scale					Items to be Used When Selecting an Adapted Scale				
	(%)1	(%)2	(%)3	(%)4	(%)5	(%)1	(%)2	(%)3	(%)4	(%)5
1	0	0	10	10	80	0	0	20	10	70
2	0	0	20	10	70	0	0	20	20	60
3	0	0	10	0	90	0	0	0	10	90
4	0	0	0	20	80	0	0	0	10	90
5	0	0	20	0	80	0	0	0	0	100
6	10	0	0	30	60	10	0	0	30	60
7	0	0	0	20	80	0	0	0	40	60
8	0	0	30	0	70	0	0	0	10	90
9	0	0	30	10	60	0	0	10	10	80
10	0	0	30	10	60	0	10	0	20	70
11	10	0	10	0	80	10	0	0	20	70
12	0	0	10	0	90	0	0	20	20	60
13	0	0	0	10	90	0	0	0	20	80
14	0	0	0	20	80	0	10	10	10	70
15	0	0	0	10	90	0	10	0	0	90
16	0	0	10	10	80	0	0	10	20	70
17	0	0	0	10	90	10	0	0	20	70
18	0	10	20	10	60	0	10	30	0	60
19	0	0	20	0	80	0	0	20	10	70
20	0	0	10	0	90	0	0	10	10	80
21	0	0	20	0	80	0	0	10	10	80
22	0	0	20	0	80	0	0	20	0	80
23	0	0	10	10	80	0	0	20	0	80
24	0	0	10	10	80	0	0	10	10	80
25	0	0	10	20	70	10	0	20	10	60
26	0	0	30	20	50	0	10	10	10	70
27	0	0	10	20	70	0	0	10	0	90
28	0	10	10	10	70	0	0	10	10	80
29	0	0	10	0	90	0	10	10	20	60
30	0	0	10	10	80	0	0	30	0	70
31	0	10	10	10	70	0	0	20	10	70

32	0	0	10	10	80	0	0	20	10	70
33	0	0	10	10	80	0	0	10	20	70
34	0	0	10	10	80	0	0	10	20	70
35	0	0	10	10	80	0	0	20	20	60
36	0	0	20	0	80	0	0	30	20	50
37	0	0	10	10	80	0	0	10	10	80
38	10	0	20	20	50	0	0	10	10	80
39	0	10	10	10	70	0	0	10	0	90
40	0	0	10	10	80	0	0	10	20	70
41	0	0	10	10	80	0	10	10	10	70
42	0	10	20	10	60	0	10	0	30	60
43	0	10	0	10	80	0	20	30	0	50
44	10	10	0	10	70	0	10	0	0	90
45	0	10	0	20	70					
46	0	20	10	20	50					
47	0	0	10	0	90					

*1: Not appropriate, 2: Inappropriate, 3: Neutral, 4: Appropriate, 5: Completely appropriate

When Table 1 is examined, it can be seen that the majority of experts considered the items related to the selection of a developed scale as appropriate or completely appropriate. Considering the scores of 4 and 5 given to the items, the lowest percentage is found to be 70%. This indicates that at least 70% (n=7) of the experts found each item suitable for selecting a developed scale. For some items, this rate reaches up to 100%. Therefore, it can be stated that the items in the checklist to be used in selecting a developed scale are sufficient according to the experts. The Gwet's AC1 coefficient, calculated to evaluate the agreement among experts, was found to be 0.54 for the checklist evaluating the developed scale. Based on this, it can be stated that the agreement among experts is at a moderate level (Wongpakaran et al., 2013). Additionally, when the average of the experts who scored the items 4 and 5 was examined, it was observed to be 8.55 for the checklist items to be used in selecting a developed scale. This suggests that the items were generally considered appropriate by the experts.

For the items to be used when selecting an adapted scale, similarly to the developed scale checklist, the majority of experts considered the items appropriate or completely appropriate. When the scores of 4 and 5 given to the items were considered, the lowest percentage was found to be 50% (item 43). For some items, this rate reaches up to 100%. Therefore, it can be stated that the items in the checklist to be used in selecting an adapted scale are sufficient according to the experts. The Gwet's AC1 coefficient, calculated to evaluate the agreement among experts, was found to be 0.50 for the checklist evaluating the adapted scale. Based on this, it can be stated that the agreement among experts is at a moderate level (Wongpakaran et al., 2013). Furthermore, when

the average of the experts who scored the items 4 and 5 was examined, it was observed to be 8.55 for the checklist items to be used in selecting the adapted scale. This suggests that the items were generally considered appropriate by the experts.

Determining the Weights of the Items in the Checklist to Be Used When Selecting a Developed/Adapted Scale According to Experts

In order to determine the weights of the items planned to be used when selecting a developed or adapted scale, the weight averages suggested by the experts and the researcher conducting this study (considered as one person) were calculated. The results obtained in this regard are presented in Tables 2 and 3.

Table 2

Average Weight Recommendations from Experts for the Control List to Be Used in Selecting a Developed Scale

No	Items	Weights
1	A literature review on the topic has been conducted.	1
2	The characteristic to be measured is clearly defined.	1
3	A pool of items was developed with reference to the literature.	1
4	During the development of the item pool, composition writing or interviews were conducted with the target group.	1
5	Evidence for content validity has been provided (e.g., content validity ratio-index).	1
6	Inter-expert agreement during content validity determination has been reported (e.g., Fleiss' Kappa, Gwet'in AC1, and Cohen's Kappa).	1
7	The number of items written during the item pool development is two to three times the target number.	1
8	At least two subject-matter experts were consulted during the item pool development phase.	1
9	At least two measurement experts were consulted regarding the prepared items.	1
10	At least two language experts were consulted regarding the prepared items.	1
11	Necessary revisions were made based on the feedback from experts regarding the item pool.	1
12	The response categories (e.g., five-point Likert scale) align with the item statements.	1
13	Each item measures a single behavior.	1
14	Grammar and spelling rules have been adhered to in item writing.	1
15	Items do not contain ambiguous expressions that could be interpreted differently by different individuals.	1
16	A pilot test of the items was conducted on a smaller group (15-20 people) outside the target group (sample) to assess clarity.	2
17	Revisions were made based on the results of the pilot test.	2

18	The sample size for the target group is determined to be 400 or more (for EFA).	3
19	Maximum diversity was considered when determining the sample size.	3
20	Exploratory Factor Analysis (EFA) was conducted during validity studies.	20
21	Missing data were controlled according to the assumptions of EFA.	1
22	Multivariate outliers were controlled according to the assumptions of EFA.	1
23	Multivariate normality was tested according to the assumptions of EFA.	1
24	Multicollinearity was tested according to the assumptions of EFA.	1
25	Objective methods (e.g., optimal parallel analysis, HULL method, MAP test) were used to determine the number of factors in EFA.	6
26	A method other than principal components was used for estimation in EFA.	5
27	The explained variance ratio from the EFA is above 30% for a unidimensional scale.	5
28	The explained variance ratio from the EFA is reasonable (above 10% for each factor) for a multidimensional scale.	5
29	The factor loadings obtained from the EFA are above 0.30.	5
30	The process of removing cross-loading items during EFA is reported.	6
31	If the variables consist of two categories (1-0), tetrachoric correlation; if they consist of 3-6 categories, polychoric; and if they consist of 7 or more categories, Pearson correlation matrix was used during EFA.	5
32	Confirmatory Factor Analysis (CFA) was conducted during validity studies.	13
33	CFA was conducted on a different sample from the one used in the EFA.	5
34	Missing data were controlled according to the assumptions of CFA.	1
35	Multivariate outliers were controlled according to the assumptions of CFA.	1
36	Multivariate normality was tested according to the assumptions of CFA.	2
37	Multicollinearity was tested according to the assumptions of CFA.	2
38	The estimation method used in CFA was DWLS, WLSMV or ULSMV.	6
39	If the variables consist of two categories (1-0), tetrachoric correlation; if they consist of 3-6 categories, polychoric; and if they consist of 7 or more categories, Pearson covariance matrix was used in CFA.	5

40	The CFA results, including χ^2 , degrees of freedom, p-value, RMSEA, 90% confidence interval for RMSEA, CFI, TLI, and SRMR indices, were reported.	5
41	The factor loadings obtained in CFA are 0.30 or higher.	5
42	The number of modifications made in CFA is below 3.	4
43	Item analysis has been conducted.	2
44	Evidence for criterion validity has been provided.	2
45	At least two internal consistency reliability methods (e.g., Cronbach's Alpha and McDonald's Omega) have been used.	5
46	Evidence for test-retest reliability has been provided.	2
47	The scale's administration guidelines have been prepared.	1
Total		150

Table 2 shows 47 criteria that can be used when selecting a developed scale. Accordingly, researchers can evaluate the scales they wish to use within the framework of the criteria presented in Table 2. Checklists provide information about the presence or absence of a characteristic. In this context, a coding system was applied where 1 was assigned to the criteria reported in the scale development study, and 0 was assigned to the criteria that were not reported. The suggested weight of each criterion can then be multiplied by these codes to determine whether the scale will be used. As seen in Table 2, the total score that can be obtained from the 47 items is 150. Accordingly, researchers who are selecting a developed scale can decide to use the scale if the score obtained from this checklist is 75 or higher.

For example, suppose that in a scale development study, the criteria between 1 and 25 presented in Table 2 are met, while the other criteria are not. In this case, the score for criteria 26 to 47 would be 0, and for criteria 1 to 25, the score would be $1 \times (\text{weight of the item}) = 59$ points. Since $59 < 75$, it can be concluded that the scale should not be selected. For more detailed classification, the following score ranges can be used: 0-74 (not selected), 75-90 (acceptable), 91-116 (good), and 117-150 (excellent). In this classification, the importance of EFA and CFA in the scale development process has been considered, and the classification has been developed based on the scores in these categories. Table 3 presents the criteria and their weights that can be used in the selection of adapted scales.

Table 3

Average Weight Recommendations from Experts for the Control List to Be Used in Selecting an Adapted Scale

No	Items	Weights
1	A literature review related to the topic has been conducted.	2
2	The feature to be measured has been clearly defined.	2

3	The need for the adaptation of the scale has been thoroughly reported.	1
4	It has been discussed whether the characteristic in the original culture of the scale is equivalent to the characteristic in the adapted culture.	3
5	The necessary permissions for the adaptation of the scale have been obtained	2
6	Opinions from field experts regarding the adaptability of the scale have been sought.	1
7	Opinions from measurement experts regarding the adaptability of the scale have been sought.	1
8	Translators who are proficient in both cultures have been used for the translation of the scale from the original language to the target language.	2
9	Translators who are proficient in both cultures have been used for the reverse translation (from the target language to the original language) of the scale.	2
10	It has been checked whether semantic, experiential, conceptual, and idiomatic equivalence has been ensured between the two forms of the scale.	3
11	Language equivalence of the scale has been ensured.	1
12	The response categories (such as a five-point Likert scale) are consistent with the item statements.	1
13	Each item measures a single behavior.	1
14	Grammar and spelling rules have been adhered to in the items.	1
15	There are no ambiguous expressions that can be interpreted differently by different individuals.	1
16	A pre-test of the items has been conducted with a smaller group (15-20 people) outside the target group (sample) to assess clarity.	2
17	Necessary revisions have been made after the pre-test.	2
18	The target group size has been determined to be 400 or more (for CFA).	3
19	Maximum diversity has been considered when determining the target group size.	3
20	Confirmatory Factor Analysis (CFA) has been conducted in the validity studies.	24
21	Missing data assumptions for CFA have been checked.	1
22	Multivariate outliers have been checked in accordance with CFA assumptions.	1
23	Multivariate normality has been tested according to CFA assumptions.	1
24	Multicollinearity has been tested according to CFA assumptions.	1
25	DWLS, WLSMV or ULSMV estimation methods have been used in CFA.	6
26	If the variables in the CFA process consist of two categories (1-0), tetracoric correlation has been used; if they consist of 3-6 categories, polychoric correlation has	5

	been used; if they consist of 7 or more categories, Pearson's covariance matrix has been used.	
27	The results of CFA have been reported, including χ^2 , degrees of freedom, p-value, RMSEA value with its 90% confidence interval, CFI, TLI, and SRMR indices.	7
28	Factor loadings obtained from CFA are at least 0.30.	5
29	The number of modifications applied in CFA is below three.	5
30	If the model fit was not achieved in CFA, the factor structure has been checked using EFA.	13
31	Missing data assumptions for EFA have been checked.	1
32	Multivariate outliers have been checked in accordance with EFA assumptions.	1
33	Multivariate normality has been tested according to EFA assumptions.	1
34	Multicollinearity has been tested according to EFA assumptions.	1
35	Objective methods (e.g., optimal parallel analysis, HULL method, MAP test) have been used when determining the number of dimensions in EFA.	6
36	An estimation method other than Principal Component Analysis has been used in EFA.	5
37	The variance explained by the scale in EFA is above 30% for a unidimensional scale.	5
38	If the scale in EFA is multidimensional, the explained variance is reasonable (above 10% for each dimension).	5
39	The factor loadings obtained in EFA are at least above 0.30.	5
40	The procedure for handling crossloading items in EFA has been reported.	4
	If the variables in the EFA process consist of two categories (1-0), tetracoric correlation has been used; if	
41	they consist of 3-6 categories, polychoric correlation has been used; if they consist of 7 or more categories, Pearson's correlation matrix has been used.	5
	At least two methods of reliability determination for	
42	internal consistency (e.g., Cronbach's Alpha and McDonald's Omega) have been used.	6
43	Evidence for test-retest reliability has been provided.	1
44	The scale's administration manual has been prepared.	1
	Toplam	150

Table 3 shows 44 criteria that can be used when selecting adapted scales. Accordingly, researchers can evaluate the scales adapted to Turkish culture within the framework of the criteria presented in Table 3. Checklists provide information about the presence or absence of a feature. After coding the criteria with 1 for those reported in the scale adaptation study and 0 for those not reported, the decision to use or

not use the scale can be made by multiplying each criterion's suggested weight with these codes. As seen in Table 3, the total possible score from the 44 items is 150. Therefore, researchers selecting an adapted scale can decide to use it if the score obtained from the checklist is 75 or higher.

For an example calculation, let us assume that in a scale development study, criteria 1 to 25 from Table 3 have been met, but the other criteria have not. In this case, criteria 26 to 44 would score 0, and criteria 1 to 25 would score $1 \times (\text{weight of the item}) = 68$ points. In this case, since $68 < 75$, the decision could be made not to select the scale. For more detailed classification, the following score ranges can be used: 0-74 (not selected), 75-90 (acceptable), 91-116 (good), and 117-150 (very good). In this classification, the importance of CFA in the scale adaptation process has been considered, and classifications have been developed based on these scores. The cutoff score of 75 has been considered to ensure that scales where at least 50% of the criteria are not met are not selected. The establishment of cutoff scores is usually carried out through standard-setting procedures, and this point has been discussed in the limitations section of the study.

Discussion, Conclusion, and Suggestions

Researchers conduct analyses relevant to their research problems by using the developed/adapted scales in the literature. However, despite the existence of scale development/adjustment studies, no concrete list of suggestions has been found in the literature regarding how to select these scales. Therefore, this study aims to develop a checklist that researchers can practically use. In line with this goal, two checklists have been developed for selecting both developed and adapted scales.

When examining the checklist and weights for selecting developed scales, it can be noted that the list consists of 47 items and that the primary focus of the criteria is on EFA. This should not be misunderstood by researchers. The purpose is not to reduce the scale development processes to the level of EFA. However, researchers selecting existing scales may face issues when trying to demonstrate structural validity if the scale does not work as expected. Some researchers may not even include evidence for structural validity, reporting only the findings from the scale development study. This situation may undermine the validity of the results obtained. As is well known, reliability and validity are characteristics of the measurement results, not the measurement instrument (American Educational Research Association [AERA] et al., 2014). Therefore, scales developed from a properly conducted EFA process will yield results that are largely similar to those obtained in the original scale development study. For this reason, more weight has been given to EFA processes in the checklist for selecting a developed scale. On the other hand, it should

be noted that these suggestions are specifically for already developed scales. If the only existing measurement tool for the characteristic of interest is the developed scale, researchers may not have many alternatives. In this case, the checklist proposed in the current study can be used to decide whether the existing scale should be used. If a decision is made not to use it, the option to develop a new scale may be preferred. In cases where exploratory factor analysis is not conducted (for example, in projective personality tests or when there is insufficient data to conduct factor analysis), criterion validity studies may have been conducted. The checklist proposed in the current study may suggest that these measurement tools should not be preferred, although this does not necessarily reflect a low quality of the scale. In some cases, criterion validity may be more meaningful and applicable. Researchers using such a scale should consider collecting additional data with similar scales, even if they are not specifically interested in criterion validity. The checklist addressed in this study aims to guide researchers who, after using a scale, realize that it does not work in their own sample. This is because administering projective personality tests, for example, requires expertise, and researchers without this expertise cannot apply such tests. Therefore, the focus of this study is on researchers who plan to collect data using published scales. In this respect, the emphasis on EFA seems justifiable.

When considering adaptation studies, more emphasis has been placed on CFA processes since the equivalence of the structure in the original culture is examined in the target culture (Kılıç, 2023a). Similarly, this does not aim to reduce adaptation studies to the level of CFA. There may be cases where the correct processes have not been properly implemented or reported in the adapted scales that have been developed and published. For this reason, more emphasis has been placed on CFA processes to provide researchers who select adapted scales with useful insights and help them achieve replicable structures in their own studies. If these processes are not conducted, and only linguistic validity, content validity, and criterion validity are studied, the only evidence that researchers can use for validity (unless additional scales have been used for criterion validity) is factor analysis. Researchers using adapted scales usually collect data with one or more other scales. In this case, collecting data with four or five scales for criterion validity means data from multiple sources, which can lead to issues such as fatigue and decreased reliability, making it less useful. Considering such cases, more emphasis has been placed on factor analysis methods. This increases the likelihood that researchers using adapted scales will obtain similar results.

Making comments about the importance of validity evidence based on the weights of the criteria would be inappropriate. Because in scale development/adaptation studies, it is recommended to collect

evidence for all types of validity, as far as possible (AERA et al., 2014). The perspective of the current study is to help researchers avoid surprises in their own sample by paying attention to the recommended criteria when selecting scales.

Although the primary focus of the study is to guide researchers in selecting scales, the checklist can also serve as a guide for researchers who will review scale development/adaptation studies. Furthermore, researchers who wish to develop or adapt scales can plan their studies according to the checklist presented here. In this case, it would be more appropriate to focus on the criteria themselves rather than their weights. It can be said that this study will guide researchers who will use, develop, adapt scales, and review studies. Future studies may examine the scores obtained from the checklist for developed/adapted scales. If a sufficient number of studies are reviewed using the checklist, the structure of the checklist can be analyzed through factor analysis, and the weights of the criteria can be determined statistically.

Limitations

As with any study, this study has some limitations. The study is limited to the opinions of 10 experts in measurement and assessment. Additionally, while the information provided in the checklist is up-to-date, it may change in the future depending on developments in the literature. For instance, the suggestion to use WLSMV or ULSMV as estimation methods in CFA may become outdated if better methods are developed in the future. This is another limitation of the study.

Another limitation that may be considered is the lack of a standard-setting procedure. However, the goal of standard-setting studies is to establish a cutoff score for the measured characteristic (Çetin, 2023). To do this, a study needs to be conducted on the minimum level of the measured characteristic. However, it may not make much sense to seek expert opinions on which steps of scale development are essential, as the steps of scale development are already clear, and researchers conducting scale development studies are expected to follow these steps. Experts are likely to agree that all these steps are necessary. Therefore, this study did not use standard-setting methods to establish a standard. While it is essential that all steps of scale development are correctly followed, there are known issues in the literature regarding this, and this study was conducted to guide researchers in selecting developed scales. The suggested cutoff scores are intended as recommendations and do not represent absolute thresholds. The reason for giving more emphasis to properly conducted CFA or EFA is to increase the likelihood of researchers obtaining valid results in their own samples. From this perspective, the cutoff score of 75 has been proposed.

Ethics Committee Approval: *This research was conducted with the approval granted by the Trakya University Social and Human Sciences Research Ethics Committee in its meeting on May 29, 2024, under decision number 05/09.*

Conflict of Interest: *The authors declare that they have no conflict of interest.*

Authors' Contribution: *MAG: Study design, writing the introduction, and data collection; AFK: Study design, writing the introduction, data collection, and analysis; YÖÖ: Writing the introduction and discussion sections.*

Acknowledgments: *We would like to express our gratitude to the experts from whom we obtained their opinions on the steps of scale development and adaptation and asked to assign appropriate weights for each step: Dr. Başak Erdem Kara, Dr. Durmuş Özbaşı, Dr. Ergül Demir, Dr. Funda Nalbantoğlu Yılmaz, Dr. Gül Güler, Dr. Hatice Kumantaş Öztürk, Dr. İlhan Koyuncu, Dr. Levent Ertuna, Dr. Mustafa İlhan, Dr. Özge Altıntaş, Dr. Seher Yalçın ve Dr. Sema Sulak for their valuable contributions.*

References

- Acar Güvendir, M., & Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Electronic Journal of Social Sciences*, 14(52), 23-33. <https://doi.org/10.17755/esosder.54872>
- Acar Güvendir, M., & Özer Özkan, Y. (2022). *Tüm yönleriyle ölçek geliştirme süreci*. Pegem Akademi.
- Alper, A., & Gülbahar, Y. (2009). Trends and issues in educational technologies: A Review of recent research in TOJET. *The Turkish Online Journal of Educational Technology- TOJET*, 8(2), 124-135.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173. <https://doi.org/10.1007/BF02294170>
- Arrindell, W. A., & Van der Ende, J. (1985). Cross-sample invariance of the structure of self-reported distress and difficulty in assertiveness: Experiences with the scale for interpersonal behaviour. *Advances in Behaviour Research and Therapy*, 7(4), 205-243. [https://doi.org/10.1016/0146-6402\(85\)90013-x](https://doi.org/10.1016/0146-6402(85)90013-x)

- Aybek, E. C., & Toraman, C. (2022). How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory. *International Journal of Assessment Tools in Education*, 9(2), 534-547. <https://doi.org/10.21449/ijate.1132931>
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), 77-85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal Of Exercise Science*, 8(3), 297.
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In: Joreskog, K.G., Wold, H. (Eds.), *Systems under indirect observation: causality, structure, prediction* (pp. 149-173). Elsevier.
- Boztunç Öztürk, N., Eroğlu, M.G., & Kelecioğlu, H. (2014, June 9-13). Eğitim bilimleri alanında yapılan ölçek uyarlama makalelerinin incelenmesi, IV. *Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi*, Ankara.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford.
- Bryman, A., & Cramer, D. (2001). Quantitative data analysis with SPSS Release 10 for Windows: A guide for social scientists. Routledge. <https://doi.org/10.4324/9780203471548>
- Büyüköztürk, Ş. (2020). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (28. baskı). Pegem Akademi.
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2020). *Bilimsel araştırma yöntemleri* (29. baskı). Pegem Akademi.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315757421>
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Cattell, R. B. (2012). *The scientific use of factor analysis in behavioral and life sciences*, Springer Science & Business Media.
- Chen, L. T., & Liu, L. (2020). Methods to analyze likert-type data in educational technology research. *Journal of Educational Technology Development and Exchange (JETDE)*, 13(2), 39-60. <https://doi.org/10.18785/jetde.1302.04>
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. L. NJ Erlbaum.

- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (2. baskı). Pegem Akademi.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (Second Edition). Lawrence Erlbaum Associate
- Cooper, C. (2019). *Psychological testing: Theory and practice*. Routledge. <https://doi.org/10.4324/9781315391342>
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 27-29. <https://doi.org/10.7275/jyj1-4868>
- Çetin, S. (2023). Standart belirleme. İçinde A. F. Kılıç (Ed.), *R programlama diliyle A'dan Z'ye ölçek geliştirme* (ss. 257-289). Nobel.
- Çüm, S., & Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.
- De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147-181. <https://doi.org/10.1080/00273170902794206>
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.
- Doğan, H., & Tok, T. N. (2018). Türkiye'de eğitim bilimleri alanında yayınlanan makalelerin incelenmesi: Eğitim ve Bilim Dergisi örneği. *Current Research in Education*, 4(2), 94-109.
- Erdem, D. (2011). Türkiye'de 2005–2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Journal of Measurement and Evaluation in Education and Psychology*, 2(1), 140-147.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar, *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2019). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (4nd ed.). Pegem Akademi.
- Everitt, B. S. (1975). Multivariate analysis: the need for data and other problems. *British Journal of Psychiatry*, 126, 237-240. <https://doi.org/10.1192/bjp.126.3.237>
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University. <https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Sage.

- Goodwin, L. D. (1999). The role of factor analysis in the estimation of construct validity. *Measurement in Physical Education and Exercise Science*, 3(2), 85-100. https://doi.org/10.1207/s15327841mpee0302_2
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510-3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Gorsuch, R. L. (1974). *Factor analysis*. Saunders.
- Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(2), 251-270. https://doi.org/10.1207/S15328007SEM0702_6
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19, 149-161. <https://doi.org/10.1007/bf02289162>
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Pearson.
- Helm, J. L., Castro-Schilo, L., & Oravecz, Z. (2017). Bayesian versus maximum likelihood estimation of multitrait-multimethod confirmatory factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 17-30. <https://doi.org/10.1080/10705511.2016.1236261>
- Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393-416. <https://doi.org/10.1177/0013164405282485>
- Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988. [https://doi.org/10.1016/0149-2063\(95\)90050-0](https://doi.org/10.1016/0149-2063(95)90050-0)
- Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Chapman & Hall.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51-62. <https://doi.org/10.1080/10447318.2015.1087664>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 37-41. <https://doi.org/10.1080/10705519909540118>

- İşçi, S. (2013). *Türkiye’de eğitim yönetimi alanında yapılmış lisansüstü tezlerin tematik, metodolojik ve istatistiksel açıdan incelenmesi* [Yüksek lisans tezi, Eskişehir Osmangazi Üniversitesi]. Ulusal Tez Merkezi. Erişim No: 344302
- Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, 8, 205-223. https://doi.org/10.1207/s15328007sem0802_3
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403. <https://doi.org/10.9734/BJAST/2015/14975>
- Karadağ, E. (2009). Türkiye’de eğitim bilimleri alanında yapılmış doktora tezlerinin tematik ve metodolojik açıdan incelemesi; Bir durum çalışması. [Doktora tezi, Marmara Üniversitesi]. Ulusal Tez Merkezi. Erişim No: 250896
- Kaiser, H.F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401-415. <https://doi.org/10.1007/bf02291817>
- Kalkbrenner, M. T. (2021). Alpha, Omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, 1-12. <https://doi.org/10.1080/21501378.2021.1940118>
- Kılıç, A. F. (2022a). Açıklayıcı faktör analizinde boyut sayısına karar verme: Yöntemlere kısa bir bakış. *Pamukkale University Journal of Social Sciences Institute*, 51 (Özel sayı 1), Ö305-Ö318. <https://doi.org/10.30794/pausbed.1095936>
- Kılıç, A. F. (2022b). Ölçek geliştirme sürecinde açıklayıcı faktör analizi. İçinde M. Acar-Güvendir & Y. Özer-Özkan (Ed.), *Tüm yönleriyle ölçek geliştirme süreci* (ss. 69-129). Pegem Akademi.
- Kılıç, A. F. (2023a). Geçerlik ve kanıtları. *R programlama diliyle A’dan Z’ye ölçek uyarlama* (ss. 61-113) içinde. Nobel.
- Kılıç, A. F. (2023b). Ölçek geliştirme süreci. *R programlama diliyle A’dan Z’ye ölçek geliştirme* (33-59) içinde. Nobel.
- Kılıç, A. F., & Doğan, N. (2021). Comparison of confirmatory factor analysis estimation methods on mixed-format data. *International Journal of Assessment Tools in Education*, 8(1), 21-37. <https://doi.org/10.21449/ijate.782351>
- Kılıç, A. F., & Koyuncu, İ. (2017). Ölçek uyarlama çalışmalarının yapı geçerliği açısından incelenmesi. *Küreselleşen dünyada eğitim* (ss. 1202-1205) içinde. Pegem Akademi.
- Kılıç, A. F., Koyuncu, İ., & Uysal, İ. (2023). Scale development based on item response theory: A systematic review. *International Journal of Psychology and Educational Studies*, 10(1), 209-223. <https://doi.org/10.52380/ijpes.2023.10.1.982>

- Kılıç, A. F., & Uysal, İ. (2023). Ölçek puanlarının geçerliği. *R programlama diliyle A'dan Z'ye ölçek geliştirme* (95-150) içinde. Nobel.
- Kline, R. B. (2016). *Principle and practice of structural equation modeling* (4th ed.). The Guilford.
- Kline, P. (1994). *An easy guide to factor analysis*. Routledge.
- Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In *Applied quantitative analysis in the social sciences* (pp. 171- 207). Routledge.
- Koyuncu, İ. ve Kılıç, A. F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Eğitim ve Bilim*, 44(198), 361-388. <https://doi.org/10.15390/EB.2019.7665>
- Lei, P. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, 43(3), 495-507. <https://doi.org/10.1007/s11135-007-9133-z>
- Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21(3), 369-387. <https://doi.org/10.1037/met0000093>
- Liang, X., & Yang, Y. (2014). An evaluation of WLSMV and Bayesian methods for confirmatory factor analysis with categorical indicators. *International Journal of Quantitative Research in Education*, 2(1), 17-38. <https://doi.org/10.1504/IJQRE.2014.060972>
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364. <https://doi.org/10.1080/00273171.2011.564527>
- Maccallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling of fit involving a particular measure of model. *Psychological Methods*, 13(2), 130-149. <https://doi.org/10.1037/1082-989X.1.2.130>
- MacCallum, R. C. Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61-88. <https://doi.org/10.1037/met0000425>
- 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.
- Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2017). Scale development: Ten main limitations and recommendations to

- improve future research practices. *Psicologia: Reflexão e Crítica*, 30(3), 1-20. <https://doi.org/10.1186/s41155-016-0057-1>
- Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, 10(2), 60-70. <https://doi.org/10.1027/1614-2241/a000068>
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5, 159-168. https://doi.org/10.1207/s15327574ijt0502_4
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599-620. https://doi.org/10.1207/s15328007sem0904_8
- Navarro-Gonzalez, D., & Lorenzo-Seva, U. (2021). EFA.MRFA: Dimensionality assessment using minimum rank factor analysis. <https://cran.r-project.org/package=EFA.MRFA>
- Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, 39(3), 439-478. https://doi.org/10.1207/s15327906mbr3903_3
- Niemand, T., & Mai, R. (2022). Flexible cutoffs for model fit evaluation in covariance-based structural models. *Journal of the Academy of Marketing Science*, 46, 1148-1172. <https://doi.org/10.1007/s11747-018-0602-9>
- Nunnally, J.C. (1978). *Psychometric theory*. McGraw-Hill.
- O'Connor, B. P. (2022). EFA.dimensions: Exploratory factor analysis functions for assessing dimensionality. <https://cran.r-project.org/package=EFA.dimensions>
- Oranje, A. (2003). *Comparison of estimation methods in factor analysis with categorized variables: Applications to NEAP data*.
- Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment Research & Evaluation*, 20(2), 1-7. <https://doi.org/10.7275/hb2g-m060>
- Osborne, J.W., & Costello, A.B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, 9(11), 1-9. <https://doi.org/10.7275/ktzq-jq66>
- Ozan, C., & Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, 4(1), 116-136. <https://doi.org/10.19126/suje.76547>
- Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics*, 32, 153-161. <https://doi.org/10.1023/a:1015210025234>
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates.

- Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* (2.2.9) [Software]. Northwestern University. <https://cran.r-project.org/package=psych>
- Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research*, 20(4), 235–249. <https://doi.org/10.1002/mpr.352>
- Sapnas, K. G., & Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement*, 10, 135–154. <https://doi.org/10.1891/jnum.10.2.135.52552>
- Schönrock-Adema, J., Heijne-Penninga, M., van Hell, E. A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher*, 31(6), 226–232. <https://doi.org/10.1080/01421590802516756>
- Seçer, İ., Ay, İ., Ozan, C., & Yılmaz, B. Y. (2014). Research Trends in the Field of Guidance and Psychological Counseling: A Content Analysis. *Turkish Psychological Counseling & Guidance Journal*, 5(41), 49–60.
- Shevlin, M. & Miles, J. N. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, 25(1), 85–90. [https://doi.org/10.1016/s0191-8869\(98\)00055-5](https://doi.org/10.1016/s0191-8869(98)00055-5)
- Sönmez, V., & Alacapınar, G. (2016). *Sosyal bilimlerde ölçme aracı hazırlama*. Anı Yayıncılık.
- Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53), 2521. <https://doi.org/10.21105/joss.02521>
- Steven, J. (1996). *Applied multivariate statistics for social sciences* (third edition). Lawrence Erlbaum Associates, Inc.
- Şeker, H., & Gençdoğan, B. (2006). *Psikolojide ve eğitimde ölçme aracı geliştirme*. Nobel.
- Şencan, H. (2005). Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik. Seçkin Yayıncılık.
- Şenyurt, S., & Özer Özkan, Y. (2017). Eğitimde ölçme ve değerlendirme alanında yapılan yüksek lisans tezlerinin tematik ve metodolojik açıdan incelenmesi. *İlköğretim Online*, 16(2), 628–653.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209–220. <https://doi.org/10.1037/a0023353>
- Tanujaya, B., Prahmana, R. C. I., & Mumu, J. (2022). Likert scale in social sciences research: Problems and difficulties. *FWU Journal of Social Sciences*, 16(4), 89–101.

- Tavşancıl, E., Güler, G., & Ayan, C. (2014). 2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreci açısından incelenmesi. In *IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı)* Hacettepe Üniversitesi, Ankara.
- Tosuntaş, Ş. B., Emirtekin, E., & Süral, İ. (2019). Eğitim ve öğretim teknolojileri konusunda yapılan tezlerin incelenmesi 2013-2018. *Yükseköğretim ve Bilim Dergisi*, (2), 277-286.
- Uyumaz, G., & Sırgancı, G. (2020). Doğrulamalı faktör analizi için gerekli örneklem büyüklüğü kaç kişidir?: Bayes yaklaşımı ve maksimum olabilirlik kestirimi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 16(32), 5302-5340. <https://doi.org/10.26466/opus.826895>
- Uysal, İ. (2023). Ölçek uyarlama aşamaları. *R Programlama Diliyle A'dan Z'ye Ölçek Uyarlama* (ss. 1-9) içinde. Nobel.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327. <https://doi.org/10.1007/BF02293557>
- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61. <https://doi.org/10.1186/1471-2288-13-61>
- Yu, C.-Y., & Muthén, B. O. (2002, April 4). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. In *American Educational Research Association Annual Conference*. New Orleans, LA.
- Zygmunt, C., & Smith, M. R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology*, 10(1), 40-55. <https://doi.org/10.20982/tqmp.10.1.p040>

Appendix A. Checklist to be Used When Selecting a Developed Scale

No	Items	Weights	Yes	No
1	A literature review on the topic has been conducted.	1		
2	The characteristic to be measured is clearly defined.	1		
3	A pool of items was developed with reference to the literature.	1		
4	During the development of the item pool, composition writing or interviews were conducted with the target group.	1		
5	Evidence for content validity has been provided (e.g., content validity ratio-index).	1		
6	Inter-expert agreement during content validity determination has been reported (e.g., Fleiss' Kappa, Gwet'in AC1, and Cohen's Kappa).	1		
7	The number of items written during the item pool development is two to three times the target number.	1		
8	At least two subject-matter experts were consulted during the item pool development phase.	1		
9	At least two measurement experts were consulted regarding the prepared items.	1		
10	At least two language experts were consulted regarding the prepared items.	1		
11	Necessary revisions were made based on the feedback from experts regarding the item pool.	1		
12	The response categories (e.g., five-point Likert scale) align with the item statements.	1		
13	Each item measures a single behavior.	1		
14	Grammar and spelling rules have been adhered to in item writing.	1		
15	Items do not contain ambiguous expressions that could be interpreted differently by different individuals.	1		
16	A pilot test of the items was conducted on a smaller group (15-20 people) outside the target group (sample) to assess clarity.	2		
17	Revisions were made based on the results of the pilot test.	2		
18	The sample size for the target group is determined to be 400 or more (for EFA).	3		
19	Maximum diversity was considered when determining the sample size.	3		
20	Exploratory Factor Analysis (EFA) was conducted during validity studies.	20		
21	Missing data were controlled according to the assumptions of EFA.	1		
22	Multivariate outliers were controlled according to the assumptions of EFA.	1		
23	Multivariate normality was tested according to the assumptions of EFA.	1		
24	Multicollinearity was tested according to the assumptions of EFA.	1		
25	Objective methods (e.g., optimal parallel analysis, HULL method, MAP test) were used to determine the number of factors in EFA.	6		
26	A method other than principal components was used for estimation in EFA.	5		
27	The explained variance ratio from the EFA is above 30% for a unidimensional scale.	5		
28	The explained variance ratio from the EFA is reasonable (above 10% for each factor) for a multidimensional scale.	5		
29	The factor loadings obtained from the EFA are above 0.30.	5		

30	The process of removing cross-loading items during EFA is reported.	6
31	If the variables consist of two categories (1-0), tetrachoric correlation; if they consist of 3-6 categories, polychoric; and if they consist of 7 or more categories, Pearson correlation matrix was used during EFA.	5
32	Confirmatory Factor Analysis (CFA) was conducted during validity studies.	13
33	CFA was conducted on a different sample from the one used in the EFA.	5
34	Missing data were controlled according to the assumptions of CFA.	1
35	Multivariate outliers were controlled according to the assumptions of CFA.	1
36	Multivariate normality was tested according to the assumptions of CFA.	2
37	Multicollinearity was tested according to the assumptions of CFA.	2
38	The estimation method used in CFA was DWLS, WLSMV or ULSMV.	6
39	If the variables consist of two categories (1-0), tetrachoric correlation; if they consist of 3-6 categories, polychoric; and if they consist of 7 or more categories, Pearson covariance matrix was used in CFA.	5
40	The CFA results, including χ^2 , degrees of freedom, p-value, RMSEA, 90% confidence interval for RMSEA, CFI, TLI, and SRMR indices, were reported.	5
41	The factor loadings obtained in CFA are 0.30 or higher.	5
42	The number of modifications made in CFA is below 3.	4
43	Item analysis has been conducted.	2
44	Evidence for criterion validity has been provided.	2
45	At least two internal consistency reliability methods (e.g., Cronbach's Alpha and McDonald's Omega) have been used.	5
46	Evidence for test-retest reliability has been provided.	2
47	The scale's administration guidelines have been prepared.	1
Total		150

*The rows marked "Yes" are considered as 1 and multiplied by their respective weights to obtain the total score. For example, if item 47 is marked "Yes," the score will be $1 \times 1 = 1$ point; if item 46 is marked "Yes," the score will be $1 \times 2 = 2$ points. The scores for each item are then summed to calculate the total score.

Appendix B. Checklist to be Used When Selecting an Adapted Scale

No	Items	Weights	Yes	No
1	A literature review related to the topic has been conducted.	2		
2	The feature to be measured has been clearly defined.	2		
3	The need for the adaptation of the scale has been thoroughly reported.	1		
4	It has been discussed whether the characteristic in the original culture of the scale is equivalent to the characteristic in the adapted culture.	3		
5	The necessary permissions for the adaptation of the scale have been obtained	2		
6	Opinions from field experts regarding the adaptability of the scale have been sought.	1		
7	Opinions from measurement experts regarding the adaptability of the scale have been sought.	1		
8	Translators who are proficient in both cultures have been used for the translation of the scale from the original language to the target language.	2		

9	Translators who are proficient in both cultures have been used for the reverse translation (from the target language to the original language) of the scale.	2
10	It has been checked whether semantic, experiential, conceptual, and idiomatic equivalence has been ensured between the two forms of the scale.	3
11	Language equivalence of the scale has been ensured.	1
12	The response categories (such as a five-point Likert scale) are consistent with the item statements.	1
13	Each item measures a single behavior.	1
14	Grammar and spelling rules have been adhered to in the items.	1
15	There are no ambiguous expressions that can be interpreted differently by different individuals.	1
16	A pre-test of the items has been conducted with a smaller group (15-20 people) outside the target group (sample) to assess clarity.	2
17	Necessary revisions have been made after the pre-test.	2
18	The target group size has been determined to be 400 or more (for CFA).	3
19	Maximum diversity has been considered when determining the target group size.	3
20	Confirmatory Factor Analysis (CFA) has been conducted in the validity studies.	24
21	Missing data assumptions for CFA have been checked.	1
22	Multivariate outliers have been checked in accordance with CFA assumptions.	1
23	Multivariate normality has been tested according to CFA assumptions.	1
24	Multicollinearity has been tested according to CFA assumptions.	1
25	DWLS, WLSMV or ULSMV estimation methods have been used in CFA.	6
26	If the variables in the CFA process consist of two categories (1-0), tetracoric correlation has been used; if they consist of 3-6 categories, polychoric correlation has been used; if they consist of 7 or more categories, Pearson's covariance matrix has been used.	5
27	The results of CFA have been reported, including χ^2 , degrees of freedom, p-value, RMSEA value with its 90% confidence interval, CFI, TLI, and SRMR indices.	7
28	Factor loadings obtained from CFA are at least 0.30.	5
29	The number of modifications applied in CFA is below three.	5
30	If the model fit was not achieved in CFA, the factor structure has been checked using EFA.	13
31	Missing data assumptions for EFA have been checked.	1
32	Multivariate outliers have been checked in accordance with EFA assumptions.	1
33	Multivariate normality has been tested according to EFA assumptions.	1
34	Multicollinearity has been tested according to EFA assumptions.	1
35	Objective methods (e.g., optimal parallel analysis, HULL method, MAP test) have been used when determining the number of dimensions in EFA.	6
36	An estimation method other than Principal Component Analysis has been used in EFA.	5
37	The variance explained by the scale in EFA is above 30% for a unidimensional scale.	5
38	If the scale in EFA is multidimensional, the explained variance is reasonable (above 10% for each dimension).	5
39	The factor loadings obtained in EFA are at least above 0.30.	5
40	The procedure for handling crossloading items in EFA has been reported.	4
41	If the variables in the EFA process consist of two categories (1-0), tetracoric correlation has been used; if they consist of 3-6 categories, polychoric correlation has been used; if they consist of 7 or more categories, Pearson's correlation matrix has been used.	5
42	At least two methods of reliability determination for internal consistency (e.g., Cronbach's Alpha and McDonald's Omega) have been used.	6
43	Evidence for test-retest reliability has been provided.	1

44	The scale's administration manual has been prepared.	1
	Toplam	150

*The rows marked with "Yes" can be considered as 1 and multiplied by their corresponding weights to calculate the total score. For example, if item 44 is marked as "Yes," the score will be 1x1 = 1 point; if item 42 is marked as "Yes," the score will be 1x6 = 6 points. The points obtained for each item can be summed to obtain the total score.