

Psikolojide Ölçme Araçlarının Geliştirilmesi ve Uyarlanması Kılavuzu

Duygu Güngör
Dokuz Eylül Üniversitesi

Özet

Psikoloji araştırmalarının çoğunda geçerli ve güvenilir ölçme araçlarına ihtiyaç vardır. Bu yazının temel amacı ölçek geliştirme veya uyarlama çalışması yapan uygulayıcılar ve değerlendiriciler için bir rehber oluşturmaktır. Bu kapsamda önce ölçek geliştirme ya da uyarlamada dikkat edilecek hususlardan bahsedilecektir. Sonra örneklem seçimine kısaca değinilecektir. İstatistiksel analizler kısmında ise kullanılan paket program, verinin analize hazırlanmasından bahsedildikten sonra geçerliği inceleme yöntemlerine kısaca değinilecek ve açıklayıcı ve doğrulayıcı faktör analizi yöntemleri açıklanacaktır. Son olarak güvenilirlik bulgularının raporlanmasına değinilecektir. Araştırmacılar için ek olarak bir kontrol listesi sunulacaktır.

Anahtar kelimeler: Ölçek geliştirme, ölçek uyarlama, geçerlik, güvenilirlik

Abstract

The aim of this study is to present a guide to researchers who develop or adapt measurement instruments. First, fundamentals of scale development and adaptation will be introduced. Then, I will mention sampling briefly. Statistical analyses part includes software, preparing data for analysis, and validation methods. Next exploratory and confirmatory factor analysis will be presented. Lastly, I will mention on reliability methods and how to report findings. Additionally, a checklist for researchers will be presented.

Key words: Scale development, scale adaptation, validity, reliability

Psikolojide araştırmalara konu olan tutum, kişilik, zeka, psikopatoloji gibi değişkenler genellikle doğrudan gözlenemez. Örtük değişkenler adı da verilen bu tür yapıların araştırılabilmesi için kuramsal olarak sağlam temellere dayanan, geçerli ve güvenilir ölçme araçlarına ihtiyaç vardır. Dolayısıyla sosyal bilim araştırmalarının azımsanamayacak kadar büyük bir kısmını ölçme araçlarının geliştirilmesi ya da uyarlanması oluşturmaktadır. Ebscohost veritabanında 2000-2016 yılları arasında yayınlanan çalışmalarda “psikometrik” anahtar kelimesi için 1214, “ölçek geliştirme” anahtar kelimesi için 1255 sonuç bulunmaktadır. Bununla birlikte zahmetli ve zor bir süreç olduğu kadar teknik ve sınırları belli olan (Kline, 2005) ölçek geliştirme ya da uyarlama çalışmalarının yürütülmesi ve raporlanması konusunda az sayıda çalışma bulunmaktadır (örn., Çüm ve Koç, 2013; DeVellis, çev. 2014; Erkuş, 2014; Kline, 2005; Sousa ve Rojjanasirrat, 2011).

Bu yazının temel amacı ölçek geliştirme veya uyarlama çalışması yapan uygulayıcılar ve değerlendirmeniler için bir rehber oluşturmaktır. Geçerleme çalışmalarında en sık başvurulan yöntem, madde varyans ve kovaryans yapılarının incelenmesi olduğundan bu yazının büyük bir kısmı geçerlik çalışmalarında kullanılan faktör analitik tekniklere ayrılmıştır. Fakat bu yöntemlerin biricik yöntemler olmadığını belirtmekte fayda vardır.

Çalışmaya uygun yöntemin belirlenmesi ölçmek istenen yapının niteliği, kullanılan gösterge ve örtük değişkenlerin düzeyleri ile doğrudan ilişkilidir. Örneğin örtük yapının süresiz dağıldığı düşünülüyorsa araştırmacılar örtük profil analizi ya da örtük sınıf analizinden, örtük yapının sürekli fakat gösterge değişkenlerin süresiz olduğu durumlarda ise madde tepki kuramından yararlanmalıdır. Buna ek olarak elbette, yakınsama geçerliği, kriter geçerliği benzeri yöntemler kullanılması da mümkün hatta bazen kaçınılmazdır. Okuyucular yöntemlere ait detaylı bilgileri yazı içerisinde sunulan referanslardan edinebilirler. Bu nedenle yazının hedefi yöntemlere aşına olan araştırmacılara bir kılavuz sunmaktır.

Yukarıda belirtilen sınırlılıklar çerçevesinde önce ölçek geliştirme ya da uyarlamada dikkat edilecek hususlardan bahsedilecektir. Sonra örneklem seçimine kısaca değinilecektir. İstatistiksel analizler kısmında ise kullanılan paket program, verinin analize hazırlanmasından bahsedildikten sonra geçerliği inceleme yöntemlerine kısaca değinilecek ve açıklayıcı ve doğrulayıcı faktör analizi yöntemleri açıklanacaktır. Son olarak güvenilirlik bulgularının raporlanmasına değinilecektir. Araştırmacılar için ek olarak bir kontrol listesi sunulacaktır. Bu yazı yöntemlerin kuramsal anlatımından çok bir makalede dikkat edilmesi gereken noktalar ve raporlama kısmına odaklanmıştır.

Ölçek Uyarlamak veya Geliştirmek

Psikoloji alanyazımında halihazırda kullanılan çokça ölçek bulunmaktadır. Bu nedenle bir ölçek geliştirmeye ya da uyarlamaya karar vermeden önce bunun bir gereklilik olduğu tespit edilmeli ve araştırmanın giriş kısmında bu gereklilik vurgulanmalıdır. Aksi halde tıpkı uzunluğu ölçerken metre, inç, mil, arşın, adım ve benzeri farklı birimler kullanırken karşılaşılan sorun gibi, aynı konuyu çalışan araştırmaları karşılaştırmak için ölçüm sonuçlarının aynı metriğe oturtulması gerekliliği ortaya çıkacaktır. Araştırmacılar için alanyazın taramasının ardından eğer araştırmaya konu olan kuramsal yapıyı ölçen bir ölçek yoksa ölçme araçlarının araştırmaya konu olan örneklemin diline uygun şekilde oluşturulması için iki temel alternatif bulunmaktadır; araştırmacılar ölçeği kendileri geliştirilebilir ya da başka dillerde var olan ölçeği kendi dillerine uyarlayabilirler. Her iki alternatifin de kendine özgü güçlü yönleri ve güçlükleri bulunmaktadır. Örneğin, geçerliği ve güvenilirliği başka dilde kanıtlanmış bir ölçeğin uyarlanması daha ekonomik görünse de çeviriden kaynaklı ya da kültür farklarından kaynaklı ölçme eşdeğerliğinin olmaması gibi sorunlar ortaya çıkabilir.

Eğer ölçek geliştirilmeye karar verilirse ölçülmek istenen yapının belirlenmesi, literatür taraması, uzmanlarla yapılan görüşmeler gibi adımlardan sonra ölçek geliştirme süreci madde havuzu oluşturma adımıyla başlamaktadır (ayrıntılı bilgi için bkz. Erkuş, 2014; Kline, 2005). Bu noktada makalede *madde havuzunun* nasıl oluşturulduğu (kaç uzmanla görüşüldü, toplam kaç madde oluşturuldu, kapsam nasıl belirlendi gibi) yöntem kısmında kısaca belirtilmelidir. Uygulama örneğine ulaşmadan önce yapılan pilot çalışma hem ölçek içeriğinin net şekilde anlaşılıp anlaşılmadığı hem de ölçek maddelerini cevaplamayı tamamlamanın ne kadar sürdüğünü belirlemek açısından oldukça önemlidir. Genellikle bu süre kayıt edilerek ölçek yönergesine eklenmektedir. Ayrıca yöntem kısmında cevaplamanın ortalama kaç dakika sürdüğünü ölçek hakkında bilgilerle yazıldığı örnekler bulunmaktadır. Böylece ölçeği kullanmak isteyen diğer araştırmacılar da araştırma planlarını bu süreyi temel alarak yapabilmektedir. Bir ölçek geliştirirken karar verilmesi gereken önemli noktalardan biri de tepki kategorileridir. Kapalı uçlu sorulardan oluşan ölçeklerde ölçek, doğru cevapları bulunan çoktan seçmeli bir yapıda olabilir. Bu durumda çeldirici olarak adlandırılan seçeneklerin bu görevi yerine getirdikleri istatistiksel yöntemlerle incelenmelidir. Maddelerin cevapları evet-hayır şeklinde iki uçlu olabileceği gibi Likert tipi derecelendirme yöntemi de kullanılabilir. Beşli derecelendirme kullanılan tip olsa da madde tepki kategorilerinin değişmesi ölçeğin yapısını da değiştirmektedir. Var olan bir ölçek kullanılırken eğer kültüre ve hedef örnekleme de

uygunsa derecelendirmeye sadık kalmak uygun olacaktır. Zaman derecelendirirken Türkçede ara sıra, bazen, nadiren gibi kelimelerin çok da ayırt edici olmadığını dikkate almak gerekmektedir. Araştırmacılar yayınlarında hem hangi dereceleme tipini kullandıklarını hem de eğer telif hakkı, maddelerin yayılmaması gerekliliği gibi özel durumlar yoksa kullandıkları maddelerden örnekler sunmalıdırlar (Bandalos ve Finney, 2010).

Ölçeği uyarlamaya karar veren araştırmacılar için ise ölçeği geliştiren yazarlardan alınan izinden sonra, madde havuzu oluşturma basamağını çeviri süreci almaktadır. Çeviri yaparken sıklıkla hedef dile çeviri ve orijinal dile geri çeviri basamakları izlenmektedir. Çeviriyi yapan uzman ekip dilbilimciler ve kurama hakim uzmanlardan oluşmalıdır ve makalede bu sürecin nasıl yapıldığı belirtilmelidir. Çeviri süreci ile ilgili örnek bir araştırma olarak Sousa ve Rojjanasrirat (2011) verilebilir. Bu noktada ölçme eşdeğerliği kavramına vurgu yapmakta yarar vardır. Ölçme eşdeğerliği gerçek puanları aynı olan bireylerin alt grup üyeliğinden bağımsız olarak aynı gözlenen puana sahip olması şeklinde tanımlanabilir (Millsap, 2011). Örnek olarak farklı kültürlerde yer alan aynı hoşgörü düzeyine sahip katılımcıların aynı gözlenen puana sahip olacağı düşünülebilir. Eğer kültür farkı aynı seviyede hoşgörü olan kişilerde farklı puanlar gözlenmesine neden oluyorsa ölçme eşdeğerliğinin varlığından söz edilemez. Bu durumda uyarlanan ölçeğin faktör yapısı farklı kültürlerde farklı çıkabilir. Bunun ölçme aracından mı yoksa kültürel farklardan mı kaynaklandığını incelemek ise araştırmacının yükümlülüğündedir. Avrupa Değerler Çalışması verilerindeki ölçme eşdeğerliğini inceleyen araştırmalarda, özellikle Türkiye'den toplanan verilerdeki tepki örüntülerinin diğer Avrupa ülkelerinden farklı olduğu yönünde vurgular mevcuttur (Kankaras ve Moors, 2009). Bu nedenle bir ölçeği uyarlayan araştırmacılar ölçeğin kültüre uygun şekilde çevrildiğinden emin olmalıdır. Bu noktada ölçek uyarlamada da pilot çalışma önemli yer tutmaktadır. Araştırmacılar bu deneme uygulamasında çevirinin anlaşılabilirliğini test edebilirler. Yapılabiliyorsa ölçeği geliştiren kişi ile işbirliği içerisinde kültürler arası karşılaştırmalar yapmak ve ölçme aracının eşdeğerliğini kanıtlamak uluslararası alanyazına önemli katkılar sağlayacaktır.

Örneklem Seçimi

Her araştırmada olduğu gibi ölçek geliştirme ya da uyarlama çalışmasında da en önemli noktalardan biri uygun örneklem seçimidir. Örneklem büyüklüğü güç, örnekleme yöntemi temsil edicilikle ilişkilidir. En sık karşılaşılan sorulardan biri örneklem büyüklüğünün ne kadar olması gerektiği yönündedir. Bu konuyla ilgili altın bir kural yoksa da güç analizleri ile örneklem genişli-

ğinin temsil edicilik gücü hesaplanabilir. Bunun dışında açılımlayıcı faktör analitik yöntemler kullanacak araştırmacılar için gözlenen değişken başka bir deyişle madde başına 10 katılımcıdan veri toplanması gerektiği yani 20 maddelik bir ölçek için 200 kişilik bir örnekleme ihtiyaç duyulduğu yönünde çalışmalar vardır. Fakat Costello ve Osborne (2005) yaptıkları simülasyon çalışmasında bu oranı kullanarak yapılan faktör analizi sonuçlarının çoğunun yanlış olduğunu göstermişlerdir. Bu nedenle madde başına en az 20 katılımcı kullanmak daha sağlam bir yöntem olarak görünmektedir (Kline, 2013). Doğrulayıcı faktör analitik yöntem kullanacak araştırmacılar için ise tahmin edilecek parametre sayısına uygun örneklem genişliği ve gözlenen madde sayısı seçmenin önemini vurgulamakta yarar vardır. Aksi takdirde yeterli serbestlik derecesine sahip olunamayacak ve model test edilemeyecektir.

Örneklem büyüklüğü kadar önemli olan bir diğer husus örneklemin temsil ediciliğidir; örneğin birinci sınıf öğrencilerinin okuma becerilerini ölçmek için geliştirilen bir ölçeğin psikometrik özelliklerini üçüncü sınıf öğrencilerinde incelemek yanlış sonuçlara ulaşmamıza neden olacaktır. Benzer şekilde üniversite öğrencileri üzerinde yapılan bir araştırmayı yetişkin örnekleme, tanı almamış bireylerle yapılan bir araştırmayı bir psikopatoloji grubuna genellemek uygun değildir. Özellikle klinik ve sağlık araştırmaları gibi araştırmaya konu olan durumun yaygınlığı ile ilgili çalışmalar mevcutsa ulaşılması hedeflenen örneklem büyüklüğü Lemeshow, Hosmer, Klar ve Lwanga'nın (1990) önerdiği $n = N \cdot t^2 \cdot pq / d^2$ ($N-1$) + $t^2 \cdot pq$ formülü ile (N = evrendeki birey sayısı, n = örnekleme alınacak birey sayısı, p = incelenen olayın görüş sıklığı, q = incelenen olayın görülme sıklığı, t = belirli serbestlik derecesinde ve saptanan yanılma düzeyinde t tablosundan bulunan kuramsal değer, d = olayın görülüş sıklığına göre yapılmak istenen \pm sapma) hesaplanabilir. Daha detaylı örneklem sayısı tespiti için yazarların yayına ulaşmak mümkündür.

İstatistiksel Analizler

Kullanılan İstatistik Paket Programı

Gelişen teknoloji ile birlikte sosyal bilim araştırmalarında kullanılabilecek istatistik paket programlarının sayısı da artış göstermiştir. Teknik olarak kullanılan iterasyon yöntemi, replikasyon sayısı, yakınsama kriteri gibi özellikle faktör analitik yöntemler kullanılırken önem arz eden tanımlamalar programdan programa hatta programın versiyonları arasında da değişmektedir. En çok kullanılan paket program SPSS olmakla birlikte bu programında pek çok versiyonu bulunmaktadır. Bu nedenle *araştırmacılar mutlaka analizlerini yaptıkları programı ve versiyonunu makalelerinin yöntem ya da bulgular bölümünde belirtmelidir.* Yine pek çok araş-

tırmada metin içinde geçen istatistik paket programının kaynakçada yer almadığını görmek mümkündür. APA (2010) yazım kılavuzuna göre, SPSS gibi hemen herkes tarafından bilinen programların kaynakçada yer alma zorunluluğu yoktur; fakat istatistik paket programları “Telif sahibi, A. A. (Yıl). Programın adı (Versiyon numarası) [Bilgisayar yazılımı]. Yer: Yayımcı.” şekilde kaynaklara eklenebilir (APA, 2010).

Verinin Analize Hazırlanması

Verilerin toplanması ve analize hazırlanması süreci aslında araştırmanın en çok zaman harcanan bölümü olmasına rağmen makalelerde bu sürece çok az değinilmektedir. Ancak istatistikleri iyi okuyabilmek için okuyucu, aykırı verilerin nasıl temizlendiğini, *eğer kayıp veriler varsa bu verilerle nasıl başa çıkıldığı bilgisine ihtiyaç duymaktadır*. Bununla birlikte analizlerden önce mutlaka kullanılacak yöntemin varsayımlarına uygunluk test edilmelidir. Ayrıca raporlanma kısmında *verilere ait ortalama, standart sapma gibi betimleyici istatistiklerin* yer alması gerekmektedir. Bununla birlikte çarpık ya da basık dağılımlar faktör analizinde sahte faktörler çıkmasına neden olabilmektedir. Örneğin içerik olarak birbirinden bağımsız iki maddenin katılımcılar tarafından popüler maddeler olarak başka bir deyişle pek çok katılımcı tarafından “tamamen katılıyorum” şeklinde cevaplanması bu iki maddenin birleşerek ayrı bir faktöre yük vermesine neden olacaktır. Böylesi bir durumda elbette faktörleri yorumlamak güçleşecektir. Bu nedenle betimleyici istatistikler verilirken basıklık ve çarpıklık katsayılarının da raporlanması okuyucuların faktör yapısının nasıl yorumlandığını anlamalarına yardımcı olacaktır.

Doğrulamalı faktör analitik teknikler kullanacak araştırmacılar için özellikle çoklu normallik varsayımı oldukça önemlidir. Bu varsayımın ihlali ile tahminlenen parametreler yanlı olacaktır. Bu nedenle normal dağılımı sağlamayan veri setleri için varyans-kovaryans matrisi yerine örneğin asimptotik varyans-kovaryans matrisi gibi uygun bir yöntem kullanılmalı ve makalede belirtilmelidir.

Geçerlik

Bir ölçeğin geçerliği çeşitli yöntemler kullanılarak araştırılabilir. Bir ölçme aracının ölçülmek istenen yapıyı ölçtüğünün kanıtlanması süreci olarak tanımlanabilecek geçerlik çalışmaları kapsam, kriter ve yapı geçerliği olarak üç temel yöntemle incelenebilir.

Yeni geliştirilen ya da başka bir kültüre uyarılanmakta olan ölçme aracının öncelikle ölçülmek istenen yapının tüm bileşenlerini kapsayan maddelerin olup olmadığı uzmanlarca değerlendirilir. Örneğin istatistik dersine yönelik tutumların ölçülmek istendiği bir durumda, sadece matematiğe ilişkin korkuların ölçüldüğü

sorular yeterli olmayacaktır. Bilgisayar kullanımı, istatistik ve sosyal kaygılar bile bu ölçme aracının kapsamına girmesi beklenen boyutlar olabilir.

Bir ölçme aracının geçerliğini kanıtlamanın bir diğer yolu da, bu yeni ölçme aracıyla benzer yapıları ölçen başka bir ölçme aracı ya da ölçüm sonucu ile ilişkisinin incelenmesidir. Kriter geçerliği olarak da adlandırılan bu süreçte araştırmacı geliştirdiği ya da uyarladığı ölçme aracının ilişkili olduğu diğer ölçümlere (kriter) ait verileri de elde eder. Kriter ölçümünün sonuçları ile geliştirdiği/uyarladığı ölçme aracının sonuçları arasındaki korelasyon katsayısı ya da birden fazla kriter kullanıldığında regresyon denklemi geçerliğin kanıtı olarak kullanılmaktadır. Kapsam ve kriter geçerliğini incelemek isteyen araştırmacılar Kline(2005), Nunnally (1978) gibi kaynaklarda ayrıntılı bilgiye ulaşabilirler.

Ölçek geliştirme veya uyarılama çalışmalarında araştırmacılar sıklıkla yapı geçerliğini faktör analitik yöntemlerle incelemektedir. Bir örtük değişken modeli olan faktör analizi gözlenen değişkenler arasındaki ilişkinin bir örtük değişken varlığında ortadan kalkması ya da tesadüfi düzeye inmesi şeklinde tanımlanabilir (McCutcheon, 1987). Yapı geçerliği çalışmalarında söz konusu örtük değişken incelemeye konu olan yapı olarak varsayılmaktadır.

Açımlayıcı ve doğrulamalı faktör analizi yöntemlerinden hangisinin kullanılması gerektiği ve hangi değerlerin raporlanması gerektiği ise en sık karşılaşılan sorulardandır. Genel olarak önsel bir bilginin olmadığı durumlarda açımlayıcı, var olan bir kuramın (örn., Faktör yapısı) test edilmesinde ise doğrulamalı faktör analitik teknikler kullanılmaktadır. Açıkça iki yöntem de gözlenen değişkenlerden örtük yapıyı çıkarsamaya yönelik işlemleri kapsamaktadır ve örneğin açımlayıcı faktör analizinde faktör sayısına kısıt koyarak doğrulamalı bakış açısıyla, doğrulamalı faktör analizini ise tüm parametreleri serbest bırakarak açımlayıcı bakış açısıyla kullanmak mümkündür. Doğrulamalı faktör analizinin (DFA) açımlayıcı faktör analizinden(AFA) belki de en önemli farkı klasik test kuramının ölçmeye karışan hatalar arasında korelasyon olmadığı yönündeki varsayımını esnetebilmesidir. Bu noktadan bakıldığında test geliştirme süreci için AFA, adaptasyon için ise DFA daha uygun görünmektedir. Fakat unutulmamalıdır ki ister uyarılama ister geliştirme çalışması olsun bir ölçek yaratma sonsuz döngüye sahip bir süreçtir (Kline, 2005). Bu süreçte farklı farklı örneklerde defalarca yapılan çalışmalar sonucu elde edilmektedir. Bu noktada özellikle AFA ile belirlenmiş bir yapının DFA ile de test etmek isteyen araştırmacılar ya alternatif modeller stratejisi ile -örneğin bir faktörlü yapıya karşı iki faktörlü modelin test edilmesi ya da bir maddenin birden fazla faktöre yük verdiği modelin tek faktöre yük verdiği modele karşı test edilmesi gibi- en iyi modeli aramak için bu yöntemi

kullanmalı ya da *farklı bir örnekleme* aynı yapıyı test etmelidirler.

Bu kısımda AFA ve DFA uygulamalarında izlenmesi gereken adımlar ve raporlanırken dikkat edilmesi gereken noktalara değinilecektir.

Açımlayıcı Faktör Analizi. Pek çok diğer analiz gibi AFA sıklıkla SPSS paket programı ile yürütülmektedir ve bu programda fabrika ayarı temel bileşenler analizi şeklindedir. Temel bileşenler analizi adından da anlaşılacağı üzere bileşen sayısının tespitinde kullanılabilir bir yöntem olmakla birlikte aslında bir faktör analizi yöntemi değildir. Bu yöntemde paylaşılan varyans hata varyansını da barındırmaktadır (Kline, 2013). Bu nedenle bu yöntem kullanıldığında açıklanan varyansın daha şişkin çıkacağını akılda tutmakta fayda vardır. Bununla birlikte güvenilirlik arttıkça faktör analizi ve temel bileşenler analizi benzer sonuçlar vermektedir (Thompson, 1992). Pek çok alternatif faktör çıkarma yöntemi arasında en popüler olan iki yöntem ise temel eksen faktörleme (principal axis factoring) ve en-çok-olabilirlik (maximum likelihood) yöntemleridir. Bu yöntemleri detayları ile öğrenmek isteyen okuyucular için Kline (1994), Kline (2013) ve Thompson'ın (1992) çalışmaları kaynak olarak önerilebilir. Bu noktada araştırmacıların yapması gereken paket programın fabrika ayarına güvenmek yerine kendi veri setleri ve araştırma sorularına uygun yöntemi belirleyip kullandıkları *faktör çıkarma yöntemini* makalelerinde belirtmeleridir.

Faktör Sayısının Belirlenmesi. Faktör analizinde faktör sayısının belirlenmesi oldukça önemlidir. Faktör sayısının belirlenmesi için de pek çok yöntem önerilmiştir. Bandalos ve Finney (2010) bu yöntemleri istatistiksel temelli, matematiksel temelli ve sezgisel olarak üç grup şeklinde sınıflandırmaktadır. Paralel analizler ve Barlett anlamlılık testi istatistiksel temelli yöntem, özdeğerin birden büyük olması koşulu matematiksel modele ve saçılma diyagramını kullanmak sezgisel yaklaşıma örnek olarak verilebilir. Bu yöntemlerle ilgili detaylı bilgi Thompson'da (1992) bulunabilir. Pek çok bilgisayar yazılımının fabrika ayarı olarak bulunan ve en yaygın kullanılan yöntemin özdeğeri birin üzerinde olan faktörlerin modele dahil edilmesidir. Faktör sayısına karar verirken birden çok yöntemi bir arada kullanarak en uygun sonucun seçilmesinde fayda vardır. *Raporlarken dikkate alınan kriterler ve bu kriterlerin işaret ettiği faktör sayısı belirtilmeli ve karar verilen faktör sayısı kuram ile uyumlu şekilde temellendirilmelidir.*

Faktör sayısına karar verdikten sonra, AFA sonuçlarında döndürülmemiş faktör yükleri genellikle yorumlanabilir sonuçlar vermezler. Açımlayıcı faktör analizi daha önce de belirtildiği üzere herhangi bir önsel beklentinin olmadığı bir yöntemdir. Bu nedenle verinin daha iyi uyum sağladığı faktör yapısını bulmak için doğrulayıcı faktör analizinde bulunmayan bir basamak olan faktör

döndürme yöntemi basamağı önemlidir. Geometrik bir yöntem olan faktör döndürme yöntemi ile daha kolay yorumlanabilir çıktılara ulaşmak mümkündür. Pek çok alt versiyonu bulunmakla birlikte faktör döndürme iki temel şekilde yapılmaktadır; dik ya da eğik. Temel fark faktörler arasında korelasyon varsa eğik, faktörler ilişkisiz ise dik döndürme yöntemini kullanmaktır. Önsel bir bilginin olmadığı durumlarda elbette araştırmacı iki yöntemi de deneyip hangi sonucun kuramına daha uygun ve yorumlanabilir olduğuna karar verebilir. Bu noktada *kullanılan döndürme yöntemi makalede belirtilmelidir.*

Modelin Son Halinin Analizi ve Yapı Katsayıları (Faktör Yüklerinin) Sunulması. Faktör analizi ilk sonuçları yorumlanabilir olmayabilir. Fakat döndürme işlemi yapıldıktan sonra ister geliştirme ister uyarılama çalışması olsun bazı maddelerin amaçlanan doğrultuda çalışmadığı tespit edilecektir. Faktör yükleri ve madde-toplam puan korelasyon incelemeleri sonucunda ölçekten çıkarılması hedeflenen maddeler belirlendikten sonra model tekrar test edilmelidir. Bu işlem model son halini alana kadar devam edebilir. Bu noktada *ölçekten kaç madde çıkarıldığı ve çıkarılma gerekçesi* belirtilmelidir. Genellikle faktör yükü .30'un altında olan maddeler ölçekten çıkarılsa da daha yüksek faktör yüküne sahip maddelerin de kurama uygun olarak ölçekten çıkarıldığı örnekler mevcuttur. Faktör yükleri genel olarak kullanılan bir terim olsa da faktör analizi sonucunda yapı katsayıları (structure coefficients) ve örüntü katsayıları (pattern coefficients) olarak adlandırılan iki farklı çıktı karşımıza çıkmaktadır. Dik döndürme yöntemleri kullanıldığında ya da faktörler arasındaki korelasyon düşük olduğunda iki çıktı birbirine yakın sonuçlar vermektedir. Ancak eğik döndürme yöntemlerinin kullanıldığı ve faktörler arasındaki ilişkinin yüksek olduğu durumlarda bu iki çıktı farklılaşacaktır. Böylesi bir durumda her iki katsayının da raporlanması önerilmektedir (Bandalos ve Finney, 2010). Son halini alan ölçeğe ilişkin *madde faktör yükleri ve açıklanan varyans* mutlaka raporda yer almalıdır. Birden çok faktörlü ölçekler için her bir faktörün açıkladığı varyans raporlanır.

Faktör analizi bulgularında son adım bu katsayılar göz önünde bulundurularak faktör yapısının tespiti ve *isimlendirilmesidir.*

Doğrulayıcı Faktör Analizi. Doğrulayıcı faktör analizi önsel bir bilgi ile kurulan hipotetik modelin test edilmesini kapsamaktadır. Temelde açımlayıcı faktör analizinde tahmin edilen parametrelere kısıt konarak yeni bir model oluşturulmasıdır. Ölçek geliştirme ya da uyarılama çalışmalarında doğrulayıcı modeller hangi maddelerin hangi faktörle ilişkili olduğu, kaç faktör olduğu, faktörlerin ilişkili olup olmadığı gibi bilgilerin araştırmacı tarafından sunulmasını kapsamaktadır. Doğrulayıcı faktör analizi yöntemini kullanırken en az bir maddenin çapa (anchor) madde olarak belirlenmesi ge-

rekmetedir. Hangi maddenin neden çapa madde olarak kullanıldığı açıklanmalıdır.

Bu noktada tıpkı açılıcı faktör analizindeki gibi gözlenen değişken sayısı ve kuram sunulmalıdır. Doğrulatoryıcı faktör analizinde de model oluşturulduktan sonraki adım hangi tahminleme yönteminin kullanılacağına karar verilmesidir. LISREL, AMOS, Mplus gibi çokça kullanılan programlarda en-çok olabilirlik yöntemi fabrika ayarı olsa da özellikle varsayımların ihlalinde farklı çıkarsama yöntemleri kullanılabilir. Araştırmacılar hangi yöntemi kullandıklarını eğer normallik ihlali nedeniyle asimptotik varyans-kovaryans matrisini kullandırlarsa Satorra-Bentler Ki-kare istatistiğini raporlamalıdır.

Model Uyumu ve Alternatif Modellerin Test Edilmesi. Açılıcı faktör analizinin aksine doğrulatoryıcı faktör analizi kullanıldığında araştırmacılar bir model uyum indeksi elde etmektedir. Model uyumunun değerlendirilmesi DFA da önemli ve karmaşık bir süreçtir. Bu karmaşıklığın en önemli nedenlerinden birisi çok sayıda farklı istatistik ve bilgi kriterinin çıktı olarak karşımıza çıkmasıdır. Tıpkı diğer analizlerde olduğu gibi burada da tek bir kriter göz önünde bulundurularak karar verilmesi doğru olmayacaktır. En çok kullanılan ve raporlanması beklenen model uyum istatistikleri şu şekildedir:

Ki-kare istatistikleri doğrulatoryıcı modellerde büyük örneklem kullanıldığı, modelin ki-kare dağılımı göstermemesi gibi nedenlerle çok fazla eleştiri olsa da serbestlik derecesi ile birlikte mutlaka raporlanması gereken istatistiklerin başında yer almaktadır. Model uyum göstergesi olarak ki-kare'nin anlamlılığı yerine ki-kare'nin serbestlik derecesine bölünmesiyle elde edilen sayıların kullanılması genel kabul görmüştür. Ki-kare ile benzer özellikler taşıyan bir diğer kriter SRMR (standardized root mean square residual) değeridir. İyi uyum için beklenen SRMR değeri .08 ya da daha küçük değerlerdir. RMSEA (root mean square error of approximation) değeri yine en çok raporlanan bilgi kriterlerinden biridir. Bu değerin ya da değerin %90 güven aralığında alabileceği değerlerin .05'den küçük olması iyi model uyumuna işaret etmektedir. Bununla birlikte .08 ile .05 arasındaki değerler kabul edilebilir uyum olarak raporlanmaktadır. CFI (comparative fit index) ve NNFI (non-normed fit index) ise .95 ve üzerinde olmalıdır (Thompson, 2004).

Alternatif modellerden hangisinin daha iyi uyum gösterdiği test etmek için ise eğer modeller iç-içe geçmiş (nested) ise Ki-kare fark istatistiklerinden yararlanmak mümkündür. Karşılaştırılacak iki model için hesaplanan Ki-kare değerleri arasındaki fark hesaplanarak Δ Ki-kare değeri elde edilir. Bu değerin istatistiksel olarak anlamlılığı ise söz konusu iki modelin serbestlik dereceleri arasındaki farkın kritik değeriyle karşılaştırılarak kontrol edilir. Ayrıca yukarıda değinilen bilgi kriterlerinin bir modelde diğerinden daha iyi olması durumunda bu mode-

lin daha iyi uyum gösterdiği söylenebilir. İç-içe geçmiş modellerde ise alternatif olarak AIC (Akaike bilgi kriteri), BIC (Bayesian bilgi kriteri) gibi kriterlerden yararlanılabilir. Bu bilgi kriterlerinde daha düşük değer daha iyi modeli işaret etmektedir. Model karşılaştırmalarında test edilen modellerin en iyisinin tespit edildiğini, tespit edilen bu iyi modelin sadece diğerlerinden iyi olduğunun kanıtlandığını not etmekte fayda vardır.

DFA uygularken sıklıkla ilk tanımlanan model en iyi alternatif olmamaktadır. Tahminlenmesi istenen bazı parametreler anlamsız olarak tespit edilmekte ya da model kurulurken tanımlanmamış bazı ek parametrelerin tahminlenmesi için modele eklenmesi ihtiyacı duyulmaktadır. Örneğin bir maddenin birden fazla faktörle ilişkilendirilmesi, bazı maddeler arasındaki korelasyonların serbest bırakılması, bazı maddelerin hata varyanslarının birbirleriyle ilişkilendirilmesi ve benzeri pek çok düzeltme modelin iyileşmesine katkı sağlıyor olabilir. Bu tür düzeltme indisleri sonucunda modele eklenen ya da çıkartılan parametreler belirtilmeli, kuramsal olarak tartışılmalı ve bu son modele ait uyum istatistikleri de mutlaka raporlanmalıdır.

Faktör Yüklerinin Raporlanması. Son olarak modele karar verildikten sonra faktör yükleri raporlanmaktadır. Faktör yükleri grafik üzerinde raporlanabilmektedir. Fakat ölçek uyarılama ya da adaptasyon çalışmalarında grafik üzerinde gösterilen yol katsayılarına ek olarak, standardize olmayan sonuçlar ve maddelere ait standart hatalar da raporlanmaktadır. Bu kadar bilgiyi bir görsele sığdırmak çok kolay olmayabilir. Bu nedenle bir tabloda bu katsayıların ve ek olarak faktörlerin açıklanma yüzdelerinin belirtilmesinde fayda vardır.

Güvenirlilik

Bir ölçeğin tamamına ve alt boyutlarına ait güvenirlik bulguları sıklıkla Cronbach's Alfa güvenirlik katsayısı ile hesaplanarak raporlanmaktadır. Bu noktada hangi güvenirlik yönteminin araştırmalarına uygun olduğunu belirlemek araştırmacıların inisiyatifindedir. Her güvenirlik yönteminin kendine özgü avantaj ve dezavantajları bulunmaktadır. İç-tutarlık güvenirlik katsayısı adından anlaşılacağı üzere maddelerin birbirleri ile olan tutarlılığının bir ölçüsünü vermektedir. En genel haliyle ölçmenin tesadüfi hatadan arınıklığı olarak tanımlanabilen güvenirliği iç-tutarlıktan ibaret olarak yorumlamak hatalı sonuçlar elde etmemize neden olacaktır. İç tutarlık raporlanırken en sık karşılaşılan hatalardan biri ise iki uçlu veriler için raporlanan alfa katsayısıdır. Sijtsma'nın (2009) tartışma makalesinde Cronbach alfa katsayısının yanlış kullanımı ve güvenirlik göstergesi olarak sınırlarına ait detaylı bilgiye ulaşmak mümkündür. Araştırmacılar ölçeğe ait güvenirlik değerlerini sıklıkla klasik kuram kapsamında raporlasalar da modern kuram başka bir deyişle madde tepki kuramı ve son yıllarda örtük sınıf

modelleri (Van der Ark, Van der Palm ve Sijtsma, 2011) kapsamında da güvenilirlik hesaplamaları yapılabilmektedir. Bu noktada çalışılan konu ve hedefe yönelik uygun yöntemi seçmek önemlidir. Eğer bir ölçek uyarlama çalışması yapılıyorsa ölçeğin diğer çalışmalarda elde edilmiş güvenilirlik değerleri belirtilmelidir. *Bulgular kısmında ölçeğin bütününe ve varsa alt boyutlarına ait güvenilirlik katsayısı kullanılan yöntem de belirtilerek raporlanmalıdır.* Sık karşılaşılan sorulardan birisi güvenilirlik katsayısı değerinin en az kaç olması gerektiği yönündedir. Lance, Butts ve Michels (2006) pek çok makalenin Nunnally'yi (1978) kaynak göstererek .70 katsayısının yeterli olduğu yönündeki bilginin eksikliğine vurgu yapmaktadır. Aksine Nunnally (1978), sadece araştırmanın başında zamandan tasarruf etmek isteyen araştırmacıların kabul edilebilir düzeyde en az .70 güvenilirlik katsayısına ulaşmaları gerektiğini ancak ölçek kullanılarak bireyler hakkında örneğin işe giriş, tanı ve benzeri kararlar verilecekse .90 hatta .95 güvenilirlik katsayısının hedef olması gerektiğini belirtmektedir. Yazarlara göre ayrıca Nunnally güvenilirlik kestirme yöntemi olarak spesifik olarak alfa katsayısını işaret etmediğini belirtmektedir. Psikometri alan yazınında Cronbach's Alfa katsayısının madde varyanslarına dayanan güvenilirliği kestirme yöntemi olarak belirtildiği ve iç tutarlılığa yönelik bilgiler verdiğini not etmekte fayda vardır. Bununla birlikte örneğin ölçek maddelerinin homojen maddelerden oluşmadığı durumlarda ya da hız testi gibi ölçek maddelerinin tamamının katılımcılar tarafından cevaplanamayacağı durumlarda bu katsayıyı kullanmak uygun değildir. Araştırmacıların ölçeğin iç tutarlılığından çok zaman içindeki kararlılığını test etmek istedikleri durumlarda ise özellikle öğrenme, olgunlaşma gibi etkilerin görünmeyeceği, kişilik gibi daha durağan psikolojik yapıların incelenmesinde test-tekrar test güvenilirlik yönteminin kullanılması uygun olacaktır. Bu yöntem kullanıldığında uygulama sonucu elde edilen puanlar arası korelasyonun yüksekliği ölçme aracının güvenilir olduğuna işaret etmektedir.

Kline'a (2005) göre güvenilirlik çalışmalarında özetle;

(i) Hız testi dışındaki ölçeklerde homojen maddeler varsa iç-tutarlılık katsayısı raporlanabilir.

(ii) Alternatif formlar kullanılarak güvenilirlik kestirimi hız testleri için uygundur.

(iii) Test-tekrar test güvenilirlik yöntemi, öğrenme, olgunlaşma gibi etkilerin görülmeyeceği yapılarda kullanılmaya daha uygundur. Test-tekrar test yöntemi ölçeğin zamana karşı kararlılığının ölçüsünü vermektedir.

(iv) Test geliştirme aşamasının başında .70 kriteri kullanılabilir ancak, kritik kararlar için kullanılacak ölçeklerde .70 güvenilirlik katsayısının %49 hata varyansı içerdiği unutulmamalıdır.

Makalelerde sıklıkla önce geçerlik ve sonra güvenilirlik bulguları sunulmaktadır. Aslında güvenilirlik geçer-

liğin ön koşuludur fakat raporlanan güvenilirlikler ölçek son halini aldıktan sonraki katsayılardır.

Sonuç

Yazının genelinden anlaşılacağı üzere ölçek geliştirme ve uyarlama çalışmaları teknik olduğu kadar öznel öğeler de içermektedir. Tüm süreç sağlam bir kuramsal bilgi ve teknik donanım gerektirmektedir. Psikometri alanında son yıllarda özellikle teknolojik gelişmelere paralel olarak klasik kuramın ötesinde yeni model önerileri ile hızla gelişmektedir. Borsboom (2006) çalışmasında psikometristler tarafından geliştirilen modern tekniklerin uygulama alanında yer bulmama nedenleri tartışmıştır. Yazar, fazla teknik yazılan çalışmaların anlaşılabilmesi ve yaygınlıkla kullanılan istatistik paket programlarında yeni gelişmelerin kullanıcı dostu olarak henüz eklenmemiş olmasının etkilerini vurgulamıştır. Bir diğer etken psikoloji eğitiminde genellikle klasik kurama ve diğer analizlere yer verilmesi olarak belirtilmiştir. Bu çalışma uygulama alanında sıklıkla tercih edilen yöntemlerle sınırlı olsa da araştırmacılar ölçme eşdeğerliği, tepki yanlılığı, yöntem yanlılığı gibi kavramları da göz önünde bulundurarak sorularına en uygun yöntemi bulabileceklerdir.

Kaynaklar

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6. baskı). Washington, DC.
- Bandalos, D. L. ve Finney, S.J. (2010). Factor analysis: Exploratory and confirmatory. G. R. Hancock ve R. O. Mueller, (Ed.), *Reviewer's guide to quantitative methods* içinde.
- Borsboom, D. (2006). The attack of psychometricians. *Psychometrika*, 71(3), 425-440.
- Costello, A. B. ve 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).
- 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.
- DeVellis, R. F. (2014). *Ölçek geliştirme kuram ve uygulamalar* (3. baskı). (T. Totan, Çev. Ed.). Ankara: Nobel Akademi Yayınları. (Orijinal çalışma basım tarihi 2012).
- Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme-I temel kavramlar ve işlemler* (2. baskı) Pegem Akademi: Ankara.
- Kankaras, M. ve Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe. Insights from a multiple group latent class factor approach. *International Sociology*, 24(4), 557-579.
- Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.
- Kline, R. B. (2013). Exploratory and confirmatory factor analysis, Y. Petscher ve C. Schatsschneider, (Ed.), *Applied quantitative analysis in the social sciences* içinde (171-207). New York: Routledge.

- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks: Sage Publications Inc.
- Lance, C. E., Butts, M. M. ve Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods*, 9(2), 202-220.
- Lemeshow, S., Hosmer, D. W. Jr, Klar, J. ve Lwanga, S. K. (1990). *Adequacy of sample size in health studies*. Chichester: Wiley/World Health Organization.
- McCutcheon, A. L. (1987). *Latent class analysis*. Beverly Hills: Sage Publication.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor & Francis Group.
- Nunnally, J. C. (1978). *Psychometric theory (2. baskı)*. New York: McGraw-Hill.
- Öztürk, N. B., Eroğlu, M. G. ve Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim*, 40(178), 123-137.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107-120.
- Sousa, V. D. ve Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*. doi:10.1111/j.1365-2753.2010.01434.x.
- Thompson, B. (1992). A partial test distribution for cosines among factors across samples. B. Thompson, (Ed.), *Advances in social science methodology* (Bölüm: 2) içinde (81-97). Greenwich, CT: JAL.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis understanding concepts and applications*. Washington, DC: APA.
- Van der Ark, L. A., van der Palm, D. W. ve Sijtsma, K. (2011). A latent class approach to estimating test-score reliability. *Applied Psychological Measurement*, 35(5), 380-392.

EK 1. Ölçek Geliştirme ve Uyarlama Çalışmaları İçin Kontrol Listesi

Kuramsal yapının işlevuruk tanımı yapıldı.	
Alanyazındaki eğer varsa diğer ölçekler belirtildi, bu ölçek çalışmasının gerekliliği açıklandı.	
Madde havuzu oluşturma basamakları açıklandı. (Kapsam geçerliği tartışıldı) Çeviri süreci açıklandı.	
Ölçekle ilgili bilgiler sunuldu. (Madde sayısı, puanlama yöntemi, örnek madde)	
Örnekleme yöntemi belirtildi ve araştırmanın amacına uygunluğu tartışıldı. (temsil edicilik, büyüklük)	
Örnekleme ilişkin bilgiler sunuldu.	
Araştırmada kayıp veri varsa kayıp veri ile başa çıkma yöntemi belirtildi.	
İstatistiksel analizlerin yapıldığı paket program ya da programlar ve versiyonları belirtildi.	
Araştırma verilerine ait betimleyici istatistikler sunuldu.	
Faktör analizi ile incelenecek araştırmalarda açımlayıcı ya da doğrulayıcı faktör analitik tekniklerden hangisinin, neden kullanılacağı belirtildi.	
AFA için; i. Faktör çıkarma yöntemi belirtildi. ii. Faktör döndürme yöntemi belirtildi. iii. Faktör sayısının hangi kriterlere göre belirlendiği ve kaç olduğu belirtildi. iv. Ölçekten çıkarılan maddeler varsa belirtildi. v. Faktör yükleri sunuldu. vi. Açıklanan varyans (toplam ve varsa alt boyutlar için) sunuldu.	
DFA için; i. Konsept model sunuldu. ii. Parametre tahminleme yöntemi belirtildi. iii. Modifikasyonlar belirtildi. iv. Model uyum istatistikleri sunuldu. v. Faktör yükleri sunuldu. vi. Açıklanan varyans sunuldu.	
(Kriter geçerliğini inceleyecek araştırmacılar için) Seçilen kriter ölçümü ve neden bu ölçümün kriter olarak belirlendiği açıklandı.	
Kriter olarak alınan başka bir ölçümle, geliştirilen/uyarlanan ölçme aracının ilişkisinin derecesi raporlandı.	
Kullanılan güvenilirliği kestirme yöntemi ve bu yöntemi seçme nedeni belirtildi.	
Güvenirlik bulguları sunuldu.	
Ölçeğin son haline ilişkin bilgiler özetlendi.	
Elde edilen ölçeğin literatür ile uyumu tartışıldı.	
Elde edilen ölçeğin katkıları ve kültüre uygunluğu tartışıldı.	

Summary

A Guide to Scale Development and Adaptation in Psychology

Duygu Güngör
Dokuz Eylül University

In psychology most of research variables cannot be observed. To study with this kind of unobserved constructs also named as latent variables, we need valid and reliable measurement instruments. The aim of this study is to present a guide to researchers who develop or adapt measurement instruments. First, fundamentals of scale development and adaptation will be introduced. Then, I will mention sampling briefly. Statistical analyses part includes software, preparing data for analysis, and validation methods. Next exploratory and confirmatory factor analysis will be presented. Lastly, I will mention on reliability methods and how to report findings. Additionally, a checklist for researchers will be presented.

To Adapt or To Develop A Scale

There are numbers of measurement instruments in psychology. Before deciding to adapt or develop a scale, it is important to find out and clearly write in the introduction part of research that this is a necessity. After literature review if a researcher cannot find an appropriate scale for her study she has two alternatives; to develop or adapt a scale. Both alternatives have their own advantages and disadvantages. For instance, although adaptation seems easier while adapting a scale due to translation process some meanings might be lost or some items might not be valid in the target culture.

If researcher decides to develop a scale, she should follow following steps: Literature review, interview with experts, preparing item pool. Adaptation includes: translation, back-translation and pilot study.

Sampling

One of the questions that needs an answer is about sample size in a study. We should note that sample size is about power and sampling method is about representativeness. Checking power tables gives a clue about

sample size. There is not a golden rule in the adaptation or development scale studies but 20 participants per item seems favorable (Kline, 2013).

Statistical Analysis

Statistical Software

There are numbers of statistical packages that can be used in psychology research. Different packages use different iteration methods, replication numbers etc., moreover, same package different versions can be varying and give different results. *It is important to mention which statistical software and which version is used.*

Preparing Data

Although collecting data and preparing data for analysis are one of the time consuming part of studies this process is rarely mentioned in manuscripts. To understand results more clearly audience needs to see; *how missing data is handled, descriptive statistics* of data like average, standard deviation, skewness and kurtosis.

Validity

Validity of an instrument can be explored with three main methods: content validity, criterion validity and construct validity (Kline, 2005; Nunnally, 1978). Most of the adaptation and development studies contain construct validation with factor analysis. One of the main questions is which factor analytic approach should be used; exploratory or confirmatory. In adaptation studies confirmatory and development studies exploratory approach seems more suitable.

Exploratory Factor Analysis. Researchers using exploratory factor analysis should report following steps; extraction method, number of factors and which criteria used to decide number of factors. Rotation method and explained variance after rotation. Lastly structure and/or pattern coefficients and the naming procedure of

factors should be reported (Bandalos & Finney, 2010).

Confirmatory Factor Analysis. In confirmatory factor analysis studies following steps should be reported. Estimation method, model fit statistics, factor coefficients and explained variance.

Reliability

Cronbach's Alpha coefficient is one the most used indices of reliability however, Sijtsma (2009) discussed misuse of Alpha in his article. Lance, Butts, and Michels (2006) also discussed using .70 alpha referring Nunnally (1978). Actually Nunnally pointed for scales that used

for decision alpha coefficient should be at least .90.

In summary, sample representativeness is more important than sample size and if scales contain homogenous items internal consistency reliability methods are suitable excluding speed tests.

Conclusion

In conclusion although this paper includes more classical methods, researchers should consider measurement invariance, response bias and method bias during their adaptation or development process.