

# Ölçek Geliştirme ve Uyarlama Çalışmalarında Karşılaşılan Sorunlar

## Adnan Erkuş\*

Mersin Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme ABD  
adnanerkuspsi@mersin.edu.tr

Psikoloji alanındaki bilimsel çalışmaların büyük çoğunluğu, özgün veya uyarlama sonucu kültürümüze kazandırılan psikolojik ölçme araçlarıyla elde edilen verilere dayanmaktadır. Bu çalışmalar sonucunda elde edilen bulguların genellenebilirliği, işlevselliği ve sağlamlığı da, kullanılan bu ölçme araçlarının güvenilirliği ve geçerliğiyle paralellik göstermektedir. Bu bakımdan, araştırma problemi istenildiği kadar yenilik getirici ve önemli, araştırma düzeneği ve yöntemi istenildiği kadar sağlam, istatistiksel analizler de istenildiği kadar doğru yapılmış olsun; kullanılan ölçme aracı sağlam olmadığında, elde edilen bulgular da önemini yitirmektedir. Bu nedenle; araştırma sürecinin her aşaması önemli olmasına karşın, doğrudan gözlenemeyen değişkenleri ölçme hüneri gösteren meslektaşlarımızın kullandığı ölçme araçlarının niteliği hepsinden de önemli olmaktadır. Son yıllarda, akademisyenlerin üzerindeki yayın baskısının da itici gücüyle, alanımızdan veya yakınız alanlardan pek çok akademisyen, ne yazık ki bu engeli aşmada “en kolay” yol olarak ölçek geliştirme veya ölçek uyarlama çalışmalarını görmektedir ve bu da pek çok hatayı ve eksikliği beraberinde getirmektedir. Aşağıdaki kısımlarda bu hata ve eksiklikler üzerinde durularak bir farkındalık yaratılmaya çalışılacaktır.

### I. Karşılaşılan Sorunlar

#### 1. Çalışmanın Başlığı

Her türlü bilimsel makalenin başlığı, çalışmanın içeriğini özetleyici, görgül bir çalışmaysa değişkenlerini -yöntemini- hatta örneklemi niteleyici, ifade bozukluğu taşımayan, ne uzun ne de

kısa olmalıdır. Ölçek geliştirme ve uyarlama çalışmaları da, öncelikle, özgün bir ölçek “geliştirme” mi, yoksa “uyarlama” mı olduğu bilgisini başlıkta yansıtmalı; ölçeğin ismini içermeli; önemliyse, kimler için geliştirildiği yer almalı; “... Ölçeğinin Güvenirliği ve Geçerliği” gibi bir başlıktan (her türlü ölçek geliştirme çalışması zaten ölçeğin güvenirliliği ve geçerliğini içerir) kaçınılmalıdır.

Sıklıkla karşılaşılan bir sorun ise uyarlama çalışmalarında gözlenmektedir. “Questionnaire” sözcüğü dilimize, doğru bir şekilde, “sormaca (anket)” olarak çevrilmiştir. Özellikle Fransa ve Kanada çıkışlı ölçek çalışmalarının veya “ilgisiz” alanlarda yapılan çalışmaların çoğunda, “scale” yerine, “questionnaire” kullanılmaktadır. Uyarlama çalışması yapan araştırmacı da, ölçme aracının içeriğine ve diğer özelliklerine bakmadan, birebir çeviriyle “questionnaire”i “anket (sormaca)” olarak çevirmektedir. Sormaca, doğası gereği belirli bir psikolojik yapı değildir; pek çok demografik ve/veya diğer değişkenler hakkında genellikle “bilgi” toplamaya yarayan bir araçtır, toplam puanı anlam taşımaz, teknik anlamda madde analizi-güvenirliliği ve geçerliği de söz konusu değildir. Bu nedenle, psikolojik bir yapıyı ölçmeye yönelik olarak hazırlanmış, alt veya toplam ölçek puanları bulunan, madde ve tepki formatları aynı maddelerden oluşmuş bir ölçme aracının “anket” olarak nitelenmesi doğru değildir. Öte yandan, sınır yeterlik (maksimum performans) dışındaki değişkenleri (kişilik, ilgi, tutum vb.) ölçen ölçme araçlarına “test” denmesi de doğru değildir.

## 2. Ölçülmek İstenen Örtük Değişkenin Kavramsal Tanımı

Bir ölçek geliştirme çalışması, ölçülecek psikolojik değişkenin, çok detaylı ve kapsayıcı bir şekilde, kuramsal ve kavramsal bir tanımlama girişimi ile başlar; çünkü geliştirilecek olan ölçeğin kendisi, bu doğrudan gözlenemeyen örtük değişkenin işevuruk tanımının ta kendisidir. Bu kavramsal tanım, bir veya birkaç kurama dayalı olabileceği gibi, gözlemler ve diğer kavramlardan hareketle de yapılabilir; ancak önemli olan, kavramsal tanımın, o kavramı niteleyen ortak özellikleri kapsamaya ve o kavrama yakın kavramlardan farkını da nitelemesidir.

Ölçeğin kapsam geçerliği bu aşamada temellendiği ve sonraki her türlü davranışsal gösterge ve o örtük değişkeni uyaran maddeler bu tanım üzerinde yapılandırıldığı için, bu aşama oldukça önemlidir. Aslında, bu aşama o örtük değişkene ilişkin bir kuram (veya denence) geliştirmeye eşdeğerdir ve ölçek de onun test edilmesinden başka bir şey değildir. Bu bakımdan, ölçek geliştirici, ölçülecek değişken ile ilgili sağlam bir kuramsal ve kavramsal temele sahip olmalıdır; bir ölçme-değerlendirmeci veya psikometrist, bilmediği, hakim olmadığı bir psikolojik değişkeni ölçme girişiminde bulunmamalıdır; elbette, tersi de (ölçmeden anlamadan) olmamalıdır.

Bu sorun, daha çok psikoloji alanı dışındakilerin ölçek geliştirme çalışmalarında sıkça gözlenmektedir: Bir turizmci veya bir işletmeci, hiç tanımadığı-bilmediği bir psikolojik değişkene ilişkin ve hemen hiç bilmediği ölçek geliştirme "temelsizliğiyle" (etik boyutunu bir yana bırakalım) ölçek geliştirme işine soyunabilmektedir. Çalışmanın bir ölçek uyarlaması olması da, kavramsal tanımlamanın "atlanacağı" kolaycılığına yol açmamalıdır.

## 3. Madde Üretilmesi

Madde üretim aşaması -örtük değişkenin uyarılarının üretilmesi aşaması- çok önemlidir. Madde üretimi kavramsal yapıya uygun durum-bağlam-ortam belirlendikten sonra, bunlara uygun davranışsal göstergeler belirlenmesiyle başlar, önceki kavramsal çerçeve ve içeriğe uygun bir madde formatının seçilmesi-belirlenmesiyle devam eder. Madde üretim süreci, üretilen maddelerin kavramsal yapı ve onun üzerinde temellenen davranışsal göstergelerin gözetilmesini gerektirir. Madde üretilmesi sürecinde, madde yazarı bir yandan kafasının bir yanında sürekli ölçülecek değişkenin içeriğini ve maddenin uyardığı tek şeyin ilgili değişken olduğunu düşünmeli; öte yandan potansiyel yanıtlayıcılarla empati kurarak "ben şöyle birisi olsaydım, nasıl cevap verirdim" diye kendine sormalıdır.

Örneğin, "kıskançlık" ölçeği geliştireceksiniz ve varsayın ki, bu konuda yapılmış hiçbir kavramsal tanım ve ölçek yok. Kıskançlığı, "kendinde olanı başkalarıyla paylaşmama, başkasında olana imrenme" olarak tanımladınız; bu iki ayrı boyuta da, iki bileşene de karşılık gelebilir (henüz bilmiyorsunuz). Bu tanıma uygun davranışsal göstergeler neler olabilir? Bu karar, "ölçeğin kimler için geliştirileceği" sorusuna karşılık gelir: Evli bireyler mi, bekarlar mı? Diyelim ki, "evli bireyler" olarak belirledik; sadece erkek veya kadınlar mı, yoksa her ikisi mi olduğuna karar vermeniz gerekir. Diyelim ki, evli kadınlar (görüldüğü gibi ölçeğin hedef kitlesi belirginleşmeye başladı); evliliği devam eden veya sonlananlar mı, yoksa her ikisi mi karar vermeliyiz. Görüldüğü gibi, bu kararlar "durumu" (situation/setting) belirlemeyi gerekli kılıyor. Diyelim ki, evliliği devam eden ve her iki cins... O zaman, yukarıdaki kavramsal tanıma uygun, evli bireyler için davranışsal göstergeler bulmak durumundayız: Bir davette dans, giyim-kuşam gibi... Bu

göstergeler çok değişik biçimlerde (ölçek ve madde formatı da dikkate alınarak) maddelere dönüştürülebilir: (Örn.; bir Likert tipi ölçek için) “Eşimin başkalarıyla dans etmesine dayanmam”, “Benim eşim de niye X gibi dans etmiyor”, “Eşimin dikkat çekici giyinmesi beni çılgına çevirir”, “Benim eşimin de X gibi giyinmesini isterim” gibi. Çok önemli olan bu aşama ve süreç, ne yazık ki pek çok çalışmada ya yeterince önemli görülmemekte ya da “atlanmaktadır”. Öte yandan bu süreçte kullanılan yöntem de önemlidir: Araştırmacının sadece kendisi tarafından mı, bir grup madde yazarı tarafından mı yazıldı? Yazılan maddeler yargıcılara verilmiş ve maddelerin anlam, boyut, ifade ve benzeri açılardan yargıcılar arası uyumuna bakılmış ve buna göre maddeler atılmış veya düzeltilmiş midir? Uygulamada, sadece bir grup (2-3, o da hocası-kendisi) yargıcıya maddeler verilir, başka hiçbir istatistiksel işlem yapmadan ölçeğin kapsam geçerliğinin sağlandığı ileri sürülmektedir. Bir uyarılma çalışması ise, doğrudan araştırmacı tarafından veya bir grup (tek veya iki dilli) tarafından mı çevrilmiş, çevirenler ilgili kavramsal yapıya aşina mı, çeviriden sonra ters çeviri yapılmış mı, çeviri geçerliği için neler yapılmış (iki dili bilen bir gruba uygulanıp Türkçe ve diğer dille yazılmış form arasındaki korelasyona bakılmış mı) gibi soruların yanıtları çok önemli olmaktadır (syntax ile semantics arasındaki fark).

Uygulamada bu sürece yeterince önem verilmediği (birileri masaya oturup, kendi çalıp kendi oynuyor!) için uyarıcı madde köklerinin ifadelerinde tutum ölçülecekse, kişiliğin veya görüşlerin bulaştığına çokça tanık olunmaktadır. Büyükçe bir örneklemden ayakkabı numaraları, yaşları, kaç yıl okudukları, boy uzunlukları, Türkçe final puanları gibi ilişkisiz pek çok değişkene ilişkin veri toplansa, onlara madde analizi yapılırsa, hiç kuşkunuz olmasın pek çoğu bir “ölçek” oluşturur!

#### 4. Deneme Uygulaması

Deneme uygulamasının niceliği ve niteliği tüm madde analizini etkiler. Bu nedenle deneme uygulamasının kimler üzerinde yapılacağı, ölçülecek değişkene ve hedef grubun kimler olacağına bağlıdır. Deneme uygulaması örneklemindeki grubun yaş aralığı, eğitim düzeyi, cinsiyeti gibi pek çok özelliği ölçeğin kullanılacağı asıl hedef grubun özellikleriyle aynı olmalıdır. Deneme uygulaması örnekleminin seçkisiz (random) seçilmesi doğru değildir; amaçlı örnekleme (non-random) söz konusudur. Bu konudaki amaçlı örneklemede ise, amaç sadece ilgili değişken aralığının temsil edilmesidir. Türbana ilişkin geliştirilen bir tutum ölçeğini sadece türbanlılara veya sadece türbana karşı olanlara uygularsanız; hem ölçeğin uygulanacağı hedef kitleyi sınırlandırmış olacağımızdan ölçeği herkese uygulayamazsınız, hem de varyans daralacağından madde analizleriniz sizi yanıltır (atılmayacak madde atılır, atılacak madde bırakılır). Deneme uygulaması örnekleminin büyüklüğü de madde analiz tekniklerine bağlı olmakla birlikte, ne kadar büyük n sayısı o kadar doğru analiz ve o kadar sağlam kararlar anlamı taşır; “faktör analitik tekniklerle madde analizi yapılacaksa, kx10 birey” gibi ölçütlerin doğru olmadığı anlaşılmıştır (MacCallum, Widaman, Zhang ve Hong, 1999). Uygulamada, asıl hedef örneklemden başka örneklemler üzerinde -yeterli olmayan birey sayıları gibi- çalışmalar yapıldığına tanık olunmaktadır. Uyarılma çalışmaları için de yukarıdaki açıklamalar geçerlidir; orijinal çalışmanın örneklemine sadık kalınmalıdır.

#### 5. Madde Analizi

Madde analizi, kabaca, maddelerin başlangıçta öngörülen psikolojik yapıya uygunluğunun test edilmesini içerir. Maddeler, ilgili özelliğe sahip olanlar ile olmayanları ayırt ediyor mu; ilgili boyut içinde yer alıyor mu; ilgili boyut üzerin-

deki yerleri ne gibi sorulara yanıt alınmak için madde analizi yapılır. Madde analizinde kullanılacak teknikler ölçülen psikolojik yapıya, madde formatına, maddelerin puanlanma biçimine vb. bağlı olarak değişir. İşlem sonunda da uymayanlar atılır veya tekrar gözden geçirilir (işlem tekrar edilir). İster özgün ister uyarılama çalışması olsun, madde analizi diğer niteliklerin irdelenmesinden önce gelir: Uygulamada görülen bir hata, özellikle SPSS paket programında "reliability" işlemi içinde madde analizi (ki, o da sadece madde-toplam ölçek korelasyonları ve maddeler arası korelasyon matrisidir) verildiği için, pek çok araştırmacı tarafından madde analizinin "güvenirlik" alt başlığı altında verilmesidir ve bu doğru değildir. Güvenirlik ve geçerlik, madde analizinden sonra, üstelik başka örneklemeler üzerinde kanıt aranması gereken psikometrik niteliklerdir ve bu nedenle zamansal olarak madde analizinden sonra yer alırlar. Madde analizi süreci, sadece istatistik bilgisi ile gerçekleştirilemez; istatistiksel sonuçların, ilgili yapıya bağlı olarak "okunmasını-koklanmasını" da gerektirir.

Madde geçerliği için (her bir madde ölçeğin bir parçası, bir başka deyişle ölçek onu oluşturan maddelerin bir fonksiyonudur; dolayısıyla tek tek ölçme aracı neyi ölçmek için hazırlanmışsa, onu ölçüp ölçmediği gösterilmelidir) madde-toplam ölçek korelasyonları kullanılacaksa, madde puanı ile ölçek puanının hangi düzeyde ölçmeye (sınıflama, sıralama vb.) uygun olduğu bilinmelidir: Madde 1-0'lı puanlanıyorsa, bu gerçek süreksiz mi, yapay süreksiz mi olduğuna bağlı olarak nokta çift serili veya çift serili korelasyon tekniğine başvurulur ve özellikle SPSS paket programında bu korelasyon teknikleri olmamasına rağmen ne yazık ki Pearson Momentler Çarpımı Korelasyon tekniğinin kullanıldığı sıklıkla gözlenmektedir. STATISTICA gibi başka pek çok paket program bu ko-

nuda kullanılabilir. Öte yandan klişe ve birilerine atfen ve özellikle başarı testleri için kullanılan .30 madde-toplam ölçek korelasyon katsayısı ölçütüne sık sık sığılmaktadır. Oysa orta düzeydeki bir korelasyon, çok boyutlu bir yapıyla karşı karşıya olunduğunun işaretini vereceği gibi, başarı dışındaki değişkenler için işe yaramaz bir maddenin göstergesi de olabilir. 'Ne kadar yüksek madde geçerliği, o kadar iyi ayırt edici madde' demek olduğuna göre, ölçek geliştirici olarak, niye .30 ölçütüne sığınılır anlamak çok (aslında işin kolaylığı) zor.

Öte yandan, madde analizi için başvuru faktör analizi sonuçlarının "geçerlik" veya "yapı geçerliği" başlığı altında verildiğine de tanık olunmaktadır. .30 - .35 faktör yükü ölçütü de, ne yazık ki dikkatle yaklaşılması gereken bir ölçüttür. .30 - .35'lik faktör yükü, o maddenin başka faktörlere de benzer yük verebileceğinin ve bu nedenle kompleks bir maddeyle karşılaşıldığının işaretini verir; böyle bir durumda ya madde çıkarılır ya da çeşitli döndürme işlemlerine başvurularak ölçeğin yapısı incelenir. Uygulamada çeşit çeşit faktör analizi olmasına rağmen, hangisinin yapıldığının belirtilmemesi; verilerin faktör analizi yapmaya uygun olup olmadığının test (Kaiser-Meyer-Olkin ve Barlett's testleri) edilmemesi; açıklanan varyans ve özdeğerlerinin rapor edilmemesi; faktör sayısına doğru karar verilememesi; gereksiz yere döndürme yapılması veya yapıya uygun döndürme işleminin yapılmaması; faktör ile bileşenin ayırt edilememesi gibi pek çok eksiklik ve hataya rastlanmaktadır. Ayrıca, "Temel Bileşenler Analizi ile ilk ortaya çıkan faktör yapısı başlangıçtaki kavramsal öngörüye ne kadar uymaktadır, uymuyorsa bu ne anlama gelmektedir, kavramsal yapıyı, ölçek misorunludur"un yanıtı, alan bilgisi ile istatistik ve ölçme bilgisinin harmanlanmasını gerektirir; bunlardan biri eksik olduğunda verilecek kararlar da yanlış olur. Her ölçek, madde

atımıyla kolaylıkla tek faktöre indirgenebilir; ancak, ölçülen yapı gerçekten tek faktörlü müdür sorusu yanıtlanmadan kalmış olur. Bazen, öyle olur ki, özdeğeri ve açıklama gücü düşük en son birkaç maddelik bir faktör, ölçeğin gerçek yapısına ilişkin çok fikir verebilir; bu durumda belki yeni baştan yeni madde desteğiyle o faktörü güçlendirme yoluna gitmek gerekebilir. Görüldüğü gibi, sadece istatistik bilgisiyle veya istatistiksel sonuçlara bakarak (özellikle faktör analitik tekniklerle madde seçilecekse) ölçeğin faktör yapısına ve maddelerin atılıp atılmayacağına karar vermek yanlışlıklara yol açabilir; ölçek geliştiricinin o madde ve test istatistiklerini “okuması-koklaması” gerekmektedir.

Özellikle uyarılma çalışmalarında faktör analitik tekniklerle maddelerin incelenmesi çok daha fazla özen gerektirir. Psikolojik değişkenlerin çoğu kültüre bağımlıdır ve bu nedenle kültürlere (sadece başka kültürlere değil, aynı kültür içinde farklı hedef gruplara da uyarılma çalışılması yapılır) “çeviri değil” “uyarlama” çalışmaları yapılır. Bu durumda, ölçek geliştiricinin yukarıdakilere ek olarak ayrıca kültürel yapıyı dikkate alması gerekir; o kültürde 3 faktörlü olan bir psikolojik yapının bizde 4 faktörlü veya 2 faktörlü çıkması önemli bir bilgi taşır. ‘Zorlama’ bir şekilde faktör sayısını ‘denkleştirmeye’ çalışma ise uyarılma esprisini ortadan kaldırır; bu gibi durumlarda orijinal çalışmadaki faktör (alt ölçek) yapısı değil, bizim kültürümüzdeki yapı dikkate alınmalı, isimlendirme ve puanlama ona göre önerilmelidir. Piyasada ne yazık ki, uyarılma çalışmasında farklı faktör yapısı bulunmasına rağmen, orijinalindeki gibi kullanılan ve yorumlanan pek çok ölçek mevcuttur; böyle bir durumda uyarılma çalışması da esprisini yitirmektedir.

Alandaki potansiyel ölçek kullanıcılarının, ölçekler ile ilgili yapılan çalışmaları da takip etmesi gerekir. XIII. Ulusal Psikoloji

Kongresi’ndeki bir oturumda birbirinden ayrı ve habersiz, biri Geleneksel Kurama, diğeri Modern Kurama dayalı olarak yapılan iki çalışmada Beck Umutsuzluk Ölçeği’nin 12. maddesinin yapılan analizlerde sorunlu olduğu (büyük olasılıkla çeviride kültürümüze özgü “umuyorum” ile “ummuyorum” yüklemelerinin ayırt edilememesinden kaynaklanan), bu madde çıkarıldığında ölçeğin psikometrik niteliklerinin yükseldiği belirtilmişti. Ancak ne yazık ki, bu ölçeğin hala eskisi gibi puanlanmaya devam edildiği görülmektedir.

İstatistik paket programlarının yaygınlaşması ve analiz çeşitliliğinin zenginleşmesinden sonra, madde analizinde birden çok yöntemin kullanılması; bunun yanında tüm ölçek veya alt faktörler için iç tutarlık, toplanabilirlik gibi özelliklerin de karar vermede göz önünde bulundurulması önerilebilir.

## 6. Güvenirlilik

Güvenirlilik belirlemede en sık yapılanın SPSS paket programının “reliability” sini tıklamak ve maddeleri (bazen, toplam puanları ve hatta demografik değişkenleri de!) doldurduktan sonra OK’e tıklamak olduğu gözlenmektedir. Hangi güvenirlilik belirleme yolunun seçileceği maddelerin puanlanmasına, ilgili yapının tek ya da çok boyutlu olup olmamasına, ilgili değişkeni kararlılık durumuna vb bağlıdır. Uygulamada, ölçek çok boyutlu olmasına karşın, yarıya bölme güvenirliliğine bakıldığı; maddeleri ikili puanlanan ölçek için Cronbach Alfa hesaplandığı; sadece (hazır) iç tutarlık katsayılarından birine ya da en fazla ikisine bakıldığı, ölçeğin test-tekrar test güvenirliliğinin çok incelendiği; tek tek puanların yorumunda önemli olan ölçmenin standart hatasının hemen hiç verilmediği; SPSS çıktısı diye, madde-toplam ölçek korelasyonları ile korelasyon matrisinin güvenirlilik içinde verildiği vb. görülmektedir. Çoğu kez



de, deneme uygulaması örneklemeden elde edilen veriler üzerinde güvenilirliğe bakılıp bırakıldığı gözlenmektedir; oysa yapılması gereken, seçilmiş maddelerden oluşan nihai ölçeğin tekrar bir örnekleme üzerinde uygulanarak güvenilirliği sınamaktır. Çünkü denemelik ölçekteki çok sayıdaki maddeye verilen tepkilerle, azaltılmış-seçilmiş maddelerden oluşan ölçeğe verilen tepkiler çok farklı olabilir. Bir bilimsel toplantıda, bir ölçmecinin ölçek geliştirme sunumu yapan bir akademisyene, “tamam da, bu ölçeğin güvenilirliğine nasıl baktınız” sorusuna, sunum yapan akademisyenin, “yahu beni tanımıyor musun, bana güvenmiyor musun; biz birbirimize güvenmeyeceğiz de kime güveneceğiz” türü trajikomik açıklamalar artık geride kalmalıdır.

### 7. Geçerlik

Geçerlik, güvenilirlikten daha önemli olduğu halde, ne yazık ki ‘gerekten ilgiyi’ görmemektedir. Oysa ki, bir ölçek güvenilir olabilir, ama geçerli olmayabilir. Geçerlik, bir ölçeğin “işe yararlılığının” kanıtı olduğundan, üzerinde önemle durulmalı; ölçek geliştirici kendi geçerlik çalışmalarının ötesinde, ölçeğiyle yapılan tüm görgül çalışmaları izlemelidir, çünkü hemen hemen tümü ölçeğinin geçerliğine kanıt oluşturur. Uygulamada, madde analizinde kullanılan Temel Bileşenler Analizi sonuçlarının “yapı geçerliği” olarak adlandırılmasına ne yazık ki çok sık rastlanmaktadır. Oysa, adı üstünde yapının geçerliği için yeniden veri toplanıp ve başka istatistiksel analizlere (Doğrulayıcı Faktör Analizi, SEM gibi) başvurulması gerekir. Üstelik yapı geçerliği, sadece faktör analitik tekniklerle de yapılmaz. Ölçeğin geçerlik kanıtları için başvurulacak işlem ve değişkenlerin ipuçları, kavramsal tanımlamanın ele alındığı alan yazında büyük ölçüde yer alır. Örneğin depresyonun, olumsuz otomatik düşüncelerle ilişkili olduğu biliniyorsa, geliştirilen depresyon ölçeği-

nin ölçütü olumsuz otomatik düşünceler (ya da tersi) olur; zekanın (belirli) yaşa bağlı olarak arttığı biliniyorsa, geliştirilen zeka ölçeği yaş farklılıklarını ortaya çıkarabilmelidir, gibi. Ama bir bakıyorsunuz, araştırmacı hiçbir kuşku veya bulgu olmamasına rağmen, “bir de cinsiyetler arası farka bakalım” (bir şey yapmış görünmek, sayfayı doldurmak için olsa gerek!) diyerek bir t-testi yapıyor. Niye yaptın? Ölçülen değişkenin cinsiyete bağlı olarak değiştiği bekleniyor mu ki?! Geçerlik için gerekli pek çok değişkenle ve pek çok yolla kanıt toplanmalı, ilişkisiz değişkenler kullanılmamalı ve işlemler yapılmamalıdır.

### 8. Puanların Yorumlanması

Ölçek geliştirme çalışmasının ve raporlarının en çok ihmal edilen kısmı bu olmaktadır. Ölçek kullanıcısı, ölçekten elde edilecek bir ölçüme sahip birini nasıl değerlendirecek? Puanların artışı değişkendeki artışa ya da azalışa mı karşılık geliyor? Toplam puan elde edilebilir mi, yoksa alt ölçek puanları ayrı ayrı mı ele alınmalı? Peki toplam puanlar hangi ölçüte göre yorumlanacak? Normlar geliştirilmiş midir, geliştirilmişse nasıl? En çok görülen yanılgılardan biri de, “normların mutlaka içinde bulunulan grubun ham puanlarının standartlaştırılması” olduğu şeklindeki bilgidir; oysa, klinik psikoloji alanındaki pek çok ölçek bir kesme noktasına göre yorumlanır. Peki normlar nasıl oluşturulacak? Normlar, mutlaka ölçek geliştirme örneklerinden çok daha büyük örneklem üzerinde geliştirilmelidir; elde edilen dağılım istatistiklerine göre bir yöntem izlenmelidir. *Norm değişkeni, ilgili değişken ölçümlerini farklılaştıran potansiyel karıştırıcı değişkenlere göre belirlenir, gelişigüzel değil.* Ölçülen değişken kadın ve erkeğe göre değişmiyorsa, cinsiyet normları yapmanın anlamı yoktur. Bir diğeri ise, özellikle uyarılma çalışmalarında, sürekli değişkenin kategorileştirilerek

norm grupları oluşturulmasında gözlenmektedir. Örneğin, yurtdışında "reaksiyon süresi" ölçeğinin yaş normları (eşit aralıklı veya değil) diyelim ki, 18-25, 26-35, 36-40, 41-55, 56 ve üstü olarak belirlenmiş ve kullanılmış. Niye bu yaş dilimleri kullanılmış, çünkü bu gruplar arasında reaksiyon süresi açısından fark çıkmış da ondan; bir başka deyişle, görgül bir çalışma yapılmış da (ANOVA gibi) sonra bu gruplar norm grupları olarak saptanmış. Bizim "ölçekçimiz" de tutuyor aynı grupları (n sayısının düşüklüğünü de bir yana bırakırsak) norm grubu olarak alıyor; ancak yaş grupları arasındaki farkı da test ediyor veya etmeden örneğin ortalamaları da veriyor: Bir bakıyorsunuz, yaş gruplarının üçünün (örneğin, 26-35, 36-40 ve 41-55) reaksiyon süreleri ortalamaları hemen hemen aynı ve (yapılmışsa) ANOVA'da da bu gruplar arasında fark yok! Ancak "ölçekçimiz" bunu göz ardı edip (ya da bilmediğinden), tüm yaş grupları için yurtdışındaki gibi norm tabloları hazırlamış! Bizim kültürümüzde bu yaş grupları arasında fark yok, niye bu normları kullanıyorsunuz?

### 9. Çalışmanın Raporlaştırılması

Bir bilimsel çalışmanın raporlaştırılması, o çalışmayı yürütüş aşamalarına göre yapıldığında bir bütünlük oluşturur, okuyucu daha iyi kavrar. Ancak ne yazık ki, psikometri alanında bir dergimizin olmamasından da kaynaklanan bir şekilde, ölçek geliştirme çalışmaları kendi özgün yürütülüş aşamaları dışında, zorlama bir şekilde başka kalıplarda yazılmak zorunda kalıyor. Raporlaştırma konusundaki bu temel sorun bir yana, "ölçekçimiz" (yukarıdaki alt başlıklar ona göre sıralandı), rapor yazımında hiç madde analizine değinmeden güvenilirlik ve geçerliğe 'atlayıyor'; tablolarda, sadece kendi öngördüğü faktör altındaki faktör yüklerini vererek, bir maddenin diğer faktörlere verdiği yükleri 'gizliyor'; özdeğerleri ve/veya açıklanan varyansları

vermiyor vb. En sık karşılaşılan kapanış cümlesi ise, "Görüldüğü gibi ölçek güvenilir ve geçerlidir." Hiçbir ölçek tam olarak güvenilir ve geçerli değildir; hele hele yukarıdaki gibi hatalı, eksik bir çalışmadan sonra bunun iddia edilmesi son derece tehlikeli olmaktadır. Birçok habersiz kişi, tez veya makalenin sonundaki bu ölçekleri (yeterli ve doğru kabul edip) alıp araştırmalarında kullanıyor, bu ölçeğe dayanarak insanlar hakkında karar veriyor...

## II. Bu Sorunların Olası Nedenleri ve Çözüm Önerileri

Aşağıdaki kısımda (yine kişisel gözlemlere dayanarak), denemelik bir nedenler listesi ve olası çözüm önerileri sıralanacaktır.

### 1. Olası Nedenler

- a) Bilgisayarda istatistiksel paket programların yaygınlaşması ve bu programları bir şekilde kullanabilenin, kendisini istatistikçi (istatistikçinin de kendini ölçmeci) olarak algılaması ve o programlarla ölçek geliştirebileceği yanlışlığına düşmesi;
- b) Akademisyenler üzerindeki yayın baskısının artması ve özellikle psikoloji dışındaki alanlar için en kolay yol olarak ölçek geliştirme veya uyarılama çalışması olduğu yanlışlığına düşülmesi;
- c) Uygulama alanlarında, özellikle klinik ortamlarda çalışanlar tarafından bireyi tanımlamada en kolay yol olarak test ve ölçek geliştirmenin (veya uyarılamanın) görülmesi;
- d) Ölçme, ölçek geliştirme, istatistik ve araştırma teknikleri alanlarında ya hiç eğitim alınmaması veya çok yetersiz alınması;
- e) İlgili kavramsal alt yapıya sahip olunca o konuda ölçek geliştirilebileceği veya

ölçek geliştirme bilgi ve becerisine sahip olunca hakim olunmayan bir değişkene ilişkin ölçek geliştirilebileceği yanılıgı;

f) Ülkemizde her türlü ölçeğin *hiçbir denetime uğramadan* geliştirilebilmesi ve rahatlıkla kullanılabilmesi;

g) Bu konuda çalışanların ve özellikle akademik ortamda yapılan tez vb çalışmaların danışmanlarının sanrı düzeyindeki benlik algılarına sahip olması gibi.

Yukarıdaki nedenlerin tümü, aslında *ölçme ve ölçek geliştirmeyi hafife almak* şeklinde bir tek nedene indirgenebilir. Bir zamanlar, üstelik çok önemli bir psikolog hocamız, bir kongremizde, "YÖK'ün (o komisyondaymış) neden Psikometriyi doçentlik alanlarından çıkardığı" konusundaki bir soruya, "eee her psikoloğun kendi ölçme aracını zaten kendisinin geliştirme yeterliğine sahip olması gerektiği" şeklinde (böyle düşünenlere, APA onayı almış birçok dergiyi incelemelerini ve bu konudaki kitap ve yeni yöntem-teknikleri okumalarını öğütleyebiliriz) yanıtlamıştı! Psikometri bir (ağırlıkla) temel ve uygulamalı alandır; üstelik psikolojinin bir bilim olması ölçmenin, dolayısıyla psikometrinin gelişimiyle olanaklı olmuştur; üstelik psikometri alanındaki üretilenler kendi dışındaki istatistik alanına bile önemli katkılar sağlamıştır. Psikometristlerin asli görevi ölçek geliştirmek değil, ölçek geliştirmek için yeni model, yöntem ve teknikler geliştirmektir. Öte yandan, psikometri, eleman yetişmesi en zor; ancak eleman eksikliği en çok hissedilen alanlardan biridir; bunun için ulusal ve uluslar arası bir tarama okuyucuya çok şey kazandırabilir.

## 2. Olası Çözüm Önerileri

a) Yasamızın bir an önce çıkarılarak;

i) Test ve Ölçek Denetleme Merkezi'nin kurulması,

ii) Bu merkeze bağlı Test ve Ölçek Bankası'nın oluşturulması,

iii) Böylelikle, geliştirilen ve uyarlanan her ölçeğin denetlenmesi, ölçeklerin bir tek merkezde toplanması, ölçek kullanıcılarının eğitilmesi, izinsiz ölçek kullanılmaması, her ölçek için küçük de olsa bir bedel ödetilerek ölçek geliştiricilerinin özendirilmesi-ölçek kullanıcılarının caydırılması vb. sağlanmalıdır.

b) Zaman geçirilmeden, bu alanda eleman yetiştirmek için;

i) yurtiçi bütünleştirilmiş lisanüstü programların açılması,

ii) yüksek lisans, doktora ve doktora sonrası eğitim için yurtdışı eğitim olanaklarının yaratılması,

iii)ulusalhizmetiçi programlarının yaşama geçirilmesi,

iv) Türkçe yayın gereksinimini karşılamak için destek sağlanması gerekmektedir.

c) Ulusal ve sonra da giderek (en azından) bölge ülkelerini kapsayacak şekilde uluslar arası bir "Psikoloji ve Eğitimde Ölçme Değerlendirme Dergisi"nin bir an önce yayına başlaması zorunludur. Bu önerinin önemsenmesi ve yaşama geçirilmesi sayısız yarar sağlayabilir: Böylelikle, giderek başka alanlardakiler de olmak üzere ölçek geliştirme çalışmaları disipline olmaya başlayacak; başka dergilerin yazım formatından ve gevşekliğinden kurtulunulacak ve asıl önemlisi, yukarıda önerilen merkezlerin doğuşu için zemin hazırlanmış olacaktır.

Bilimsel amaçlarla ve ciddi kararlar vermek için kullanılacak ölçekler, hafta sonu magazinlerindeki gibi üç soruda "partnerinizi ne kadar kıskanıyorsunuz test edin" türü eğlencelik zırvalardan çok farklıdır; ölçek geliştirmek çok zor, çok ciddi, ama çok keyifli bir iştir. "Çok zor" olması bir ölçüde yukarıda yazılanlardan çıkarsanabilir; "çok ciddi" olması, bu



ölçme araçlarına dayanarak araştırmalar yapılması ve bireyler hakkında ciddi yaşamsal karar vermekte yatmaktadır; “çok keyifli” olması ise, ölçek geliştiricinin kimsenin doğrudan gözleyemediği bir şeyi gözleyebilme becerisi göstermesidir.

### Kaynaklar

MacCallum, R. C., Widaman, K. F., Zhang, S. ve Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4 (1), 84-99.

\* Bu sınırlı uzam taşıyan yazıda ele alınanlar, yazarın dergi ve diğer konulardaki hakemlik ve danışmanlıkları ile kişisel yaşantılarına dayanmaktadır; bu nedenle, bu yazıda örneklere hiç atıf yapılmaması uygun görülmüştür ve daha pek çok konu, eleştiri ve öneri bu yazıya eklenebilir. Yazının hedefi hiçbir kurum veya kişi değildir, sadece bu konuda bir farkındalık yaratmak ve bu yönde bazı girişimleri başlatabilmektir. Bu alanda bir özel sayı çıkardıkları ve bana böyle bir fırsatı verdikleri için Türk Psikoloji Bülteni Yayın Kurulu'na teşekkür ederim.