

KÜLTÜRLERARASI ÖLÇEK UYARLAMASI İÇİN REHBER II: Psikometrik Özellikler ve Kültürlerarası Karşılaştırma

Doç. Dr. Sebahat GÖZÜM*
Prof. Dr. Seçil AKSAYAN**

ÖZET

Ölçek uyarlama çalışması yürütecek olan araştırmacılara rehber olması amacıyla hazırlanan makalenin bu bölümünde, uyarlama çalışmalarının aşamalarından psikometrik özelliklerin-güvenirlilik ve geçerlik-sınama yöntemleri ve kültürlerarası karşılaştırma üzerinde durulmuştur.

Anahtar kelimeler: Ölçek uyarlama, psikometrik özellikler, güvenirlik, geçerlik, kültürlerarası karşılaştırma

A GUIDE FOR TRANSCULTURAL ADAPTATION OF THE SCALE II: Psychometric Characteristics and Cross-Cultural Comparison

ABSTRACT

In this section of the article, analyses of reliability and validity of an adapted scale, and cross-cultural comparison between the original and adapted scales are reviewed.

Key words: Adaptation of a scale, psychometric characteristics, reliability, validity, cross-cultural comparison

GİRİŞ

Ölçek geliştirmede olduğu kadar uyarlama çalışmalarında da, her ölçme sonucunda en azından iki temel psikometrik özelliğe ilişkin bilgi aranmaktadır. Bunlar **güvenirlilik** ve **geçerliktir**. Öner'in (1994) belirttiği üzere, Türkiye'de geliştirilen ya da uyarlanan testlerin çoğunda güvenirlik, geçerlik ve kültüre/topluma özgü norm verilerini bir arada bulmak güçtür. Psikometrik nitelikleri sınanmış olsa da, eğer testlerin güvenirliği ve geçerliği düşük düzeyde ise kullanılması sakıncalıdır. Kaçınılmaz olan, ölçme hatasını en aza indirebilmek için test güvenirliğinin yüksek olmasının yanı sıra, ölçülmek istenilen tutumun/davranışın doğru ölçüldüğünden emin

olmak için de test geçerliğinin saptanmış olması gerekir. Bunlar sağlanmadığı zaman, uygulamada bireyler hakkında karar vermek (seçme, tanı, sağaltım, yönlendirme vb.) ya da araştırma bulgularını geçerli kabul etmek, geriye dönülmesi olanaksız ve zararlı sonuçlar yaratabilir (Öner, 1994).

Bir araştırmada ölçüm hataları iki biçimde görülebilir (Portney, Watkins, 1993; Wood, Haber, 1994; Peirce, 1995);

- **Sistemik** (ölçek iyi kalibre edilmediğinde oluşur. Örneğin, bir tartı aletinin kişinin gerçek ağırlığından daha az/çok göstermesi) ya da

*Atatürk Üniversitesi Hemşirelik Yüksekokulu

** Kocaeli Üniversitesi Sağlık Yüksekokulu

• **Rastlantısal** (değişen kişisel koşullar, çevre koşulları ya da ölçüm aracındaki değişiklikler nedeniyle oluşabilir). Bu tür ölçüm hataları güvenilirlik için daha önemlidir.

Güvenilir bir ölçek, geçerli olabilir ya da olmaz. Ancak, güvenilirliği olmayan bir ölçek hiçbir zaman geçerli değildir (Peirce, 1995; Portney, Watkins 1993).

GÜVENİRLİK/GÜVENİLİRLİK (Reliability)

İlgili kaynaklarda güvenilirlik için verilen tanımlar, birbirine eşdeğer ya da birbirini tamamlayan kavramlardır. Güvenirlik; bir ölçme aracının duyarlı, birbiriyle tutarlı ve kararlı ölçme sonuçları verebilmesi, diğer deyişle, aynı değişkenin bağımsız ölçümleri arasındaki kararlılıktır; ölçülmek istenen belli bir değişkenin, sürekli olarak aynı sembollerini almasıdır; aynı süreçlerin izlenmesi ve aynı ölçütlerin kullanılması ile aynı sonuçların alınmasıdır; ölçmenin rastlantısal yanılğılardan arınmış olmasıdır. Daha açık bir ifadeyle, bir gruba ya da bireye uygulanan testten bireylerin aldıkları puanların, testin her uygulamasında kararlı, benzer olması beklenir. Test her defa aynı koşullarda uygulandığında elde edilen puanlar önemli düzeyde farklılık gösteriyorsa, testin güvenilirlik derecesinin düşük olduğu anlaşılır (Karasar, 1995; Özgüven, 2000; Tezbaşaran, 1996). Özetle güvenilirlik; değişmezlik, yeterlilik, kestirim, eşdeğerlik ve tutarlılığın sağlanmasını anlatır (McLaughlin, Marasculio, 1990; Wood, Haber, 1994). Diğer deyişle, ölçüm aracının yinelenebilir sonuç verme yeteneğidir. Kavramsal ve istatistiksel olarak güvenilirlik, bir testteki gerçek farklılıkların, toplam farklılığa oranıdır (Peirce, 1995).

Ölçeğin (testin) Güvenirliğini Bulma Yöntemleri

Bir ölçeğin güvenilirlik katsayısının bulunmasında çeşitli yöntemlerden yararlanılmaktadır. Bu yöntemler; hata kaynaklarından hangisinin dikkate alınıp alınmadığına, içinde bulunulan

koşullara, ölçeğin tekrar uygulama olasılığının bulunup bulunmamasına, ölçeğin paralel formunun varlığına ya da yokluğuna, ölçeğin bir güç ya da hız testi oluşuna göre seçilip kullanılmaktadır (Özgüven, 2000; Salkind, 2000; Tezbaşaran, 1996; Portney, Watkins, 1993; Peirce, 1995).

Temelde güvenilirlik hesaplaması, bir ölçeğin kendi kendisiyle (içindeki maddelerin birbiriyle) olan ilişkisidir. Bu ilişkiyi sınamada kullanılan korelasyon formülleri; Pearson Momentler çarpımı, Spearman-Brown ve Kuder Richardson 20 ya da Cronbach alfa (alfa katsayısı) ve çift seri (bi-serial) teknikleriyle beirlenmiştir (Öner, 1997; Peirce, 1995; Portney, Watkins 1993; Wood, Haber, 1994). İlgili kaynaklarda, birçok güvenilirlik ölçütünden söz edilmekle birlikte, bunlar temelde üç başlık altında toplanmaktadır (Öner, 1987; Öner, 1997; Özgüven, 2000; Karasar, 1995; Polit, Hungler, 1997; Portney, Watkins, 1993; Wood, Haber, 1994):

I. *Değişmezlik (Stability)*

II. *Bağımsız gözlemciler arası ve içindeki uyum (Inter-rater and intra-rater consistency)*

III. *İç tutarlılık (Internal consistency)/ Tutarlılık (Homogeneity) (Tablo 1.)*

Herhangi bir ölçek uyarlama çalışmasında bunlardan hangisinin kullanılması gerektiğine kolayca verilen bir yanıt yoktur. Aynı durumda kullanılacak çeşitli güvenilirlik katsayıları olabilir. Araştırmacı; araştırma problemi, kullanılacak ölçek, toplanacak veri türü ve yanıtlardaki olası objektifliğe göre birden fazla, çeşitli güvenilirlik sınamalarını gerçekleştirmelidir (Portney, Watkins, 1993).

I. *Değişmezlik (Stability)*

1. *Test-tekrar test (test-retest) güvenilirliği*

Test-tekrar test güvenilirliği, bir ölçme aracının uygulamadan uygulamaya tutarlı sonuçlar verebilme, zamana göre değişmezlik gösterebilme gücüdür (Karasar, 1995; Öner, 1987; Öner,

1997; Özgüven, 2000; Polit, Hungler, 1997; Tezbaşaran, 1996). Bu yöntem, ölçülen niteliğin değişmez olduğu durumlarda uygulanır. Testin ölçtüğü niteliğin sürekli değişkenlik gösterdiği durumlarda kullanılmaz. Örneğin; bilgi, tutum, ruh hali ve fiziksel durum gibi özellikler iki ölçüm arasındaki zaman aralığında herhangi bir nedenle değişebilir. Test-tekrar test güvenilirliği genel bilişsel yetenekler ve kişilik gibi zaman içinde çok hızlı değişiklik göstermeyen nitelikleri ölçen testler için daha uygundur (Özgüven, 2000; Polit, Hungler, 1997; Peirce, 1995).

Testin tekrarı yönteminde, iki farklı yol izlenir.

- **Aralıksız Yöntem:** Test bir gruba aralıksız ya da kısa bir dinlenmeden sonra uygulanır.

- **Aralıklı Yöntem:** Test iki ile dört hafta gibi bir zaman aralığı ile iki kez uygulanır.

İki uygulamada da gözlemci/ölçümcü, işlemler, saat ve çevre faktörleri aynı olmalıdır (Portney, Watkins, 1993; Peirce, 1995).

Test-tekrar test güvenilirliğinin en kritik yönü, iki ölçme arasında bırakılması gereken zaman aralığıdır. Bırakılan zaman aralığının çok kısa olması, yeniden anımsamayı kolaylaştıracağından, güvenilirliğin yapay olarak yüksek çıkmasına neden olabilir. Zaman aralığı uzun tutulduğunda ise, iki ölçme için "aynı koşullar"ın sağlanması olanaksız olabileceğinden, ölçülen özellikte bazı değişimlerin oluşması sonucu güvenilirlik ölçütünün yorumu güçleşir. Ayrıca bireyin, testin içeriği ile ilgili olarak başka kaynaklardan öğrendikleri, ikinci uygulamadaki puanını etkileyebilir (Karasar, 1995; Portney, Watkins, 1993; Peirce, 1995; Wood, Haber, 1994).

Bu yöntemin getirdiği diğer bir problem de, birinci uygulamada test üzerinde pratik yapmış olmanın ikinci uygulamadaki puanı yükseltici etkisidir (Özgüven, 2000; Portney, Watkins, 1993; Peirce, 1995). Bu olası sakıncalar göz önüne alınarak, aralıklı yöntemde bırakılacak

zaman aralığı iki haftadan az, dört haftadan fazla olmamalıdır (McLaughlin, Marasculio, 1990). Ayrıca, özellikle objektif ölçümler yapıldığında ikinci ölçüm sırasında ölçümcü/gözlemcinin yan tutma olasılığı da göz önüne alınmalıdır (Portney, Watkins, 1993; Peirce, 1995).

Olası sınırlılıklarına karşın, bir testin zamana göre değişmezliğinin en önemli ölçütü olan test-tekrar test tekniği, en sık kullanılan ve önerilen güvenilirlik göstergesidir. Ülkemizdeki hemşirelik alanında ölçek uyarlama çalışmalarının bir çoğunda kullanılmıştır (Çam, 1991; Esin, 1999; Gözüm, Aksayan, 1999; Nahcivan, 1993).

Geliştirilen ya da uyarlanan ölçeğin test-tekrar test güvenilirliğini bulmak için iki uygulamadan elde edilen puanlar arasındaki korelasyon hesaplanır. Test puanları sürekli bir değişken ve eşit aralıklı (interval) ölçek nitelikleri taşıdığı için güvenilirliğin bulunmasında **Pearson Momentler Çarpımı Korelasyon** eşitliği kullanılmalıdır (Karasar, 1995; Öner, 1987; Öner, 1997; Özgüven, 2000; Peirce, 1995). Bu hesaplama için, SPSS gibi istatistik paket programları kullanılabilir (Munro, Page, 1993; Salkind, 2000; McLaughlin, Marasculio, 1990; Peirce, 1995; Portney, Watkins, 1993). Sonuçlar (korelasyon katsayısı) doyurucu olsa bile, ileri analiz yapılmalıdır. Bunun için her iki testin ortalamaları ve standart sapmaları incelenmelidir. Değişken gerçekten stabil ise, her iki ölçeğin ortalama ve standart sapma değerleri yakın olmalıdır. İkinci ölçümdeki varyans artışı kişilerin duyarlılığının arttığını (test etkisi) gösterir (Peirce, 1995).

2. Paralel form güvenilirliği

Alternatif ya da eşdeğer form güvenilirliği olarak da anılan bu güvenilirlik ölçütü genellikle ölçek oluştururken kullanılan bir yöntem olup; iki form halinde ve eşdeğer nitelikte geliştirilmiş bir ölçeğin aynı gruba bir ya da iki oturumda uygulanmasından elde edilen puanlar arası korelasyonlar **Pearson Momentler Çarpımı** formülüyle hesaplanır. Test-tekrar test sınavındaki

test etkisini azaltmak için kullanılır. Testlerden alınan sonuçların ortalama ve standart sapmaları olabildiğince benzer ve ayrıca gruptan elde edilen üçüncü bir değişkenin değerleri ile olabildiğince ilişkili, yani hesaplanan korelasyon katsayısı yüksek olmalıdır (Peirce, 1995). Eşdeğer test formları geliştirmek güç olduğu için fazla kullanılmayan bir güvenilirlik ölçütüdür (Karasar, 1995; Öner, 1997; Özgüven, 2000; McLaughlin, Marasculio, 1990; Peirce, 1995; Wood, Haber, 1994).

Ölçek uyarlamasında bu yöntem; ancak, daha önce aynı amaçlar için geliştirilmiş bir ölçeğin (referans test) varlığında kullanılabilir ve bu iki ölçek arasındaki korelasyonun düzeyine bakılır. Bu güvenilirlik katsayısı aynı zamanda ölçeğin eşzaman (concurrent) geçerliğine yönelik veri sağlar.

Ülkemizde, hemşirelik alanında, Norton Bası Yarısı Değerlendirme Ölçeği'nin güvenilirliği, daha önceden dilimize uyarlanan Braden ölçeğiyle karşılaştırılarak, iki ölçek arasındaki ilişkinin düzeyi incelenmiş (Pınar, Oğuz, 1998), böylece hem paralel form güvenilirliği hem de eş zaman geçerliğine ilişkin bulgular elde edilmiştir.

II. Ölçümcü Güvenirliği (Bağımsız gözlemciler arası ve içindeki uyum)

Gözlemciler arasındaki uyumu belirleyen güvenilirlik ölçütü, özellikle verilerin gözleme dayalı olarak toplandığı ve birden çok gözlemcinin, önceden eğitilerek ve birbirinden bağımsız olarak, aynı durumu, aynı zamanda, aynı ölçüm aracı ile ölçmeye çalıştıkları durumlarda uygulanır. Bu tür ölçmelerde gözlemcilerin ayrı ayrı yaptıkları ölçümlerin ortalaması alınarak, her bir durum için bir tek değer bulunur. Aynı ayrı gözlem sonuçları birbirine ne kadar yakın ise, sonuçta elde edilen ortalama değer güvenilirliği de o kadar yüksek olur (Karasar, 1995; Özgüven, 2000; Portney, Watkins, 1993; Percel, 1995). Birden fazla ölçümcü arasında %70 ve daha yüksek tutarlılık, güvenilirlik sınaması için

uygundur. Ancak, tutarlılık yüzdesinin, şansa bağlı tutarlılığın yüksek olması nedeniyle yüksek çıkabileceği dikkate alınmalıdır. Bunun için yalnızca şansa bağlı olmayan beklenen uyumu gösteren Cohen Kappa istatistiği kullanılır (Hayran, Özdemir, 1996; Peirce, 1995; Portney, Watkins, 1993).

Bağımsız gözlemciler arası uyumu hesaplamak için korelasyon, t-testi ya da ikiden fazla gözlemcinin olduğu durumlarda özel varyans çözümlenmeleri ve Cronbah alfa kullanılır (Karasar, 1995; Peirce, 1995; McLaughlin, Marasculio, 1990).

Bu güvenilirlik ölçütü Pınar ve Oğuz (1998) ile Oğuz ve Olgun'un (1997) çalışmalarında, verilen gözleme dayalı olarak toplandığı Norton ve Braden Bası Yarısı Değerlendirme Ölçekleri'nin güvenilirliğinin sınanmasında kullanılmıştır.

Gözlemciler içi uyum sınamasında, iki ya da daha fazla gözlemin aynı gözlemci tarafından yapılarak puanlanması söz konusudur (Peirce, 1995; Portne, Watkins, 1993). Örneğin, anne-çocuk etkileşimine ilişkin bir video kasetinin aynı kişi tarafından farklı zamanlarda izlenmesi ve puanlanması. Burada da ölçümler arası tutarlılığın yüzdesi, güvenilirlik sınaması için en çok kullanılan istatistik yöntemidir. Ayrıca, yine iki ya da fazla ölçümcünün kendi içlerindeki uyumu göstermek için Cronbach Alfa ve şansa bağlı tutarlılığın etkisini azaltmak için ise Kappa istatistiği kullanılabilir (Peirce, 1995).

III. İç Tutarlılık

İç tutarlılığın dayandığı temel görüş, her ölçme aracının belli bir amacı gerçekleştirmek (bir bütünü oluşturmak) üzere, birbirinden deneysel olarak bağımsız ünitelerden oluştuğu ve bunların bir bütün içinde, bilinen ve birbirine eşit ağırlıklara sahip olduğu varsayımdır (Karasar, 1995). Bir ölçeğin iç tutarlılık güvenilirliğine sahip olduğundan söz edebilmek için, ölçeğin tüm alt bölümlerinin aynı özelliği ölçtüğünü kanıtlamak gerekir (Polit, Hangler, 1997). Aynı

özelliği ölçen maddelerin ayıklanması amacıyla yapılır (Portney, Watkins, 1993).

Bir ölçeğin iç tutarlılığını değerlendirmek üzere birçok eşitlik formülü geliştirilmiştir. Bunlardan en eskisi; *Testi yarılama yöntemi*'dir (split-half technique). Bu yöntem motivasyon, yorgunluk ya da zaman içerisinde performansı etkileyebilecek diğer psikolojik etkenler konularında ve özellikle de uzun testlerde kullanılır. Aynı fizik, mental ve çevresel etkenler söz konusu olduğundan test-tekrar test ve paralel formlara göre daha güçlüdür (Portney, Watkins, 1993). Bu yöntemde, test gruba bir kez uygulanmakta ve puanlamadan önce iki yarıya ayrılmaktadır. Bunun için en çok tercih edilen yarılama yöntemi, tek numaralı (1-3-5 gibi) soruları bir grup, çift numaralı soruları (2-4-6 gibi) bir grup yaparak ayırmaktır (Karasar, 1995; Özgüven, 2000; Polit, Hungler, 1997). Daha sonra, iki yarım test arasındaki korelasyon hesaplanır. Testin bütününe ilişkin güvenilirlik katsayısını elde etmek için *Spearman-Brown* tarafından geliştirilen bir eşitlikten (düzeltme formülü) yararlanılmaktadır. Testin iki yarısına ilişkin varyansın (standart sapmanın karesi) eşit ya da çok yakın olmaması durumunda bu eşitlik kullanılmamalıdır (Karasar, 1995; Öner, 1997; Özgüven, 2000).

Homojen bir yapıyı ölçtüğü varsayılan ve benzer maddelerden oluşan ölçme araçlarının bir tek uygulama ile güvenilirliğinin belirlenmesi, o ölçme aracının iç tutarlılığı hakkında bilgi verir. Bir tek uygulamayla güvenilirlik belirlemede çeşitli teknik ve formüller bulunmakla birlikte, en çok kullanılanları *Kuder Richardson 20 (KR-20)* ve *Cronbach Coefficient Alfa teknikleri*'dir. Kuder Richardson 20 ve Cronbach Alfa formülleri birbirlerinden türetilmiş olmasına karşın, hangisinin kullanılacağı, madde puanlarının ölçeklenme biçimine bağlıdır. Madde puanları süreksiz ise (var-yok, evet-hayır, doğru-yanlış veya 1-0 şeklinde) KR-20; sürekli ise (Likert tipi) Cronbach Alfa hesaplanması gerekir. Ayrıca *Kuder-Richardson 21 (KR-21)* eşitliği ise, ölçekteki her sorunun güçlük derecesinin aynı olduğu varsayımında kullanılabilir. Uygulamada bu varsayım nadiren

gerçekleştiği için KR-21 çok fazla kullanılmaz (Erkuş, 1999; Öner, 1997; Özgüven, 2000; Tezbaşaran, 1996; Wood, Haber, 1994).

SPSS programındaki "reliability-alpha" seçeneği Cronbach Alfa katsayısını hesapladığı için bu işlem, 1-0 şeklinde puanlanan maddeli testlere uygulanmamalıdır (Erkuş, 1999).

Likert tipi bir ölçeğin iç tutarlılığını değerlendirmek için, Cronbach Alfa ya da Alfa Katsayısı olarak bilinen güvenilirlik ölçütünün kullanılması gerekir. Birbiriyle yüksek ilişki gösteren maddelerden oluşan ölçeklerin alfa (α) katsayısı yüksek olur. Cronbach Alfa katsayısı ölçek içinde bulunan maddelerin iç tutarlılığının, homojenliğinin ölçüsüdür. Ölçeğin alfa katsayısı ne kadar yüksek olursa, bu ölçekte bulunan maddelerin o ölçüde birbirleriyle tutarlı ve aynı özelliğin öğelerini yordayan maddelerden oluştuğu varsayılır. Ancak, alfa katsayısı güvenilirlik katsayısını vermez, yalnızca güvenilirliğin bundan çok daha yüksek olacağına işaret eder. Ölçeğin güvenilirlik katsayısı alfa katsayısından daha yüksektir (Özgüven, 2000). Likert tipi bir ölçekte yeterli olabilecek alfa katsayısı, olabildiğince 1'e yakın olmalıdır (Peirce, 1995; Tezbaşaran, 1996).

Gerek geliştirilen, gerekse uyarılan bir ölçeğin iç tutarlılığını değerlendirmek üzere başvurulan bir diğer yöntem ise; *Madde istatistikleridir*. Madde istatistikleri ya da madde güvenilirliği olarak bilinen bu yöntemde, her bir test maddesinin varyansı, toplam test puanının varyansı ile karşılaştırılarak arasındaki ilişkiye bakılır. Bu ilişki *Pearson Momentler çarpımı korelasyonunun düzeltilmiş formülü* ile hesaplanır. Eğer test yanıtları evet-hayır, var-yok ya da 1-0 gibi kategorik iki seçenekli ise, hesaplamada bi-serial; seçenekler sürekli derecelendirmeli ya da likert tipi ise point-bi-serial teknikleri uygulanır (Öner, 1998; Öner, 1997; Portney, Watkins, 1993).

Ölçekteki maddeler, eşit ağırlıkta ve bağımsız üniteler şeklinde ise, her madde ile toplam değerler arasındaki korelasyon katsayısının

yüksek olması beklenir. Eğer bir maddenin toplam puana olan korelasyonu düşük ise, bu o maddenin testteki diğer maddelerden farklı bir niteliği ölçtüğünü gösterir (Karasar, 1995; Özgüven, 2000; Portney, Watkins, 1993; Wood, Haber, 1994).

Madde toplam puan korelasyon katsayısının hangi ölçütün altına düşünce güvenilirliğinin yetersiz sayılacağı konusunda belirli bir standart olmamakla birlikte, Karasar'a (1995) göre 0.50'den küçük katsayısı olan maddelerin güvenilirliğinden kuşku duyulmalı iken, Öner'e (1987) göre, bu katsayı 0.30'un üzerinde olmalıdır. Uygulamada ise çoğu araştırmacı 0.20 alt seviyesini kullanmaktadır.

Kültürlerarası ölçek uyarlaması yapılırken orjinal ölçekte kabul edilen katsayı sınırına, araştırmacının amacı ve yapılan ölçmenin niteliğine göre karar verilebilir. Yukarıdaki kaynaklara göre, belli bir düzeyin altında güvenilirliği olan maddelerin ölçekten çıkarılması ve toplam puan hesabına katılmaması gerekir. Uyarlanan ölçekteki birçok maddenin güvenilirlik katsayısı çok düşük çıkıyor ve atılması gerekiyor ise araştırmacının iki soruyu sorması ve yanıtlaması gerekir:

"Bu maddenin orijinal ve hedeflenen dildeki anlamı aynı mı?, Çeviri-uyarlama uygun mu?"

"Bu maddeleri çıkarırsam, orijinal ölçekle ve toplam puanla olan karşılaştırmaları nasıl yapabilirim?"

Güvenirlik Katsayısının Yorumu

Güvenirlik katsayısı korelasyon hesaplamalarıyla bulunur ve korelasyon katsayısı "r" olarak ifade edilir. Korelasyon katsayısı, iki değişken arasındaki ilişkinin "derecesi" ve "yönü" hakkında bilgi vermekte, -1 ile +1 arasında değerler almaktadır. Bir ölçmenin güvenilirlik katsayısı değerlendirilirken, bu değer pozitif sınırlar içinde ve oldukça yüksek olması arzu edilmektedir. Değer +1'e yaklaştıkça, güvenilir-

liğin yüksek olduğu kabul edilir (Karasar, 1995; Özgüven, 2000; Polit, Hungler, 1997; Portney, Watkins, 1993; Tezbaşaran, 1996).

Burada, "güvenirlik ölçütü olarak elde edilen korelasyon katsayısı, hangi düzeyde olduğu zaman yeterlidir?" sorusu önemlidir. Bunun yanıtı, ölçeğin hangi amaçla kullanılacağına bağlıdır. Çoğu amaç için 0.70 ve üzeri doyurucu olmakla birlikte, 0.85-0.95 aralığının seçilmesi gerektiği; (Polit, Hangler, 1997) ancak, bireylerin geleceği hakkında önemli kararlar verilmesine yönelik ölçmelerde güvenilirliğin 0.95 ve üzerinde olmasının arzu edildiği belirtilmektedir (Öner, 1987; Özgüven, 2000). Fizyolojik ölçümlerde 0.90 ve üzeri, tutum ölçeklerinde 0.70 kabul edilebilir düzeydedir. Gözlemciler arası uyum için korelasyon katsayısının en az 0.80 olması, eğer katılım yüzdesi kullanılıyorsa, katılımın 0.70 ve 0.90 arasında olması doyurucu olabilir. Ayrıca, yeni geliştirilmiş ölçekler için test-tekrar test yönteminde korelasyon katsayısı 0.70 kabul edilebildiği halde, önceden geliştirilen ve çalışmalarda sınınanlar için 0.80 olmalı ve aynı zamanda tüm alt ölçeklerin güvenilirliği de hesaplanmalıdır (Peirce, 1995; Portne, Watkins, 1993).

Güvenirlik katsayısı 0.78 olan bir testten elde edilen puanlar %78 oranında gerçek farkları %22 oranında da hatayı yansıtır. Güvenirlik katsayısı yükseldikçe, testle ölçülen farkların gerçek farkları yansıttığı düşünülür. Bu durumda test puanındaki hata payı az olur. Böylece, güvenilirlik ve hata kavramları arasında ters yönde bir ilişkinin varlığı belirlenir (Öner, 1997).

Korelasyon kullanımının sınırlılıkları nedeniyle, birçok araştırmacı bir çalışmada birden fazla güvenilirlik sınaması yapar. Örneğin; test-tekrar test ya da gözlemci güvenilirliği çalışmalarında hem korelasyon hem de t-testi (iki yüzde ya da iki ortalama arası farkı sınamak üzere) kullanılabilir. Gözlemci güvenilirliğinde ayrıca Cohen Kappa değeri hesaplanabilir (Hayran, Özdemir, 1996; Portney, Watkins, 1993).

Katsayı, her araştırmada ayrıca hesaplanmalı, herhangi bir çalışmanın bulguları ile *genellemeye gidilmemelidir* (Pierce, 1995; Portney, Watkins, 1993).

GEÇERLİK/GEÇERLİLİK (Validity)

Geçerlik, “bir ölçme aracının ölçülmek üzere hazırlandığı amacı, değişkeni ölçme derecesidir”. Bir ölçeğin “neyi”, ne denli “isabetli/doğru” olarak ölçtüğüyle ilgili bir kavramdır (Karasar, 1995; Öner, 1997; Özgüven, 2000; Peirce, 1995; Tezbaşaran, 1996).

Geçerlik bilimin özü olarak kabul edilen, bilimsel olduğu kadar felsefi bir sorun olan ve güvenilirliğe oranla çok daha karmaşık bir kavramdır. Bir ölçmenin geçerli sayılablmesinin ilk koşulu güvenilirlik olmasına karşın, güvenilirlik hiçbir zaman geçerliği garantileyemez. Bu nedenle, bir ölçme aracının güvenilirliği ile birlikte üzerinde önemle durulması gereken konu geçerliktir. Geçerlik bir ölçüm aracı için yapılması zorunlu olan ancak, ölçümün her zaman ve her durum için geçerli olduğunu söylemeyi

engelleyen ve asla sonu olmayan bir süreçtir, yani ölçeğin her kullanımında yeniden sınanması demektir (Karasar, 1995; Özgüven, 2000; Peirce, 1995; Polit, Hungler, 1997; Portney, Watkins, 1993; Tezbaşaran, 1996).

Bir ölçme aracının geçerliğini sınamaya yönelik birçok ölçüt bulunmakla birlikte, bunlar genel olarak üç başlık altında toplanmaktadır (Karasar, 1995; Özgüven, 2000; Polit, Hungler, 1997; Tezbaşaran, 1996):

- I. İçerik/kapsam geçerliği (content validity)
- II. Ölçüt-bağımlı geçerliği (criterion-related validity)
- III. Yapı geçerliği (construct validity)

Bunlardan başka, belirli hastalık ve fonksiyon bozukluklarının erken tanısı amaçlı tarama çalışmalarında seçicilik (specifity) ve duyarlılık (sensitivity) ile prediktif (yordayıcı) değer sınamaları önem kazanmaktadır (Dawson-Sounders, Trapp, 1990; Portney, Watkins, 1993). Bu sınamalar, ileride değinilecek olan hemzaman/eşzaman geçerliliği içinde de geçerlidir (Portney, Watkins, 1993).

Tablo 1. Güvenirlik Sınama Yöntemleri, Kullanılabileceği Durumlar ve İstatistiksel Yöntemler

Yöntem	Durum	İstatistiksel Yöntemler
Değişmezlik Test-tekrar test	Ölçülen nitelik kararlı ise	Pearson Momentler Çarpımı Korelasyonu
Paralel Form	Genellikle ölçek geliştirmede, referans test varlığında	Pearson Momentler Çarpımı Korelasyonu
Ölçümcü Güvenirliği Gözlemciler arası uyum	2 ya da çok gözlemci aynı niteliği ölçerse	Korelasyon t-testi Varyans analizi Kappa İstatistiği
Gözlemciler içi uyum	Tek gözlemci aynı niteliği farklı zamanlarda ölçerse	Korelasyon t-testi
İç Tutarlılık Testi yarılama	İki yarıda varyans eşit-çok yakın ise Madde puanları süreksiz ise Sürekli (Likert tipi) ise Her madde aynı güçlük derecesinde ise	Spearman-Brown KR-20 Cronbach Alfa KR-21
Madde istatistikleri	Maddeler süreksiz, iki seçenekli ise Sürekli (Likert tipi) ise	Pearson Momentler çarpımı korelasyonu düzeltilmiş formülü (bi-serial) Pearson Momentler çarpımı korelasyonu düzeltilmiş formülü (point-bi-serial)

I. İçerik/kapsam geçerliği

Intrinsic ya da domain geçerlilik olarak da bilinen içerik geçerliği, testin ölçüm amacıyla ilgisiz olan faktörlerin etkisinden arınmış olmasını ifade eder. Örneğin, kaba motor yeteneğini test eden bir araç, ince motor yeteneğini değerlendiren maddeleri içermemeli, ya da hastanın kaygı düzeyi vb. etkilenmemelidir. Ölçeğin hedeflerinin tümüyle mantıksal değerlendirilmesini ve test içindeki parçaların teorik temeli nasıl yansıttığını gösterir (Portney, Watkins, 1993). Daha açık bir ifadeyle, içerik geçerliği ölçeğin, bütünü ve alt boyutlarının ölçülmek istenen alanı ölçüp ölçmediğini ve ölçülecek alan dışında farklı kavramları barındırıp barındırmadığını değerlendirmek amacıyla yapılır.

Bu geçerlik sınavının amacı, ölçme aracında bulunan maddelerin ölçülmek istenen alanı temsil edip etmediğini bir uzman gruba inceleyerek, anlamlı maddelerden oluşan bütünü oluşturmaktır. Burada sözü edilen uzman kişi hem ölçeğin hazırlandığı bilim alanını iyi bilen ve hem de ölçek sorusu hazırlama teknik ve yöntemlerini bilen bir kişidir. Uzmanların öneri ve eleştirileri doğrultusunda ölçek yeniden yapılandırılır (Karasar, 1995; Özgüven, 2000; Polit, Hungler, 1997; Tezbaşaran, 1996).

İçerik geçerliği uzmanların yargılarına dayanan bir ölçüttür. Ölçeğin içeriğinin yeterli olduğunu garanti altına alacak objektif kriterleri yoktur. Uzmanların çoğunluğunun aynı fikirde olması bir gösterge olabilir ve bu aşagıda açıklandığı gibi hesaplanabilir; fakat, sonuçta uzmanların öznel yargılarına güvenilmelidir (Polit, Hungler, 1997; Portney, Watkins, 1993).

İçerik geçerliğini saptama, özellikle ölçek geliştirme çalışmalarında yapılması gereken bir aşamadır. Ancak, herhangi bir dilde geliştirilen bir ölçme aracını Türkçeye uyarlamak isteyen araştırmacı da bu mantıksal ölçütü kullanmalı, yani içerik geçerliğini sınamalıdır. Dilde eşdeğerliliği sağlanan ölçek, konuyla ilgili olan uzmanların

görüşüne sunulacak daha rafine, anlaşılır ve anlamlı maddelerden oluşan bir uyarlama sağlanmaya çalışılmalıdır. Süreç, testin defalarca düzenlenmesini gerektirir. Maddelerin hem dil ve kültür eşdeğerliği hem de içerik geçerliğini sınamak üzere uzman görüşüne başvurmada, her madde için uzmanın maddelerin uygunluğunu 1-4 ya da 1-5 arası puanlar vererek değerlendirmesini sağlayacak bir form kullanılır. Burada öncelikle uzmana çalışmanın amacı, ilgili değişkenleri, örnek grup özellikleri vb. açıklamalar yapılması çok önemlidir. Her madde için uzmanların katılım yüzdeleri karşılaştırılır (Peirce, 1995).

II. Ölçüt-bağımlı geçerlik

En objektif ve en pratik olan bu geçerlik sınavında ölçek puanlarının bazı dış ölçütlerle ilişkisi aranır. Bir testin diğer bir testle elde edilen sonuçları verme yeteneğidir (Pierce, 1995; Portney, Watkins, 1993; Wood, Haber, 1994). Araştırmacı, ölçüt bağımlı geçerliliğin yüksek olma eğilimi olduğunu bilmesi gerektiğinden, ölçeğin kullanımından öncelikle sonuçları dikkatle gözden geçirmelidir. Bu geçerlik sınavında en önemli faktör, örneklemin temsil yeteneğidir. Diğerlerine göre örneklemin özelliklerine en çok bağımlı olanı budur. Temsil yeteneği ne kadar az ise, ölçüt geçerliği o kadar problemlidir (Peirce, 1995).

Bu geçerlik ölçütünü değerlendirmek üzere iki yaklaşım vardır:

Yordama-kestirim geçerliğinde (predictive validity), ölçekten elde edilen bir “yordayıcı puan” ile gelecekteki durumlarla ilgili bir “ölçüt”e ilişkin değerler arasındaki korelasyon katsayısı belirlenir (Öner, 1997; Özgüven, 2000; Peirce, 1995; Portney, Watkins, 1993; Tezbaşaran, 1996). Bu bir anlamda, yapılan ölçme ile ölçülmeye çalışılan şeyin gerçek hayattaki yansımalarının karşılaştırılmasındaki uyumu gösteren uygulama geçerliğidir (Karasar, 1995). İki ölçüm

arasında verilmesi gereken aralık, ölçüte ulaşma için gereken zamandır (Portney, Watkins, 1993). Örneğin, hemşirelik öğrencilerinin kateter uygulama becerisi ölçümünden aldıkları test puanı, uygulamadaki kateter uygulama performansı ile karşılaştırıldığında, ilişki yüksek ise ölçeğin yordama geçerliği sınanmış olur. Ya da öz-bakım gücü ölçeğinden yüksek puan alan öğrencilerin gerçek yaşamda kendilerine bakma davranışları yeterli ise, ölçeğin yordama geçerliğinin yüksek olduğu sonucuna varılır. Burada önemli olan, dış ölçütün uygun biçimde belirlenebilmiş olmasıdır.

Ancak burada bazı güçlüklerle karşılaşılır. Bunlardan en önemlisi uygulamadaki beklentilerin, ölçütlerin ve kavramların, gözlenebilir değişkenlerle ifade edilmesindeki güçlülüdür (Karasar 1995, Portney, Watkins, 1993). Fizik ya da fizyoloji bilimlerinde ölçütler hazırdır. Örneğin, birey “öz bakımını yapabiliyor” ifadesinden ne anlaşılmalıdır? Fiziksel bakım mı, sağlık bakımı mı, ya da hepsi mi? Ölçüt belirlendiğinde geçerlik kolaylıkla saptanabilir. Örneğin, kadınların doğum kontrolü çabalarının etkililiğini ölçen bir puan (ölçek) ile gebelik sayıları (ölçüt) arasındaki ilişkinin katsayısı 0.83 ise bu ölçeğin, ölçüt-bağıntı geçerliğinin oldukça yüksek olduğu sonucuna varılır (Polit, Hungler, 1997). Ölçütün kolaylıkla saptanamadığı durumlarda bu geçerlik sınaması uygun olmayabilir. O zaman, daha karmaşık olan yapı geçerliğini sınamak gerekir (Portney, Watkins, 1993).

Uyarlanan bir ölçeğin yordama geçerliği, aynı kültürde kullanılan ve ilgili olduğu bilinen diğer sağlam (psikometrik özellikleri saptanmış bulunan) ölçeklerden elde edilen ölçümlerle olan ilişkisinin düzeyine bakılarak da yapılabilir (Öner, 1987). Örneğin, Öz-Bakım Gücü Ölçeğinin Türkçeye uyarlamasını gerçekleştiren Nahcivan (1993), Öz-Bakım Gücü ile Aile Ortamı, Aile hayatı ve Çocuk Yetiştirme Tutumu ölçekleri arasındaki ilişkiye bakarak bu geçerlik ölçütünü kullanmıştır.

Hemzaman/eşzaman geçerliğinde (concurrent validity), ölçek puanları ölçüm anında var olan bir ölçütle karşılaştırılır. Benzer ölçek geçerliği olarak da bilinen bu yöntemde, daha önceden geçerliği saptanmış olan bir ölçeğe gereksinim vardır. Yeni uyarlanan ölçeğin geçerliğini bulmak için yeni test ve geçerliği yüksek olduğu bilinen önceki test birlikte aynı gruba uygulanır; bireylerin yeni ve eski testten aldıkları puanlar arasındaki korelasyon hesaplanır ve bu korelasyon katsayısının yüksek olması beklenir (Özgüven, 2000). Ülkemizde Pınar ve Oğuz (1998) tarafından iki bası yarısı ölçeğinin karşılaştırılması, Aslan (1998) tarafından postoperatif ağrı değerlendirme ölçeklerinin duyarlılık ve seçiciliklerini karşılaştırmak üzere yürütülen çalışmalarda, bu geçerlik (hemzaman/eşzaman geçerliğini) ölçütünü kanıtlayan bulgular da bulunmaktadır.

III. Yapı geçerliği

Ölçeğin ilgili kavram ya da kavramsal yapının tümünü ölçme yeteneğini gösterir (Portney, Watkins, 1993). Bir ölçeğin ve ondan elde edilen puanın gerçekte ne anlama geldiğini araştırma sürecidir. Bu süreç, ölçeğin ölçtüğü faktörler incelenerek ya da geçerliği araştırılan ölçeğin diğer ölçek ve ölçülerle olan ilişkisini araştırarak gerçekleştirilir. Her defasında ölçekle ilgili yeni bir parça bilgi elde edilerek, yığılmalı bir şekilde ölçeğin yapısı ve puanın anlamı hakkında bilgiler elde edilir (Özgüven, 2000). Psikolojik ölçekler için yapı geçerliği birinci derecede önem taşır (Tezbaşaran, 1996).

Bir ölçeğin yapı geçerliğini değerlendirmek üzere en fazla kullanılan iki yaklaşım; **faktör analizi ve bilinen grup ile karşılaştırmadır** (Karasar, 1995; Peirce, 1995; Polit, Hungler, 1997; Portney, Watkins, 1993). Bunların dışında en sık kullanılan bir diğer yapı geçerliği değerlendirme yolu, **hipotez sınaması** (Öner, 1997; Wood, Haber, 1994), bir diğer adıyla **mantıksal analizdir** (Polit, Hungler, 1997). Ayrıca **çok değişkenli-çok yönlü matriks,**

diğer bir sınama yoludur (Pierce, 1995; Portney, Watkins, 1993).

Faktör analizi, ölçekteki maddelerin farklı boyutlar altında toplanıp-toplanamayacağını değerlendirmek üzere yapılan bir işlemdir (Karasar, 1995; Munro, Page, 1993; Öner, 1987-1997; Özgüven, 2000; Pierce, 1995; Polit, Hungler, 1997; Salkind, 2000; Tezbaşaran, 1996; Wood, Haber, 1994). Faktör analizi, açıklayıcı (exploratory) veya doğrulayıcı/hipotez destekleyici (confirmatory) olabilir. *Ölçek uyarlamalarında* daha çok, ölçekteki maddelerin yapısı hakkında var olan bir hipotezi sınıadığı için, **doğrulayıcı faktör analizi** kullanılır. Başka bir ifadeyle, uyarlanan ölçeğin faktör yapısı orijinal ölçeğin faktör yapısı ile karşılaştırılır, benzerlik ve ayrılıklar gözlenir. Bir ölçeğin başka bir dile uyarlanması sonucu o ölçeğin faktör yapısının esasen çok fazla değişmemiş olması beklenir (Öner, 1987). Faktör analizi SPSS paket programında doğrudan yapılabileceği gibi, korelasyon analizi ile de ölçeğin faktör yapısı hakkında bir sonuca varılabilir (Munro, Page, 1993; Öner, 1987).

Bilinen grup karşılaştırmasında; araştırmacı, yapının dayandığı kurama bağlı olarak, bir testten anlamlı olarak farklı puan alabilecek grupları belirler, ölçme aracını her iki gruba uygular ve gruplararası farka bakar. Diğer deyişle, gruplardan birisi, ölçülmek istenen faktörler açısından özellikleri bilinen bir gruptur. Ölçme sonucunda, özelliği bilinen grup ile bilinmeyen grup arasında fark olması beklenir (Karasar, 1995; Polit, Hungler, 1997; Wood, Haber, 1994). Örneğin, doğum deneyimi korkusunu ölçen bir ölçme sonucunda primipar ve multiparların puanlarının farklı olması gerekir (Polit, Hungler, 1997). Ya da kaygı düzeyi için, sınav öncesi öğrenciler ile sınavı olmayan öğrenciler arasında anlamlı fark beklenir (Pierce, 1995).

Hipotez sınaması/mantıksal (logical) analizde; ölçeğin diğer ölçüm araçlarıyla ilişkisine bakılır. Bilinen grup karşılaştırmasına benzer

bir yöntemdir. Burada araştırmacı, ilgili kaynaklar (ya da gözlemler) doğrultusunda, önceden aralarında ilişki olacağı varsayımını kurduğu ilişkilerin yönünü ve düzeyini korelasyon analiziyle değerlendirerek test eder (Pierce, 1995; Polit, Hungler, 1997; Wood, Haber, 1994). Bu geçerlik ölçütü, uyarlanan ölçeğin yapısal durumunun açıklanmasına önemli katkı sağlayan bir yöntemdir. Ölçekle ilgili yapılan her korelasyonel araştırma aynı zamanda mantıksal geçerliğine yönelik bulgular da sağlar. Öz etkililik-yeterlik ölçeğini dilimize uyarlayan Gözüm ve Aksayan'ın çalışmasında (1990), yapı geçerliğini değerlendirmek üzere faktör analizinin yanı sıra, ölçeğin mantıksal geçerliğini değerlendirmek üzere hipotez edilen bir ilişki de sınanmıştır (Gözüm, 1999).

Çok özellikli-çok yöntemli matris (multi trait-multi method matrix/MMTM); Campbell ve Fiske, bir testin ölçümlerinin aynı ya da farklı yapılardaki diğer test(ler) ile nasıl ilişkili olduğunun değerlendirilmesi yoluyla yapı geçerliğinin sınanabileceğini belirtmiştir. Diğer deyişle, bir testin ölçtüğü yanında ölçmediği değişkenleri belirlemek de önemlidir (Pierce, 1995; Portney, Watkins, 1993). Araştırmada ilgilenilen iki ya da fazla, farklı kavramın (örneğin, iyilik hali ile kaygı düzeyi) ölçümleri arasındaki korelasyon katsayısı düşük ise **ayırddedici geçerlik (discriminant)**; her kavram için iki ya da fazla, farklı ölçüm yolundan (örneğin, yukarıdaki iki kavram için hem fizyolojik ölçümler, hem de anket vb. kullanılabilir) elde edilen puanlar yüksek korelasyon gösterirse, **benzer sonuçlu geçerlik (convergent)** geçerlik doğrulanmış olur. Bu ölçümlerin her birinden alınan puanlar bir korelasyon matrisinde gösterilir. Bunun için, grubun bir oturumda dört ya da fazla testi doldurmaya gönüllü olması gerekir. Geçerlik için çok bilgi vermekle birlikte güçtür (Bu konuda daha fazla bilgi için bkz. Peirce, 1995; Portney, Watkins, 1993; Wood, Haber, 1994).

Tablo 2. Geçerlik Sınama Yöntemleri, Kullanılabileceği Durumlar ve İstatistiksel Yöntemler

Yöntem	Durum	İstatistiksel Yöntemler
İçerik/Kapsam Geçerliği	Maddelerin ölçülecek alanı temsiline ilişkin uzman yargısı	Uzman puanlarının tutarlılığı
Ölçüt-Bağımlı Geçerlik Yordama-kestirim	Ölçek puanlarının, gelecekteki durumla ilgili ölçüt(ler)e ilişkin değerler ile karşılaştırması	Korelasyon
Eşzaman/hemzaman	Ölçek puanlarının, geçerliliği sınanmış önceki test puanları ile karşılaştırması	Korelasyon
Yapı Geçerliği Faktör analizi	Özellikle ölçek uyarlamasında	Doğrulayıcı (confirmatory) faktör analizi
Bilinen grup karşılaştırması	Ölçeğin bilinen grubu ayırabilme gücü	t-testi
Hipotez sınaması/ mantıksal analiz	Önceden varsayılan ilişkilerin sınanması	Korelasyon
Çok özellikli çok yöntemli matriks Benzer sonuçlu	Ölçümler arası uyumun sınanması	Korelasyon
Ayırdedici	Farklı değerleri/kavramları ayırd edebilme gücü	Korelasyon

KÜLTÜRLERARASI KARŞILAŞTIRMA

Ölçek uyarlama çalışmasının bu aşamasında uyarlanan ölçeğin normları saptanır ve diğer dillerdeki ölçek normları ile karşılaştırılır. Bu aşamada aşağıdaki soruların yanıtları aranır (Öner, 1987);

“Uyarlanan ölçeğin puan ortalamaları, standart sapmaları ve kesme noktası gibi özellikleri orijinal ölçeğin norm değerlerine benziyor mu?”

“Uyarlanan ölçeğin ölçme hatası orijinal ölçeğin ölçme hatasına yakın mı?”

“Uyarlanan ve orijinal ölçeğin faktör yapıları benziyor mu?”

“Faktör-madde yükü her iki ölçekte de benzer değerlerde mi?”

Yukarıda sorulan soruların birçoğu, uyarlanan ölçeğin psikometrik özellikleri değerlendirilirken kendiliğinden yanıtlanır. Bu sorulara

istendik yanıtlar alınamıyor ise, örneğin, ölçeklerin faktör yapısı örtüşmüyorsa bunun olası nedenleri araştırılmalı ve açıklanmalıdır. Pratikte faktörlerin büyük oranda benzeşik olması yeterli görülmekte ve orijinal ölçekteki faktörler benimsenmektedir.

Bir ölçeği farklı dillere uyarlarken, kültürlerarası karşılaştırma (cross-cultural) yapmak ayrı bir araştırma konusu olup, özel koşulları vardır. Bu tür araştırmalarda ölçeği geliştiren ve uyarlayan araştırmacılar genellikle birlikte çalışır. Her iki kültürde benzer özellikleri olan bireylere uygulanan iki ölçeğin tüm psikometrik özellikleri, ölçekteki maddelere verilen yanıtların yüzdeleri ve benzeri birçok özellik karşılaştırılır. Kültürel özelliklerden kaynaklanan ayrılıklar tartışılır. Batı kaynaklı psikoloji ve hemşirelik literatüründe “cross-cultural research” olarak isimlendirilen bu çalışmalar, ölçeğin evrenselleşmesini sağlamaktadır.

KAYNAKLAR

- Aslan, F.E. (1998) Postoperatif Ağrı Değerlendirmesinde Görsel Kıyaslama ve Basit Tanımlayıcı Ölçeklerin Duyarlılık ve Seçiciliklerinin Karşılaştırılması. VI. Ulusal Hemşirelik Kongresi, Kongre Kitabı, Ankara, Damla Matbaacılık, 178-184.
- Çam, O. (1991) Hemşirelikte Tükenmişlik Sendromunun Araştırılması. Yayınlanmamış Doktora Tezi, İzmir, Ege Üniversitesi.
- Dawson-Saunders B., Trapp, R.G. (1990) Basic and Clinical Biostatistics. Appleton&Lange, 232.
- Esin, M. N. (1999) Sağlıklı Yaşam Biçimi Davranışları Ölçeğinin Türkçeye Uyarlanması. Hemşirelik Bülteni, 2(45); 87-96.
- Erkuş, A. (1999) İstatistik Paket Programlarını Doğru Kullanabiliyormuyuz? Türk Psikoloji Bülteni, 12; 14-17.
- Gözüm, S., Aksayan, S. (1999) Öz-Etkililik-Yeterlik Ölçeğinin Türkçe Formunun Güvenilirlik ve Geçerliliği. Atatürk Üniversitesi Hemşirelik Yüksekokulu Dergisi, 2(1); 21-34.
- Gözüm, S. (1999) Öz-Etkinlik-Yeterlik Ölçeğinin Türkçe Formunun Yapı Geçerliliği: Öz-Etkililik-Yeterlik İle Stresle Başa Çıkma Algısı Arasındaki İlişki. Atatürk Üniversitesi Hemşirelik Yüksekokulu Dergisi, 2(1); 35-43.
- Hayran, M., Özdemir, O. (1996) Bilgisayar, İstatistik ve Tıp. İkinci baskı, Ankara, Hekimler Yayın Birliği; 372.
- Karasar, N. (1995) Bilimsel Araştırma Yöntemi. 7. Basım, Sim Matbaası, Ankara.
- McLaughlin, F.E., Marasculio, L. A. (1990) Advanced Nursing and Health Care Research: Quantification Approaches. Philadelphia, W. B. Saunders Company; 35-37.
- Munro, B. H. Page, E. B. (1993) Statistical Methods for Health Case Research. Second Edition, J. B. Lippincott Company, Philadelphia.
- Nahcivan, N.Ö. (1993) Sağlıklı Gençlerde Öz Bakım Gücü ve Aile Ortamının Etkisi. Yayınlanmamış Doktora Tezi, İstanbul, İstanbul Üniversitesi.
- Oğuz, S. Olgun, S. (1997) Bası Yaralarında Braden Ölçeğinin Kullanımı. V. Ulusal Hemşirelik Kongresi Kitabı, İzmir, D.E.Ü. Rektörlüğü Matbaası, 71-75.
- Öner, N. (1987) Kültürlerarası Ölçek Uyarlamasında Bir Yöntembilim Modeli. Psikoloji Dergisi, 6(21); 80-83.
- Öner, N. (1994) Güvenirliliği ve/veya Geçerliliği Sınanmış Psikolojik Testler. Türk Psikoloji Dergisi, Özel Sayı, Psikolojik Testler, 9(33); 9-18.
- Öner, N. (1997) Türkiye’de Kullanılan Psikolojik Testler, Bir Başvuru Kaynağı. 3. Basım, İstanbul, Boğaziçi Üniversitesi Matbaası, 3-39.
- Özguven, İ. E. (2000) Psikolojik Testler. Dördüncü Baskı, Ankara, PDREM Yayınları, Sistem Ofset; 83-120
- Peirce, A. G. (1995) Measurement, Principles and Practice of Nursing Research. (ed. Laura A. Talbot), St. Louis, Mosby-Year Book, Inc, 265-290.
- Pınar, R., Oğuz, S. (1998) Norton ve Braden Bası Yarası Değerlendirme Ölçeklerinin Yatağa Bağımlı Aynı Hasta Grubunda Güvenirlilik ve Geçerliliğinin Sınanması. VI. Ulusal Hemşirelik Kongresi, Kongre Kitabı, Ankara, Damla Matbaacılık, 172-175.
- Polit, D. E., Hungler, B. P. (1997) Essentials of Nursing Research Methods Appraisal and Utilization. 4.th Lippincott Raven Publishers.
- Portney, L. G., Watkins, M. P. (1993) Foundations of Clinical Research. Application to Practice. Norwalk, Connecticut, Appleton Lange, 53-85; 505-527
- Salkind, N. J. (2000) Statistics for People Who (think they) Hate Statistics, Sage Publications. Inc. Thousand Oaks, London, New Delhi.
- Tezbaşaran, A. (1996) Likert Tipi Ölçek Geliştirme Kılavuzu. Psikoloji Derneği Yayınları, Ankara, Özyurt Matbaası.
- Wood, G. L., Haber. (1994) Nursing Research: Methods, Critical Appraisal and Utilization, St. Louis, Mosby-Year Book, Inc, 265-382.