

Psikolojik Ölçmelere İlişkin Doğru Bilinen Yanlışlar

R. Nükhet Çıkrıkçı-Demirtaşlı

Ankara Üniversitesi, Eğitimde Psikolojik Hizmetler Bölümü
nukhet66@hotmail.com

Bu yazıda, eğitimde ve psikolojideki, ölçme ve değerlendirme işlem ve uygulamalarında sıklıkla rastlanan, yanlış bilinen veya eksik yorumlanan bazı durumlar özetlenmiş ve bunlara ilişkin doğrular, kaynaklara dayalı olarak açıklanmıştır. Psikolojik ölçmelere ilişkin kavramların ve tekniklerin doğru kullanılması ve yorumlanması, bu alanda yapılacak bilimsel çalışmaların akademik ve toplumsal düzeydeki güvenilirliği için önemlidir. Şimdi bunları gözden geçirelim.

1. Her iki dili de iyi bilen herhangi biri, psikolojik test çevirisi yapabilir.

Eğitimde ve psikolojide kullanılan testlerin ve ölçeklerin uyarlanması sırasında en sık rastlanan yanlış bilinen doğrulardandır (Hambleton ve Patsula, 1999). Uyarlama çalışması sırasında testin çevirisini yapacak kişilerin, kaynak ve hedef dili iyi bilmelerinin yanında, her iki kültürü de iyi tanımaları, testle ölçülen psikolojik yapıyı ve ayrıca nitelikli test geliştirme ile ilgili temel ilkeleri ve süreçleri bilmeleri ve bununla ilgili uygulamalar yapmış olmaları gerekir.

2. Var olan bir testin uyarlanması, ikinci dilde yeni bir test geliştirmekten her zaman daha iyidir.

Test uyarlamayı geçerli kılabilecek pek çok neden vardır. Ancak, aynı biçimde uyarlama *yapmamayı* da gerekli kılabilecek önemli gerekçeler vardır. Özellikle kültürlerarası karşılaştırmalar *yapılmayacaksa*, ikinci dilde yeni bir test geliştirmek nispeten daha kolay ve uygun bir yol olacaktır. Bu durum, uygulayıcıları, uyarlama yapılacak testin yayın hakkının alınması ile sorunlardan uzak tutabilece-

ği gibi, geliştirilecek testin kültüre uygunluğunu garantiler ve test geliştirme sürecinin başında ölçülecek psikolojik yapının tanımında gereken değişikliklerin yapılmasına da olanak sağlar (Hambleton ve Patsula, 1999).

3. Anketlerden elde edilen verilere, faktör analizi ve diğer parametrik istatistiksel teknikler uygulanabilir.

Anketler, bireylerin, belli bir durum, olay, kişi ya da gruplar hakkındaki görüşlerini, eğilimlerini yoklamak üzere kullanılan ve test dışı teknikler içinde yer alan bilgi toplama araçlarıdır. Anketlerde yer alan sorulara veya maddelere verilen yanıtlar ve tepkiler, basit kodlamaları veya sınıflandırmaları gerektirir. Anket maddelerinin puanlanmasında matematiksel işlemler (dört işlem; toplama, çıkarma, çarpma, bölme) anlamlı değildir. Çünkü bu türden araçlar bize ölçülen özellik hakkında sınıflama ölçeği düzeyinde bilgi verir (Aiken, 1997, s. 56, 117). Bu ölçek düzeyinde elde edilen verilerle; frekans, yüzde, tepe değer gibi basit istatistik ölçüler hesaplanabilir.

Kategoriler arası karşılaştırmalar yapmak için oranların karşılaştırması ve ki-kare gibi parametrik olmayan istatistik testler uygulanabilir. Faktör analizi ise, sürekli değişken niteliğinde olan, toplam puan veren ve ayrıca ilgi, tutum, algı, başarı, yetenek, zeka gibi özellikleri ölçmede kullanılan araçlardan elde edilen veriler üzerinde uygulanmaya elverişlidir. Bunun dışında yine aynı nedenden dolayı, tutum, yetenek, ilgi ve algı ölçeklerinden elde edilen puanların güvenilirliğini kestirmede kullanılan teknikler; görüş, eğilim yoklamak

üzere uygulanan anketler için uygun değildir. Anketlerde daha çok, tepkilerin tutarlılığını saptamak üzere, ankete yerleştirilen kontrol maddelerine verilen yanıtların uyuşumunun hesaplanması söz konusudur. Bu amaçla, bireyleri, varlıkları veya olayları değerlendiren iki veya daha fazla sayıdaki yargıcı/hakem grubu arasındaki tutarlılığın derecesini gösteren “değerlendiriciler arası güvenilirlik katsayısı” (interrater reliability coefficient) veya “uyuşum katsayısı” (coefficient of concordance), “Kapa Katsayısı” gibi, istatistikler hesaplanır (Aiken, 1997, s. 159). Anketlerin amaca uygun yoklama yapması önemlidir ve bu nitelik, anketi geliştirme sürecinde yapılacak deneme uygulamaları ile kontrol altına alınabilir.

4. Bir psikolojik testten elde edilen verilere faktör analizi uygulandığında, ilk sonuçlar yorumlanır.

Faktör analizi çeşitli amaçlarla uygulanır. Bunlardan bazıları, değişkenler arasındaki ilişkilerin ortaya çıkarılması, verilerin özetlenmesi, verilerin ölçeklenmesi ve dönüştürülmesi, hipotez testi, sebep-sonuç ilişkileri analizidir (Baykul, 2000). Faktör analizi, test ve ölçek verilerine uygulandığında, başlangıç (initial) veya döndürülmemiş (unrotated) faktör yükleri (factor loadings) elde edilir. Tipik olarak, araştırmacılar bu döndürülmemiş faktör yüklerini *yorumlamazlar* (Crocker ve Algina, 1986, s. 292). Hiçbir faktör çıkarma tekniği (factor extraction technique), döndürme yapılmadan yorumlanabilir sonuçlar vermez (Tabachnick ve Fidel, 1996, s. 662). Başlangıç faktör yükleri, çeşitli döndürme yöntemlerinden biri kullanılarak döndürülür ve faktör analizi sonuçları bundan sonra yorumlanır.

5. Test ve ölçeklerle toplanan verilere, faktör analizi uygulamak için gereken örneklem büyüklüğünün, madde sayısının üç veya beş katı olması yeterlidir.

Faktör analizi yapabilmek için, örneklem büyüklüğünü tayin etmede yaygın olarak kabul gören kural, değişken (madde) başına 10 kişinin olmasıdır. Diğer bir ifadeyle, madde sayısının 10 katı kadar uygulama yapmaktır (Crocker ve Algina, 1986, s. 296).

6. Psikolojik ölçme (Psychological measurement) ve psikolojik değerlendirme (Psychological assessment) aynı anlamı taşır.

Psikolojik ölçme ve psikolojik değerlendirme kavramları, eğitim ve psikoloji alanında çalışan akademisyenler ve uygulayıcılar tarafından çoğu kez birbirinin yerine kullanılabilen, eksik veya yanlış yorumlanabilmektedir. Psikolojik değerlendirme (assessment), süreç içinde, bireyleri anlama ve sorunlarıyla baş etmede yardımcı olacak anlamlı bilgilerin toplanması anlamına gelir. Oysa psikolojik ölçme (measurement), önceden belirlenmiş kurallara göre ölçülecek özelliklere sayı veya sembol tayin etme sürecidir (Walsh ve Betz, 1995). Ölçme alan yazınına bakıldığında, Stevens (1951) tarafından tanımlanan ve yaygın kabul gören sınıflamada, dört tür ölçmeden söz edilmektedir (akt., Walsh ve Betz, 1995). Bunlar; sınıflama (adlandırma), sıralama, eşit aralıklı ve eşit oranlı ölçmelerdir. Bu ölçme türlerinin her biri, ölçme işleminden sağlanan ölçme bilgisinin düzeyini tanımlar. Psikolojik değerlendirme sürecinde, ölçme işlemleri birden fazla ve farklı türde ölçme araç ve yöntemleri kullanılarak yapıldıktan sonra, bireyin, sorun durumuyla ilintili olarak bir resminin, bir profilinin çıkarılması ve bu süreçten elde edilen bilgilerin yorumlanarak çözüme gidilmesi amaçlanır.

7. Cronbach alfa, ölçeğin tek boyutluluğunu gösterir.

Cronbach alfa katsayısı ile ilgili yaygın olarak yapılan yanlış yorumlardan biri,

yüksek alfa değerinin, test maddelerinin tek boyutluluğunun (testteki performansın tek bir özelliikle açıklanması) bir göstergesi olarak kabul edilmesidir. Cronbach alfa, madde kovaryanslarının bir fonksiyonudur ve maddeler arası yüksek kovaryans birden fazla ortak faktörün sonucu olabilmektedir (Cocker ve Algina, 1986, s. 142). Örneğin; matematik problem çözme testindeki maddelere verilen yanıtlardan alınan puanlarda, yanıtlayıcının hem o konudaki bilgisi hem de problemi okuyup, problemde isteneni kavrama gücü de (okuduğunu anlama gücü) etkili olabilecektir. Bu durumda, Cronbach alfa, maddelerde ve testte gözlenen performansın altında yatan, *ortak faktörlerle* açıklanan test puanlarındaki varyansın (değişkenliğin) oranının, toplam değişkenliğe katkısının bir alt sınırı olarak yorumlanmalıdır. Bunun dışında alfa, maddeler arası olumlu (pozitif) korelasyonların bütünü gösteren “iç tutarlık” indeksi olarak yorumlanabilir. Ayrıca, alfa, “duyarlık” katsayısı (coefficient of precision) olarak bilinen kuramsal güvenilirliğin alt sınırı olarak da kabul edilmektedir.

8. Cronbach alfa, sadece likert tipi dereceli puanlanan maddelerin güvenilirliğini kestirmede kullanılır.

Cronbach 1951’de, içtutarlılığı kestirmede çeşitli yöntemleri tartışarak bir senteze gitmiş, Cronbach’ın alfası adı verilen tek ve genel bir formül geliştirmiştir. Cronbach alfa formülünde, testteki toplam madde varyanslarının, test puanları varyansına oranı, madde sayısı ile orantılı olarak hesaplanmaktadır. Cronbach alfa, hem iki uçlu puanlanan (0 - 1, Evet - Hayır, Doğru - Yanlış) maddeler için hem de dereceli puanlanan, tutum ölçekleri ve açık uçlu soru türü sınavların (essay) iç tutarlılık anlamında güvenilirliğini kestirmede kullanılır (Crocker ve Algina, 1986).

9. Bir testin bir grup için elde edilen geçerlik ve güvenilirlik kanıtları istenen düzeyde ise, aynı ölçek farklı gruplarda da aynen kullanılır.

Geçerlik, bir ölçme aracının amacına uygunluk düzeyidir. Daha teknik ifade edilecek olursa; bir ölçme aracının geçerliği, o ölçme aracı ile elde edilebilecek ölçülerdeki değişkenliğin (varyansın) ne kadarının, incelenen kişilerin ölçülen özelliğe sahip oluş dereceleri arasındaki *gerçek* farklardan geldiğini gösterir (Özçelik, 1992). Güvenirlik ise, ölçme aracının tekrarlı ölçmelerde tutarlı sonuçlar vermesi ile tanımlanır. Bu tutarlılık, ölçme aracının hatalardan özellikle de seçkisiz (random) hatalardan arınık ölçme yapma derecesi ile ilgilidir. Bir testin güvenilir sonuçlar vermesi onun geçerliğini garantilemez. Bir başka ifadeyle, bir ölçme aracı, ölçtüğü özelliği ne kadar az hata ile ölçmekte olursa olsun, istediğimiz özelliği ölçmüyorsa, başka özelliklerle karıştırarak ölçüyorsa, işe yaramaz (Özçelik, 1992, s. 43).

Bir testin geçerlik ve güvenilirliği belli bir amaç ve belli bir grup içindir. Test veya ölçek, farklı amaçlarla veya farklı gruplarda kullanılacaksa, bu durumda ölçme aracının yeni amaç veya gruplar için geçerlik ve güvenilirlik ölçülerinin yeniden kestirilmesi gerekir. Bu durum, klasik test kuramı kapsamında tanımlanmış olan geçerlik ve güvenilirlik ölçülerinin, testin uygulandığı grubun özelliklerine ve testte kullanılan maddelerin niteliklerine (güçlük ve ayırıcılık) bağlı olarak değişmesinden kaynaklanmaktadır. Bir başka ifadeyle, test ve madde istatistikleri gruba bağımlıdır. Örneğin genel lise grupları için geliştirilmiş bir Standart Matematik Başarı testinin güvenilirliği, öğrencilerini belli zihinsel süreçler bakımından seçerek alan ve üstün yetenekli olduğu saptanan bireylerin eğitildiği bir lise grubunda kestirilen güvenilirlik katsayısından daha büyük olacaktır. İlk grup,

testle ölçülen özellik (matematik başarısı) bakımından daha heterojen yapıda puanlar sergilerken, diğer grup seçilmiş olmanın da etkisiyle, testten aldığı puanlar bakımından homojen, varyansı dar bir puan dağılımı gösterir. Bilindiği gibi, ölçülen özellik bakımından birbirine benzer puanların alındığı bir gruptan kestirilen güvenilirlik katsayısı, aynı özellik bakımından bireylerarası farklılıkların daha büyük olduğu heterojen bir gruptan kestirilene göre daha düşük değerler verir. Bu durumda, genel liseler için geliştirilmiş Standart Matematik Testi, yeteneğine göre seçilmiş lise öğrenci grubu için yeterince güvenilir değildir. Aynı durum, seçme amaçlı testler için kestirilen yordama geçerlik katsayıları için de söz konusudur. Yordama geçerliği katsayısı, geniş bir başvuran grubundan seçilmiş daha küçük bir grup üzerinden kestirildiği için, katsayı gerçeğin değerinin altında bir değer gösterir.

Buna göre, test ve ölçeklerin hali hazırdaki geçerlik ve güvenilirlik ölçüleri ne denli yüksek olursa olsun, ölçme araçları geliştirildiği gruplardan farklı gruplarda çeşitli amaçlarla (seçme, tanılama, araştırma) uygulanacaksa, geçerlik ve güvenilirlik ölçülerine ilişkin kanıtların, yeni gruplar için tekrar kestirilmesi gerekir.

10. Araştırma yapmak için önce ölçme aracını bulup, ona göre araştırma sorusu oluşturulur.

Araştırma süreci, meraklı bir sorgulama sürecidir. Bu süreç, bireyi rahatsız eden bir durum veya yanıtı merak edilen bir soruyla başlar. Bilimsel yöntemin aşamaları izlenerek soruların yanıtları bulunmaya çalışılır. Bu süreçte, araştırmaya konu olan olgu, durum veya davranışlar hakkında bilgi toplamaya ihtiyaç duyulur. Bu amaçla çeşitli ölçme araç ve yöntemleri kullanılır. Bazen araştırmacılar önce ölçme aracını bulup, ardından araştırma

problemini tanımlamaya girişir. Özellikle lisansüstü eğitim kapsamında yapılan tez çalışmalarında bu türden durumlara rastlanabilmektedir. Bu yaklaşım, ölçme araçlarının, bilimsel bilgi toplama araçları olarak değil, amaç olarak algılandığı göstermektedir. Kuşkusuz, araştırma sürecinde ölçme araçlarının yeri önemlidir. Ancak süreçte, ölçme araçları karar verici olarak değil, kararlara dayanak sağlayıcılar olarak değerlendirilmelidir. Araştırmacının temel sorumluluğu, araştırdığı problemle ilgili yanıtları bulmasına yardımcı olacak, geçerliği ve güvenilirliği kanıtlanmış ölçme araçlarıyla bilgiyi toplamak ve araştırma sürecini yansız bir tavırla tamamlamaktır.

Kaynaklar

Aiken, L. R. (1997). *Questionnaires and inventories: Surveying opinions and assessing personality*. NY: John Wiley & Sons, Inc.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.

Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. NY: CBS College Publishing.

Hambleton, R. K. ve Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Association of Test Publishers, August 1* (1), 1-30.

Özçelik, D. A. (1992). *Ölçme ve değerlendirme*. Ankara: ÖSYM Yayınları.

Tabachnick, B. G. ve Fidel, L. S. (1996). *Using multivariate statistics* (3. baskı). NY: HarperCollins College Publishers.

Walsh, W. B. ve Betz, N. E. (1995). *Tests and assessments*. NJ: Prentice Hall.