

Varolan Ölçek Geliştirme Yöntemleri ve Ölçme Kuramları Psikolojik Ölçek Geliştirmede Ne Kadar İşlevsel: Yeni Bir Öneri¹

Adnan ERKUŞ*

Mersin Üniversitesi

Özet

Bugüne kadar geliştirilmiş olan ölçek geliştirme yöntem ve teknikleri ile ölçme kuramları, ölçülecek psikolojik özelliğin ne ve nasıl olduğu konusunda herhangi bir fikir vermemektedirler. Bunun yanında, ölçek maddeleri ister ikili ister çoklu olsun eşit ağırlıklıymış gibi işlemlere tabi tutulmakta, özellikle ulusal çaptaki testlerde yapay alttestler sanki tekboyutluymuş gibi ele alınmakta ve ağırlıklandırılmaktadırlar. Bu çalışmada, psikolojik ölçek geliştirmede olmazsa olmaz yöntemin faktör analitik teknikler olduğu ileri sürülmekte ve bunun gerekçeleri ele alınmaktadır. Ayrıca, çokbileşenli yapılarda, her alt bileşenin önce ayrı ayrı tekboyutlu hale getirilmesi, başlangıçta yer alan 1-0 matrisi yerine faktör puanlarının konulması ve sonra hepsinin faktör analizine sokularak bileşenli yapının oluşturulması önerilmektedir. Bileşenli yapı oluşturulduktan sonra, bireysel puanların bulunmasında da, geleneksel alttest toplamlarının bulunup çarpılıp toplanmasından farklı olarak, bir maddenin tüm bileşenlere katkısının dikkate alındığı bir yöntem önerilmektedir.

Anahtar Sözcükler: çok bileşenli psikolojik yapılar, faktör analizi, faktör puanı

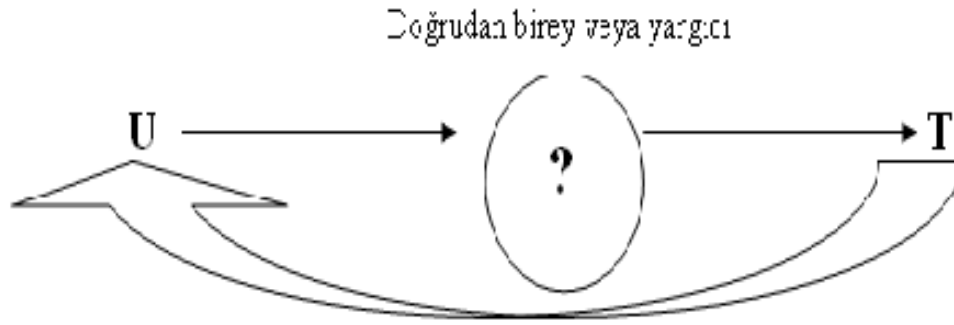
Psikolojik değişkenler doğrudan gözlenemediği için ölçmecilerin karşısına çeşitli sorunlar çıkar: Ölçülen değişken gerçekte ne, nasıl, ölçme sonuçları hatalardan nasıl arındırılabilir, gibi. Psikolojik değişkenlerin ölçümü, bireylere ölçülecek ilgili özelliği uyuracak maddeler takımı sunulup, bireylerin bu uyarıcılara verdikleri tepkilerden hareketle dolaylı bir şekilde gerçekleştirilir. Bu anlamda, psikolojik ölçme, örtük özelliğe ilişkin öngörülen bir denencenin test edilmesine dönüşür ve ölçek geliştirme süreci de ilgili özelliğin işevuruklaştırılması ‘çabası’ anlamına gelir. Bu bakımdan, “gerçek” hep bilinmez kalır; ancak gerçeğin arayışı veya ölçmenin fizik ölçmeler düzeyine getirilme çabası psikometri tarihi boyunca irdelenir durur. Bu çabalar sonucu, tarihsel olarak epey yol katedilmiştir; çok çeşitli ölçek geliştirme yöntem ve teknikleri geliştirilmiş, “gerçek puanı bulma” anlamında da çok çeşitli kuramlar gündeme gelmiştir (Thorndike, 1982; Hambleton ve Swaminathan, 1985). Ancak, süreç içinde, *ölçek geliştirmede* yeni yöntem ve teknikler aramak yerine, epey bir süredir “gerçek puan” arayışının egemen hale geldiğini ileri sürmek yanlış olmasa gerektir. Bir anlamda, sanki, “ormana bakarken ağacın, ağaca bakarken ormanın” görülememesi durumu yaşanmaktadır.

Psikolojide ölçek geliştirme tarihi, algısal büyüklüklerle fiziksel büyüklükler arasındaki bireysel farka dayanan psikofizik ölçme ile başlamıştır (Guilford, 1954; Torgerson, 1958). Elbette,

¹ Bu araştırma, Abant İzzet Baysal Üniversitesi’nde düzenlenen III. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi’nde (19-21 Eylül 2012) sunulmuştur.

psikofizik ölçmeden günümüzdeki psikolojik ölçmeye geçişte pek çok çalışma yapılmıştır; psikolojik ölçmenin klasik anlamda ilk formülasyonları Spearman (1904), Stevens (1946), Gulliksen (1950), Guilford (1954), Torgerson (1958), vd tarafından geçtiğimiz yüzyılın ilk yarısında yapılabilmektedir. Thurstone (1927) ise, bu tarihsel süreçte, döneminin kuramsal alt yapısının etkisiyle psikofizik ölçmeden, özellikle geliştirdiği ölçek geliştirme yöntem ve teknikleriyle de kuramcılardan ayrı ele alınmalıdır.

İster yargıcı, ister bireyin kendisi kullanılsın; ölçek geliştirme, gerçekte, ölçülmesi amaçlanan değişkenin işevuruk olarak tanımlanması (operational definition) işleminin ta kendisidir. Bu işevuruk tanımlamayı yapmak da, yine ister yargıcılar isterse doğrudan bireylerin kendileri olsun, sunulan uyarıcılara (maddelere) verilen tepkilere dayanır:



Uyarıcılar ve tepkiler çok çeşitli biçimlerde ve ölçme düzeylerinde olabilir; ancak, *psikolojik ölçek geliştirmede tek ve nihai amaç, bireyin doğrudan gözlenemeyen ölçülen özelliğinin ne ve nasıl olduğunu anlamaya, ortaya çıkarmaya çalışmaktır*. Uyarıcı ve tepkilerin ölçme düzeyi, başvurulacak olan yöntemi değil, sadece istatistiksel teknikleri değiştirebilir.

Bugüne Kadarki Ölçek Geliştirme ve Ölçme Kuramlarına Genel Bir Bakış

Varolan ölçek geliştirme yöntem ve teknikleri

Tarihsel süreç içinde psikolojik ölçek geliştirme çabalarına bakıldığında, Binet ve Simon'dan (1905-1908; akt., Dennis, 1949) sonra, bugünkü anlamda olmasa da Bogardus'un (1925) "sosyal uzaklık ölçeği"nin başlangıç kabul edilmesi uygun olur. Psikolojik ölçek geliştirmede, bugünkü anlamda "madde analizi" yöntemi olarak ilk öncünün ise, Thurstone (1927) olduğu görülmektedir. Thurstone'un psikofizikten uyarladığı "yargıcı yaklaşımı"na karşılık, Likert'in (1932) "denek tepkilerine dayalı ölçek geliştirme" yaklaşımı bugün de yaygın olarak kullanılan bir ölçek geliştirme yaklaşımı ve tekniği olmuştur. Daha sonra, Bogardus'un 'ilkel' ölçek geliştirme yöntemini madde ve ölçek analizi bağlamında geliştiren Guttman'ın (1944) "birikimlilik ölçeği" önemli bir aşama oluşturmuştur. Daha az önemli olmakla birlikte Osgood, Suci ve Tannenbaum'un (1957) "duygusal anlam ölçeği" ölçek geliştirme tarihinde yerini almıştır. Ölçek geliştirme tarihi, ne yazık ki, kabaca bu kadardır. Tüm bu ölçek geliştirme yöntemlerine bakıldığında, aslında, *ölçülen özelliğin ne ve nasıl*

olduğunu ortaya çıkarmanın ikinci planda kaldığı, genellikle de ölçülen özelliğin tekboyutlu olduğu temeline dayandıkları kolaylıkla ileri sürülebilir.

Varolan ve hâlâ geliştirilmekte olan ölçme kuramları

Spearman'ın çalışmaları şimdilik bir yana konulacak olursa, Gulliksen ve Guilford'un öncülüğünü yaptığı Klasik Test Kuramıyla (KTK), "gerçek puan"ın arayışı ve bu bağlamda, geliştirilmiş kabul edilen ölçeğin hata ile bağlantılı güvenilirlik çalışmalarına yoğunlaşmış, ölçülecek özelliğin ne ve nasıl olduğuna odaklanılmamış, bir başka deyişle ölçek geliştirme görmezden gelinmiştir. Klasik Test Kuramı'nın "gerçeği" aramadaki sayıltılarının ve bunlara dayanan formülasyonlarının zaafı yanlarından hareketle, Modern Test Kuramı (MTK) denilen (Lord, 1952; Lord ve Novick, 1968) ve bugün artık tamamen teknik bağlama sıkışan, "yeni bir gerçek puan arayışı" gündemimize oturmuştur. İster klasik isterse modern olsun tüm ölçme kuramlarının, yine ölçülen özelliğin ne ve nasıl olduğuna değil, madde analizine yoğunlaştığı ve ölçülen özelliğin tekboyutlu olduğu temeline dayandığı rahatlıkla ileri sürülebilir. Son yıllarda, ortaya çıkan "Çokboyutlu Madde-Tepki Kuramı (ÇBMTK)" (Reckase, 2009) ise, kendinden önceki tüm kuramlardan farklı olarak, en azından madde bazında da olsa, çokboyutluluğu ele alması bakımından önemli bir gelişme olarak görülmelidir. Ancak, ÇBMTK'nın da, ölçülen özelliğin ne ve nasıl olduğu konusunda bir şey getirmedigi rahatlıkla görülebilir. ÇBMTK, bir maddenin çokboyutlu olabileceğini öngörerek, yeni madde analizi formülasyonları geliştirmiş, bu da bugün için "ikiboyut" ile sınırlı kalmıştır (McDonald, 1999; Ackerman, 2005).

Varolan ölçek geliştirme yöntemleri ve ölçme kuramlarına genel bir eleştiri

1. Varolan ölçek geliştirme yöntem ve teknikleri ile varolan ölçme kuramları, ölçülen psikolojik yapının ne ve nasıl olduğunun ortaya çıkarılması hakkında bir fikir vermemektedir. Fen bilimlerinden "ödünc alınan" biçimiyle veya etkisiyle, ölçülecek yapının "tanımlı ve tekboyutlu olduğu" yanılgısı, ne yazık ki psikolojik ölçek geliştirme çabalarına sinmiş durumdadır.

2. Yine fen bilimleri bakış açısının etkisiyle, bir psikolojik ölçek maddesinin " bir birim ve eşit ağırlıklı olduğu" ön kabulü psikolojik ölçek geliştirme sürecine hakim durumdadır. Uyarıcılara verilecek tepkilerin değerleri (puanlar) iki türdür: İkili ve çoklu. Bu önsel (a priori) puanlamayla ölçek geliştirmeye başlamak kaçınılmaz olmasına rağmen, nihai ölçeğin puanlarının (MTK'de bile) aynı şekilde kalması doğru görünmemektedir. Çünkü, hiçbir psikolojik değişkene ilişkin herhangi bir madde eşit ağırlıklı olamaz, her madde aynı büyüklükte bir psikolojik değişken parçasını ölçemez.

3. Ölçek geliştirme sürecinde, bilgisayarların hızla yaşamımıza girmesiyle birlikte, faktör analitik tekniklere başvurulmasına rağmen, psikolojik alt yapıları ayrıklaştırma zorlaması söz konusu olmaktadır. Ulusal çapta uygulanan testlerde ise, konu alanları, sanki her biri ayrı alt psikolojik boyut ve kendi içinde tekboyutluymuş gibi toplam puanlar elde edilmekte, yine önsel ağırlıklarla bu alttestler ağırlıklandırılarak bireyler hakkında kararlar verilmektedir.

Sonuç olarak, her maddeye eşit ağırlıklı puan atfederek, yapay alttestler ve ağırlıklar kullanarak ölçme sonucuna karışan hata miktarı artmaktadır ve ölçek geliştirme sürecine uygun olmayan test ve ölçekler kullanılmaktadır. Likert, Osgood, Thurstone vb yöntem ve teknikler veya KTK ile MTK'larına dayanarak "yapılanlar" bize ölçülen yapının ne ve nasıl olduğu hakkında hiçbir bilgi vermemektedir. Oysa ki *ölçek geliştirmede tek amaç ölçülen özelliğin ne olduğunun ve ona en uygun madde ve yapıların belirlenip yapılandırılmasıdır*. Çeşitli ölçme kuram ve modellerine dayanılarak geliştirildiği düşünülen ölçekler, pratikte ölçek kullanıcılarına yine "şu şu maddeleri üstelik yine önsel puanlarla toplayıp bireyler hakkında karar verebilirsin" denmektedir. Bu bakımdan MTK ile de madde parametreleri kestirilse yine aynı işlem yapılmaktadır; MTK daha çok Bireyselleştirilmiş Ölçme (Computerized Adaptive Testing) türü ölçmelerde işlevsel kalmaktadır. Ölçek geliştirilirken ister doğrudan özbildirimli, isterse dolaylı (yargıcılar aracılığıyla) veri toplama yollarına başvurulmuş, amaç bu yapının anlaşılıp işevuruklaştırılmasıdır. Bugüne kadar geliştirilen ölçek geliştirme yöntemleri ile ölçme kuramlarına bakıldığında, psikolojik ölçek geliştirmede, ölçülen özelliğin ne ve nasıl olduğunun göz ardı edildiği, bunun da sadece eğitimde ve özellikle de başarı ölçümüne odaklanılmasından kaynaklandığı ileri sürülebilir.

Yeni Yöntem Önerisi

Psikolojik ölçme veya ölçek geliştirmede, ilk ve temel amaç, ölçülecek olan psikolojik değişkenin ne ve nasıl olduğunun ortaya çıkarılmasıdır. Çünkü, *psikolojik ölçek geliştirme, ölçülmesi öngörülen örtük özelliğin işevuruklaştırılmasından başka bir şey değildir*. Bu işevuruklaştırma için başvurulan ve bugüne kadar yerine başkası konulamayan yegane yöntem ise faktör analitik yöntemlerdir. Ölçülen bazı özellikleri dışta tutarsak, psikolojik değişkenlerin büyük çoğunluğunun yapısını ortaya çıkarmanın yolu, bireylerin verdiği tepkilerin maddeler bazında, "aynılar aynı yere" mantığıyla madde öbeklenmelerinin ortaya çıkarılmasıdır ki bu temel Spearman (1904) tarafından temeli atılmış ve Thurstone'un, Thorndike'in vb özellikle zekanın yapısını tanımlamada kullanılmış faktör analitik teknikler gerçek ölçek geliştirme işlemlerinin başlangıcı olarak görülmelidir. O zamanlarda, bir ölçek geliştirme yöntemi olarak önerilmemiş olsa da, ölçek geliştirme psikolojik değişkenin işevuruklaştırılması olarak ele alındığında, bugün ölçek geliştirmede en temel, biricik ve olmazsa olmaz yolun faktör analizi olduğu ileri sürülebilir. Bu iddiayı temellendirmek için aşağıdaki bazı sayıltıları yapmak yararlı ve gerekli olacaktır.

Bazı önermeler ve bunlara dayalı işlemler

İnsan özellikleri ile ilgili olanlar:

İnsan özellikleri doğası gereği birbiriyle ilişkilidir; basit bir yapı (pür tekboyutlu) olmaları hemen hemen mümkün değildir. Bu nedenle, belirli bir özellik, diğerlerinden ayrı olmak yerine daha çok bir bileşen (composite) yapısındadır; maddelerin büyük bölümü de çokboyutlu yapıdadır. Bir özelliğin bileşen olması, özelliklerin bir kısmının kendi içinde tekboyutlu olmasına engel değildir; önemli olan bileşen yapısındaki kesişim alanlarının yok sayılmamasıdır.

Ölçme ile ilgili olanlar:

a) İnsan özelliklerini uyaracak maddeler takımının yapılandırılması (ölçek geliştirme) aşaması, başlı başına ciddi bir keşif işlemi ve sürecidir. Önemli olan aşama bu aşamadır. Bu aşama sakat olduğunda, hangi kuram hangi kestirim yöntemi olursa olsun ölçme sakat doğar. Ne MTK, ne de başka bir kuram ölçek yapısı hakkında bize bilgi vermez; bu aşamada en önemli aracımız faktör analitik tekniklerdir.

b) Bir ölçek maddesi asla bir birim değildir; birim örtük özelliği temsil edecek hayali (üzerinde anlaşmaya varılmış) en küçük parçadır ve bunun da her maddede farklı büyüklükte olma olasılığı yüksektir. Bu nedenle her maddenin ölçtüğü örtük özellik parçası ve toplama katkısı farklıdır. Bu anlamda, psikolojik değişkenlere ilişkin maddeler, hemen hemen büyük ölçüde çokboyutludur (cross loading items).

c) Her maddenin ölçtüğü özelliğin farklı olması, onların toplanabilirliğini engellemez. Bu, değişik uzunluklarda kırılmış parçalardan oluşan bir metrenin toplanabilirliğine benzer.

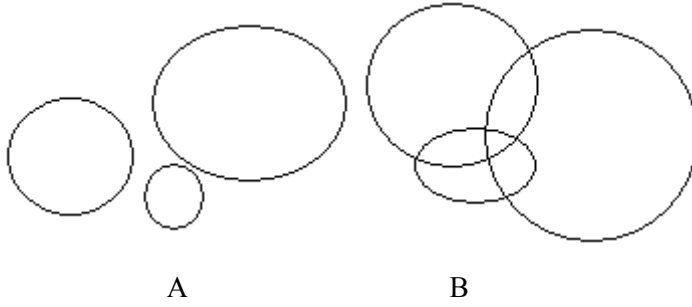
d) Psikolojik özellikler basit yapıda olamayacağına göre, onları basit yapıya indirgemeye çalışma, sadece gerçeği çarpıtmaya yarar. Gerçekte bileşeni oluşturan her alt bileşen farklı büyüklükte ve genel toplama katkıları da bu büyüklüklerle orantılı olmak durumundadır ve bileşenler arası kesişim alanlarındaki maddelerin de temsil ettikleri alt bileşenlere ayrı ayrı büyüklükteki katkıları hesaba alınmalıdır. Tüm bunlar ölçek geliştirme sürecinde ve özellikle de açıklayıcı faktör analitik tekniklerle ortaya çıkarılmalıdır.

e) Temel olan madde ve alt bileşenler ile bunlar arasındaki ilişkileri ve katkıları ölçek geliştirme sürecinde görgül yollarla saptamaktır. Bu yapılmadan elde edilecek bireysel puan (veya yetenek/özellik) kestirimleri doğru olmayacak; hep gerçeğin çarpıtılmış hali elde edilecektir.

Ölçmede ve psikolojik değişkenlerin yapısında ayrıklık ve bileşiklik

Faktör yükleri, maddelerin (değişkenin) ilgili faktörle korelasyonunu gösterir; bu bakımdan ölçek geliştirme sürecinde maddelerin hangi faktör altında yer aldığını belirlemede bu değerler dikkate alınır. Herhangi bir faktör altında “yeterince” yüksek faktör yüküne sahip maddelerin “benzer” yapıyı ölçtüğü sonucu çıkarılır. Tabachnick ve Fidell’e (2001) göre, bunun için alt sınır 0.32’dir; bu da, o maddenin açıkladığı varyansın %10 olduğu anlamına gelir. Ölçek geliştirmede bir maddenin minimum ortak varyansı 0.50 olmalı ise, geriye kalan %40 ne ile açıklanmaktadır? Bu tür maddeler de “kompleks” (Thorndike, 1982) veya çapraz yük (cross-loading) veren maddelerden biri olur ki, bu durumda madde hakkındaki kararımız ne olacaktır? Bundan başka, diğer faktör analitik tekniklere başvurmadan ve hele hele çeşitli döndürmeler denemeden 0.32 kuralına dayanılıp Thorndike’in (1982) ve başkalarının önerdiği gibi ölçekten çıkarılırsa ne olur? Basit! *Ölçülen kavrama ilişkin yapı hatalı oluşturulmaya başlanır.* Bir başka açıdan bakılırsa, böyle bir maddenin çokboyutlu olduğu da göz ardı edilmiş olur. Görüldüğü gibi, sadece öne sürülen bazı ölçütlere dayanmak, özellikle **faktör analitik tekniklerle ölçme geliştirmede** sakıncalı sonuçlara yol açabilir. Kaldı ki, herhangi bir maddenin hangi

faktör içinde yer alacağına karar vermek, yukarıdaki işlemlerin yanında, örneklem büyüklüğüne de bağlıdır (MacCallum, Widaman, Zhang & Hong, 1999). Bu durumu, çerez örneğine benzetirsek, ‘az bir çerez yığnında biraz farklı olan göze çarpar, yığın büyüdükçe o farklılık göze çarpmaz olur’ ilkesiyle örtüşür. Öte yandan, tüm denemelerden sonra bile böyle bir madde, iki-üç faktöre de benzer-yakın değerlerde yük veriyorsa, onu atmak mı gerekir? Birbirinden ayrık ve birbiriyle ilişkili iki psikolojik yapıyı Venn diyagramlarıyla gösterecek olursak;



Şekil 1: Ayrık ve bileşik yapıların Venn diyagramıyla gösterimi

B gibi bir yapıya sahip olan bir psikolojik özellik için kesişim alanlarındaki maddelerin hepsini, “kompleks diye” atarsak ne olur? Basitçe, A’daki gibi bir yapı ortaya çıkar. Bu şekilde hareket ettiğimizde tüm *psikolojik özellikleri yanlış tanımlıyor, ölçüyor ve kararlar veriyoruz demektir*. Böyle bir maddeye ilişkin *puan*, niye ortak olduğu her iki yapıda da toplama katılmasın? Ayrı ayrı toplanabilen ve aynı zamanda birbiriyle toplanan özelliklere **bileşen** ve farklı özelliklerinden dolayı toplanamayan ayrık özelliklere de **faktör** demek bu açıdan çok önemli görünmektedir. *Bu, bileşenli bir yapı için Temel Bileşenler Analizi yapılacağı anlamına gelmemelidir*. Bir yapının ayrık veya bileşenli olup olmadığının göstergeleri, deneme uygulaması sonrası hesaplanan madde-ölçek korelasyonları ile döndürmesiz ortaya çıkarılan 1. faktör yüklerinden anlaşılabilir. Eğer madde-ölçek korelasyon katsayılarının tümü pozitif, ancak düşük veya orta düzeyde ise; 1. faktörün özdeğeri ve açıkladığı varyans çok yüksek ve 1. faktöre maddelerin verdiği yükler pozitif ise, altbileşenlerin tüm ölçekle korelasyonu birbirleriyle olandan daha yüksekse, ölçeğin tümünün içtutarlığı yüksekse bileşenli bir yapıdan söz edilebilir. Elbette, diğer, özellikle eğik döndürme tekniklerine de başvurup, faktörler arası korelasyon katsayılarını da inceleyerek bileşenli yapı sonucuna ulaşılabilir. Bileşenli yapılarda, maddeler çok boyutlu olduğundan, “kompleks diye” madde attıkça, daha önce belirgin bir şekilde bir faktörde görünen maddeler de diğer faktörlere yük vermeye başlar. Elbette, her türlü çoklu yapı, madde atarak sonuçta tekboyutlu duruma getirilebilir. Ancak, amaç psikolojik yapıyı en iyi temsil eden maddelerden oluşan bir ölçek geliştirmek olduğundan, faktör analizi sonuçlarını sadece belirli ölçütlere göre ele almak doğru sonuç vermeyecektir.

Tekboyutluluk toplam puan elde edebilmenin koşuludur; ancak çokboyutlu ve bileşik yapıda da her altölçek ve tüm ölçek tekboyutluluğu sağlayabilir ve bu nedenle de hem altölçeklerin maddeleri toplanarak ayrı ayrı, hem de tümü toplanarak tüm ölçek bazında kararlar verilebilir. “Gerçek-ideal” tekboyutluluğun göstergesi, bir faktörde yer alan maddelerin her birinin faktör yüklerinin 1.00

olmasıdır. Bu durumda 1.00'den ne kadar sapma olursa, (üstelik kaç maddeden kaç tanesi) tekboyutluluktan da sapılır? Bu açıdan bakıldığında, madde bazında 0.32 veya 0.40 gibi ölçütler anlamını yitirir. Denilebilir ki, her faktörün özdeğeri veya açıkladığı varyans dikkate alınır; alınmaz. Özdeğere veya açıklanan varyansa, isterse 'çok çok küçük' olsun, her maddenin faktör yükü katılır, oysa ki aynı madde başka bir faktöre çok büyük yük veriyor olabilir; bu nedenle bu öneri geçerli olamaz. Bilindiği gibi, KTK ve MTK tekboyutluluk sayılına dayanır; bunun için de 1. faktörün açıkladığı varyans dikkate alınır. 1. faktörün açıkladığı varyans %20'lere kadar bile düşse, her iki kuramın işlediği gösterilmiştir (Sünbül, 2011). Bu durumda, geriye kalan %80'in ve diğer faktörlerin ne olduğu, ölçülen psikolojik yapının nasıl göz ardı edildiğini sorgulamak gerekmektedir. Aksi halde, psikolojik yapıları hep yanlış veya eksik tanımlamaya ve ölçmeye (ve de buna bağlı olarak bireyler hakkında karar vermeye) devam edeceğiz demektir.

Öte yandan, bugün özellikle eğitimde ölçmede yapılan şudur: Bileşik bir puan elde edileceğinde, bu bileşenlerin ham puanlarının "konu alanına göre" saptanması ve "her konu alanı içindeki maddelerin de toplanabilirliği" sayılına dayanarak, bulunan toplam puanların yine önsel ağırlıklarla çarpılıp toplanmasıdır.

Tablo 1: Ulusal çapta uygulanan sınavlarda kullanılan yapay alttestlerin simgesel gösterimi

Türkçe	Sosyal Bilgiler	Coğrafya
1.	1.	1.	1.
2.	2.	2.	2.
3.	3.	3.	3.
.	.	.	.
.	.	.	.
k	k	k	k
$(\sum k) w$	$(\sum k) w$	$(\sum k) w$	$(\sum k) w$

Bileşke puan = $(\sum k) w + (\sum k) w + (\sum k) w$ eşitliği doğru görünmemektedir.

Bu durum temel ölçme ilkelerine terstir; oysa ki, asıl yapılması gereken, görgül-istatistiksel yollarla ölçülen yapının altında yatan örtük özelliklere göre alt bileşenlerin bulunması ve yine o bileşenlerin "toplam içindeki paylarını" (örneğin açıkladıkları varyans vb) ağırlık olarak dikkate almak olmalıdır. Aksi halde, "madde puanlarının aynı büyüklükteki örtük özelliği ölçtüğü" gibi yanlış bir temelle başlayan ölçme, yukarıdaki gibi doğru olmayan bir işlemle daha da hatayı artırmakta ve bu ölçmeye dayanarak bireyler hakkında verilecek kararlar da hatalı olmaktadır.

Yeni yöntemin açıklanması

Genel işlemlerin sırası

a) Ölçülecek özelliğin kuramsal-kavramsal yanı çok iyi irdelenmeli, denemelik kavramsal yapının alt kavramlarla ilişkisi vb çok iyi tanımlanmalıdır; bunun için varolan açıklamalardan hareket

etmekle birlikte, sadece bu açıklamalarla yetinilmemelidir; gözlemler, ilgili çalışmaların sonuçları, kompozisyon yazdırma, yargıcı değerlendirme süreçleri vb her türlü yola başvurulmalıdır. Bu süreç bir gömülü kuram (grounded theory) sürecidir.

b) Özelliği ortaya çıkaracak en uygun madde türü ve tepki kategorileri seçilmelidir.

c) *Öncelikle her bileşen/veya faktör için ayrı ayrı ve aynı sayıda madde üretilmelidir*: Alt bileşen ağırlıklarının gerçeğe yakın ortaya çıkarılması için bu bir gerekliliktir. Açıkladıkları varyansın dışında, toplam puan bulmada madde sayısı önemli olduğundan, tüm maddeleri “çorba misali” cevaplayıcılara vermek (hem tepkilerin kirlenmesine (contamination) de yol açar), alt bileşenlerin genel yapı içindeki ağırlıklarını daha gerçekçi saptamada bu yol önemlidir.

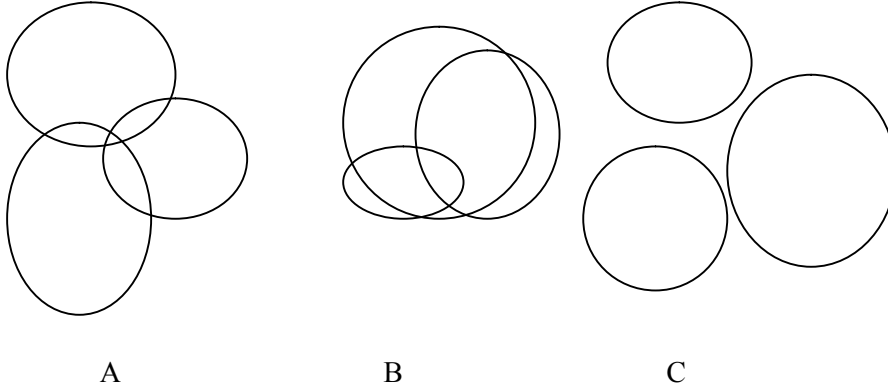
d) *Her alt bileşen ayrı ayrı (aynı) hedef kitleye sunulmalı ve ayrı ayrı açıklayıcı faktör analitik tekniklerin ve döndürme tekniklerinin hemen hepsi uygulanıp en iyi çözüm veren yapı bulununcaya kadar, her bir alt bileşen olabildiğince yaklaşık basit yapıya indirgenmelidir*.

e) *Madde puanlarının yeniden hesaplanması gerekir*: Maddelerin yerleri ve söz konusu bileşenlere katkısını daha iyi saptayabilmek için, her alt bileşen analizinde, önce-geçici olarak 1-0 veya 0-5 gibi önsel ve eşit ağırlıklı analize başlamak kaçınılmaz olmasına rağmen, ilk faktör analizinden (tümü denendikten sonra, en iyi çözümü veren yapı ortaya çıktıktan sonra) elde edilen faktör puanları (veya p , $1-p$, her bir maddenin faktörle regresyon katsayısı, ortakvaryans(communality) veya b parametresi de olabilir; ancak burada faktör puanlarının (faktör yükü \times madde puanı) kullanılması önerilecektir), 1-0 veya 0-5 matrisiyle yer değiştirilip tekrar aynı faktör analizi tekrar edilir; bu işlem Newton-Raphson yöntemi gibi yöntemlerle, madde puanları değişmez oluncaya kadar yinelenir. Böylelikle artık başlangıçtaki önsel eşit ağırlıklı madde puanları, madde-faktör puanlarıyla yer değiştirmiş olur. Tüm alt bileşenler için aynı yöntem izlenir. Bu aşamada, madde güçlük veya ayırt edicilik parametreleri (zaten iyi konumlandırılmış maddeler, dayandıkları istatistiksel temel ve işlemler gereği ayırt edici olacaklardır-hangi yetenek grubu için olduğu da önemli değil) göz ardı edilir; önemli olan ölçeğin ölçtüğü psikolojik yapıyı anlamaya ortaya çıkarmaya çalışmak ve bu yapılar içindeki maddelerin kendi bileşenlerine ve diğer bileşenlere verdikleri katkıyı-yükleri (madde konumları) saptaktır.

f) Ayrı ayrı yaklaşık basit yapıya indirgenmiş alt bileşenlere topluca aynı açıklayıcı faktör analitik teknikleri uygulayarak, alt bileşenler ve altbileşen kesişim alanlarına giren maddeler ile genel yapı içindeki büyüklükleri saptanmalıdır. En son değişmez madde-faktör puanlarıyla tüm alt bileşenler bir araya getirilerek tekrar açıklayıcı faktör analizine tabi tutulur. Bu şekilde, ölçeğin yapısı çok daha gerçeğe yakın elde edilir; madde-faktör puanları, ait oldukları her bir bileşene katkıyı daha net temsil eder; madde-faktör puanları diğer bileşenlere ne kadar katkı sunuyor, bu daha iyi açığa çıkar.

f) Madde-altbileşen haritası çıkarılarak, ölçülen psikolojik yapı çok daha iyi resmedilir ve ölçek kullanıcılara, hangi madde hangi bileşene ne kadar katkı sunuyor, toplamda artık bu madde-faktör puanları kullanılıp değerlendirme yapmak üzere sunulur. Bu bağlamda tüm psikolojik yapıları

Venn diyagramlarıyla resmetmek ve her psikolojik yapıya uygun (tıpkı kimyadaki bileşenler gibi) görsel ve kullanışlı yapılar sunmak avantajlı olacaktır. Bu şekilde, bileşenli yapılarda, bireysel toplam puanları bulmak için, ne altbileşenleri standardize etmek, ne de ağırlıklarıyla çarpıp toplamak söz konusu olacak; bireylerin maddeler bazında, örn., faktör puanlarının toplanması yeterli olacaktır.



Şekil 2: Çeşitli yapıların Venn diyagramıyla gösterimi

Her yapının kesişim alanlarındaki maddeler her iki yapı ile de ilişkili demektir ve doğrusu bu nedenle iki ayrı yapının toplamına katkıları ölçüsünde katılmaları gerekir; ancak dışında kalan alanlardaki maddeler de sadece ilgili yapıya katkı sunuyor anlamı çıkar ve diğer yapılar ile ilişkili olmayan maddelerdir; ancak her birinin o yapıya katkısı faktör puanı kadardır. *Eğer, karmaşık maddeler diye kesişim alanlarındaki maddeler atılırsa, kaçınılmaz olarak ayırık yapılara doğru gidilir ki, bu da ilgili psikolojik yapının yanlış tanımlanması ve ölçülmesi anlamına gelir.*

g) Bu yapı, en son olarak doğrulayıcı faktör analiziyle de test edilir.

Tüm bu aşamalar, geçerlik çalışması değil, ölçek geliştirme (ölçek ve madde analizi) çalışmasıdır. Ölçek geliştirme aşamasında faktör analizi, ölçeğin geçerliği için değil, madde analizi için yapılır. Geçerlik için ayrıca başka büyük örneklemeler üzerinde başka değişkenlerle ilişkisi-farkı, yapının doğrulanması gibi bilindik işlemlerle ayrı çalışmalar yapılması gerekir.

Yeni yöntemde madde puanı ve bireysel puan hesaplama

Üç bileşenli yapı için önerilen kümesel yeni eşitlik;

$$\text{Bileşke puan} = Aw_A + Bw_B + Cw_C = \{A=[A+(A \cap \bar{B})+(A \cap \bar{C})] \} + \{B=[B+(B \cap \bar{A})+(B \cap \bar{C})]\} + \{C=[C+(C \cap \bar{B})+(C \cap \bar{A})]\} + [(A \cap B \cap C)]$$

3 bileşenli, 10 maddeden oluşan bir testin faktör yükü toplamlarını hipotetik veriler üzerinde şu şekilde gösterelim ve madde puanlarını hesaplayalım:

Tablo 2: 3 Bileşenli 10 maddelik testin faktör yüklerine ilişkin hipotetik örnek

Madde	1. Bileşen faktör yükleri	2. Bileşen faktör yükleri	3. Bileşen faktör yükleri	Faktör yükü toplamı (veya ortakvaryans da olabilir)

1	0,76	0,23	0,15	1,14
2	0,87	0,15	0,10	1,12
3	0,65	0,43	0,20	1,28
4	0,18	0,20	0,78	1,16
5	0,12	0,34	0,56	1,02
6	0,20	0,13	0,66	0,99
7	0,10	0,20	0,70	1,00
8	0,24	0,80	0,16	1,20
9	0,20	0,77	0,10	1,07
10	0,19	0,70	0,15	1,04

Görüldüğü gibi, her maddenin puanı farklı olabilmektedir. Buna göre, 6 maddeyi de doğru cevaplamış olan üç farklı kişinin geleneksel yolla testten aldıkları toplam puan aynı olacakken, yeni yöntemle göre farklı farklı olabilecektir. Örneğin;

1. Kişi 1, 3, 4, 6, 7 ve 10. maddeleri doğru cevaplamış olsun: **6,61**
2. Kişi 2, 3, 4, 5, 9 ve 10. Maddeleri doğru cevaplamış olsun: **6,69**
3. Kişi 1, 4, 7, 8, 9 ve 10. Maddeleri doğru cevaplamış olsun: **5,41**

Görüldüğü gibi, aynı sayıda maddeyi doğru cevaplayan üç kişinin aldıkları puanlar birbirinden farklıdır; çünkü her maddenin test içindeki gücü farklıdır. Bu sonuçlara göre doğaldır ki, geleneksel yolla ve yeni yöntemle 3 birey hakkında kesme puanlarına bağlı olarak değişik kararlar verilebilecektir. Bu yöntemle, her alttestteki ham puanı bulup, kendi ağırlığıyla çarpıp toplayarak toplam puanları bulmaya da gerek kalmamaktadır. Bu yöntem ile geleneksel yöntemlerin karşılaştırıldığı bir çalışmada (Yüce, 2012), yöntemin işlediği gösterilmiştir.

Sonuç ve Öneriler

Bu çalışmada, varolan tüm ölçek geliştirme yöntem ve teknikleri ile ölçme kuramlarının, ölçülecek psikolojik yapıyı ortaya çıkarmaktan uzak olduğu savına dayanarak; hem ölçeğin ölçmeyi amaçladığı psikolojik yapıyı ortaya çıkarmada, hem de madde ve birey puanlanmasında kullanılabilecek yeni bir öneri getirilmiştir. Bu yöntemle, özellikle bileşenli yapılarda her maddenin ortak yapıya katkısı da dikkate alındığından bilgi kaybının en aza indirildiği ve bu nedenle bireyler hakkında verilecek kararlarda da önemli değişiklikler olacağı ileri sürülebilir. Bu yöntemin ilk sınaması Yüce (2012) tarafından yapılmış ve oldukça işlevsel olduğu gösterilmiştir.

Önerilen yöntemin, MTK'daki bireysel yetenek kestirimlerine benzediği ileri sürülebilir; bu bakımdan, önerilen yöntemle ve MTK kestirimleriyle elde edilecek bireysel puanların, özellikle bir dış ölçüte göre verilecek kararlar açısından bir karşılaştırılması yapılabilir. Ancak, yeni yöntemin,

psikolojik değişkenlerin yapısını ortaya çıkarmaya uygun bir yaklaşım içerdiği, özellikle çokbileşenli yapılar için işlevsel olacağı; oysa MTK'nın tekboyutluluk sayılısıyla bireysel yetenek kestiriminde bulunduğu ve bu nedenle ölçülecek yapıyı açıklamakta herhangi bir şey önermediği unutulmamalıdır. Çokboyutlu MTK'nın ise şimdilik iki boyutlu maddelerle sınırlı kaldığı ve bu yeni yaklaşımın da ölçülecek psikolojik yapının ne ve nasıl olduğundan çok, ikiboyutlu madde parametrelerini kestirmekle ilgilendiği de dikkate alınmalıdır. Bu bakımdan, KTK veya MTK gibi kuramsal çerçevelerle sınırlandırılmadan, ölçülecek yapının ne ve nasıl olduğu açısından bakıldığında yeni yöntemin daha işlevsel olacağı ileri sürülebilir.

Yeni yöntem hem ikili hem de çoklu puanlanan maddelerden oluşan ölçeklere uygulanabilir. Faktör analitik yöntemlerle ortaya çıkarılacak yapının faktör yükü toplamları yukarıda ele alındığı gibi bulunduktan sonra, bireysel madde puanları, her maddenin faktör yükü toplamıyla bireyin verdiği cevap kategorisi çarpılarak (faktör puanı) hesaplanabilir. Öte yandan, burada önerilen faktör yükü toplamları yerine, regresyon katsayıları, Bartlett puanları, Anderson-Rubin puanları gibi (DiStefano, Zhu & Mindrila, 2009) çeşitli ham ve eşit aralıklı dönüştürmelerin karşılaştırılması da yapılabilir.

Önerilen yeni yöntemin, yeni çalışmalarla üstün ve eksik yanlarının ortaya çıkarılması yararlı olacaktır.

Kaynaklar

- Ackerman, T. A. (1996). Graphical representation multidimensional item response theory analyses. *Applied Psychological Measurement, 20*, 311-329.
- Binet, A. & Simon, T. (1905-1908). The development of the Binet-Simon scale. W. Dennis (ed) içinde, *Readings in the history of psychology*. New York: Appleton-Century-Crofts, 1948.
- Bogardus, E. S. (1925). Measuring social distances. M. Fishbein (ed)'de *Attitude theory and measurement*. New York: John Wiley and Sons, Inc., 1967.
- DiStefano, C., Zhu, M. & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment Research & Evaluation, 14*(20), 1-10.
- Guilford, J. P. (1954). *Psychometrics methods (2nd ed.)*. New York: McGraw-Hill Book Co.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons Inc.
- Guttman, L. (1944). A basis for scaling qualitative data. M. Fishbein (ed)'de *Attitude theory and measurement*. New York: John Wiley and Sons, Inc., 1967.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory; Principles and applications*. Boston: Kluwer Academic Pub.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140, New York.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, No. 7.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). *Sample size in factor analysis. Psychological Methods, 4*(1), 84-99.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. New Jersey: Lawrence Erlbaum Assoc.
- Osgood, C., Suci, G. & Tannenbaum, P. (1957). *The measurement meaning*. Urbana, University of Illinois Press.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer Dordrecht Heidelberg.
- Spearman, C. (1904). General intelligence, objectively determined and measurement. *American Journal of Psychology, 15*, 201-293.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-688.
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, tekboyutlu madde tepki kuramı ve çokboyutlu madde tepki kuramı çerçevesinde incelenmesi*. Yayınlanmamış Doktora Tezi, Mersin Üniversitesi Eğitim Bilimleri Enstitüsü.
- Tabachnick, B. G. & Fidell, S. L. (2001). *Using multivariate statistics (4th ed)*. Need Heights MA: Allyn & Bacon.

- Thorndike, L. R. (1982). *Applied psychometrics*. Boston: Houghton Mifflin Co.
- Thurstone, L. L. (1927) A law of comparative judgement. *Psychological Review*, 34, 273-286.
- Torgerson, W. S. (1958). *Theory and method scaling*. New York: John Wiley and Sons Inc.
- Yüce, G. (2012). *Seviye Belirleme Sınavı alttestlerinin ve ağırlıklarının önsel (a priori) ve sonsal (a posteriori) olarak belirlenmesinin bireylerin sınıflama ve sıralamaları açısından incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Mersin Üniversitesi Eğitim Bilimleri Enstitüsü.