



Tek Boyutluluktan Çok Boyutluluğa Boyutluluğun Çerçevesi

The Framework of Dimensionality From Unidimensionality to Multidimensionality

Fulya BARIŞ PEKMEZCİ

Dr. Öğr. Üyesi ◆ Yozgat Bozok Üniversitesi ◆ fulyabaris@gmail.com ◆ OrcID: 0000-0001-6989-512X

Özet

Boyutluluk kavramı geçmişten günümüze tartışılmalı bir konu olmuştur. Test puanlarının boyutluluğu hakkında en genel anlayış evrende bir testi alanlar arasında testle ilgili tüm farklılıkları tam olarak tanımlamak için gereken minimum boyut veya ölçülen yapıyla ilişkili istatistiksel yetenek sayısının boyut sayısını oluşturduğudur. Bir testin boyutluluğu sadece test maddelerine bağlı değildir. Aynı zamanda, evrende testi alanların maddelerle etkileşimi de testin boyutluluğuna kaynaklık etmektedir. Eğer boyutluluk değerlendirmelerinde test puanları güçlü bir çok boyutluluk gösterirse, diğer bir ifadeyle, bazı boyutların istatistiksel olarak birbirinden bağımsız olduğu görülürse, testin kapsam genişliği değişmeyecek şekilde daha homojen iki veya daha fazla alt test oluşturulması çözüm olabilir. Buna rağmen, birçok test planında maddelerin mantıksal açıdan birbirinden bağımsız olması bir gereklilik olarak görülür. Test geliştiriciler, bazı maddeler arasındaki mantıksal bağlılığın bazı karmaşık yeterliklerin ölçülebilmesi için gerekli olduğunu düşünmektedir. Test puanlarının kesinliği ve doğruluğunun sağlanabilmesi için bu tür mantıksal açıdan birbiriyle ilişkili maddelerin tek bir madde gibi puanlanması gerekir. Eğer bu maddeler bağımsız puanlanacaksa maddeler arası koşullu kovaryansların incelenmesine dayanan istatistiksel teknikler kullanılarak bu maddelerin gerçekten bağımsız bilgi sağlayıp sağlamadığına karar verilmeli ve en azından bazı maddeler puanlamada birleştirilmelidir. Çok boyutluluk ve tek boyutlulukla ilgili net bir ayrım olmamakla birlikte çok boyutluluğun planlanan test yapısından mı yoksa yapıdan bağımsız istenmeyen faktörlerden mi ortaya çıktığı incelenmelidir. Boyutluluğun belirlenmesinde birçok yöntem bulunmaktadır. Bu araştırma kapsamında boyutluluğun belirlenmesinde Paralel analiz ve Velicer'in MAP testinin kullanımına ilişkin bir örnek sunulmuştur.

Anahtar Kelimeler: Boyutluluk, Homojenlik, İstenmeyen (nüans) faktör

Abstract

The concept of dimensionality has always been a controversial subject. The most general understanding of the dimensionality of a test is the minimum number of dimensions or statistical abilities required to fully describe all test-related differences in the universe between test takers. The dimensionality of a test depends on both the test items and the result of the interaction between the test takers and test items in the main population. If a strong multidimensionality is detected in dimensionality assessments, that is, if some dimensions are found to be almost statistically independent, the solution may be to create two or more homogeneous subtests without changing the content of the test. However, in many test plans it is seen as a requirement that items be logically independent of each other. Test developers consider that logically interrelated items are necessary to measure complex abilities. These items should be scored as a single item to ensure the precision and accuracy of test scores in such logically interrelated items. If these items are to be scored independently, it should be decided whether these items provide truly independent information using empirical methods based on the conditional item covariance, and at least some items should be combined in scoring. Although there is no clear distinction between multidimensionality and unidimensionality, it should be examined whether multidimensionality arises from the planned test structure or from the undesirable factors irrelevant targeted construct. There are many ways assessing dimensionality. In the scope of this research, Velicer's MAP test and parallel analysis were used to assess dimensionality.

Keywords: Dimensionality, Homogeneity, Nuisance factor

Giriş

Test puanlarının bir özelliği olarak görülen boyutluluk tartışmalı bir konu olarak geçmişten günümüze taşınmıştır. Boyutluluğun elde edilen veriye mi yoksa ölçülen psikolojik yapıya mı ilişkin olduğuna dair tartışmalar hala devam etmektedir. Bu noktada ne ölçülen yapının ne de verinin, temelde maddelerin çok boyutlu olduğunu savunan görüşler de ortaya çıkmıştır. Özetle, boyutluluğun net bir tanımı yapılamamış ve bu durum günümüze kadar ulaşmıştır.

Bu çalışma kapsamında boyutluluğun geçmişten günümüze Klasik Test Teorisi (KTT) ve Madde Tepki Kuramı (MTK) bağlamında incelemesi yapılmıştır. Ayrıca çok boyutluluk ve tek boyutluluk ayrımında çok boyutluluğun ölçülmek istenen özelliklerden mi yoksa ölçülmek istenmeyen yani karıştırıcı özelliklerden mi kaynaklandığına ilişkin tartışmalara odaklanılmıştır. Son olarak ise boyutluluğun belirlenmesinde sıklıkla kullanılan iki istatistiksel yöntem için örnek bir uygulama sunulmuştur.

Klasik Test Kuramında Boyutluluk

Klasik Test Kuramı'ndan (KTK) başlayarak geçerliğin içerikle ilgili yönü göz önüne alındığında, testin boyutluluğu hakkında formal bir varsayım yoktur. Yalnızca, testi oluşturan maddelerin test planını (teste yer alacak maddelerin konu alanının ve bilişsel düzeyinin belirlenmesi) uygun şekilde yansıtması gerekliliği söz konusudur. Bu gereklilikteki amaç testle ölçülmek istenen yetenek/becerilerin kestirilebilmesine olanak sağlayacak şekilde test planının oluşturulmasıdır. Buna karşılık, KTK'de standart puan güvenilirliğinin belirlenebilmesi için maddelerin homojen olması varsayımı söz konusudur ki bu, tek boyutluluk varsayımına eşdeğer bir varsayımdır (McDonald, 1999).

Temel olarak homojenlik ve onun yakın-eş anlamlısı iç tutarlılık, ölçülen özelliğin iyi tanımlanmadığı; ama madde gruplarının bir şekilde tek boyutluluğa işaret ettiği durumda kullanılmaktadır. Burada bilinen en iyi örnek Cronbach'ın Alfa katsayısıdır. Alfa katsayısının ilgili madde setinin tek boyutlu bir yapıya işaret etmesine ilişkin iki önerme bulunmaktadır. Bunlardan ilki, eğer ilk ortak faktöre ilişkin varyans oranı yüksek ise madde seti tek boyutludur ya da en azından tek boyutluluğa yakındır. İkincisi ise Alfa'nın alt ve üst sınırının birinci ortak faktörden kaynaklanan varyansa eşit olduğudur. İkinci önermenin hiçbir durumda doğru olması söz konusu değildir (McDonald, 1981). İkinci önerme ile ilgili olarak Green vd. (1977), beş faktörden elde edilen varyansın 0,90 olduğu ve ilk faktörün, varyansın yalnızca 0,18'ni açıkladığı bir durumda dahi Alfa'nın yüksek sayılabilecek düzeyde (0,81) olabileceğini göstermişlerdir. Yaygın olarak ilk temel bileşenin açıkladığı varyans tek boyutluluğa karar vermede bir ölçüt olarak kullanılmaktadır. Bununla birlikte ilk ortak faktörün açıkladığı varyans oranı ile diğer ek ortak faktörlerin varlığı ve yokluğu arasında doğrudan bir ilişki söz konusu değildir. Ayrıca ilk faktörü yüksek varyansa sahip çok boyutlu madde setlerini hazırlamak tek boyutlu veri hazırlamaktan daha kolaydır (McDonald, 1981).

Alfa ile ilgili tartışmalar uzun bir tarihe sahiptir. Gulliksen'e (1950) göre eğer bir grup madde, ortak faktör modelinde (common factor model) tek faktörlü yapıya uyum gösteriyorsa o madde grubu homojendir. Bununla birlikte Lord ve Novick (1968) Gulliksen'in bu varsayımının maddeler *tau-eşitliğine* sahipse, başka bir ifadeyle madde kovaryansları eşitse mümkün olacağını söylemiştir. Faktör analitik anlayışa göre bu durum ancak maddelerin aynı yeteneği ölçtüğü durumda söz konusudur. Bununla birlikte tek boyutluluğun sağlanmadığı durumda da alfa katsayısı olduğundan yüksek çıkma eğilimindedir. Feldt ve Qualls (1996) başarı ve zekâ testlerinde, *tau-eşitliğinin* ya da tek faktör varsayımının Alfa'nın kullanımını sınırlandırdığını ve bu testlerde *tau-eşitliği* varsayımının bir dereceye kadar ihlal edilebildiğini belirtmiştir. Cortina (1993) ise araştırmasında yeterli maddeye (örn. 20'den

fazla) sahip ölçeklerde maddeler arası korelasyonların düşük olması durumunda dahi Alfa'nın 0,70'den büyük çıkabildiğini göstermiştir. Araştırmasında 14 maddelik bir ölçeğin iki boyutlu ve boyutlar arası korelasyonun orta düzeyde olduğu durumda Alfa'nın 0,70 ve üstü değerler alabildiğini; üç boyutlu ve boyutların dik olduğu durumda da yeterli madde sayısı ile Alfa'nın yeterli değere ulaşabildiğini göstermiştir. Green vd. (1977) ise çalışmalarında Alfa'nın madde sayısındaki artıştan etkilendiğini ve yüksek çıkma eğiliminde olduğunu belirtmişlerdir. Belirtilen çalışmalardan da anlaşılacağı üzere tek boyutluluk indeksi olarak yaygın kullanımına rağmen, Alfa son derece şüphelidir. Ayrıca Alfa'nın testin uzunluğuna bağlı olması önemli bir sorundur. Bunun nedeni ise kavramsal olarak, bir testin tek boyutluluğu, uzunluğundan bağımsız olması gerektiğidir.

Alfa, testin homojenliğinin nicel bir ölçüsüdür. Ne homojenlik ne de iç tutarlığın açık ve evrensel olarak kabul görmüş bir tanımı bulunmamaktadır. Daha önce de belirtildiği gibi homojenlik ve tek boyutluluk eş anlamlı kavramlardır. Bu nedenle homojenlik madde gruplarının sahip olduğu ya da olmadığı bir özellik olarak görülebilir (McDonald, 1981). Bununla birlikte Alfa katsayısı, güvenilirlik ya da genellenebilirlik katsayısı olarak sadece güvenilirliğin alt sınırı olabilir. Tek boyutluluk için bir karar ölçütü olamaz. İlk ortak faktörün ya da ilk temel bileşenin açıkladığı varyanstan tek boyutluluk yorumu yapılması yetersiz ve kabaca bir değerlendirmedir (McDonald, 1981).

Alfanın dışında tek boyutluluğun belirlenmesi için madde yanıt örüntülerini kullanan ve mükemmel tek boyutlu testin ideal yanıt örüntüsünden farklılaşan yanıtların bir fonksiyonu olduğu temeline dayanan indeksler de bulunmaktadır. Bunlardan biri Gutmann'ın geliştirdiği yeniden üretilebilirlik katsayısıdır. Gutmann'ın katsayısı, madde güçlüklerinin bir fonksiyonu olduğu için bu katsayının alt sınırı 0,50 olarak belirlenmiştir (Hattie, 1985).

Madde yanıt örüntüsü üzerinden hesaplanan tek boyutluluk indeksleri üzerine birçok eleştiri bulunmaktadır. En ciddi eleştiri ölçeklenebilirliğin (örn. mükemmel ölçek) güçlü varsayımı karşılanabildiğinde bu yöntemlerin kendi üst sınırlarına ancak ulaşabileceğidir (Lumbsden, 1950, akt. Hattie, 1985). Diğer büyük eleştiri ise yalnızca bir özelliğin test edilmesi ve olası yetenek bileşimlerinin testte ayırt edilememesidir. Bu gibi durumlarda Loevinger (1944), ölçülen birden fazla yeteneğin mi yoksa sadece bir yeteneğin mi bulunduğunu belirlemeye yardımcı olabilecek faktör analitik yöntemlerin olduğunu öne sürmüştür (Akt. Hattie, 1985). Guilford (1965) ise iki yeteneği ölçen bir testin, her bir maddenin her iki yeteneği de ölçmesi koşuluyla tek boyutlu bir test olarak kabul edilebileceğinin mümkün olduğunu savunmuştur. Başka bir itiraz da, mükemmel bir ölçek oluşturan ancak tek boyutlu görünmeyen bir madde grubu oluşturmanın mümkün olmasıdır. Örneğin, farklı yetenekleri test eden 10 maddelik bir madde grubu düşünün. Bu madde setinde 1. sınıftan 10. sınıfa kadar her sınıfa uygun zorlukta maddeler olsun. Bu test 1. sınıftan 10. sınıfa kadar her sınıf düzeyi için ortalama başarıya sahip 10 öğrenciye verilsin. Bu senaryoya göre mükemmel ölçeğin oluşması oldukça olasıdır. Bu nedenle mükemmel yeniden üretilebilirlik maddelerin tek boyutlu olduğunun ölçüsü için yeterli değildir (Hattie, 1985). Bir testin tek boyutlu olduğunu söylemek, o boyutu tanımlamaz. Aynı şekilde, bir testin güvenilir olduğunu söylemek, neyi güvenilir olarak ölçtüğünü belirlemez.

Cronbach, düşük bir ortalama korelasyonun homojen olmayan bir testin göstergesi olabileceğine dikkat çekmiş ve ortalama korelasyon düşük olduğunda, bir testin daha homojen alt testlere bölünüp bölünemeyeceğinin belirlenmesinde yalnızca maddeler arası korelasyonların incelenmesini önermiştir (Hattie, 1985). Armor (1974), düşük veya negatif korelasyon örüntüleri için maddeler arası korelasyonların genellikle incelenmediğini bununla birlikte bu incelemenin boyutluluğa karar vermek için gereken tüm bilgileri içeren olası en önemli adım olduğunu savunmuştur. Ayrıca maddeler arası korelasyonların sıfıra yakın olduğu durumlarda bu maddelerinin sayısının değerlendirilerek, tek boyutluluk yanlısına düşmekten kaçınılabileceğini de belirtmiştir.

Buraya kadar tek boyutluluğun KTK bağlamında tanımından, homojenlik kavramından ve boyutluluk için hesaplanan bazı indekslerden bahsedilmiştir. Bu indekslerin dışında temel bileşenlere dayanan indeksler ve faktör analizine dayalı indeksler de bulunmaktadır. Tüm bu indekslerin yanı sıra Madde Tepki Kuramı'na (MTK) göre de boyutluluk farklı açılardan ele alınmaktadır.

Madde Tepki Kuramında Tek Boyutluluk

Madde Tepki Kuramı için tek boyutluluk ölçülen özelliği diğer bir ifadeyle yeteneği kontrol altına aldıktan sonra tüm madde yanıtlarının yerel olarak bağımsız olduğunun varsayılmasıdır. Uygulamada, katı tek boyutluluk varsayımının her zaman bir dereceye kadar ihlal edildiği genel olarak kabul görmektedir. Yerel bağımsızlık ilkesinin tam olarak anlamı, yalnızca faktör puanları kısmen çıkarıldığında test puanlarının kısmi korelasyonlarının sıfır olması değil; aynı zamanda farklı test puanlarının istatistiksel olarak tamamen bağımsız olmasıdır.

Çoğu örtük özellik modelinde temel varsayım testle ölçülen varsayımsal değişkenin tek boyutlu örtük uzayda tanımlanabildiğidir (Lord & Novick, 2008; Whitely & Dawis, 1974). Bu varsayım geleneksel madde analizi ve faktör analizi süreçleriyle birçok psikolojik test için karşılanmaktadır. Bununla birlikte, eğer örtük özellik teorileri standart ve sınıf içi başarı testi alanına genişletilirse, bu varsayımın ihlal düzeyi belirlenmelidir. Reckase'e (1979) göre başarı testleri genellikle yalnızca faktör ölçümleri vermesi için tasarlanan bir metodolojiyle geliştirilmez. Bunun yerine göreceli öneme göre oluşturulan özellikler tablosuyla davranışsal hedeflerin ölçülmesi için geliştirilir. Daha sonra maddelerin bu özelliklerle eşleşmesi sağlanarak yazımı yapılır. Bu yolla üretilen testler nadiren tek bir örtük değişkeni ölçer ve sıklıkla karmaşık bir faktör yapısına sahip olur (Reckase, 1979).

Örtük uzayın tamamının doğası ve boyutluluğunun tanımı psikometristin ilgilendiği belirli evrene ve belirli özelliğe dayanmaktadır. Psikometrist, verilen madde setine etki edecek tüm önemli psikolojik boyutları ekleyip “ölçme hatasını” oluşturan diğer değişkenleri dışlayarak kendi örtük uzayını tanımlamayı ister. Bununla birlikte mantıksal olarak bu değişkenleri kolayca “ölçme hatasından” arındırmak imkânsız gözükmektedir. Tam örtük uzayın boyutluluğu ne dağılımsal varsayımlara ne maddeler arası korelasyonun hesaplanması için seçilen ölçüye ne de örtük değişkenlerin dönüşümüne bağlıdır. Bu nedenle, tam örtük uzayın boyutluluğu ortak faktör sayısından daha temel bir kavramdır (Lord & Novick, 2008). Örneğin başarı değişkeni hem bilişsel hem de duyuşsal boyutu olan psikolojik bir değişkendir. Başarıya etki eden motivasyon, tutum, ilgi vb. gibi ölçemediğimiz; ortam, test, puanlayıcı vb. gibi ölçemediğimiz birçok değişkenlik kaynağı da söz konusudur. Ölçmediğimiz bu değişkenlik kaynakları ölçme hatasına katkı sağlamaktadır. Çünkü bu değişkenlik kaynaklarını, ölçülen özellik olan başarıdan tamamen arındırmak mümkün değildir. Bu nedenle Lord ve Novick'in (2008) bahsettiği gibi ölçülen özelliklerin yanı sıra ölçme hatası da boyutluluğa hizmet etmektedir.

Neden Tek Boyutluluk?

Thurstone (1931) en kullanışlı ölçümlerin, sadece tek bir şeyin ölçüldüğü durumlar olduğu fikrini ortaya atmıştır. Bu durumu “Herhangi bir nesnenin veya varlığın ölçümü ölçülen nesnenin yalnız bir özelliğini tanımlar. Bu ölçmenin evrensel bir karakteristiğidir” (Thurstone, 1931, s.259) şeklinde ifade etmiştir. Bu görüş McNemar (1946) ve Stout (1987) tarafından da desteklenmiştir.

Stout (1987) bir testin neden tek boyutlu olması gerektiği ile ilgili üç önemli gerekçe sunmuştur:

- i. Ölçülmek istenen yapının düzeyleri bir ya da birden fazla başka yapıların değişen düzeylerinden anlamlı olarak etkilenmemelidir.

- ii. Bir değerlendirmenin bireysel farklılıkları belirlemede kullanılabilmesi, aynı puana sahip iki bireyin ölçülen özellik bağlamında aynı olduğunun belirlenmesi için yapılan değerlendirmenin tek boyutlu yapıyı ölçmesiyle mümkündür.
- iii. Tek boyutluluk toplam puan hesaplamadan önce sağlanmalıdır yoksa madde ya da birey parametreleri geçerli olmayabilir.

İki varlık arasında doğrudan karşılaştırma yapmak için karşılaştırmanın yalnız tek bir özellikten elde edilmiş ölçümlere dayanması gerekir (Horton, 2013). Diğer bir deyişle bu karşılaştırma için tek boyutlu ölçümler gereklidir.

Stout'a (1987) göre temel boyutluluk, bir ya da daha fazla küçük (minör) boyutun varlığında baskın bir boyutun olması ve baskın boyuta dayalı yapılan kestirimlerin küçük boyutların varlığından etkilenmeyecek kadar güçlü olmasıdır. Bu çerçevede, tek boyutlu bir model uygulandıktan sonra maddeler arası artık kovaryans ortalamasının madde sayısı arttıkça sifıra yaklaşması yanıtların temel tek boyutluluğu sağladığının göstergesi olarak öne sürülmüştür. Temel boyutluluk, MTK analizlerinde tek boyutluluk varsayımının karşılanması gerektiği durumlarda tek boyutluluk için kanıt olarak kullanılmaktadır (Horton, 2013). Benzer bir anlayışla Smith (2002), tek boyutluluğu ikili bir "evet" veya "hayır" kararı olarak kabul etmemektedir. Bunun aksine tek boyutluluğun bir süreklilik arz ettiğini düşünmektedir. Bu nedenle, bir ölçeğin tek boyutlu olup olmadığını sormak yerine; çok boyutluluğun sürekliliğin hangi noktasında madde ve kişi tahminlerinin yorumlanmasını tehdit ettiği sorusunu sormanın daha uygun olabileceğini belirtmiştir.

Boyutluluğu Etkileyen Faktörler ve Çok Boyutluluk

Bir testin boyutluluğu sadece test maddelerine bağlı değildir, testi alanlara (ör. aday, öğrenci vb.) ve testi alanların maddelerle etkileşimlerine de bağlıdır (Ackerman, 1994). Testi alan popülasyonun testin boyutluluğunu belirlemedeki rolünü göstermek için her biri farklı bir cebir yeteneği ve okuduğunu anlama becerisinin bileşimini ölçen birden fazla test maddesi olduğu örnek bir durumdur; hedef kitle hem cebir yeteneği hem de okuma becerisinin belirli düzeylerine göre değişkenlik gösteriyorsa yani her ikisinde de farklı düzeyde yetkinlik ve beceriye sahipse test iki boyutlu olacaktır. Buna karşılık, tüm maddelerin gerektirdiği okuma becerisi düzeyine göre popülasyon homojen ise test tek boyutlu olacak ve bireyler arası farklılıklar sadece cebir yeteneği üzerinden yapılabilecektir. Teste yer alan maddelerin tamamının aynı cebir yeteneği ve okuma becerisini gerektirmesi durumunda ise ilgili beceri veya yeteneklerin sayısının iki olacağı ancak testin istatistiksel boyutunun bir olacağı unutulmamalıdır (Reckase vd., 1988). Testte birden fazla baskın faktörün görgül kanıtı varsa (veya eşdeğer olarak bir yeteneği kontrol ettikten sonra yerel madde bağımlılığı kanıtı) KTK ve MTK'de tek boyutluluk varsayımını ihlal etmenin sonuçlarını dikkate almak gerekir.

Çok boyutluluk, planlanan test yapısından kaynaklanabilir ya da yapıdan bağımsız istenmeyen faktörlerden ortaya çıkabilir. Bazı koşullar altında (örneğin, çok boyutluluk neredeyse tamamen planlanan test yapısından kaynaklandığında ve ilgili yetenek bileşenleri en azından orta derecede ilişkili olduğunda), test puanları tek boyutluluk varsayımının ihlaline karşı dayanıklıdır. Diğer koşullar altında (örneğin, test tarafından ölçülen bileşen yetenekleri zayıf bir şekilde ilişkilendirildiğinde veya yapıyla alakasız güçlü faktörler olduğunda) ihlalin sonuçları test geçerliğini, güvenilirliğini, yansızlığını ve puanların karşılaştırılabilirliğini olumsuz yönde etkilenecektir (Tate, 2012).

Testin oluşturulması aşamasında, test içerik alanını, madde format(lar)ını ve madde oluşturma sürecini belirlerken, çok boyutluluğun ilgili kaynaklarının farkında olmak ve ilgili varsayım ihlallerinin ne zaman ciddi sonuçlara yol açabileceğini anlamak önemlidir. Örneğin, MTK tabanlı bir test için içerik

alanının kapsamını belirlerken, test geliştiricisi kapsamla ilgili geçerliğin yalnızca kapsamla ilişkili yetenek veya beceriler arasındaki korelasyonlar orta düzeyde veya daha güçlü olduğunda tek boyutluluk varsayımının ihlallerine karşı dayanıklı (robust) olduğunu unutmamalıdır (Tate, 2012). O halde amaç, öğrenci performansının değerlendirmesini sağlayacak kadar geniş, ancak yine de kapsamı oluşturan yetenek veya beceriler arasında önemli korelasyonların öngörülmesine izin verecek kadar homojen olmayan bir alan belirlemektir (Luecht & Miller, 1992).

Geliştirme ve değerlendirme aşamalarında, geniş bir alan testinden (birden fazla derse ait test) elde edilen verileri kullanan test geliştiricileri, bu beklentiye test boyutunun ampirik değerlendirmesiyle doğrulamak isterler. Ancak, değerlendirmede beklenen boyutu yansıtan güçlü çok boyutluluk tespit edilmesine karşın bazı boyutların neredeyse istatistiksel olarak bağımsız olduğu görülebilir. Bu durumda olası bir çözüm, bir alanı hala nispeten geniş olan ancak daha homojen iki veya daha fazla alana bölerek test alanını yeniden gözden geçirmek olabilir. Daha sonra her yeni alan için bir test puanı tahmini yapılabilir (Luecht & Miller, 1992).

Bir yeteneğin bilişenlerinin yani teste ait alt puanların (test kapsamının alt düzeylerini oluşturan yetenekler örneğin: cebir, geometri, olasılık alanlarından alınan puanlar) olası tanı amaçlı kullanımı, yalnızca testin istatistiksel yapısının planlanan alt puanlarla tutarlı olması beklendiğinde uygundur (Tate, 2012). Burada kastedilen genel bir yetenek ve onu oluşturan alt yeteneklerdir. Örneğin, test geliştiriciler, ölçülen yeteneği oluşturan alt yeteneklere ilişkin puanlarının birbirleriyle yüksek düzeyde ilişkili olmasını bekliyorlarsa test sonuçlarının raporlanmasında alt yeteneklere ilişkin puanların değerini sorgulamalıdır. Tate'e (2012) göre alt yetenekler planlı oluşturulmuş ancak deneysel değerlendirme istatistiksel yapının kısmen mevcut olduğunu gösteriyorsa, daha yüksek düzeyde ilişkili alt yeteneklerin bazılarını birleştirmek, daha az sayıda yeniden adlandırılmış alt yetenekler oluşturmak mümkün olabilir. Eğer değerlendirme testin tek boyutlu olduğunu gösteriyorsa, alt yeteneklere ilişkin puanlar rapor edilmemelidir.

Çok boyutluluğun yukarıda geçen planlı kapsam yapısına ek olarak farklı kaynakları da bulunmaktadır. Örneğin, bir maddenin doğru cevabı çoğu kez bir ya da birden fazla istenmeyen (nuisance-nüans) ya da kapsam-ilişkisiz yetenekleri (ölçülen kapsam dışında test puanına etki eden yetenekler) içerebilir (Tate, 2012). Örneğin sadece matematik yeteneğinin ölçülmek istendiği bir testte, testi alanların (aday, öğrenci vb.) okuma yeteneği ve kararlılık durumları bireyin yeteneğinin kestirilmesi üzerinde etkili olacaktır. Bu iki kaynak (okuma yeteneği ve kararlılık) kapsam-ilişkisiz kaynaklardır. Bu tür testle ölçmek istenen özellik üzerinde istenmeyen (nuisance-nüans) faktörlerin önemi arttıkça (yani, tahmin edilen test bileşimindeki ağırlıkları arttıkça) geçerliğin içerikle ilgili yönüne olan tehdit daha ciddi hale gelir. Diğer istenmeyen faktörler, testin uygun olmayan uygulama koşullarından kaynaklanabilir. Örneğin, bir oturumda uzun bir test vermek istenmeyen (nuisance) faktörlerin oluşmasına neden olur. Ayrıca madde formatları da istenmeyen (nuisance) faktörlerin oluşmasına kaynaklık edebilir. Özellikle okuma parçalarında yer alan ortak köklü maddeler yerel bağımsızlık varsayımını ihlal ederek çok boyutluluğa neden olmaktadır. Yerel bağımlılık varsayımının kanıtları, altta yatan birden fazla boyut olduğu durumlarda ortaya çıkar. Bununla birlikte, yerel bağımlılık otomatik olarak çok boyutluluğun kanıtı olarak yorumlanmamalıdır. Ayrıca, yanlılık da istenmeyen (nuisance) faktör nedeni olabilir. Başka bir ifadeyle değişen madde fonksiyonu içeren maddeler de yapı ilişkisiz faktörlere sebep olabilmektedir (Tate, 2003).

Boyutluluk Belirleme Yöntemleri

Boyutluluğun diğer bir deyişle boyut/faktör sayısının belirlenmesinde birçok yöntem bulunmaktadır. Özdeğeri 1'den büyük olan faktörlerin boyutluluk açısından değerlendirilmesi (Kaiser-

Guttman yöntemi), Yamaç Birikinti Grafiği'nin (Scree Plot) incelenmesi, son yıllarda sıklıkla söz edilen Horn'un Paralel Analizi ve Velicer'in MAP (Minimum Average Partial) testi bunlardan bazılarıdır (Fabrigar vd., 1999; Ford vd., 1986).

Alanyazındaki birçok çalışmada Kaiser-Guttman yönteminin ve Yamaç Birikinti Grafiği yönteminin faktör sayısını belirlemede yetersiz olduğu ve gerekenden fazla sayıda faktör önerdiği belirtilmiştir (Fabrigar vd., 1999; Ford vd., 1986; Horn, 1965; Silverstein, 1977, 1987; Zwick ve Velicer, 1986). Horn'un Paralel Analizi yönteminin ise diğer yöntemlerle karşılaştırıldığında faktör sayısını belirlemede ideal sonuçlar ürettiği aynı şekilde Velicer'in MAP testinin de faktör sayısının belirlenmesinde doğru sonuçlar ortaya çıkardığı görülmüştür (Dinno, 2009; Revelle, 2007; Silverstein, 1977, 1987; Zwick ve Velicer, 1986).

Horn'un Paralel Analizi yöntemi, boyut sayısının belirlenmesi aşamasında simülatif olarak paralel bir veri seti oluşturup bu veri setinin beklenen özdeğerinin hesaplanarak gerçek veriye ait özdeğer ile karşılaştırılması esasına dayanmaktadır. Horn (1965), evrende p kadar değişkenin oluşturduğu korelasyon matrisine ait özdeğerlerin 1 olacağını, bununla birlikte örneklem üzerinde yapılan analizlerde ilk özdeğerin 1 daha sonraki özdeğerlerin örneklem hatasının korelasyon matrisine eklenmesi nedeniyle 1'den küçük olacağını belirtmiştir. Bu nedenle gerçek veriden elde edilen özdeğerlerin paralel olarak üretilen veriden üretilen özdeğerle karşılaştırıldığında daha yüksek özdeğer veren bileşen ya da faktörler, önemli faktörler olarak belirlenir (Zwick ve Velicer, 1986).

Faktör sayısını belirlemede kullanılan diğer bir yöntem ise Velicer'in MAP testidir. Velicer'in MAP testi ise bir dizi kısmi korelasyonlar matrisinin ve temel bileşen analizlerinin incelenmesine dayanmaktadır. Velicer'in yöntemi korelasyon matrisini oluşturan bileşenlerin tek tek dışarıda tutularak kısmi korelasyonların hesaplanıp bu korelasyonların kareler ortalamasının alınması esasına dayanır. Bileşenlere ait korelasyon matrisinin oluşturulması ve ardından ilk basamakta birinci bileşenin matristen ayrılması ile analiz başlar. Kalan bileşenlere ait kısmi korelasyonların kareler ortalaması hesaplanır ve bir sonraki aşamada ilk iki bileşen korelasyon matrisinden çıkarılır. Bu adımlar bileşen sayısının bir eksiği kadar tekrarlanır. Bu adımlarda en düşük kısmi korelasyon ortalamasının elde edildiği basamak faktör sayısının karar verildiği noktadır (O'Connor, 2000; Velicer, 1976). Özetle kısmi korelasyon değerlerinin kareleri için en düşük özdeğer faktör sayısını verir (Velicer, 1976).

Çalışmanın devam eden kısmında paralel analiz (Horn, 1965) ve MAP (Velicer, 1976) analizi kullanılarak boyut sayısının belirlenmesine ilişkin birer uygulama, erişimi açık, ücretsiz ve son yıllarda sıklıkla kullanılan R ortamında yapılarak sunulmuştur.

Örnek Uygulama

Uygulama verisi olarak Dennis ve Wal (2010) tarafından geliştirilen Bilişsel Esneklik Ölçeği'nin farklı sınıf düzeylerinden ve farklı lisans programlarından 1232 lisans öğrencisine uygulanması sonucunda elde edilen veriler kullanılmıştır. Bilişsel Esneklik Ölçeği *Alternatifler* ve *Kontrol* olmak üzere iki boyuttan oluşmaktadır. *Alternatifler* boyutunda 13 *Kontrol* boyutunda ise 7 madde bulunmaktadır. Ölçeğin uygulanması için gerekli etik kurul izni 08.03.2022 tarihinde yapılan toplantıdaki 2022/39 sayılı karar ile Recep Tayyip Erdoğan Üniversitesi Etik Kurulundan alınmıştır.

Bilişsel Esneklik Ölçeği'nin boyut sayısını belirlemek amacıyla Velicer'in MAP testi ve Horn'un Paralel analizi kullanılmıştır. Analizlerde ücretsiz olması ve her iki analizi de yapabilmesi nedeniyle RStudio 1.4 programı kullanılmıştır. Horn'un paralel analizi için *paran* (Dinno, 2018) paketi kullanılmış olup R Studio'da kullanılan kodlar ve çıktıları Şekil 1'de verilmiştir.

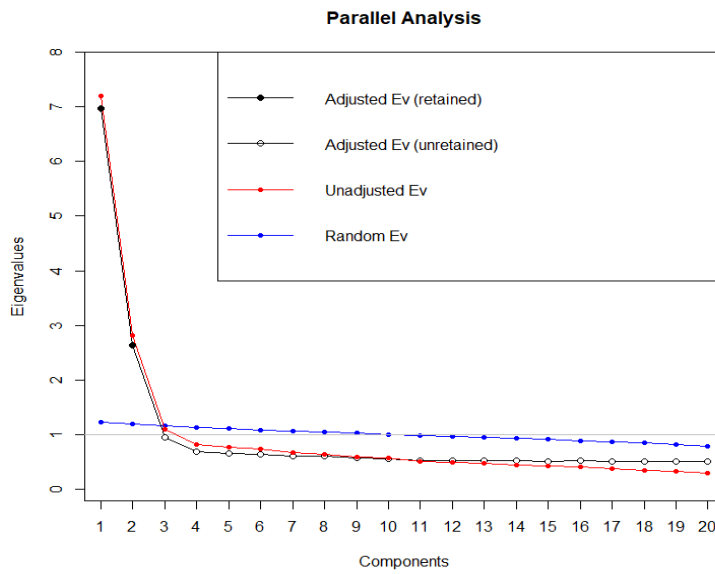
Şekil 1. Paralel Analiz Kodları ve Çıktıları

```

paran(data=BEE, cfa=FALSE, graph=TRUE,
      color=TRUE, col=c("black","red","blue
Using eigendecomposition of correlation matrix.
Computing: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Results of Horn's Parallel Analysis for component retention
600 iterations, using the mean estimate
-----
Component   Adjusted   Unadjusted   Estimated
            Eigenvalue Eigenvalue   Bias
-----
1           6.964998  7.195172    0.230173
2           2.630992  2.820352    0.189360
-----
Adjusted eigenvalues > 1 indicate dimensions to retain.
(2 components retained)

```



Paralel analiz için Şekil 1’de verilen R çıktıları incelendiğinde özdeğeri 1’den büyük iki boyut olduğu görülmüştür. Paralel analize ilişkin grafikler incelendiğinde, paralel veri için çizilen grafiğin mavi renkle (Random Ev), gerçek veriye ilişkin grafiğin siyah (Adjusted Ev) ve kırmızı (Unadjusted Ev) ile gösterildiği görülmektedir. Grafik yorumlanırken paralel analiz için çizilen grafiğin yani mavi çizginin üstünde kalan gerçek veriye ait siyah grafiğin üzerindeki özdeğere ilişkin noktalar sayılır. Bu noktalar ölçeğin boyut sayısını vermektedir. Bu örnek uygulamada iki faktör noktası görülmektedir yani ölçeğin iki boyuttan oluştuğu söylenebilir. Velicer’in MAP testi için RStudio 1.4’de *EFA.dimensions* (O’Connor, 2022) paketi kullanılmış olup ilgili kodlar ve çıktılar Şekil 2’de verilmiştir.

Şekil 2. Velicer'in MAP Testi Kodları ve Çıktıları

```

map_BEE<-MAP(data=BEE,corkind="pearson",1232)
MINIMUM AVERAGE PARTIAL (MAP) TEST
Number of cases = 1232
Number of variables = 20
Specified kind of correlations for this analysis: Pearson
Total Variance Explained (Initial Eigenvalues):

      Eigenvalues    Proportion of Variance
Factor 1           7.20                0.36
Factor 2           2.82                0.14
Factor 3           1.10                0.06
Factor 4           0.81                0.04
Factor 5           0.77                0.04
Factor 6           0.73                0.04
Factor 7           0.67                0.03
Factor 8           0.64                0.03
Factor 9           0.60                0.03
Factor 10          0.57                0.03
Factor 11          0.51                0.03
Factor 12          0.50                0.02
Factor 13          0.48                0.02
Factor 14          0.45                0.02
Factor 15          0.42                0.02
Factor 16          0.41                0.02
Factor 17          0.37                0.02
Factor 18          0.35                0.02
Factor 19          0.33                0.02
Factor 20          0.29                0.01

      Cumulative Prop. Variance
Factor 1           0.36
Factor 2           0.50
Factor 3           0.56
Factor 4           0.60
Factor 5           0.64
Factor 6           0.67
Factor 7           0.70
Factor 8           0.74
Factor 9           0.77
Factor 10          0.79
Factor 11          0.82
Factor 12          0.85
Factor 13          0.87
Factor 14          0.89
Factor 15          0.91
Factor 16          0.93
Factor 17          0.95
Factor 18          0.97
Factor 19          0.99
Factor 20          1.00

Velicer's Average Squared Correlations

      root    Avg.Corr.Sq.    Avg.Corr.power4
0         0.12101         0.02444
1         0.03002         0.00245
2         0.01088         0.00032
3         0.01317         0.00082
4         0.01685         0.00118
5         0.02335         0.00439
6         0.02867         0.00510
7         0.03689         0.00855
8         0.04441         0.01091
9         0.05323         0.01356
10        0.06394         0.01848
11        0.08009         0.02625
12        0.09993         0.03274
13        0.12286         0.04953
14        0.15440         0.06113

```

15	0.20411	0.09081
16	0.27025	0.14503
17	0.36344	0.23511
18	0.50432	0.38450
19	1.00000	1.00000

The smallest average squared correlation is 0.01088
The smallest average 4rth power correlation is 0.00032
The number of components according to the original (1976) MAP Test is = 2
The number of components according to the revised (2000) MAP Test is = 2

Öncelikle Şekil 2’de verilen ortalama kısmi korelasyonlarının kareleri (Avg.Corr.Sq) ve ortalama kısmi korelasyonların dördüncü kuvvetine (Avg.Corr.power4) ilişkin öz değerler incelenmelidir. MAP testi sonuçlarına göre kısmi korelasyonun karesine ait öz değerler sırasıyla 0.12101, 0.03002, **0.01088**, 0.01317, 0.01685... biçimindedir. Kısmi korelasyonların karelerine göre özdeğerler incelendiğinde en düşük özdeğere ikinci basamakta ulaşıldığı görülmüştür. O’Connor (2000) boyutluluğun belirlenmesinde kısmi korelasyonun dördüncü kuvvetinin kullanılmasının daha etkili sonuçlar verdiğini belirtmiştir. Kısmi korelasyonların dördüncü kuvvetleri sırasıyla 0.02444, 0.00245, **0.00032**, 0.00082, 0.00118... olarak bulunmuştur. Buradan da görüldüğü üzere en küçük özdeğer ikinci basamakta elde edilmiştir.

Sonuç olarak kuramsal olarak iki boyutlu olan Bilişsel Esneklik Ölçeği'nin boyut sayısının iki olduğu hem Paralel analiz hem de MAP analizi ile de desteklenmiştir. Elde edilen sonuçlar alan yazında belirtilen Horn’un Paralel Analizi ve Velicer’in MAP analizinin faktör sayısını belirlemede ideal sonuçlar ürettiği (Dinno, 2009; Silverstein, 1977, 1987; Revelle, 2007; Zwick ve Velicer, 1986) bulgusunu da desteklemektedir.

Sonuç ve Öneriler

Test boyutluluğu, bir testi alan grubun testle ölçülen özelliğe ilişkin bireyler arası tüm farklılıkları evrende tam olarak tanımlamak için gereken minimum örtük değişken sayısıdır (Tate, 2012). Bir testin boyutluluğunu etkileyen birçok faktör bulunmaktadır. Testin boyutluluğunu testi alanlardan ve testi alanların maddeler ile etkileşiminden ayrı görmek doğru değildir. Testin içeriğine ve amacına dayanan önemli noktalar, uygun bir nihai istatistiksel model seçimine rehberlik etmektedir.

Tek boyutluluk için KTK kapsamında net bir tanım olmamakla birlikte maddelere verilen yanıtların homojenliği tek boyutluluk varsayımına eş değer görülmektedir (McDonald, 1999). MTK kapsamında ise tek boyutluluk ölçülen özellik kontrol altına alındığında madde yanıtlarının birbirinden bağımsız olması durumudur. Ölçülen özellik çoğunlukla örtük uzayda tanımlanmaktadır. Örtük uzayın tamamının doğası ve boyutluluğunun tanımı psikometristin ilgilendiği belirli evrene ve belirli özelliğe dayanmaktadır. Bununla birlikte genel olarak çok boyutluluk ve tek boyutluluk ile ilgili net ve keskin bir ayırım olmamakla birlikte çok boyutluluğun planlanan test yapısından mı yoksa yapıdan bağımsız istenmeyen faktörlerden mi ortaya çıktığı incelenmelidir.

Çok boyutluluk tamamen planlanan test yapısından kaynaklandığında ve ölçülen özelliğin ilişkili olduğu yetenek bileşenleri en azından orta derecede ilişkili olduğunda test puanları tek boyutluluk varsayımının ihlaline karşı dayanaklıdır. Diğer durumda örneğin test ile ölçülen özellik ilişkili yetenekler birbiriyle zayıf ilişkilere sahipse tek boyutluluğun ihlali test geçerliği ve güvenilirliğini, ölçümlerin yansızlığını olumsuz etkileyecektir (Tate, 2012). Küçük (minör) yani açıkladığı varyans oranının çok düşük olduğu faktörlerin önemi tek boyutluluk-çok boyutluluk karar noktasında önemli belirleyicilerden biridir.

Kaynaklar

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278. https://doi.org/10.1207/s15324818ame0704_1
- Armor, D. J. (1973). Theta reliability and factor scaling. *Sociological methodology*, 5, 17-50. <https://doi.org/10.2307/270831>
- Cortina, J., M. (1993). What is coefficient Alpha? An Examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Dinno, A. (2009). Exploring the sensitivity of Horn's paralel analysis to the distributional form of random data. *Multivariate Behavioral Research*, 44, 362-388. <https://doi.org/10.1080/00273170902938969>
- Dinno, A. (2018). *Horn's test of principal components/factors*. Package Paran.
- Drasgow, F., & Parsons, C. (1983). Applications of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199. <https://doi.org/10.1177/014662168300700207>
- Enders, C. K., & Bandalos, D. L. (1999). The effects of heterogeneous item distributions on reliability. *Applied Measurement in Education*, 12(2), 133-150. https://doi.org/10.1207/s15324818ame1202_2
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Feldt, L. S., & Qualls, A. L. (1996). Bias in coefficient alpha arising from heterogeneity of test content. *Applied Measurement in Education*, 9(3), 277-286. https://doi.org/10.1207/s15324818ame0903_5
- Ford, J. K., MacCallum, R. C. & Tait, M. (1986). The applications of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39, 291-314. <https://doi.org/10.1111/j.1744-6570.1986.tb00583.x>
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37(4), 827-838. <https://doi.org/10.1177/001316447703700403>
- Gulliksen, H. (2013). *Theory of mental tests*. Routledge.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied psychological measurement*, 9(2), 139-164. <https://doi.org/10.1177/014662168500900204>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Lord, F. M. & Novick M. R. (2008). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company.

- Luecht, R. M., & Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement*, *16*(3), 279-293. <https://doi.org/10.1177/014662169201600>
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117. <https://doi.org/10.1111/j.2044-8317.1981.tb00621.x>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, *43*(4), 289–374. <https://doi.org/10.1037/h0060985>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402.
- O'Connor, B. P. (2022). *Exploratory factor analysis functions for assessing dimensionality*. Package EFA dimensions.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*(3), 207-230. <https://doi.org/10.2307/1164671>
- Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, *25*(3), 193-203. <https://doi.org/10.1111/j.1745-3984.1988.tb00302.x>
- Revelle, W. (2007). Determining the number of factors: The example of the NEO-PI-R. <http://personality-project.org/r/book/numberoffactors.pdf>.
- Silverstein, A. B. (1977). Comparison of two criteria for determining the number of factors. *Psychological Reports*, *41*, 387-390. <https://doi.org/10.2466/pr0.1977.41.2.387>
- Silverstein, A. B. (1987). Note on the paralel analysis criterion for determining the number of common factor or principal components. *Psychological Reports*, *61*, 351-354. <https://doi.org/10.2466/pr0.1987.61.2.351>
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589-617. <https://doi.org/10.1007/bf02294821>
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293-325. <https://doi.org/10.21236/ada207301>
- Tate, R. (2012). Test dimensionality. In Tindal, G., & Haladyna, T. M. (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 181-213). Routledge.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, *27*(3), 159-203. <https://doi.org/10.1177/0146621603027003>

- Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, 26(3), 249–269. <https://doi.org/10.1037/h0070363>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327. <https://doi.org/10.1007/bf02293557>
- Whitely, S. E., & Dawis, R. V. (1974). The nature of objectivity with the Rasch model. *Journal of Educational Measurement*, 11(3), 163-178. <https://doi.org/10.1111/j.17453984.1974.tb00988.x>
- Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432-442. <https://doi.org/10.1037/0033-2909.99.3.432>

Extended Abstract

Introduction

Dimensionality is a frequently debated subject. The source of dimensionality is a part of this discussion. Some views argue that dimensionality arises from the measured construct, while others purport that dimensionality is based on data. In addition, other views are on dimensionality can be attributed to items. Consequently, a clear definition of dimensionality could not be made so far. In this study, dimensionality was examined in the context of Classical Test Theory (CTT) and Item Response Theory (IRT). In addition, the concepts of multidimensionality and unidimensionality were discussed.

Dimensionality in Classical Test Theory

There is no definite assumption about the dimensionality of a test in CTT. According to CTT, the items that constitute the test should reflect the test plan appropriately. This necessity aims to create a test plan to measure the intended construct. In CTT, homogenous items are equivalent to the unidimensionality assumption (McDonald, 1999).

Put simply, *homogeneity* and its close-synonym *internal consistency* are used when the measured trait is not well defined; but item groups somehow indicate unidimensionality. The best-known example here is Cronbach's Alpha coefficient. Discussions about alpha have a long history. According to Gulliksen (1950), if a group of items constitute a single factor in the common factor model, it indicates that item group is homogeneous. However, Lord and Novick (1968) stated that this assumption is possible if the items have tau-equivalence. According to the factor analytic view, this is observed only when the items measure the same ability. However, in cases where unidimensionality cannot be achieved, the alpha coefficient tends to be overestimated. Feldt and Qualls (1996) stated that in achievement and intelligence tests, the tau-equivalence or single-factor assumption restricts the use of alpha and that the tau-equivalence assumption can be violated to some extent in these tests. On the other hand, Cortina (1993) showed that with an adequate number of items, alpha can be greater than .70 even when the inter-item correlations of scales are low (e.g., more than 20). In his research, if scale is two-dimensional and the correlation between dimensions is moderate, alpha can be .70 and over; it has been shown that alpha can reach an adequate value with a sufficient number of items.

Despite its widespread use as an index of unidimensionality, alpha is problematic. The unidimensionality of a test should be independent of its length. However, alpha is dependent on the length of the test.

Apart from alpha, some indices use item responses to determine unidimensionality. The perfect unidimensional test is a function of responses that differ from the ideal response pattern. One of these is the reproducibility coefficient developed by Gutmann. Because Gutmann's coefficient is a function of item difficulties, the lower limit of this coefficient was determined as .50 (Hattie, 1985). There are many criticisms of unidimensionality indices estimated over the item response pattern. The most serious objection is that these methods can only reach their upper limits when the strong assumption of scalability (e.g., perfect scale) can be met (Lumsden, 1950, cited in Hattie, 1985). The other major criticism is that only one ability is tested, and possible combinations of abilities cannot be distinguished in the test.

So far, the definition of unidimensionality in the context of CTT, the concept of homogeneity and some indices for dimensionality have been mentioned. There are also indices based on principal components and factor analysis. Additionally, dimensionality is discussed from different perspectives in Item Response Theory (IRT).

Dimensionality in Item Response Theory

The dimensionality in IRT is an assumption of local independence of item responses after controlling the ability. In practice, it is generally accepted that the strict unidimensionality assumption is always violated to some degree. The actual meaning of the local independence is not only that the partial correlations of test scores are zero when factor scores are partially subtracted; at the same time, different test scores are entirely statistically independent. The basic assumption in most latent construct models is that the hypothetical variable measured by the test can be described in unidimensional latent space (Lord & Novick, 2008; Whitely & Dawis, 1974).

The definition of the nature and dimensionality of the entire latent space is based on the particular universe and particular property with which the psychometrist is concerned. Psychometrists want to define their own latent space by adding all the important psychological dimensions that will affect the given item set, and excluding other variables that constitute "measurement error". However, it seems logically impossible that these variables are free from "measurement error" (Lord & Novick, 2008).

The dimensionality of the full latent space is a more fundamental concept than the number of common factors (Lord & Novick, 2008). For example, success is a psychological variable that has both cognitive and affective dimensions. Variables such as motivation, attitude, interest that affect success can be measured. There are also many sources of variability, such as environment and rater, that cannot be measured. This unmeasurable variability leads to measurement error, because removing these sources of variability entirely from the measured characteristic of success is impossible. For this reason, as Lord and Novick (2008) mentioned, measurement errors like measured properties serve dimensionality.

Why Unidimensionality is Required?

When comparing two entities directly, the comparison must be based on measurements obtained from only one feature (Horton, 2013). In other words, unidimensional measurements are required for this comparison. According to Stout (1987), basic dimensionality is the existence of a dominant dimension in addition to the presence of one or more small (minor) dimensions. Also, the estimations based on the dominant dimension are strong enough not to be affected by the existence of small dimensions. Basic dimensionality is used as evidence for unidimensionality in IRT analyses when the unidimensionality assumption has been met (Horton, 2013).

Multidimensionality and Factors Affecting Dimensionality

Multidimensionality may arise from the planned test structure or may arise from unplanned factors, which are independent of the structure. Under certain conditions (e.g., when the abilities concerning to test were moderately correlated) test scores are resistant to violations of the unidimensionality assumption. Under other circumstances (e.g., when the component abilities measured by the test are weakly correlated or there are strong factors unrelated to the construct) the results of the violation will negatively affect test validity, reliability, bias, and comparability of scores (Tate, 2012).

When examining dimensionality, planned and unplanned sources may lead to multidimensionality. Planned sources are usually due to the nature of the measured construct. On the other hand, unplanned sources, for example the correct answer to an item, often includes one or more unintended (nuisance-nuance) factors or unintended abilities (abilities that affect the test score outside the measured scope) (Tate, 2012). For example, in a test that is intended to measure only mathematical ability, an examinee's reading ability and stability will impact the estimation of the ability. These two resources (readability and stability) are content-unrelated sources. The greater the importance of the nuisance factors on the trait that is intended to be measured by this type of test (e.g., the greater their weight in the predicted test composition), the more serious the threat to the content-related aspect of validity.

Detecting Number of Dimension

There are many methods for determining dimensionality. Evaluation of factors with an eigenvalue greater than 1 (Kaiser-Guttman method), examination of Scree-plot, Horn's Parallel Analysis and Velicer's MAP (Minimum AveragePartial) test, for example, are methods that have been used and mentioned frequently in recent years (Fabrigar et al., 1999; Ford et al., 1986). In the literature, it has been stated that the Kaiser-Guttman method and the Scree-plot method are insufficient in determining the number of factors (Fabrigar et al., 1999; Ford et al., 1986; Horn, 1965; Silverstein, 1977, 1987; Zwick and Velicer, 1986). Compared to other methods, Horn's Parallel Analysis and Velicer's MAP test produced accurate results in determining the number of factors (Silverstein, 1977, 1987; Zwick and Velicer, 1986; Dinno, 2009; Revelle, 2007).

Horn's Parallel Analysis method is based on the principle of creating a simulative data at the stage of determining the number of dimensions, calculating the expected eigenvalue of this data and comparing it with the eigenvalue of the real data (Zwick & Velicer, 1986). Another method to determine the number of factors is Velicer's MAP test. Velicer's MAP test is based on examining a set of partial correlation matrices and principal component analyses (Velicer, 1976; O'Connor, 2000).

Conclusions and Recommendations

In this study, dimensionality has been examined in the context of CTT and IRT. Many factors lead to the dimensionality of a test. It is not proper to see the dimensionality of the test separately from the test takers and the interaction of the test takers with the items. Although there is no clear line between multidimensionality and unidimensionality, it should be examined whether multidimensionality arises from the planned test structure or unintended factors independent of the structure. As a result of this examination, the validity and reliability of the test scores will be negatively affected when the content-unrelated factors have strong relationships with the measured structure.

Yayın Etiği Beyanı

Bu arařtırmada kullanılan veriler için Recep Tayyip Erdoğan Üniversitesi Sosyal ve Beşeri Bilimler Bilimsel Arařtırma ve Yayın Etiği Kurulu tarafından 08.03.2022 tarihinde yapılan toplantıda 2022/39 sayılı kararıyla verilen etik kurul izni bulunmaktadır. Bu arařtırmanın planlanmasından, uygulanmasına, verilerin toplanmasından verilerin analizine kadar olan tüm süreçte “Yükseköğretim Kurumları Bilimsel Arařtırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Arařtırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbirini gerçekleştirilmemiştir. Bu arařtırmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamıştır. Bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

Çatışma Beyanı

Arařtırmanın yazarı olarak herhangi bir çıkar/çatışma beyanım olmadığını ifade ederim.