

Kula Kartal, S., Mor Dirlik, E. (2016). Geçerlik kavramının tarihsel gelişimi ve güvenilirlikte en çok tercih edilen yöntem: Cronbach Alfa Katsayısı. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(4), 1865-1879.

Geliş Tarihi: 25/03/2016

Kabul Tarihi: 17/11/2016

## GEÇERLİK KAVRAMININ TARİHSEL GELİŞİMİ VE GÜVENİRLİKTE EN ÇOK TERCİH EDİLEN YÖNTEM: CRONBACH ALFA KATSAYISI

Seval KULA KARTAL\*  
Ezgi MOR DİRLİK\*\*

### ÖZ

Geçerlik ve güvenilirliğe ilişkin daha işe vuruk tanımlar oluşturmak amacıyla bu kavramlar üzerinde yapılan tartışmalar sürmektedir. Messick (1995) tarafından yapı geçerliğinin merkeze alındığı, diğer geçerlik türlerinin yapı geçerliğinin alt boyutları olarak kabul edildiği bir birleştirilmiş geçerlik görüşü oluşturulmuştur. Ancak, özellikle eğitim alanındaki testlerde kapsam geçerliğinin daha önemli olduğunu savunan bazı araştırmacılar birleştirilmiş geçerlik görüşüne karşı çıkmıştır. Bazı araştırmacılar da yapının anlamlandırılması sürecinde içsel (kapsam, güvenilirlik) ve dışsal (yapısal ya da sonuçsal geçerlik) geçerlik kanıtlarının birbirine üstünlüğü olmadığını ve hepsinin geçerleme sürecinde incelenmesi gerektiğini savunmuştur. Klasik test kuramı kapsamında paralel testler arasındaki korelasyon olarak tanımlanan ve tesadüfi hatalarla ilişkilendirilen güvenilirlik ise ölçme alan yazınında geçerlik kadar tartışmalı bir kavram olmamıştır. Cronbach alfa katsayısı tüm bilim dallarında en sık tercih edilen güvenilirlik belirleme yöntemi olmuştur. Her ne kadar kullanım alanı geniş ve araştırmacılar için kullanışlı bir katsayı olsa da bu katsayının hesaplanması için bazı varsayımların karşılanması gerekmektedir. Varsayımlarının karşılanmadığı durumlarda ise Cronbach alfa katsayısı ile elde edilen değer gerçek güvenirlüğün alt sınırını vermektedir. Böyle durumlarda kullanılması daha uygun olan farklı güvenirlilik katsayıları bulunmaktadır.

**Anahtar Kelimeler:** geçerlik, birleştirilmiş geçerlik, güvenilirlik, Cronbach alfa katsayısı

## HISTORICAL DEVELOPMENT OF THE CONCEPT OF VALIDITY AND THE MOST PREFERRED TECHNIQUE OF RELIABILITY: CRONBACH ALPHA COEFFICIENT

### ABSTRACT

To create more operational definitions for the terms of validity and reliability, the discussions on these terms have been continuing. Messick (1995) constituted the unified validity approach, in which the construct validity has been put into the center and other types of validity have been considered as the sub-dimension of the construct validity. However, some researchers, supporting the idea that the content validity is more important especially for the educational tests, disagree with the unified validity approach. Furthermore, some researchers support the idea that internal (content, reliability) and external (construct or consequential validity) evidences of validity are not superior to each other and all of them are required to be analyzed during the validation process. The term of reliability, defined as correlation between the parallel tests and associated with the random errors, has not been as questionable as the term of validity. Cronbach alpha coefficient has been mostly preferred in all scientific fields. Despite the fact that Cronbach alpha has extensive usage and it is a practical coefficient for researchers, some assumptions are needed to be satisfied so as to be able to calculate this coefficient. In the situations that assumptions do not met, the value is obtained by using Cronbach alpha formula corresponds to the lower limit of true reliability. In these cases, there are various formulas, which one can take benefit from and more convenient to use to estimate reliability.

**Key Words:** validity, unified validity, reliability, Cronbach alpha coefficient.

---

\* Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitimde Ölçme ve Değerlendirme,  
seval.kula@hotmail.com

\*\* Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitimde Ölçme ve Değerlendirme,  
ezgimor@gmail.com

### **1.Geçerlik Tanımı ve Geçerleme Sürecine İlişkin Farklı Görüşler**

En klasik tanımıyla geçerlik, bir testin ölçülmesi istenilen özelliği başka özelliklerle karıştırmadan ölçebilme derecesi olarak ele alınmaktadır (Ebel ve Frisbie, 1991; Murphy ve Davidshofer, 1998; Baykul, 2010). Bu tanımına dayalı olarak yapılan ilk araştırmalarda geçerliğe ilişkin en iyi kanıtın korelasyonel çalışmalarla sağlanacağı düşünülmüş ve ölçüt bağıntılı geçerlik çalışmaları ön plana alınmıştır (Loveinger, 1957). Ancak uygun ölçütün bulunması ve ölçütün geçerliğine ilişkin kanıtların sağlanması gibi güçlükler nedeniyle hem kuramsal hem de uygulama boyutunda geçerlik kavramına ilişkin çalışmalar sürmüştür. Klasik geçerlik tanımı zamanla uygulamada kullanılamayacak kadar geniş bulunmuştur. Bu doğrultuda, tek bir geçerlik katsayısı ile geçerliğe ilişkin yorum yapmak yerine; farklı geçerlik kanıtı sağlama yolları ortaya çıkmıştır. Rulon (1946; akt. Sireci, 1998) ölçme araçlarının geçerliği incelenirken ölçme amacının ve araçta ele alınan kapsamın da incelenmesi gerektiğini belirtmiştir. Cronbach ve Meehl (1955) ise yapı geçerliği kavramına vurgu yaparak kapsam geçerliği, ölçüt bağıntılı geçerlik ve yapı geçerliği olmak üzere üç farklı geçerlik türünden bahsetmiştir.

Geçerlik üzerine yapılan çalışmalar ve tartışmaların sürmesi geçerlik sorununu ölçülmek istenilenin ölçülüp ölçülemediğinden, test puanları arasındaki ilişkilerin kuramsal boyuttaki ilişkilerle tutarlı olup olmadığı sorusuna dönüştürmüştür (Borsboom, Mellenbergh ve van Heerden, 2004). Dolayısıyla, yapı geçerliği kavramı önem kazanmıştır. Cronbach ve Meehl (1955) yapı geçerliğinin de geçerlik kanıtı sağlamak amacıyla kullanılabileceğini belirtmiştir. Yazarlara göre yapı geçerliği, sadece incelenmekte olan yapının ölçümünü değil ayrıca bu nitelikleri birbirleriyle ilişkilendiren ağın doğasını içermektedir. İlgilenilen yapı ve ilişkili olan diğer yapılar arasındaki ilişkileri belirleyen kuram olarak tanımlanabilecek nomolojik ağ, yapı geçerliği çalışmalarında oldukça önemli görülmektedir. (Lissitz ve Samuelson, 2007).

Geçerlik kavramı üzerinde yapılan tartışmalar sonucunda en son olarak ise geçerlik; test puanları kullanımının ve bu puanlara dayalı olarak verilen kararların test kullanımının etik, sosyal ve ampirik açıdan sonuçları ile doğrulanabilir olup olmadığı sorusuna dönüşmüştür (Borsboom, Mellenbergh ve van Heerden, 2004). Bu görüşü ortaya atan Messick (1989; 1995; Thordike ve Thorndike-Christ, 2014) geçerlik tanımında test puanlarına ilişkin yapılan yorumlara, onların doğurduğu sonuçlara, deneysel ve kuramsal boyutun bu sonuçları destekleyip desteklemediğine odaklanmaktadır. Messick, Cronbach ve Meehl'in (1955) çalışmasındaki yapı geçerliği vurgusunu daha da katılaştırarak, kapsam ve ölçüt bağıntılı geçerliğin yapı geçerliğinin bir alt boyutu olarak ele alındığı bir taksonomi oluşturmuştur. Birleştirilmiş geçerlik kavramına vurgu yaparak, kapsam ve ölçüt bağıntılı geçerlik yöntemleri ile sağlanan kanıtların da temelde yapı geçerliğini sağlamaya hizmet ettiğini belirtmiştir.

Messick'in (1989; 1995) tanımlamasına göre, geçerliğin merkezinde yapı geçerliği bulunmakta ve diğer geçerlik türleri, yapı geçerliğinin alt boyutları olarak tanımlanmaktadır. Tanımlanan tüm boyutlara ilişkin kanıtların toplanmasıyla yapı geçerliğinin bütünlüklü bir geçerlik olarak yorumlanması söz konusu olmaktadır. Bu kanıtlar ölçme aracının geliştirilmesine karar verildiği andan itibaren toplamaya başlanmakta ve aracın uygulanması sonrasında da devam etmektedir. Dolayısıyla, geçerlik kanıtları geçerleme süreci olarak adlandırılan bir süreç boyunca toplanmaktadır. Bu süreçte, kanıtların tümü bütünlüklü olarak ortak bir yapı geçerliği çatısı altında toplanıp, ölçme aracının yapı geçerliği olarak yorumlanmaktadır.

Messick (1995) kapsam, ölçüt bağıntılı, yapı geçerliği gibi geçerlik türlerinin yanında, sonuçsal geçerliği de bir geçerlik türü olarak kabul etmektedir. Messick'e göre bir testin ölçülmek istenilen özelliği ne derecede ölçebildiğini değerlendirmek için, testten elde edilen puanların, bu puanlara dayalı olarak yapılan yorumların ve bu yorumların yarattığı sonuçların, testin hedeflenen amaçlarıyla ve aynı zamanda diğer sosyal değerlerle olan tutarlılığının sorgulanması gerekmektedir. Ancak, Messick tarafından ortaya atılan sonuçsal geçerlik kavramına ilişkin araştırmacılar arasında bir uzlaşma sağlanabilmiş değildir. Mehrens (1997) yapılar arasındaki ilişkilerin nomolojik ağ içerisinde sonuçsal geçerliğe gerek kalmadan ortaya koyulabileceğini söylemektedir. Belirli bir test kullanımının sonuçlarının yapının anlamına ya da yapının ölçülmesi sürecinin yeterliğine ilişkin bilgi vermediğini düşünmektedir. Geçerlik kavramının daha fazla kavramı kapsamaması için genişletilmesi yerine; kavramın daraltılması ve daha açık hale getirilmesi gerektiğini belirtmektedir.

Sonuçsal geçerlikte olduğu gibi birleştirilmiş geçerlik görüşüne ilişkin tartışmalar da sürmektedir. Anastasi'ye (1986) göre geçerliğin ölçüt bağıntılı, kapsam ve yapı geçerliği olmak üzere üç farklı tür olarak ele alınması başlangıçta geçerleme süreçlerini daha anlaşılır kılsa da; sonrasında testin ne ölçtüğüne ilişkin olarak veri toplama süreçlerinin basite indirgenmesine neden olmuştur. Test geliştirme ya da kullanma sürecinde toplanan tüm veriler geçerlikle ilgilidir ve testin ölçtüğü özelliğe ilişkin önemli bilgi sağlamaktadır. Eğitim alanında en yaygın olarak kullanılan başarı testlerinin amacı, hazırbulunuşluğu, öğrenme eksikliklerini ve öğrenme düzeylerini ölçmektir. Dolayısıyla, bu testler için ölçülebilir davranışsal göstergelere dönüştürülmüş eğitim hedefleri ölçülmek istenen yapıları oluşturmaktadır. Bu testlerde temelde uzman yargılarına dayalı olarak incelemeler yapılmaktadır. Daha belirgin düzeylerde öğretim programı, ders kitapları ve uzman yargılarından elde edilen verilere dayalı olarak analiz gerçekleştirilmektedir. Ancak, Anastasi (1986) yapılan bu analizleri kapsam geçerliği değil; yapı geçerliği olarak tanımlamaktadır. Benzer şekilde, Haynes, Richard ve Kubani (1995) de kapsam geçerliğini, bir ölçme aracının bileşenlerinin belirli bir ölçme amacına uygun olarak ele alınan yapıyı temsil etme derecesi olarak tanımlamakta ve kapsam geçerliğinin temel hedefinde yapının temsil edilmesinin yer aldığını düşünmektedir. Yazarlar, kapsam geçerliğinin, ölçme aracının yapı geçerliğine ilişkin kanıt sunduğunu belirtmekte ve yapı geçerliğini odak olarak alan birleştirilmiş geçerlik görüşünü savunmaktadır.

Birleştirilmiş geçerlik görüşünün farklı yönlerini ele alan ve eleştiren pek çok çalışma da mevcuttur. Borsboom, Mellenbergh ve van Heerden (2004), testle ilişkili her konunun geçerlikle bağlantılı olarak düşünülmesinin ve tüm bu konuların tek bir başlık altında toplanmasının doğru olmadığını savunmaktadır. Birleştirilmiş geçerlik görüşünün, geçerliğe ilişkin anlayışı, hem kuram odaklı hem de uygulama odaklı çalışan araştırmacılar için güçleştirdiğini belirtmektedir. Yazarlar, test geliştiren kişilerin birleştirilmiş geçerlik görüşünden uygulanabilir kavramsal bir şema oluşturmalarının mümkün olmadığını düşünmektedir. Psikolojik ölçme ile ilgili olan her şeyin ilgili, önemli ve psikolojik ölçmelerde kullanılması gerektiği görüşünü ortaya atan bir geçerlik yaklaşımının araştırmacılara yol göstermediğini savunmaktadır.

Bir testin geliştirilmesi sürecinde ve sonrasında birleştirilmiş geçerlik yerine; kapsam geçerliğinin temel alınmasını belirten bir yazar da ölçme alanına büyük katkılar getiren Ebel (1956) olmuştur. Ebel'e göre, kapsam geçerliği hiçbir test geliştiricinin göz ardı

edebileceği bir geçerlik türü değildir. Test geliştirme aşamasında, ölçülecek özellik üzerinde yeterince düşünülmeyişi takdirde, ileriki aşamalarda ortaya çıkacak problemlerin kaçınılmaz olduğunu düşünmektedir. Ebel, testin geçerliğinin incelenmesi için öncelikli olarak testin kapsam geçerliğine ilişkin kanıtların toplanmış olması gerektiğini, ancak bundan sonra, test dışında dolaylı bir ölçüte başvurulabileceğini ifade etmektedir.

Lissitz ve Samuelsen (2007) de birleştirilmiş geçerlik kavramına karşı çıkan yazarlar arasında yer almaktadır. Yazarlara göre, bir testin geçerliği incelenirken diğer testler ile ilişkisi yerine; öncelikle testin kendisinin incelenmesi gerekmektedir. Yazarlar, özellikle eğitim alanındaki testler için, kapsam geçerliğinin incelenmesinin daha önemli olduğunu vurgulamaktadır. Özellikle eğitim alanında, kapsam ve yapı geçerliğinin birlikte ele alınmasının, her iki geçerlik türünün de anlaşılmasını güçleştirdiği belirtilmektedir. Eğitim alanındaki testler için öncelikli hedef, bileşenlerin bir nomolojik ağ içerisinde tanımlanmasından çok; testin tanımlanan alanı yeterli bir biçimde yansıtmayı yansıtmadığının incelenmesidir. Yazarlar, yapı geçerliğini dışsal kanıt olarak ele almaktadır.

Sireci (1998) ve Embretson (2007), geçerleme sürecinde sadece yapı geçerliğinin ya da kapsam geçerliğinin önemli görüldüğü görüşleri eleştirmektedir. Embretson, sadece kapsama olan uygunluğu açısından değerlendirilen bir testte kapsam geçerliği açısından uygun maddelerin yer alacağını; ancak ölçülmek istenen özellikle ilişkisiz değişkenlik kaynakları içeren maddelerin tespit edilemeyebileceğini belirtmektedir. Bu nedenle, testte ölçülmek istenen özelliğin tanımlanması sürecinde kapsam geçerliğinin ön planda tutularak, test dışındaki geçerlik kaynaklarının da kullanılmasını desteklemektedir. Eğitim alanındaki testler de dâhil olmak üzere geçerleme sürecinde, test dışı geçerlik kaynaklarının (diğer yapılarla ilişkiler, test kullanımının sonuçları vb.) tamamen göz ardı edilmemesi gerektiğini düşünmektedir. Benzer şekilde, Sireci (1998), bir testin kapsam geçerliği incelenirken, test puanlarına dayalı olarak yapılacak yorumlar gibi test dışı geçerlik kaynaklarının da göz önünde bulundurulması gerektiğini düşünmektedir.

Türkiye’de geçerlik türlerine ilişkin yapılan çalışmalar incelendiğinde, geçerliğin kuramsal olarak tartışıldığı çalışma sayısının sınırlı olduğu görülmektedir. Berberoğlu (2012), okul içi ölçme ve değerlendirme çalışmalarında, öncelikli olarak toplanması gereken geçerlik kanıtının kapsam geçerliğine yönelik olması gerektiğini belirtmektedir. Kelecioğlu ve Göçer Sakin (2014) tarafından geçerliğin ölçme tarihindeki gelişimi ve kazandığı anlamlar ele alınmaktadır. Yazarlar, geçerlik tartışmalarının gelişimine, geçerlik kanıtlama yollarının yıllar içinde farklı görüşlere göre nasıl şekillendiğine odaklanmaktadır. Kapsam geçerliğine yönelik tartışmalara da değinilerek, eğitim alanındaki ölçmelerde kapsam geçerliğine yönelik kanıtların ön planda olması gerektiği görüşü desteklenmektedir. Ancak, odak noktası kapsam geçerliği de olsa; eğitimdeki ölçmelerde diğer geçerlik kanıtlarının da mutlaka sunulması gerektiği belirtilmektedir.

Geçerliğe ilişkin ortaya atılan farklı görüşler incelendiğinde, birleştirilmiş geçerlik görüşünde yapı geçerliği daha fazla önemsenmekte, diğer geçerlik kanıtlarının da yapı geçerliğine ilişkin bilgi sağladığı kabul edilmektedir. Geçerlik kanıtları (kapsam, ölçüt bağıntılı) birbirinden bağımsız ve testin amacına uygun olanın seçilebileceği geçerlik türleri olarak değil; bir test geliştirme sürecinde temelde yapı geçerliğini ortaya koymak için incelenmesi gereken boyutlar olarak ele alınmaktadır. Birleştirilmiş geçerlik görüşüne karşı çıkan Lissitz ve Samuelsen (2007) ise test ile ölçülen özelliği tanımlamak

için kapsam geçerliğinin incelenmesi gerektiğini belirtmektedir. Yazarlara göre, bir testin anlamının dışsal kaynaklar (testten elde edilen puanlardan yapılacak çıkarımlar ve nomolojik ağ) ile değil; testin kapsamı temsil gücü ve testin güvenilirliğine dayalı olarak incelenmesi gerekmektedir. Sireci (1998) ile Embretson (2007), Lissitz ve Samuelsen'nin özellikle eğitim alanındaki testlerde kapsam geçerliğinin ön planda tutulması görüşünü desteklemekte; ancak testin anlamını ortaya koymada kapsam ve güvenilirlik incelemelerinin yeterli olmayacağını ve test dışı kaynakların da ölçülen yapının anlamlandırılması sürecinde göz önünde bulundurulması gerektiğini düşünmektedir.

## **2.Güvenirlik ve En Çok Tercih Edilen Güvenirlik Belirleme Yöntemi: Cronbach Alfa Katsayısı**

Geçerliğe ek olarak, ölçme uygulamalarında sorgulanan bir diğer nitelik de güvenirliliktir. Güvenirlik, test puanlarının hatadan arınık olma derecesi olarak tanımlanmaktadır (Baykul, 2010; Crocker ve Algina, 2006). Geçerlik kavramı yıllar boyunca yapılan tartışmalarla değişip gelişse de; güvenirlik kavramında bu denli köklü değişiklikler gözlenmemiş, sadece farklı güvenirlik belirleme katsayılarının kullanılması gündeme gelmiştir. Bu durumun nedeni ise Klasik Test Kuramı (KTK) kapsamında, güvenirlik için geçerliğe göre daha net bir tanım yapılmış olmasıdır. Bilindiği gibi KTK bir takım varsayımlara ve temel bir denkleme sahiptir. KTK'nın temel denklemi;  $X=T+E$  ( $X$ =gözlenen puan,  $T$ =gerçek puan ve  $E$ =hata) olarak oluşturulmuş ve bu denkleme göre soyut özelliklere ilişkin yapılan her ölçmeye bir miktar hata karıştığı kabul edilmiştir. Bu temel denklemden yola çıkılarak güvenirlik kavramı paralel ölçmeler arası korelasyon olarak tanımlanmış, tesadüfi hata kavramıyla ilişkilendirilmiş ve sınırlandırılmıştır. Dolayısıyla güvenirlik kavramına geçerliğe göre daha sınırlı bir çerçeve çizilmiştir.

Paralel ölçmeler arasındaki korelasyon olarak tanımlanan güvenirliliğe ilişkin kanıt toplama sürecinde, en sık tercih edilen güvenirlik belirleme yöntemleri arasında test maddelerinin paralellik koşulunu sağlaması durumunda hesaplanan ve maddeler arasındaki tutarlılık derecesini belirten katsayılar yer almıştır. İç tutarlılık anlamında bilgi veren ve güvenirlik kanıtı olarak kullanılan katsayılar içerisinde hem eğitim hem de psikoloji alanında en sık kullanılanı Cronbach alfa katsayısıdır. Cronbach alfa katsayısı, ölçme ve değerlendirme ile ilgilenen her bireyin en az bir kez karşılaşmış olduğu bir güvenirlik katsayısıdır. Ölçme alanıyla yakından ilgili olmayanlar dahi, Cronbach alfa katsayısı hakkında fikir yürütebilmekte ve bir ölçüğe ilişkin güvenirlik katsayısı söz konusu olduğunda bir çok kişinin aklına doğrudan Cronbach alfa gelmektedir. Dolayısıyla Cronbach alfa katsayısı hemen hemen tüm ölçme araçlarının güvenirliliğini belirlemede kullanılan bir katsayı olarak kabul edilmektedir. Benzer şekilde, Sijtsma (2009), test puanlarına ilişkin Cronbach alfa katsayısından daha sık kullanılan ve bu istatistik kadar da yanlış anlaşılma başka bir istatistiğin olmadığını ifade etmektedir.

Cronbach alfa katsayısının, psikolojik ölçme araçlarından yararlanılan tüm alanlarda sıklıkla kullanılma nedenleri, Cronbach alfa katsayısına ilişkin önemli özelliklerin incelenmesi ile ortaya çıkarılabilir. Bu katsayıyı daha iyi anlamak için öncelikle adını aldığı kaynaktan başlanabilir. Sijtsma (2009), büyük buluşların onu keşfedenin adını taşımak yerine, söz konusu buluşa ait olumlu özellikleri en başarılı şekilde ortaya çıkaran kişinin adını taşıdığını, bu durumun Cronbach alfa için de söz konusu olduğunu belirtmiştir. Herhangi bir yanlış anlaşılmayı engellemek için Cronbach da bu katsayının kendi buluşu olmadığını, Guttman tarafından üretildiğini açıklamıştır. Ancak, bu katsayı

üzerinde çalışan Kuder, Richardson, Guttman ve Hoyt'tan ziyade; daha çok Cronbach'ın adıyla anılmıştır. Cronbach (1951) bu katsayıyı, Kuder Richardson katsayısının genel hali ve testin olası tüm iki yarılarından kestirilen iki yarı korelasyonlarının ortalaması olarak tanımlamaktadır. Cronbach alfa katsayısı ile testin tümünden tesadüfi olarak seçilebilecek tüm yarılarla ilişkin bilgi elde edilebilmektedir. Ayrıca güvenilirliği incelenen testin yeterli sayıda madde içermesi ve tek boyutluluğu sağlaması durumunda, bu katsayı ile hesaplanan güvenilirlik değeri, gerçek güvenilirlik değerine eşit olmaktadır (Cronbach, 1951).

Alfa katsayısı adını Cronbach'tan alsa da; aslında bu katsayı Guttman'ın çalışmaları sırasında ortaya çıkardığı katsayılardan biridir. Guttman (1945), paralel ölçmelerin dışındaki tüm ölçme kümeleri için hesaplanan güvenilirlik katsayılarının, gerçek güvenirlüğün altında değer üreteceğini belirterek "güvenirlüğün alt sınırı" olarak nitelediği altı ( $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5, \Gamma_6$ ) katsayı önermiştir. Guttman tarafından önerilen katsayılar içerisinde  $\Gamma_3$  katsayısı aynı zamanda Cronbach alfa ( $\alpha$ ) güvenilirlik katsayısı olarak bilinmektedir (Sijtsma, 2009). Guttman bu katsayıyı güvenirlüğün alt sınırı olarak tanımlarken, Cronbach (1951) paralel maddelerin yanı sıra eşdeğer maddeler için de alfa güvenilirlik katsayısının gerçek güvenirlüğü vereceğini ifade etmiştir. Feldt ve Qualls (1996) ise Cronbach alfa katsayısının testin güvenirlüğünün yeterli bir kestirimi olması için test bileşenlerine ait puanların, gerçek puana eşdeğer olması gerektiğini belirtmiştir. Bu durum da testin tek boyutlu olmasını gerektirmektedir. Bu nedenle, KR-20 ve Cronbach alfa gibi iç tutarlılık anlamında bilgi veren katsayıların, gerçek güvenilirlik değerini verebilmesi için testlerin tek boyutluluk varsayımını karşılaması gerekmektedir.

Cronbach (1951), alfa katsayısının, özellikle birden çok bileşen içeren ölçmelerin güvenirlüğünü hesaplamayı sağladığını belirtmektedir. Bu ifadedeki bileşen kelimesi, bir madde, bir testin yarısı ya da bir alt test anlamına gelebilir. Çoklu bileşenler, tek bir gizil değişkeni ölçmeye yönelik olabileceği gibi; birden çok gizil değişkeni/faktörü de ölçebilir. Bu özellikleri, Cronbach Alfa katsayısının kullanımını arttırmaktadır. Osburn (2000), Cronbach Alfa katsayısının esnekliğinin, psikoloji başta olmak üzere birçok alanda sıklıkla kullanılmasını sağladığını belirtirken; Cortina (1993), bu katsayının kullanım sıklığına ilişkin daha somut bilgiler sunmaktadır. Söz konusu çalışmada, SSCI'da indekslenen dergiler 1966 ve 1990 yılları arasında incelenmiş Cronbach'ın 1951'deki makalesinde tanıttığı bu katsayının, söz konusu dergilerdeki makalelerce her yıl ortalama 60 kez atıf aldığı ve 278 farklı dergide kullanıldığı belirlenmiştir. Sijtsma (2009) da bu katsayının çok sık kullanımına kanıt olarak, Cronbach'ın 1951 yılında bu katsayıyı tanıtmak üzere yazdığı makalesinin internette indirilme sıklığının gösterilebileceğini, Psychometrika dergisinde yayınlandığı tarihten itibaren, bu makalenin en fazla indirilen makale olduğunu belirtmiştir. Ayrıca, Web Of Science'ın raporuna göre, Cronbach'ın makalesi her yıl en az 6500 atıf almaktadır. Söz konusu katsayının sık kullanılmasının nedenlerini inceleyen başka bir çalışmada, bu katsayının en sık tercih edilen güvenilirlik belirleme yöntemi olduğu belirtilmiş ve bu durumun olası dört nedeni şu şekilde açıklanmıştır (Yang ve Green, 2009):

1. Cronbach alfa katsayısını yorumlamak oldukça kolaydır. Katsayı 1'e yaklaştıkça ölçekte yer alan maddelerin iç tutarlılığının yüksek olduğu sonucuna ulaşılır.
2. Cronbach alfa katsayısı objektiftir, diğer güvenilirlik belirleme yöntemlerine göre öznel yargıları daha az içermektedir. Örneğin, iki yarı yönteminde, testlerin yarılarına nasıl ayrılacağı araştırmacının kararına bağlıdır ya da test tekrar test

yönteminde iki uygulama arasında geçecek zaman araştırmacının müdahalesi ile belirlenmektedir. Dolayısıyla bu yöntemlere, araştırmacının etkisi dâhil olmakta ve objektiflik Cronbach alfa katsayısında olduğu kadar yüksek olamamaktadır.

3. Cronbach alfa katsayısı ile ölçeği yeniden düzenlemek diğer güvenilirlik belirleme yöntemlerine göre daha pratiktir. Herhangi bir maddenin ölçek dışına çıkarıldığı durum için Cronbach alfa katsayısı hesaplanabilir, bu sayede maddelerin güvenilirliğe olan katkıları ayrı ayrı incelenebilir.
4. Cronbach alfa katsayısının sık kullanılma nedenlerinden biri de, araştırmacıların daha önce bu katsayısı kullanmış olan binlerce kişinin yanıltılmış olma olasılığının düşük olduğuna ilişkin inancıdır. Söz konusu katsayının kullanım sıklığı sayesinde, elde edilen değeri görece olarak yorumlamak için belirli ölçütler geliştirilmiştir.

Socan (2000) bu katsayının sık kullanılma gerekçelerine, hesaplama kolaylığını (sadece madde ve toplam varyansın hesaplanmasını gerektirir) eklemiştir. Ayrıca Cronbach alfa katsayısına ilişkin örneklem dağılımının bilindiğini, bu sayede evren için güven aralıklarının hesaplanabileceğini ve Cronbach alfa katsayısının şans etkisine daha kapalı olduğunu ifade etmiştir. Özetle, Cronbach alfa katsayısı ölçme alan yazınına Cronbach tarafından tanıtıldığı günden itibaren yaklaşık 64 yıldır sosyal bilimlerde büyük ilgi görmüş ve en sık kullanılan güvenilirlik belirleme yöntemi haline gelmiştir. Söz konusu katsayının diğer güvenilirlik belirleme yöntemlerine göre daha sık tercih ediliyor olmasının nedenlerini belirleyebilmek için diğer güvenilirlik belirleme yöntemlerini de incelemek gerekmektedir.

### 3. Diğer Güvenirlik Belirleme Yöntemleri

Ölçme sonuçlarının, hatalardan arınlık derecesi olarak tanımlanan güvenilirliği belirlemenin, hataların türüne bağlı olarak hesaplanan birden çok yöntemi vardır. Bunlardan ilki paralel formlar yöntemidir. Osburn'a (2000) göre, farklı bir güvenilirlik kestirim yöntemi olan paralel formlar da aslında sık kullanılan güvenilirlik belirleme yöntemlerinden biridir. Fakat bu güvenilirlik kanıtının elde edilmesi için uygulanacak testin bir ya da daha fazla paralel formunun geliştirilmiş olması gerekmektedir. Test tekrar test güvenilirlik belirleme yöntemi ise kâğıt kalem testlerinde deneyim ve hatırlama etkisi gibi bazı sorunlara yol açmaktadır. Bu nedenle, test tekrar test yönteminin güvenilirlik kestiriminde kullanılması için çeşitli önlemlerin alınması gerekmektedir. Cronbach alfa katsayısı ve bu katsayının türetildiği iç tutarlılık anlamında bilgi veren güvenilirlik katsayıları ise oldukça esnek oldukları için birçok duruma uygun şekilde kullanılabilir.

Cronbach (1951) da kendi makalesinde bu katsayısı tanıtırken diğer güvenilirlik belirleme yöntemleriyle kıyaslamıştır. İlk olarak kıyasladığı durum, bireylerin puanlarına karışan hata miktarını kestirmek için yapılan paralel test uygulamasıdır. Bu uygulamanın dezavantajı olarak da psikologların ya da eğitimcilerin bir gruba söz konusu testi ikinci kez uygulama fırsatını çoğunlukla bulamamalarını göstermiştir. Ayrıca klinik amaçlı ya da yöneltme amacı taşıyan testlerin kalabalık gruplara uygulanması durumunda, paralel test uygulaması ile hem maliyetin hem de uygulama için gereken zamanın çok artacağını belirtmiştir. Cronbach, daha sonra paralel test uygulamasının dezavantajlarını ortadan kaldırmak üzere geliştirilen bir diğer yöntem olan iki yarı yöntemi ile Cronbach alfa

katsayısını karşılaştırmıştır. İki yarı yönteminin uzun yıllardır kullanılan bir yöntem olduğunu ve bu yöntemde kullanılabilecek Spearman Brown formülüne ek olarak farklı formüllerin de türetildiğini belirtmiştir. Uzun zamandır kullanılan bu tekniğe yöneltilen eleştirileri; “testin iki yarısı, testin bütünü kadar bilgi verici değildir” şeklinde özetlemiştir. Bunun nedeni olarak da, testin iki yarısından elde edilen katsayıların, aslında iki ayrı nitelikten elde edilmiş olmasını göstermektedir. Dolayısıyla, bu katsayıları aynı niteliğe sahip yarılarından elde edilmiş olduğunu kabul ederek bütünsel bir katsayıya ulaşmanın çok uygun olmadığını belirtmiştir. İki yarı yönteminin, önce Brownell daha sonra ise Kuder-Richardson tarafından, testin bütününe ilişkin tek bir katsayı vermediğinden dolayı eleştirildiğini belirten Cronbach, bu eleştirilerin paralel testler için de geçerli olduğunu ifade etmiştir. Dört paralel formu olan bir test için altı güvenilirlik katsayısının elde edileceğini belirterek bunları bileşik bir katsayıymış gibi yorumlamamanın oldukça güç olduğundan bahsetmiştir. Cronbach (1951) çalışmasında tanıttığı Cronbach alfa güvenilirlik katsayısını diğer güvenilirlik kestirim yöntemleriyle kıyaslayarak, bu katsayının onlardan birçok açıdan daha üstün olduğunu göstermiştir. Hem Osburn (2000) hem de Cronbach (1951) tarafından yapılan kıyaslamalar dikkate alındığında, Cronbach alfa katsayısının diğer güvenilirlik belirleme yöntemlerine göre daha az zamanda, daha az maliyetle ve daha objektif bir şekilde güvenilirlik kestirimi sağladığı sonucuna ulaşılabilir.

#### **4. Cronbach Alfa Katsayısının Kullanımının Uygun Olmadığı Durumlar**

Cronbach alfa katsayısının kullanılmasının uygun olduğu durumların yanında, bu katsayının kullanımının uygun olmadığı durumlar da mevcuttur. Bunlardan biri de testin bazı kısımlarının farklı konu ya da süreçleri ölçmeye yönelik olduğu durumlardır. Cronbach (2004), bu durumu şöyle örneklemektedir: Matematiksel akıl yürütmeyi ölçen bir testin yaklaşık %20’sinin geometrik şekillere dayalı maddelerden oluştuğu bir durum için, Cronbach alfa katsayısının tam olarak uygun olduğu söylenemez. Geometriye dayalı maddelerin seçkisiz alınarak oluşturulduğu farklı formlarda maddeler arası korelasyonlar daha yüksek çıkabilir; dolayısıyla, testlerin paralellüğünden söz etmek mümkün değildir. Bunun gibi ölçme durumlarında Cronbach alfa, gerçek güvenilirliğin altında değerler üretecektir. Testteki maddelerin içeriğe dağılımı dengeli ve planlı bir şekilde yapılmışsa, böyle bir durumda Cronbach alfa katsayısı kullanılabilir. Ancak maddelerin dağılımı deneysel ilişkiler yerine mantıksal yaklaşıma dayalı ise Cronbach alfa katsayısının kullanımı uygun olmayacaktır. Bunun gibi ölçme durumlarında Cronbach alfa, gerçek güvenilirliğin altında değerler üretecektir. Cronbach 1951’de yayımlanan Cronbach alfa katsayısını tanıtan makalesini yeniden düzenlediği 2001 yılındaki makalesinde, artık birçok verinin güvenilirliğini incelemek için en uygun yöntemin Cronbach alfa olmadığını düşündüğünü belirtmektedir.

Cronbach alfa katsayısının kullanımının uygun olmadığı durumları tartışan bir başka araştırmacı da Osburn (2000)’dur. Bu kaynakta, Cronbach alfa katsayısının güvenilirliği gerçek değerinden düşük kestirdiği durumlara ek olarak yüksek kestirdiği durumların olduğu da belirtilmiştir. Cronbach alfa katsayısının, diğer güvenilirlik belirleme yöntemlerine göre farklı test koşullarından ve bireyin performansındaki değişikliklerden daha az düzeyde etkilendiği, dolayısıyla, diğer güvenilirlik belirleme yöntemlerine göre daha yüksek değerler verebileceği ifade edilmiştir. Ayrıca maddelere verilen yanıtların hataları arasında pozitif korelasyon gözleniyorsa, Cronbach alfa katsayısı için olduğundan daha büyük bir değer elde edildiği ifade edilmiştir. Fakat, hız testleri için

alfa hesaplanacak olursa gerçek değerinden daha küçük değerler elde edileceği ifade edilmiştir. Cronbach alfa katsayısının belirtilen durumlarda olduğundan daha yüksek bir değer almasından ziyade; asıl tehlike arz eden durum, verinin çok boyutlu olması durumunda ya da testin az sayıda ve/veya heterojen maddelerden oluştuğu durumlarda, bu katsayının güvenilirliği gerçek değerinden daha düşük hesaplıyor olmasıdır. Bu durumun daha tehlikeli olarak belirtilmesinin nedeni, güvenilirliğin olduğundan düşük kestirilmesinin hem testin geçerliğinin sorgulanmasına neden olması hem de ölçme sonuçlarına duyulan güveni azaltmasıdır. Cronbach (1951), alfa katsayısının bazı durumlarda gerçek değerinden daha düşük güvenilirlik kestirimlerinde bulunduğunu belirtmiş ve farklı madde kümelerinden oluşan testler için hesaplanacak eşdeğer yarılar katsayılarının alfa katsayısından daha yüksek, dolayısıyla, daha uygun olacağını belirtmiştir.

Shevlin, Miles, Davies ve Walker (2000) tarafından yapılan çalışmada Cronbach alfa katsayısının güvenilirliğin kullanışlı bir kestiricisi olup olmadığı farklı durumlar için yapılan Monte-Carlo simülasyonu ile incelenmiştir. Simülasyon çalışmasında maddelerin faktör yükleri, sistematik hataları ve örneklem büyüklükleri değişimlenmiştir. Yapılan incelemeler sonucunda, Cronbach alfa katsayısının ölçtüğü oluşturulan maddelerin güvenilirliği dışındaki faktörlerden de etkilendiği belirlenmiştir. Özellikle sistematik hata miktarı ve tek boyutluluktan sapmanın, Cronbach alfa katsayısının değerini yükselttiği belirlenmiştir. Ayrıca madde güvenilirliklerinin yüksek olmasının da Cronbach alfa katsayısını arttıran bir başka durum olduğu ve bu bulgunun da Cortina (1993) ile örtüştüğü belirtilmiştir. Sonuç olarak Cronbach alfa hesaplanmadan önce testin boyutluluğunun ve sistematik hata içerip içermediğinin incelenmesinin gerekli olduğu ifade edilmiştir.

Cronbach alfa katsayısının ölçülen yapının tek boyutlu olmaması, maddelerin hata varyanslarının korelasyon göstermesi, bileşenlerin paralellik koşulunu sağlamaması ya da madde sayısının az olması durumunda gerçek güvenilirlikten daha düşük değer verdiği alan yazında çeşitli çalışmalarla doğrulanmıştır (Cortina, 1993; Osburn, 2000). Raykov (1997), Cronbach alfanın varsayımlarının sağlamadığı durumlarda ortaya çıkan yanlış kestirimlerin miktarını belirlemek ve bunları düzeltmek için bir formül önermiştir. Fakat bu formülün kullanılması da yapısal eşitlik modeli kullanmayı gerektirmektedir. Ayrıca gerçek ölçme uygulamalarında çoğu zaman paralel, eşdeğer ya da eş biçimli maddelerden oluşan ölçmeler yapmak çok zor olduğu için, Cronbach alfa katsayısının neredeyse her zaman güvenilirliği olduğundan düşük kestirdiği sonucuna ulaşılabileceği belirtilmektedir (Sacon, 2000).

Alanyazına dayalı olarak bir özetleme yapmak gerekirse; Cronbach alfa katsayısının kullanımının uygun olmadığı durumlar şu şekilde sıralanabilir (Cortina, 1993; Feldt ve Qualls, 1996; Raykov, 2001):

1. **Testi oluşturan maddelerin paralel, eşdeğer ya da eş biçimli olmadığı-konjenerik olduğu durumlar:** Cronbach alfa gerçek güvenilirliğin altında değerler vermektedir.
2. **Maddelere ilişkin hata terimlerinin negatif korelasyon gösterdiği durumlar:** Cronbach alfa gerçek güvenilirliğin altında değerler vermektedir.
3. **Maddelere ilişkin hata terimlerinin pozitif korelasyon gösterdiği durumlar:** Cronbach alfa gerçek güvenilirliğin üstünde değerler vermektedir

4. **Madde sayısının az olması durumu:** Bu durumda da Cronbach alfa gerçek güvenilirlikten daha düşük sonuçlar verebilmektedir; ancak bu durum her zaman geçerli değildir. Cortina (2003) bu durumu şöyle örneklendirilmiştir: Ölçeğin yeterli sayıda, (20'den fazla) maddesi varsa, maddeler arasındaki korelasyon çok düşük bile olsa Cronbach alfa katsayısı 0.70'den yüksek çıkabilir. Cortina, maddeler arası korelasyonu 0.30 civarında olan, ilişkisiz iki faktörlü ve 14 maddeden fazla maddeye sahip olan ölçeklerde, Cronbach alfa katsayısının 0.70'den de yüksek çıkabileceğini raporlamıştır. Hatta maddeler arası korelasyonun ortalama 0.70 civarında ya da daha yüksek olduğu durumlarda, üç farklı boyuta sahip ölçeklerde dahi Cronbach alfa katsayısının 0.80'den yüksek çıkabileceği belirtmiştir. Dolayısıyla, Cronbach alfa katsayısının gerçek güvenilirlikten daha düşük değer almasının temel nedeni madde sayısının az olmasından ziyade, maddeler arasındaki korelasyonun düşük olmasıdır.

Cronbach alfa katsayısının varsayımları olan paralelliğin ve hataların ilişkisizliğinin sağlanmadığı durumlarda, bu katsayının yerine kullanılabilecek alternatif iç tutarlılık katsayıları geliştirilmiştir. Bu katsayıların hangi durumlarda kullanılabileceği ve elde edilecek güvenilirlik katsayısının gerçek güvenilirlik katsayısına göre durumu karşılaştırmalar yapılarak çeşitli kaynaklarda verilmiştir. Farklı katsayıları ayrıntılı olarak tartışan makalelerden biri Osburn'un (2000) çalışmasıdır. Bu çalışmada, Cronbach alfa katsayısının hesaplanmasının uygun olmadığı durumlar tartışılmış ve bu gibi durumlarda kullanılabilecek farklı katsayılardan bahsedilmiştir. Test bileşenlerinde farklı sayıda madde bulunduğu durumlarda kullanılabilecek alternatif katsayılardan biri Raju'nun katsayısıdır. Güvenirlik kestirimi yapılacak ölçme aracının rastgele iki yarıya ayrıldığı durumlarda Angoff-Feldt katsayısı da kullanılabilir. Bir başka alternatif katsayı ise Kristof'un katsayısıdır. Bu katsayı testin konjenerik üç bölüme ayrıldığı durumlarda kullanılabilir. Bu katsayılara ek olarak, Feldt Gimmer, Lamda 2, Lamda 4, Standardize Alfa ve Tabakalı Alfa gibi güvenilirlik katsayıları, Cronbach alfa katsayısının kullanımının uygun olmadığı farklı durumlarda tercih edilebilir. Yurdugül (2006) bu katsayılara ek olarak farklı güvenilirlik katsayılarının da kullanılabileceğini belirtmiştir. McDonald tarafından geliştirilmiş Omega katsayısı da Cronbach alfaya alternatif katsayılardan biridir. Bu katsayı özellikle konjenerik ölçmeler için geliştirilmiştir ve maddelere ait faktör yükleriyle hesaplanmaktadır (McDonald, 1999). Omega katsayısı, tüm ölçmelerde Cronbach alfaya eşit ya da daha yüksek değer vermektedir. Bileşenlerin konjenerik olması durumunda kullanılabilecek bir diğer katsayı ise Revelle'nin beta katsayısıdır. Bu katsayı ise madde kovaryanslarının ortalamalarına dayalı olarak hesaplanmaktadır.

Cronbach alfaya alternatif olarak hesaplanabilecek farklı katsayılar Zinbarg, Revelle, Yoven ve Li (2005) tarafından tanıtılmış ve bu katsayılar arasında karşılaştırmalar yapılmıştır. Bu çalışmada Cronbach alfaya alternatif olarak kullanılabilecek Revelle'nin beta katsayısı, McDonalds'ın omegası ve güvenilirliğin iki ayrı kavramsallaştırılması olan hiyerarşik faktör modeli ile iki faktörlü model kıyaslanmıştır. Veri setinde de çeşitli değişimler yapılarak tek boyutlu ve çok boyutlu verilerde katsayılar karşılaştırılmıştır. Sonuç olarak Cronbach alfa'nın birçok durumda diğer katsayılardan daha kararlı sonuçlar ürettiği fakat eğer ortak faktör varyansı daha önemliyse, Omega katsayısının daha açıklayıcı olduğu rapor edilmiştir. Farklı durumlar için farklı katsayıların tercih edilmesi gerektiği belirtilmiştir.

Cronbach alfa katsayısı gerek esnekliği gerekse kullanışlılığından dolayı hemen hemen her türlü ölçme aracından elde edilen puanların güvenirliğini belirlemede sıkça kullanılan bir katsayı haline gelmiştir. Esnekliğinden, kolay hesaplanmasından ve çokça tercih edilmesinden dolayı varsayımları yokmuş gibi algılanıp her veri seti için doğrudan hesaplanıp yorumlanan bu katsayının dikkatsizce kullanımı söz konusu ölçme aracının güvenirliğine ilişkin hatalı bilgi vermektedir. Cronbach alfa katsayısının doğru ve yerinde kullanımının sağlanması için maddelerin paralellik şartını sağlıyor olması ve homojen olması gerekmektedir. Dolayısıyla ölçme sonuçlarına ilişkin güvenirliğin gerçek değerine ulaşmak için öncelikle bu varsayımlar test edilmeli ve varsayımlar karşılanıyorsa, Cronbach alfa katsayısı hesaplanmalıdır.

## 5. SONUÇ

Geçerlik üzerindeki tartışmalar birleştirilmiş geçerlik görüşü üzerinde yoğunlaşmıştır. Birleştirilmiş geçerliğin merkezinde yapı geçerliği bulunmakta ve diğer geçerlik türleri yapı geçerliğinin alt boyutları olarak ele alınmaktadır. Birleştirilmiş geçerlik görüşüne karşı çıkarak, eğitim alanındaki testlerde uygulanmasının oldukça güç olduğunu ve öncelikli olarak kapsam geçerliğine odaklanması gerektiğini savunan çalışmaların yanında, her iki görüşe de karşı çıkarak geçерleme sürecinde kapsam geçerliğinin ya da yapı geçerliğinin ön planda tutulmasından çok; testin anlamını ortaya koymayı sağlayacak test içi ve test dışı kaynaklara başvurulması gerektiğini ifade eden çalışmalar da bulunmaktadır.

Güvenirlik tanımı ve güvenirliğe ilişkin kanıt elde etme süreci üzerinde geçerliğe göre daha fazla uzlaşma sağlanmış olsa da; çok sık kullanılan bir güvenirlik katsayısı olan Cronbach alfa'ya ilişkin yapılan çalışmalarda, araştırmacıların tam anlayamadığı noktaların olduğu ve bu nedenle yanlış yorumlamalara gidildiği görülmektedir. Ayrıca Cronbach alfa katsayısının varsayımlarının karşılanmadığı durumlarda iç tutarlılık anlamında güvenirliği kestirmede kullanılabilecek alternatif katsayılar olduğu görülmüştür. Yapılan birçok simülasyon çalışmasında hangi durumlarda hangi katsayıların daha kararlı ve gerçek güvenirliğe daha yakın sonuçlar verdiği tartışılmıştır. Dolayısıyla, araştırmacılar doğrudan Cronbach alfa katsayısını hesaplamak yerine veri setlerine ve ölçme araçlarına uygun bir güvenirlik belirleme yöntemi kullanmalı ve gerçek güvenirliğe daha yakın sonuçlar elde etmelidirler.

## KAYNAKLAR

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Baykul, Y. (2010). Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması. Ankara: ÖSYM Yayınları.
- Berberoğlu, G. (2012). Kapsam geçerliği. *Cito Eğitim: Kuram ve Uygulama*, 15, 10-16.
- Borsboom, D., Mellenbergh, G.J & van Heerden, J. (2005). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Crocker, L. & Algina, J. (2006). *Introduction to classical and modern test theory*. Fort Worth, FL: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J. (2004). My current thoughts on Coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 281-302.
- Ebel, R. L. (1956) Obtaining and reporting evidence on content validity. *Educational and Psychological Measurement*, 16, 269-282.
- Ebel, R. L. & Frisbie, D.A. *Essentials of educational measurement*. New Jersey: Prentice - Hall International, Inc.
- Embretson, S.E. (2007). Construct Validity: a universal validity system or just another test evaluation procedure?. *Educational Researcher*, 36(8), 449-455.
- 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.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Haynes, S. N., Richard, D. C. S. & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.
- Kelecioğlu, H. ve Göçer Sakin, S. (2014). Geçmişten günümüze geçerlik. *Eğitimde ve Psikolojide Ölçme Dergisi*, 5(2). 1-11.
- Lissitz, R. W. & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36, 437-448.
- Loveinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.
- McDonald R. P. (1999). *Test theory: a unified treatment*. Mahwah, NJ: LEA Publisher.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16-18.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Murphy, K.R. & Davidshofer, C. O. (1998). *Psychological testing: principles and applications*. New Jersey: Prentice - Hall International, Inc.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5, 343-355.

- Pan, Y. C. (2009). Evaluating the appropriateness and consequences of test use. *Colomb. Appl. Linguist. J.*, 11, 93-105.
- Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32(4), 329-353.
- Raykov, T. (2001). Bias of coefficient  $\alpha$  for fixed congeneric measures with correlated errors. *Applied Psychological Measurements*, 25(1), 69-76
- Sacon, G. (2000). Assessment of reliability when test items are not essentially  $\tau$ -Equivalent. *Developments in Survey Methodology*, 23-35.
- Shevlin, M. Miles, J.N.V. Davies, M. N. O. & Walker, S. (2000). Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences*, 28: 229-237.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Sjtsma, K. (2009). On the use, the misuse, and the very limited usefulness on Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Thorndike, R. M. & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education*. USA: Pearson Education Limited.
- Yang, Y. & Green, S.B. (2011). Coefficient alpha: a reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4) 377-392.
- Yurdugül, H. (2006). Paralel, eşdeğer ve konjenerik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15-37.
- Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's,  $\beta$  and McDonald's  $\omega$ : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 1-11.

## **EXTENDED ABSTRACT**

### **1. Different Approaches to the Validity and Validation Process**

The objective of this study is to present various viewpoints regarding definition of validity and validation process, furthermore, to enlighten the most frequently preferred reliability coefficient with the assumptions underlying it. Historically, validity evidences have been revealed based on correlational studies. However, the classical validity definition is seen as too broad to be used in practice, so some other kinds of validity are taken into consideration together with the correlational studies. The discussion on whether test scores could be supported through the relation with the theoretical construct has started. As a result of the discussions, the concept of construct validity has been suggested. Lastly, Messick (1995) proposed a new validity type named as “consequential validity”. He focused on the deductions from the test scores, the consequences of these deductions and whether or not the empirical studies and theory are compatible with the decisions based on the test scores. He also mentioned a concept of “unified validity”. The unified validity approach in which the construct validity has been put into the center and other types of validity have been considered as the sub-dimension of the construct validity has been constituted. Under the concept of unified validity, construct validity is emphasized more than other types of validity. The evidences obtained through other types of validity can be utilized in providing construct validity. The types of validity are not separated from each other such as content or criterion based validity. Conversely, they are accepted as dimensions of validity, which should be analyzed in order to get the construct validity of the scales. However, some researchers, supporting the idea that the content validity is more important especially in the educational tests disagree with the unified validity approach. Furthermore, some researchers support the idea that internal (content, reliability) and external (construct or consequential validity) evidences of validity are not superior to each other and all of them are required to be analyzed during the validation process.

Lissitz and Samuelsen (2007) are the ones opposing the idea of unified validity. They supported that in the scenario of defining the essence of the construct measured by the test, content validity has more importance than construct validity. Embertson (2007) who criticized Messick and Lissitz and Samuelsen (2007) because of the over emphasis on the construct or content validity respectively, argued that making decisions based only on the content validity, reliability studies or construct validity to bring out the meaning of test scores is not an integrated approach. All the sources of evidence are ought to be considered in the validation process. The other researcher, Sireci (1998) also disagrees with the idea of the central role of construct validity and he argued that focusing only on the content validity cannot be sufficient to define the measured construct and both construct and content validity should be given equal importance.

### **2. The Most Preferred Reliability Coefficient Cronbach Alpha and the Assumptions Underlying It.**

The second primary goal of this article is to enlighten the facts about the most frequently used reliability coefficient, which is Cronbach alpha. It is preferred so frequently that even Cronbach has been surprised to this situation. Some researchers have studied the reasons of this frequently usage of Cronbach alpha. Yang and Green (2009) were the ones

who wondered that why Cronbach alpha had been preferred so much and they found some reasons. One of these reasons is that it is so easy to interpret. The other reason of the extensive usage of this coefficient is that it is much more objective reliability technique than the others such as split half method or test retest application. By using Cronbach alpha, the researcher is not obliged to make a decision about the time elapsed between the applications or how the test should be separated into two splits. The third reason of Cronbach alpha's common usage is that it is easier to revise the test. By using Cronbach alpha, one can get the knowledge about how the reliability coefficient will change if one of the items is deleted. Lastly, calculating this coefficient is much more practical and economical than other reliability methods. Due to these advantages, nearly for all of the tests, Cronbach alpha can be calculated and the researchers have preferred this coefficient. However there are some situations in which this coefficient gives deceptive knowledge about the reliability of the test. If the items of the tests are not parallel, essential equivalent or tau-equivalent, the coefficient gives the lower bound of the real reliability. Also under the condition of the items' errors correlate negatively and positively, this coefficient can be lower and higher than the real reliability respectively. The last situation in which Cronbach alpha can be lower than the real value of the reliability is that the test is composed of few items of the test consists of subtests, which are not correlated each other. In these situations, Cronbach alpha tends to get the value of the lower bound of the real reliability. These disadvantages of Cronbach alpha have been known and accepted by the inventor of it, Guttman (1945). Also, for these situations in which the Cronbach alpha is not suitable, Guttman has suggested to use different methods of reliability rather than this coefficient. There are some other coefficients, which are much more applicable than Cronbach alpha, such as stratified alpha, Kristof's coefficient, lambda<sup>2</sup>-4 etc (Osburn, 2000). To sum up, in spite of the extensive usage of Cronbach Alpha, in order to get the real reliability value of test, the assumptions of this coefficient should be analyzed. Instead of using Cronbach alpha as soon as collecting the test data, first of all, the analyses should be carried on whether the assumptions of this coefficient are met or not. If assumptions of Cronbach alpha do not hold, the real reliability value about the scores cannot be obtained with the usage of Cronbach alpha. In a nutshell, before using this coefficient to gather reliability evidence, the researchers have to analyze the scale, the correlations between the items and decide whether the usage of Cronbach alpha is suitable or not.