

KUDER-RICHARDSON 20, CRONBACH'IN ALFASI, HOYT'UN VARYANS ANALİZİ, GENELLENİRLİK KURAMI VE ÖLÇÜM GÜVENİRLİĞİ ÜZERİNE BİR ÇALIŞMA

A Study on the Kuder-Richardson 20, Cronbach's Alpha, Hoyt's Analysis of Variance, Generalizability Theory and Score Reliability

Vahit BADEMCİ¹

Özet

Güvenirlilik ve geçerlik, testlerin ya da ölçme araçlarının özellikleri değildir. Güvenirlilik gibi geçerlik de, ölçme aracının kendisine değil, ölçümlere işaret eder. Böylelikle, "testin güvenirliliği" veya "ölçeğin geçerliği" benzeri ifadeler kullanmak doğru değildir. Güvenilir ya da güvenilmez olan testler veya ölçme araçları değil, bir test ya da ölçme aracından elde edilmiş ölçümlerdir. Dolayısıyla, "ölçüm güvenirliliği" ya da "test ölçümlerinin güvenirliliği" benzeri ifadeler kullanılmalıdır. Cronbach'ın alfası, Kuder-Richardson 20 ve Hoyt'un varyans analizi, iç tutarlılık güvenirlilik kestirimlerinde yaygın kullanılan yöntemlerdir. Kuder-Richardson 20, sadece iki değerli [0,1] ölçümlenmiş maddeler için uygundur. Cronbach'ın alfası, hem iki değerli [0,1] ölçümlenmiş maddelerle, hem de ağırlıklı [1,2,3,4,5] ölçümlenmiş maddelerle kullanılabilir. Tüm maddeler iki değerli [0,1] ölçümlendiğinde, Kuder-Richardson 20, Cronbach'ın alfası ve Hoyt'un varyans analizi formülleri aynı sonuçları verir.

Anahtar kelimeler: Ölçüm güvenirliliği, geçerlik, Kuder-Richardson 20, Cronbach'ın alfası, Hoyt'un varyans analizi, genellenirlilik kuramı

Abstract

Reliability and validity are not the characteristics of tests or measurement instruments. Validity also refers to scores not to the measurement instrument itself, like reliability. Accordingly, it is not correct to use the terms such as "the reliability of the test" or "the validity of the scale". It is not the tests or measurement instruments which are reliable or unreliable but the scores obtained from a test or measurement instrument. For this reason, statements such as "score reliability" or "the reliability of the test scores" should be used. Cronbach's alpha, the Kuder-Richardson 20 and Hoyt's analysis of variance are commonly used methods in estimating the internal consistency reliability. The Kuder-Richardson 20 is only suitable for dichotomously scored items [0,1]. Cronbach's alpha can be both used with dichotomously scored items [0,1] and weightily scored items [1,2,3,4,5]. The formulas of the Kuder-Richardson 20, Cronbach's alpha and Hoyt's analysis of variance give identical results when all items scored dichotomously [0, 1].

Keywords: Score reliability, validity, the Kuder-Richardson 20, Cronbach's alpha, Hoyt's analysis of variance, generalizability theory.

¹ Yrd.Doç.Dr.; Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi, 06830, Gölbaşı - Ankara, vahitbademci@yahoo.com

1. GİRİŞ

Ölçme ölçümlerinin istatistiksel kuramları son 100 yıl içinde gelişmiştir (Suen & Lei, 2007:1). Teorik temelleri 1900'lü yılların başlarına, Charles Spearman'ın (1904a; 1904b; 1907; 1910; 1913) bir dizi çalışmasına dayanan ve çeşitli istatistiksel teoremler ve eşitliklere sahip ve de ölçmenin en fazla popüler kuramı olan (Crocker & Algina, 1986; Thorndike, 1982; Strube, 2000; Suen & Lei, 2007) Ölçmenin Klasik Kuramı (Suen & Lei, 2007) ya da Klasik Gerçek Ölçüm Kuramı (Allen & Yen, 1979) veya bir başka adlandırılmasıyla Klasik Güvenirlik Kuramı,-başlangıçta- bilişsel psikoloji ve eğitimde ölçme için geliştirilmiştir (Shrout, 1998); bugün ise, psikiyatriden tıbbı, işletmeden mühendisliğe kadar çok çeşitli disiplinlerde kullanılmakta veya yararlanılmaktadır. [1900'lü yılların başlarında başlayan ve çeyrek yüzyılın üstünde devam eden -ilk- süreçte, Charles Spearman'ın yanı sıra, Karl Pearson, George Udny Yule, Truman Lee Kelley, Edward Lee Thorndike, William Brown gibi isimlerin de Klasik Gerçek Ölçüm Kuramı ve gelişimine önemli katkıları unutmamak gerekir]*

Psikometri, ruhbilimsel ölçmenin bilimidir; psikometrist ise, test etme, ölçme ve bellilendirme (assessment) alanında uzmanlaşmış ruhbilimsel ya da eğitimsel bir profesyoneldir (Reynolds, Livingston & Willson, 2009). Nitelikli bir psikometrist, ilgi alanı olan husustaki temel kavramları doğru kullanabilmelidir; bir psikometrist, örneğin, yukarıda bahsedilen Klasik Gerçek Ölçüm Kuramıyla (Allen & Yen, 1979) ilgili çalışmalar sürdürüyorsa veya bilimsel etkinliklerde bulunuyorsa, söz gelimi, *güvenirlik* ve *geçerlik* kavramlarıyla ilgili çağdaş gelişmeleri iyi takip etmeli, değişen ya da anlam değiştiren bu kavramları ve içeriklerini ve ilgili yenilikleri de çalışmalarına aktarmalıdır. Ancak, tam da bu noktada çok dikkate değer bir husus gözden kaçırılmamalıdır; tutuculuk, pek çok istatistiksel kuramda bir gelenektir (Li & Wainer, 1998) ve çoğu psikometrist de tutucudur (Gardner, 1999). Bunun tam karşıtı olarak, Norton'da (2001), *iyi bir bilim adamının niteliğini, yeni yerleşik düşünce [paradigma; bilimsel devrim] icat etmeyi başaran [kimse] olarak ifade eder.*

Türkiye'de, çoğu psikometrist de var olan tutuculuk özelliğinin, tembellik, bilimsel yetersizlik, haset vd. olumsuz bazı özellikler ile de -maalesef- aşırı düzeyde harmanlanmış durumda bulunması gibi sebeplerden dolayı, [en azından son 20-25 yıldır] psikometri [bilimi] geri, hatta çağ dışı kalmıştır ve -ne acıdır ki- çağ dışı olmayı da sürdürmekte ve içine sindirmektedir. Son zamanlarda ise, Vahit Bademci'nin paradigma değişikliği ile Türk eğitim ve bilim gündemine *ilk defa* taşıdığı çok çeşitli konuları ve çalışmalarındaki hususları, kendi makale ve yayınlarına ya da çalışmalarına taşıyan akademik unvanlı bazı [sözde] ölçme ve değerlendirme uzmanlarının ya da psikometristlerin veya araştırmacıların ya da özetle bazı [sözde] bilim insanlarının, Bademci'nin çalışmalarına *atıf vermemeleri* ise, o kişilerin bilimsel ahlak dışı davranışlara sahip olduklarının ve akademik terbiye

eksikliklerinin ve bilimsel yetersizliklerinin ve bilimsel zavallılıklarının da açık bir göstergesi olarak ifade edilebilir.

2. PARADİGMA DEĞİŞİKLİĞİ: GÜVENİRLİK GİBİ GEÇERLİK DE, TESTİN YA DA ÖLÇME ARACININ KENDİSİNE DEĞİL, ÖLÇÜMLERE İŞARET EDER

Eğitim ve psikolojide üst düzey kalitede bir ölçmenin kalbinde iki kavram yatar; güvenilirlik ve geçerlik. **Güvenirlik**, belirli bir evrene veya örnekleme uygulanmış bir test ya da ölçme aracından elde edilmiş ölçümlerin tutarlılığı veya tekrarlanabilirliği; **geçerlik** ise, belirli bir evrene veya örnekleme uygulanmış bir test ya da ölçme aracından elde edilmiş ölçümlerden yapılmış belirli yorumların ve kullanımların uygunluğu ve yeterliği şeklinde tanımlanabilir (Bademci, 2001a; 2007; 2011). Güvenirlik ve geçerlikle ilgili Bademci (2001a; 2007; 2011) tarafından ortaya konulan yeni paradigma doğrultusundaki tanımlamalarda, testlerin veya ölçme araçlarının kendilerine değil, testlerden veya ölçme araçlarından elde edilen ölçümlere vurgu yapılmıştır. Çünkü güvenilirlik ve geçerlik, testlerin ya da ölçme araçlarının değil, ölçümlerin fonksiyonlarıdır (Bademci, 2007; Mji & Onwuegbuzie, 2004).

Güvenirlik gibi geçerlik de, ölçme aracının kendisine değil, ölçümlere işaret eder (Bademci, 2001a; 2007; 2010; 2011; Brookhart & Nitko, 2008; Nitko, 2001). Bir başka ifadeyle, güvenilirlik ve geçerlik, testlerin ya da ölçme araçlarının özellikleri değildir (Murphy & Davidshofer, 2001; Barnes, Harp & Jung, 2002; Worthen, White, Fan & Sudweeks, 1999). Böylelikle, güvenilirlik ve geçerlik bir testin değişmeyen, kalıcı özellikleriymiş gibi, “bir testin güvenilirliği”nden veya “bir testin geçerliği”nden söz etmek ise, *uygun değildir, doğru değildir* (Bademci, 2007; Kubiszyn & Borich, 2007). “Ölçüm güvenilirliği” (Baugh, 2002; Buhi, 2005; Ebel & Frisbie, 1991) veya “yorumun geçerliği [ya da ‘çıkarımın geçerliği’]”nden (Linn & Gronlund, 2000; McMillan, 2001) bahsetmek, çok daha doğru olur. Bu hususlarla ilgili olarak, örneğin, Cronbach (1972), bundan 38 yıl önce, “test geçerliği” hakkında konuşmanın *mantıksız* olduğunu, geçerlemenin (validation) ise, test yorumlayıcısının görevi olduğunu açıkça vurgulamıştır. Ebel (1965) ve Gronlund (1965) ise, bundan 45 yıl önce, “testin güvenilirliği”nden ziyade, “test ölçümlerinin güvenilirliği” hakkında konuşmanın çok daha uygun olduğunu altını çizmişlerdir.

2.1. “Puan” yerine “ölçüm” terimi kullanılmalıdır; “puanların güvenilirliği” ifadesini kullanmak ise, uygun değildir, doğru değildir

Oğuzkan (1981: 129) tarafından hazırlanmış *Eğitim Terimleri Sözlüğü*'nde “puan” kavramının karşısında “Bir testten, genellikle doğru olarak yanıtlanan madde sayısına göre elde edilen sonuca sayıyla biçilen değer” açıklaması bulunmaktadır. Bu tanımlamanın da işaret ettiği -ve Türk eğitim ve bilim topluluğunca da yaygın kabul edildiği- gibi, “puan” kavramı

güçlü biçimde testlere vurgu yapmakta ve testleri çağrıştırmaktadır. Ölçme işlemlerinde, sadece testler değil, *çıkın* (portfolio) ve de diğer *erişim* (performance) ölçme araçları veya sanatsal ürünlerin kalitesinin ortaya konulmasıyla ilgili çok çeşitli ölçme araçları ya da çok çeşitli gözlem *izlekleri* (*prosedürleri*) de kullanılmaktadır; “ölçüm”, “ölçme işlemleri sonunda elde edilen sayılar” (Turgut’tan aktaran Bademci, 1999) veya “ölçme sonuçlarını gösteren sayılar” (Jones, 1971, Linn & Gronlund, 2000) şeklinde de ifade edilebilmektedir. Dolayısıyla, sadece bu çok kısa açıklamalar dahi, sınırlı “puan” terimi yerine, çok daha geniş ve kapsamlı çerçeveye sahip olan “ölçüm” teriminin kullanılmasının *yerindediğini ve gerekliliğini* ortaya koyucu niteliktedir. Böylelikle de, “puanların güvenilirliği” yerine “ölçümlerin güvenilirliği” ifadesini kullanmak *çok daha uygun ve doğrudur*.

2.2. Güvenirlik, Bir Test veya Ölçme Aracından Elde Edilmiş Ölçümlerin Bir Özelliğidir

1940’lardan bu yana, yani 2010’lara, yaklaşık 70 yıldır Türkiye’de ve Türk eğitim ve bilim topluluğunda, [neredeyse] bütün araştırmacılar, psikometristler veya ‘eğitimde ölçme ve değerlendirme uzmanları’, güvenirliliğin bir testin ya da ölçme aracının özelliği olduğunu vurgulayan “testin güvenirliliği”, “ölçme aracı güvenilirdir”, “ölçeğin güvenirliliği”, “test güvenilirdir” ve benzeri ifadeleri kullanmaktadır ve *-çağdaş* bilim, bilimin varlığı, bilim insanların ve Türkiye’deki psikometristlerin ya da araştırmacıların niteliği adına ne acıdır ki- hâlâ da kullanmayı sürdürmektedirler (konuyla ilgili örnekler için bakınız, Bademci, 2005a; 2007; 2008; 2010). Ancak, bu kullanım biçimleri ve güvenirliliğe “testin ya da ölçme aracının kendisinin bir özelliği” gibi anlam yüklenmesi, *doğru değildir* ve de *büyük hatadır*. Zira, bir ölçme aracı veya “bir test, güvenilir veya güvenilmez değildir” (Crocker & Algina, 1986: 144). Güvenilir ve güvenilmez olan testler veya ölçme araçları değil, bir test ya da ölçme aracından elde edilmiş ölçümlerdir (Bademci, 2007). Çünkü *güvenirlik, sınava giren belirli bir gruba uygulanmış bir testten ya da ölçme aracından elde edilmiş ölçümlerin bir özelliğidir* (Bademci, 2001a; 2004; 2007; 2011; Barnes, Harp & Jung, 2002; Crocker & Algina, 1986). Yani güvenirlilik, test ya da ölçek sonuçlarının bir özelliğidir; dolayısıyla, test ya da ölçek ölçüm güvenirliliği de, testi alan öğrencilerin grubuna bağlı olacaktır (Bademci, 2007; 2011; Livingston, 1988).

2.3. Örneklem Özellikleri Ölçüm Güvenirliliğini Etkiler: Aynı Test veya Ölçme Aracı, 100 Farklı Örneklem Uygulansa, 100 Farklı Güvenirlilik Katsayısı Ortaya Çıkabilir

Güvenirlikle ilgili bir husus yıllardır çok açıkça göz ardı edilmiştir ve bu durum, bu kavramın doğru anlaşılmasının da önemli etkenlerinden biri olmuştur; güvenirlilik katsayıları, neredeyse her zaman, tüm evrenlerden değil, kişilerin örneklemelerinden alınmış ölçmelerden hesaplanmaktadır (Traub, 1994). Yine hatırlanmalıdır ki, örneklem özellikleri ölçüm güvenirliliğini etkilemektedir (Bademci, 2007; Henson, Kogan & Vacha-Haase, 2001).

Örneğin, bir testin veya ölçme aracının uygulandığı örneklemin bağdaşık [homojen] ya da ayrışık [heterojen] olması, ölçüm güvenilirliğinin azalmasına veya artmasına neden olmaktadır (Bademci, 2007). Bir başka ifadeyle, *aynı test*, bağdaşık veya ayrışık örneklere uygulandığı zaman güvenilirliğe ilişkin farklı sonuçlar doğurabilecektir (Bademci, 2007). Kısaca, ölçüm güvenilirliği, örneklemden örnekleme değişmektedir (Bademci, 2007; Capraro & Capraro, 2002). *Aynı ölçek veya test ya da ölçme aracı, 100 farklı örnekleme uygulansa, 100 farklı güvenilirlik katsayısı ortaya çıkabilir veya çıkar* (Bademci, 2007; 2011; Buhi, 2005). Hâl böyle iken, “test güvenilirdir” ya da “ölçeğin güvenilirliği” demek ve güvenilirliği, testin veya ölçeğin ya da ölçme aracının bir özelliği gibi ima veya ifade etmek *uygun değildir, doğru değildir* (Bademci, 2007; 2011).

2.4. Yurt Dışında ve Türk Eğitim ve Biliminde, Güvenirlik ve Geçerlik Hususları Etrafında Ölçme ve Araştırma Yöntembilimindeki Bilimsel Devrimin Öncüleri

Yurt dışında, “test ölçümlerinin güvenilirliği” ve “güvenirliğin testin kendisinin değil, test ölçümlerinin bir özelliği” olduğunu vurgulama hususunda Crocker & Algina (1986), Ebel & Frisbie (1991), Gronlund & Linn (1990) ve Suen (1990) gibi isimler ve çalışmaları önemli bir yer tutar; ancak, bu konuyu ısrarla vurgulayan ve etkin bir şekilde tartışmalara katılan ve alanın en önde gelen ve etkili süreli yayımlarından biri olan *Educational and Psychological Measurement* dergisinin 1995-2003 yılları arasında editörlüğünü de yapan Thompson’un (1994a; 1994b; 1999; Thompson & Vacha-Haase, 2000), -özellikle 1994 yılından bu yana- ölçme ve ölçmeyi yakından ilgilendiren konulardaki köklü reform hareketinin tetikleyicisi ve öncüsü olduğu söylenebilir. Geçerlik hususu ile ilgili olarak da yukarıdakilere, Cronbach (1971; 1988), Messick (1989; 1995) ve Kane (1990; 1992) gibi isimler ve çalışmaları eklenebilir.

Türkiye’de ve Türk eğitim ve bilim topluluğunda ise, 1940’lardan bu yana, 60 yılı aşkın bir süredir kullanılan “test güvenilirdir” veya “test geçerlidir” gibi hatalı ifade etme biçimlerine, *hem güvenilirliğin ve hem de geçerliğin* hatalı yorumlanış ve uygulama şekillerine -en azından 1994 yılından beridir- sürdürdüğü çalışmalarıyla Bademci (2001a; 2001b; 2002; 2004; 2005a; 2005b; 2005c; 2006a; 2006b; 2006c; 2007; 2008; 2010; 2011) *tek başına* ve açık biçimde karşı çıkmış, bir *paradigma değişikliği* gerekliliğini vurgulayarak, bir *yeni* paradigmayı da *bilimsel kanıtlarıyla* Türk eğitim ve bilim topluluğunun gündemine taşımıştır.

Paradigma değişikliği ya da yeni paradigmaya geçiş, gerçeklerin peşinde koşan, objektif ve bağımsız düşünen -ancak- *yaratıcı* bir bilim insanının gerçekleştirebildiği/gerçekleştirdiği bilimsel bir devrimdir ve bilimsel ilerleme, devrimsel bir süreçtir ve de *bilimsel devrim, nadiren ortaya çıkan olağan dışı bilimsel süreçtir* (Bademci, 2007; 2010; 2011; Güneş, 2003;

Kuhn, 1995; Serdar, 2001; Topdemir, 2002). Bademci'nin (2007) 60 yılı aşkın bir süre sonra ölçme ve araştırma yöntembilimi alanlarında gerçekleştirdiği paradigma değişikliğinin, Türkçe literatürde, “*Türk Eğitim ve Biliminde “Vahit Bademci'nin Paradigma Değişikliği ya da Vahit Bademci Markası: Testler veya Ölçekler Güvenilir ve Geçerli Değildir”* (Gazi Haber, 2010: 48; Korkmaz, 2010: 21) şeklinde ifade edildiği de görülmektedir.

Bademci'nin (2007) ortaya koyduğu *yeni paradigma* etrafındaki görüşleri, yaklaşımları ve çalışmalarında meydana vurduğu bazı bilimsel kanıtlamaları ise, bilimdeki çağdaş gelişmelerin ve yenileşmelerin yanında olan araştırmacıların bilimsel çalışmalarındaki yerini almaya başlamıştır (örneğin, bakınız, Beycioğlu, 2007; Buluş Kırıkkaya, Bozkurt, İşeri, Vurkaya & Bali, 2011; Cebeci, 2006; Hotaman & Yüksel-Şahin, 2010; Kartal, 2009; Kartal & Pekkanlı, 2011; Korkmaz, 2010; Sayın, 2008; 2010; Sever, 2008).

2.5. Korkmaz'ın (2010) Çalışmasının Ortaya Koyduğu Acı Gerçek: Güvenirlikle İlgili Olarak, 2000-2009 Yılları Arasında Yapılmış Yaklaşık Her 10 Yüksek Lisans ve Doktora Tezinin 8'inde, Bilimsel Olarak Hatalı ve Çağcıl Olmayan Kullanım İfadeleri Bulunmaktadır

Zonguldak Karaelmas Üniversitesi'nde, Saime Sayın'ın tez danışmanlığında yapılmış olan Korkmaz'ın (2010) yüksek lisans tezi, konu ve tespitleri ve topladığı verileri ve ortaya koyduğu pek çarpıcı bilimsel bulgularından dolayı, Türkiye'deki tüm üniversitelerin ilgili birimlerince mutlaka ve dikkatlice okunulmalı ve bilgilendirilmelidir. *Eğitim bilimleri ile ilgili olarak*, Korkmaz'ın (2010) tezinin 81. sayfasında olan ve aşağıya olduğu gibi aktarılmış bulunan sadece bir bulgu dahi, mevcut durumun ciddiyetini anlatmaya yeter görünmektedir:

“...sadece *güvenirlik hususu ile ilgili olarak, Ankara, Gazi ve Hacettepe Üniversitelerinde 2000-2009 yılları arasında yapılmış olan 444 yüksek lisans ve doktora tezinin 349'unda, yani yaklaşık her 10 tezin 8'inde bilimsel olarak hatalı ve çağcıl olmayan* [“testin güvenilirliği”, “ölçeğin güvenilirliği”, “aracın güvenilirliği” gibi]* **kullanım ifadeleri bulunmaktadır**. Böylelikle, tez danışmanlarının ve tez jüri üyelerinin ve bilimsel araştırma ve ölçme ile ilgili derslere giren akademisyenlerin ve Türkiye'de piyasada bulunan “bilimsel araştırma teknikleri veya yöntemleri” ve de “ölçme” ile ilgili yazılmış ders kitaplarının nitelikleri de tartışmalı hale düşmüştür; çünkü tüm bu sonuçlara dayanılarak, bu hususlardaki bilimsel niteliğin son derece düşük ya da çok zayıf kalitede olduğu da söylenebilir ” (Korkmaz, 2010: 81). [Yazarından izin alınmıştır.]

3. KLASİK GERÇEK ÖLÇÜM KURAMINDA ÖLÇÜM GÜVENİRLİĞİ KESTİRME YÖNTEMLERİ

Test-tekrar test, almaşık (alternate) form, yarıya bölme (split-half), değer biçiciler arası (interrater), Cronbach'ın (1951) alfası (α) [ya da alfa katsayısı], Kuder-Richardson 20 (KR-20; Kuder & Richardson, 1937) ve Hoyt'un (1941) varyans analizi, ölçüm güvenirliliğinin kestiriminde kullanılan bazı yöntemlerdendir; değer biçiciler arası, test-tekrar test, almaşık form, yarıya bölme ölçüm güvenirliliği kestirim yöntemlerinde, Pearson çarpım moment korelasyon katsayısından yararlanılır (Crocker & Algina, 1986; Linn & Gronlund, 2000; Mehrens & Lehmann, 1991). Yeri gelmişken, burada bir hususun belirtilmesinde ve de bazı yanlış yorumlamaların düzeltilmesinde, fayda görülmektedir: Cronbach'ın alfası, basit bir ölçümlerin korelasyonu *değildir* (Guthrie, 2000; Spector, 1992).

4. CRONBACH'IN ALFASI ÜZERİNE BİR KISA GİRİŞ

Cronbach'ın alfası, ölçüm güvenirliliğinin kestiriminde, diğerlerine göre, kat kat fazla kullanılan bir yöntem olarak görünmektedir (Hogan, Benjamin & Brezinski, 2000). Ancak bu yaygın kullanımının yanı sıra, Cronbach'ın alfasının, üzerinde en fazla mit [hayali öykü] üretilen yöntem olduğu da ifade edilebilir. Örneğin, kimi araştırmacılar ya da psikometristler, Cronbach'ın alfa katsayısının 0 ile 1 arasında değerler aldığı belirtmişlerdir (Akdeniz, Aydemir, Akdeniz, Gülseren & Kültür, 1999; Aygin & Eti Aslan, 2005; Demir & Okan, 2009; Özkan & Sevil, 2007; Tan & Erdoğan, 2004; Tan, 2008; Yılmaz & Sümbül, 2009); kimi araştırmacılar veya psikometristler ise, kullanım olarak, Cronbach'ın alfa katsayısını sadece ağırlıklı ya da çok değerli [-polytomously-; 1,2,3,4,5 gibi] ölçümlenmiş maddelere rezerv etmişlerdir (Bahar, Nartgün, Durmuş & Bıçak, 2006; Boysan, 2008; Bulduk, 2003; Büyükoztürk, 2005; Erkuş, 2003; Kan, 2006; Öncü, 1994; Tan & Erdoğan, 2004; Tavşancıl, 2005; Yurdabakan, 2008); kimi araştırmacılar ya da psikometristler de, "iç tutarlılık katsayılarını hesaplarken" iki değerli [dichotomously] ölçümlenmiş maddelerde Kuder-Richardson 20, çok değerli ölçümlenmiş maddeler de ise, Cronbach'ın alfası formüllerinin kullanılması "gerektiğini" ileri sürmüşlerdir (Erkuş, 2000); ancak, *bunların hiç biri doğru değildir*.

Bademci (2001a; 2005a; 2006a; 2006b; 2007; 2010), bilimsel kanıtlarıyla ileri sürdüğü *yeni* paradigma doğrultusunda, Cronbach'ın alfa katsayısının, bilinenlerin aksine, 0'dan ve hatta -1'den *küçük* [örneğin, -7 ; Bademci, 2005a; 2006a; 2007] değerler alabileceğini ve de Cronbach'ın alfasının, yalnızca iki değerli ya da ağırlıklı [çok değerli] ölçümlenmiş maddelerle de kullanılabilceğini *matematiksel olarak* ortaya koymuştur. Kısaca ifade etmek gerekirse, Cronbach'ın alfa katsayısı, test maddelerinin *tüm* çeşitleriyle kullanılabilir (Bademci, 2001a; 2006c; 2007; Worthen, White, Fan & Sudweeks, 1999). Ayrıca Bademci (2001a; 2005a; 2006a; 2006b; 2007) *tüm maddeler iki değerli ölçümlenmiş* olduğu zaman $\alpha = KR-20$ olduğunu,

yani tüm maddeler iki değerli ölçümlenmiş olduğu zaman KR-20 ile α katsayısının aynı [özdeş] sonuçları verdiğini, hem belirtmiş, hem de örnekleri üzerinde göstermiştir. Bu çalışmada ise, tüm maddeler iki değerli [0,1] ölçümlendiği zaman, Kuder-Richardson 20, Cronbach'ın alfasi ve Hoyt'un varyans analizi formüllerinin aynı sonuçları verdiğinin, matematiksel olarak, ortaya konulmasına çalışılacaktır (ayrı olarak, bkz., Bademci, 2005a; 2006a; 2006b; 2006c).

5. MADDELER İKİ DEĞERLİ [0,1] ÖLÇÜMLENDİĞİNDE, CRONBACH'IN ALFASI VE KUDER-RICHARDSON 20 FORMÜLLERİ AYNI SONUÇLARI VERİR

Cronbach'ın alfasi, Kuder-Richardson 20 ve Hoyt'un varyans analizi, iç tutarlılık güvenirlik kestirimlerinde yaygın kullanılan yöntemlerdir (Barchard & Hakstian, 1997). Kuder-Richardson 20, sadece iki değerli ölçümlenmiş maddeler için uygunken; Cronbach'ın alfasi, hem iki değerli [0,1] ölçümlenmiş maddelerle, hem de ağırlıklı [1,2,3,4,5] ölçümlenmiş maddelerle kullanılabilir (Bademci, 2001a; 2006b; 2006c; 2007; 2010; Crocker & Algina, 1986).

Klasik Gerçek Ölçüm Kuramının ilk formüleştirmelerinde, bir *güvenirlik katsayısı*, paralel test formlarından elde edilen ölçümler [ya da paralel ölçmeler] arasındaki korelasyon *gibi* tanımlanmıştır; bu katsayı, matematiksel olarak, gerçek ölçüm varyansının gözlenmiş ölçüm varyansına oranı biçiminde gösterilebilmekte ve $\rho_{XX'}$ [X ve X' paralel ölçmeler] sembolüyle de belirtilebilmektedir; yeri gelmişken, yine yaygın bir hatalı kullanımın ve yanlış anlamının düzeltilmesinde fayda görülmektedir: Güvenirlik katsayısı, bir korelasyon katsayısının terimlerinde *değil*, varyansların bir oranının terimlerinde tanımlanmıştır ve bu nokta asla gözden kaçırılmamalıdır (Allen & Yen, 1979; Bademci, 2004; 2007; Crocker & Algina, 1986; Feldt & Brennan, 1989). Güvenirlik katsayısını $\rho_{XX'}$ şeklinde belirtebilmekle bağlantılı olarak, bu çalışmada, Cronbach'ın alfasi, Kuder-Richardson 20 ve Hoyt ölçüm güvenirlik katsayıları, sırasıyla, $\alpha\rho_{XX'}$, $KR-20\rho_{XX'}$ ve $HOYT\rho_{XX'}$ sembolleriyle gösterilmiştir. Metin içerisinde μ , σ^2 ve ρ sembolleri üzerine konmuş ($\hat{}$) [şapka ya da düzeltme] işareti ise, bu sayı ya da niceliklerin bir evrenden ziyade bir örneklem için hesaplandığını [ya da elde edildiğini] tarif etmektedir.

5.1. Test Yapma/Geliştirme ve Ölçüm Güvenirlik Çalışmaları ile İlgili Bazı Pratik Bilgiler

Tablo 1'de *denencel veriler* sunulmuştur; buna göre, 10 kişiden oluşan bir öğrenci grubuna [ya da örnekleme], 6 maddelik bir test uygulanmış ve maddeler doğru (1) ve yanlış (0) olarak, yani *iki değerli* ölçümlenmiştir. [Tablo 1'de açıklayıcı amaçlı bir örnek verilmektedir; test-tekrar test, almaşık

form, yarıya bölme, Cronbach'ın alfası, Kuder-Richardson 20 yöntemleri kullanılarak bir test ölçüm güvenilirliğinin hesaplandığı hallerde, bir başka ifadeyle *ölçüm güvenilirlik kestirimleri veya çalışmaları için örneklem büyüklüğü en az 400 kişiden oluşmalıdır -geçerlik çalışmalarında ise, 400'den daha fazlası kişi veya deneğe gereksinim olabileceği gözden kaçırılmamalıdır-*: Daha duyarlı kestirimler için, '*test yapma/geliştirme süreci*'nin gözden kaçırılmaması kaydıyla, [test veya ölçek ya da ölçme aracı] *madde sayısının en az 25, tercihan 30 ya da daha fazlası olmasına; madde güçlük değerlerinin .30 ile .70 arasında veya .40 ile .60 arasında [ve ortalaması .50 civarında] bulunmasına; madde ayırıcılık değerlerinin ise .30 ve daha üstünde, tercihan .40 ve daha üstünde olmasına; madde analiz çalışmaları ile ilgili olarak, [-gerekliyse- öndeneme için örneklem büyüklüğünün 100 kişiden ve] maddelerle ilgili daha doğru ve duyar bilgi elde etmek üzere yürütülmüş formal deneme için ise örneklemin en az 400 kişiden veya 400 ve daha fazlası kişiden oluşmasına; temel ya da genel araştırma amaçları için de ölçüm güvenilirlik katsayılarının en az 0.80 ya da 0.80 ve daha yukarısı olmasına dikkat edilmelidir (Allen & Yen, 1979; Bademci, 2001a; 2005b; 2006c; 2007; 2010; Charter, 1999; 2001; Crocker & Algina, 1986; Ebel & Frisbie, 1991; Ellis & Mead, 2002; Guilford, 1954; Henrysson, 1971; Henson, 2001; Hopkins & Antes, 1978; McCoach, 2002; Nunnally & Bernstein, 1994; Osterlind, 1989; Traub, 1994).]*

Tablo 1. Kuder-Richardson 20, Cronbach'ın alfası ve Hoyt'un varyans analizi güvenilirlik kestirim yöntemleri için denencel veriler
[Tablo, Crocker & Algina'dan (1986) uyarlanmıştır.]

Kişiler (i = 10)	Maddeler (j = 6)						Toplam Ölçüm
	1	2	3	4	5	6	
A	0	0	0	0	0	0	0
B	0	0	0	0	1	0	1
C	1	0	1	1	1	0	4
D	1	1	1	1	1	1	6
E	1	1	1	1	1	1	6
F	0	0	1	0	0	0	1
G	0	0	1	1	1	0	3
H	0	0	0	1	0	0	1
I	1	0	1	1	1	0	4
İ	0	1	0	1	0	1	3
p_j	0.40	0.30	0.60	0.70	0.60	0.30	
$p_j = 1 - p_j$	0.60	0.70	0.40	0.30	0.40	0.70	
$p_j q_j$	0.24	0.21	0.24	0.21	0.24	0.21	
$\hat{\sigma}_T^2$							4.09
$\Sigma X_{.j}$	4	3	6	7	6	3	

$${}_{KR-20}P_{XX} = k / (k-1) * [1 - (\sum p_j q_j / \sigma_T^2)]$$

$${}_{KR-20}\hat{P}_{XX} = (6 / 6-1) * [1 - (1.35 / 4.09)] = \mathbf{0.8039}$$

$$\alpha P_{XX} = k / (k-1) * [1 - (\sum \sigma_j^2 / \sigma_T^2)]$$

$\alpha P_{XX} = k / (k-1) * [1 - (\sum p_j q_j / \sigma_T^2)]$ [Tüm maddeler iki değerli ölçümlendiğinde]

$$\alpha \hat{P}_{XX} = (6 / 6-1) * [1 - (1.35 / 4.09)] = \mathbf{0.8039}$$

$\sum p_j q_j = j$ madde ölçüm varyanslarının toplamı (Bademci, 2007)

$$\sum p_j q_j = 0.24+0.21+0.24+0.21+0.24+0.21 = 1.35$$

$\mu_X =$ toplam ölçüm[lerin] aritmetik ortalaması

$$\mu_X = \sum X / N \quad (\text{Crocker \& Algina, 1986})$$

$$\hat{\mu}_X = (0+1+4+6+6+\dots+3) / 10 = 29 / 10 = 2.9$$

$\sigma_T^2 =$ toplam ölçüm[lerin] varyansı

$$\sigma_T^2 = \sum (X - \mu_X)^2 / N \quad (\text{Crocker \& Algina, 1986})$$

$$\hat{\sigma}_T^2 = [(0-2.9)^2 + (1-2.9)^2 + (4-2.9)^2 + \dots + (3-2.9)^2 / 10]$$

$$\hat{\sigma}_T^2 = 40.9 / 10 = 4.09$$

Cronbach'ın alfa katsayısı, aşağıdaki formül yoluyla hesaplanmaktadır (Bademci, 2006a; 2006b; 2007).

$$\alpha P_{XX} = k / (k-1) * [1 - (\sum \sigma_j^2 / \sigma_T^2)] \quad (\text{Formül 1})$$

$\alpha P_{XX} =$ ölçüm güvenilirlik katsayısı (Cronbach'ın alfası)

$k =$ test üzerindeki madde sayısı

$\sigma_j^2 = j$ madde ölçüm varyansı [ya da bir madde üzerindeki bir grup bireyden elde edilen ölçümlerin varyansı]

$\sum \sigma_j^2 = j$ madde ölçüm varyanslarının toplamı

$\sigma_T^2 =$ toplam test ölçümlerinin varyansı

Kuder-Richardson 20 ise, aşağıdaki formül yoluyla hesaplanmaktadır (Bademci, 2006b; 2007)

$${}_{KR-20}P_{XX} = k / (k-1) * [1 - (\sum p_j q_j / \sigma_T^2)] \quad (\text{Formül 2})$$

${}_{KR-20}P_{XX'}$ = ölçüm güvenirlik katsayısı (Kuder–Richardson 20)

k= test üzerindeki madde sayısı

p_j = j maddesini doğru cevaplayan kişilerin oranı

[Örneğin, 1. madde için, 4/10= 0.40]

q_j = j maddesini doğru *cevaplayamayan* kişilerin oranı

[Örneğin, 1. madde için, $q = 1-p = 1- 0.40 = 0.60$]

p_jq_j = j madde ölçüm varyansı [ya da bir madde

üzerindeki bir grup bireyden elde edilen

ölçümlerin varyansı]

[Örneğin, 1. madde için, $0.40 \times 0.60 = 0.24$]

Σp_jq_j = j madde ölçüm varyanslarının toplamı

σ_T^2 = toplam test ölçümlerinin varyansı

Formül 1 ile Formül 2 alt alta tekrar yazıldığında, madde ölçüm varyansının türetimi hariç, Kuder-Richardson 20 formülü ile Cronbach'ın alfa katsayısı formülünün aynı olduğu görülebilecektir.

$$\alpha P_{XX'} = k / (k-1) * [1 - (\Sigma \sigma_j^2 / \sigma_T^2)]$$

$${}_{KR-20}P_{XX'} = k / (k-1) * [1 - (\Sigma p_jq_j / \sigma_T^2)]$$

Tüm maddeler *iki değerli* [0,1] olduğu zaman, σ_j^2 matematiksel olarak p_jq_j 'ya eşittir (Suen, 1990), yani $\sigma_j^2 = p_jq_j$ (Bademci, 2007).

Buradan, tüm maddeler *iki değerli* [0,1] ölçülenmiş olduğunda, $\Sigma \sigma_j^2 = \Sigma p_jq_j$ 'da yazılabilir (Worthen, White, Fan & Sudweeks, 1999; Bademci, 2007).

Böylelikle, tüm maddeler *iki değerli* [0,1] ölçülenmiş olduğu zaman, Cronbach'ın alfa katsayısı,

$$\alpha P_{XX'} = k / (k-1) * [1 - (\Sigma p_jq_j / \sigma_T^2)] \quad (\text{Formül 3})$$

şeklinde de yeniden yazılabilir (Bademci, 2007).

Tüm maddeler *iki değerli* [0,1] ölçülenmiş olduğunda, Formül 2 ile Formül 3'ün *eşit* olduğu, bu iki formül alt alta yazıldığında daha da açık görülebilecektir (Bademci, 2006b; 2007).

$${}_{KR-20}P_{XX'} = k / (k-1) * [1 - (\Sigma p_jq_j / \sigma_T^2)]$$

$$\alpha P_{XX'} = k / (k-1) * [1 - (\Sigma p_jq_j / \sigma_T^2)]$$

Formül 2 ve Formül 3'den görüleceği üzere, *tüm maddeler iki değerli* [0,1] ölçümlenmiş olduğu zaman $\alpha \rho_{XX'} =_{KR-20} \rho_{XX'}$ 'dir. Dolayısıyla, Tablo 1'deki veriler göz önünde tutulduğunda da, $\alpha \hat{\rho}_{XX'} =_{KR-20} \hat{\rho}_{XX'} = 0.8039$.

6. HOYT'UN ÖLÇÜM GÜVENİRLİĞİ KESTİRME YÖNTEMİ VE GENELLENİRLİK KURAMI

Hoyt'un (1941) ölçüm güvenirliliği kestirim yöntemi, varyans analizi (ANOVA; analysis of variance) üzerine temellenmiştir. Genellenirlik Kuramı (Generalizability Theory; Cronbach, Gleser, Nanda & Rajaratnam, 1972) ise, güvenirlilik katsayılarının kestiriminde [Klasik Gerçek Ölçüm Kuramı ve] varyans analizinin uygulamaları üzerine inşa edilmiştir (Thompson, 2003). Varyans analizinin güvenirlilik kestiriminde kullanılmasına ilişkin Türkiye'deki *ilk uygulamalar* da, -bundan 23 yıl önce- yine Bademci'nin (1988; 1994) çalışmalarında görülmektedir. Bademci'nin (1988; 1991; 1994; 2000; 2001a, 2004; 2005a; 2007; 2010; 2011) ölçme ve araştırma yöntem biliminde ortaya koyduğu *paradigma değişikliğinin* düşünsel arka planında da, 1990'ların ortalarından itibaren üzerlerinde yoğun çalışmalar sürdürdüğü Genellenirlik Kuramının ve varyans analiziyle güvenirlilik kestirim yöntemi ve ilişkili uygulamalarının izlerini ve etkilerini bulmak ve görmek mümkündür.

6.1. Klasik Gerçek Ölçüm Kuramı, Hoyt'un Ölçüm Güvenirliliği Kestirme Yöntemi ve Genellenirlik Kuramı

Klasik Gerçek Ölçüm Kuramında, [gözlenmiş] ölçüm varyansı, gerçek ölçüm varyansı ve [random] hata varyansı olarak iki parçaya bölünmüştür (Kieffer, 1999; Shavelson, Webb & Rowley, 1989); bu bölünmenin sonuçları, zarif ve doyurucudur (Shavelson, Webb & Rowley, 1989). Fakat, Hoyt (1941), varyans analizini kullanarak, ölçüm varyansının bireyler arasında varyans, maddeler arasında varyans ve random hata varyansı şeklinde üç bileşene bölünmesine olanak sağlayan bir yöntem geliştirdi: Bu, ölçüm varyansını sadece gerçek ölçüm varyansı ve [random] hata varyansı olarak iki parçaya bölen Klasik Gerçek Ölçüm Kuramından kökten bir ayrılıştı; Genellenirlik Kuramının (Brennan, 1992) gelişmesine, Hoyt'un (1941) güvenirlilik kestirme yönteminin önemli katkısı da, ölçüm varyansının, varyans analizi kullanılarak bileşenlerine ayrılması olmuştur (Kieffer, 1999).

7. MADDELER İKİ DEĞERLİ [0,1] ÖLÇÜMLENDİĞİNDE, KUDER-RİCHARDSON 20, CRONBACH'IN ALFASI VE HOYT'UN VARYANS ANALİZİ FORMÜLLERİ AYNI SONUÇLARI VERİR

Tablo 1'deki denencel madde yanıtlarına varyans analizi (geniş bilgi için bakınız, Hicks, 1973) uygulanmış ve bu verilerden elde edilen varyans analizi özet tablosu da, Tablo 2'de verilmiştir. Burada kullanılan varyans analizi deseni, genellikle, *tekrarlanmış ölçüler* (repeated-measures) veya

tekrarlanmış ölçmeler (repeated measurements) deseni olarak adlandırılmaktadır (Ferguson & Takane, 1989; Glass & Hopkins, 1996; Keppel & Wickens, 2004; Winer, 1971).

Tablo 2. Tablo 1'deki denencel veriler için varyans analizi sonuçları

Kaynak	Kareler Toplamı (SS)	Serbestlik Derecesi (df)	Kareler Ortalaması (MS)
Kişiler (i)	6.8166	9	0.7574
Maddeler (j)	1.4833	5	0.2966
Kalan	6.6834	45	0.1485
Toplam	14.9833	59	

Hoyt'un varyans analiziyle ölçüm güvenilirliği kestirim yöntemi için gereksinim duyulan hesaplamalar [hesaplama formülleri ile birlikte], Tablo 1'deki denencel verilerle ilgili olarak aşağıda, adım adım ve kolaylaştırılmış şekilde verilmeye çalışılmıştır (Nunnally & Bernstein, 1994).

$$1. \text{ Toplam kareler toplamı } [\sum X_{ij}^2 - (\sum X_{ij})^2 / N]$$

$$= [(0^2 + 0^2 + 1^2 + 1^2 + \dots + 0^2 + 0^2 + 1^2) - (29^2 / 60)] = [29 - (29^2 / 60)]$$

$$= [29 - (841 / 60)]$$

$$= \mathbf{14.9833}$$

$$2. \text{ Kişiler ile ilgili kareler toplamı } [\sum X_i^2 / j - (\sum X_{ij})^2 / N]$$

$$= [(0^2 + 1^2 + 4^2 + 6^2 + 6^2 + \dots + 3^2) / 6 - (29^2 / 60)] = [(125 / 6) - (841 / 60)]$$

$$= \mathbf{6.8166}$$

$$3. \text{ Maddeler ile ilgili kareler toplamı } [\sum X_j^2 / i - (\sum X_{ij})^2 / N]$$

$$= [(4^2 + 3^2 + 6^2 + 7^2 + 6^2 + 3^2) / 10 - (29^2 / 60)] = [(155 / 10) - (841 / 60)]$$

$$= \mathbf{1.4833}$$

$$4. \text{ Kalan kareler toplamı } [\sum X_{ij}^2 - (\sum X_{ij})^2 / N] -$$

$$[\sum X_i^2 / j - (\sum X_{ij})^2 / N] - [\sum X_j^2 / i - (\sum X_{ij})^2 / N] = [(\text{Toplam kareler}$$

$$\text{toplamı}) - (\text{Kişiler ile ilgili kareler toplamı}) - (\text{Maddeler ile ilgili kareler toplamı})] = [(14.9833) - (6.8166) - (1.4833)] = \mathbf{6.6834}$$

5. Kişiler ile ilgili serbestlik derecesi ($i - 1$) = $(10 - 1) = 9$; maddeler ile ilgili serbestlik derecesi ($j - 1$) = $(6 - 1) = 5$; kalan ile ilgili serbestlik derecesi $[(i - 1) * (j - 1)] = [9 * 5] = 45$; toplam [ile ilgili] serbestlik derecesi ($ij - 1$) = $[(10 * 6) - 1] = 59$.

6. Kişiler ile ilgili kareler ortalaması ($MS_{kişi}$)
= Kişiler ile ilgili kareler toplamı / Kişiler ile ilgili serbestlik derecesi
= **6.8166 / 9 = 0.7574**

7. Maddeler ile ilgili kareler ortalaması (MS_{madde})
= Maddeler ile ilgili kareler toplamı / Maddeler ile ilgili serbestlik derecesi
= **1.4833 / 5 = 0.2966**

8. Kalan kareler ortalaması (MS_{kalan})
= Kalan kareler toplamı / Kalan ile ilgili serbestlik derecesi = **6.6834 / 45**
= **0.1485**

Varyans analizinin standart simgelemlerini kullanarak Hoyt (1941) güvenilirlik kestirimini aşağıdaki şekilde [Formül 4] tanımlamış veya betimlemiştir (Crocker & Algina, 1986; Kieffer, 1999; Genellenirlik Kuramı ile bağlantılı olarak, Cronbach, Rajaratnam & Gleser, 1963).

$$HOYT \rho_{XX'} = (MS_{kişi} - MS_{kalan}) / MS_{kişi} \quad (\text{Formül 4})$$

$HOYT \rho_{XX'}$ = ölçüm güvenilirlik katsayısı (Hoyt)

$MS_{kişi}$ = Kişiler ile ilgili kareler ortalaması

MS_{kalan} = Kalan kareler ortalaması

$$HOYT \hat{\rho}_{XX'} = (0.7574 - 0.1485) / 0.7574$$

$$HOYT \hat{\rho}_{XX'} = \mathbf{0.8039}$$

Formül 1, *cebirsel olarak* Formül 4'e eşittir (Stanley, 1971; Feldt & Brennan, 1989). Bir başka ifadeyle, Cronbach'ın alfası ile Hoyt'un varyans analizi ölçüm güvenirliliği hesaplama formülleri *matematiksel olarak* özdeştir [aynıdır]. Formül 1 ile Formül 4,

$$HOYT \rho_{XX'} = (MS_{kişi} - MS_{kalan}) / MS_{kişi}$$

$$= k/(k-1) * [1 - (\sum \sigma_j^2 / \sigma_T^2)] = \alpha_{XX'}$$

şeklinde gösterilebilir. Buradan, kısaca $HOYT \rho_{XX'} = \alpha_{XX'}$ yazılabilir. Tablo 1'deki *iki değerli* ölçümlenmiş denencel madde yanıtları üzerinde yapılan

hesaplamalardan hem $\alpha \hat{\rho}_{XX'}$, hem de $_{HOYT} \hat{\rho}_{XX'}$ için, aynı sayısal değerler, yani **0.8039** elde edildiği görülmektedir. Bir başka ifadeyle, $_{HOYT} \hat{\rho}_{XX'} = \alpha \hat{\rho}_{XX'} = \mathbf{0.8039}$.

Hoyt (1941: 153-160), orijinal makalesinde, kendi formülü [Formül 4] ile Kuder-Richardson 20 formülünden [Formül 2] elde edilen sonuçların özdeş (aynı) olduğunu belirtmiştir. [*Burada, maddelerin iki değerli ölçümlenmiş olduğu unutulmamalıdır*]. Bir başka ifadeyle, maddeler *iki değerli* ölçümlendiğinde, Kuder-Richardson 20 ile Hoyt'un varyans analizi ölçüm güvenilirliği hesaplama formülleri *matematiksel olarak* [birbirlerine] eşittir; Formül 2 ve Formül 4'den yararlanılarak,

$$_{HOYT} \rho_{XX'} = (MS_{kişi} - MS_{kalan}) / MS_{kişi}$$

$$= k / (k-1) * [1 - (\sum p_j q_j / \sigma_T^2)] =_{KR-20} \rho_{XX'}$$

şekli ifade edilebilir: Buradan, kısaca $_{HOYT} \rho_{XX'} =_{KR-20} \rho_{XX'}$ biçimi belirtilebilir. Tablo 1'de sunulan *iki değerli* ölçümlenmiş denencel madde yanıtları üzerinde yapılan hesaplamalardan görüleceği üzere, hem $_{KR-20} \hat{\rho}_{XX'}$, hem de $_{HOYT} \hat{\rho}_{XX'}$ için *aynı sonuçlar*, yani **0.8039** elde edilebilmektedir: Bir başka ifadeyle, $_{HOYT} \hat{\rho}_{XX'} =_{KR-20} \hat{\rho}_{XX'} = \mathbf{0.8039}$.

8. SONUÇ

Maddeler *iki değerli* [0,1] ölçümlendiğinde, Kuder-Richardson 20, Cronbach'ın alfası ve Hoyt'un varyans analizi formülleri *aynı* sonuçları vermektedir; bu durum, Tablo 1'deki *iki değerli* ölçümlenmiş veriler üzerinde yapılan kestirimlerden de görülmektedir, yani $_{KR-20} \hat{\rho}_{XX'} = \alpha \hat{\rho}_{XX'} =_{HOYT} \hat{\rho}_{XX'} = \mathbf{0.8039}$ 'dur.

* Metin içindeki [...] arasındaki ifadeler yazar tarafından eklenmiştir.

KAYNAKLAR

- Akdeniz, C., Aydemir,Ö., Akdeniz, F., Gülseren,Ş. & Kültür, S. (1999). Sağlık Düzeyi Ölçeği'nin Türkçe'ye Uyarlanması ve Güvenilirliği. *Klinik Psikofarmakoloji Bülteni*, Cilt 9 (2), 104-108.
- Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, California: Brooks/Cole.
- Aygin, D. & Eti Aslan, F. (2005). Kadın Cinsel İşlev Ölçeği'nin Türkçeye Uyarlanması. *Türkiye Klinikleri Tıp Bilimleri Dergisi*, Cilt 25 (3), 393-399.
- Bademci, V. (2011). Türk Eğitim ve Biliminde Bilimsel Devrim: Testler ya da Ölçme Araçları Güvenilir ve Geçerli Değildir. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, Sayı16, 116-132. (http://www.zgefdergi.com/Makaleler/713177846_16_09_Bademci.pdf)

- Bademci, V. (2010). *Türk Eğitim ve Biliminde Paradigma Değişikliği: Testler veya Ölçekler Güvenilir ve Geçerli Değildir*. Konferans. Düzenleyen: Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi Dekanlığı. Ankara: G.Ü. Gazi Eğitim Fakültesi, Resim-İş Eğitimi Anabilim Dalı Konferans Salonu, 26 Nisan. [Konferansla ilgili haber için; *Gazi Haber*, Nisan 2010, Sayı 104, Sayfa 48-49.]
- Bademci, V. (2008). Araştırmalarda Ölçme ile İlgili Bazı Büyük Hataları Düzeltmek ve Eğitimde Yeniden Yapılanmayı Sürdürmek: Güvenirlik, Testlerin Bir Özelliği Değildir. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 22, 50-69.
(http://www.esef.gazi.edu.tr/html/yayinlar/22_pdf/22_5.pdf)
- Bademci, V. (2007). *Ölçme ve Araştırma Yöntemlerinde Paradigma Değişikliği: Testler Güvenilir Değildir*. Ankara: Yenyap Yayınları.
- Bademci, V. (2006a). Güvenirliği Doğru Anlamak ve Bazı Klişeleri Yıkamak: Bilinenlerin Aksine, Cronbach'ın Alfa Katsayısı, Negatif ve -1'den Küçük Olabilir. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, Cilt 7, Sayı 12, 3-26.
(<http://web.inonu.edu.tr/~efdergi/arsiv/bademci.pdf>)
- Bademci, V. (2006b). Tartışmayı Sonlandırmak: Cronbach'ın Alfa Katsayısı, İki Değerli [0,1] Ölçümlenmiş Maddeler ile Kullanılabilir. *Kazım Karabekir Eğitim Fakültesi Dergisi*, Sayı 13, 438-446.
(<http://e-dergi.atauni.edu.tr/index.php/kkefd/article/viewFile/4116/3940>)
- Bademci, V. (2006c). *Paradigma Değişikliği: Testler Güvenilir Değildir*. Konferans. Düzenleyen: Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi Dekanlığı. Ankara: G.Ü. Mesleki Eğitim Fakültesi Konferans Salonu, 28 Nisan. [Konferansla ilgili haber için; *Gazi Haber*, Nisan 2006, Sayı 66, Sayfa 64.]
- Bademci, V. (2005a). *Araştırmalarda Ölçme ile İlgili Bazı Büyük Hataları Düzeltmek ve Bir Reformu Başlatmak: Güvenirlik, Testlerin Bir Özelliği Değildir*. Eğitim Fakültelerinde Yeniden Yapılandırmanın Sonuçları ve Öğretmen Yetiştirme Sempozyumu. Ankara: Gazi Üniversitesi, Gazi Eğitim Fakültesi, 22-23-24 Eylül.
- Bademci, V. (2005b). Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 17, 33-45.
(http://www.esef.gazi.edu.tr/html/yayinlar/17_pdf/17_c.pdf)
- Bademci, V. (2005c). Hakemlerin Değerlendirmelerindeki Hatalar Üzerine: Fisher'in Z Dönüşümü ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 17, s. 46-75.
(http://www.esef.gazi.edu.tr/html/yayinlar/17_pdf/17_d.pdf)
- Bademci, V. (2004). Testin Güvenirliği" veya "Test Güvenilirdir" Diye İfade Etmek Doğru Değildir. *Türk Eğitim Bilimleri Dergisi*, Cilt 2, 367-373.
(<http://www.tebd.gazi.edu.tr/c2s3.html>)
(http://www.tebd.gazi.edu.tr/arsiv/2004_cilt2/sayi_3/367-373.pdf)
- Bademci, V. (2002). *Türkiye'deki Okullar Ne İşe Yarar? Türkiye'nin Anomi, Yabancılaşma, Ekonomik Büyüme, Demokratikleşme Sorunlarına Çözüm Önerisi*. Konferans. Düzenleyen: ESEF Öğrenci Bilimsel Faal. Org. Kom. Ankara: G.Ü. Mesleki Eğitim Fakültesi Konferans Salonu, 30 Mayıs 2002.
- Bademci, V. (2001a). *Düşünmenin Öğretilmesi ve Öğretimde Kullanılan Yöntemler-Teknikler*. Konferans. Düzenleyen: TÜRMOB. Bursa: Bursa SMMM Odası Konferans Salonu, 9 Kasım 2001.
- Bademci, V. (2001b). *Türkiye'deki Okullar Ne İşe Yarar?* Konferans. Düzenleyen: Ankara Türk Telekom Anadolu Teknik L. Ankara: Başkent Öğretmenevi Konferans Salonu, 9 Aralık 2001.
- Bademci, V. (2000). *Türkiye'deki Okullar Ne İşe Yarar? Türkiye'nin Anomi, Yabancılaşma, Ekonomik Büyüme, Demokratikleşme Sorunlarına Çözüm Önerisi: Toplam Kalite Yönetimi Temelli Bir Eğitimde Yeniden Yapılanma Stratejisi*. (Birinci Basım). Ankara: Başkent Basım Yayın.

- Bademci, V. (1999). *Hedefin Davranışlara Çevrilmesi, Davranışlardan Seçmeli Test Maddeleri Yazılması*. (Geliştirilmiş Üçüncü Baskı). Ankara: Gazi Kitabevi.
- Bademci, V. (1994). "Ürün ve Süreç Değerlendirmesi Yapılan Öğrencilerin Erişi Düzeyleri ile İlgili Bir Araştırma". Yayımlanmamış Doktora Tezi. Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.
- Bademci, V. (1991). *Varyans Analiziyle Güvenirlilik Hesaplanması*. Ankara.
- Bademci, V. (1988). "Ağaçışleri Atelyesine Gelen Lise ve Meslek Lisesi Çıkışlı Öğrencilerin Psiko-Motor Giriş Davranışları ve Psiko-Motor Erişi Düzeyleri Arasındaki Farklılık". Yayımlanmamış Yüksek Lisans Tezi. Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.
- Bahar, M., Nartgün, Z., Durmuş, S. & Bıçak, B. (2006). Bahar, M. (Ed.), *Geleneksel-Alternatif Ölçme ve Değerlendirme Teknikleri Öğretmen El Kitabı*. Ankara: Pegem A.
- Barchard, K. A. & Hakstian, A. R. (1997). The Effects of Sampling Model on Inference with Coefficient Alpha. *Educational and Psychological Measurement*, Vol.57, 893-905.
- Barnes, L. L. B., Harp, D. & Jung, W. S. (2002). Reliability Generalization of Scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement*, Vol. 62, 603-618.
- Beycioğlu, K. (2007). Alfa Güvenirliliği ve Eğitim Araştırmaları. *Çağdaş Eğitim*, 347, 37-42.
- Boysan, M. (2008). *Ölçme ve Değerlendirme. KPSS El Kitabı*. Ankara: Yargı.
- Baugh, F. (2002). Correcting Effect Sizes for Score Reliability: A Reminder That Measurement and Substantive Issues are Linked Inextricably. *Educational and Psychological Measurement*, Vol. 62, 254-263.
- Brennan, R. L. (1992). *Elements of Generalizability Theory*. (Revised Edition). Iowa City, Iowa: American College Testing.
- Brookhart, S. M. & Nitko, A. J. (2008). *Assessment and Grading in Classrooms*. Upper Saddle River, New Jersey: Pearson.
- Buhi, E. R. (2005). Reliability Reporting Practices in Rape Myth Research. *Journal of School Health*, Vol. 75, 63-66.
- Bulduk, S. (2003). *Yeni Başlayanlar İçin Deneysel Psikolojide Araştırma Yöntemleri*. İstanbul: Çantay Kitabevi.
- Buluş Kırıkkaya, E., Bozkurt, E., İşeri, Ş., Vurkaya, G. & Bali, G. (2011). Tubitak Supported Science Summer School for Primary School Students: Happiness of Learning by Exploring and Enjoying. *Procedia Social and Behavioral Sciences*, 15, 2219-2227.
- Büyüköztürk, Ş. (2005). *Sosyal Bilimler İçin Veri Analizi El Kitabı*. (Gözden Geçirilmiş 5. Baskı). Ankara: PegemA.
- Capraro, R. M. & Capraro, M. M. (2002). Myers-Briggs Type Indicator Score Reliability Across Studies: A Meta-Analytic Reliability Generalization Study. *Educational and Psychological Measurement*, Vol.62, 590-602.
- Cebeci, S. (2006). "The Examination of Guidance and Research Centers' Administrators' Conflict Management Strategies with the Perceptions of Self and Teachers". Unpublished Master's Thesis. Ankara: Middle East Technical University, The Graduate School of Social Sciences.
- Charter, R.A. (2001). Damn the Precision, Full Speed Ahead with the Clinical Interpretation. *Journal of Clinical and Experimental Neuropsychology*, Vol. 23, 692-694.
- Charter, R.A. (1999). Sample Size Requirements for Precise Estimates of Reliability, Generalizability, and Validity Coefficients. *Journal of Clinical and Experimental Neuropsychology*, Vol. 21, 559-566.
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Fort Worth: Holt, Rinehart and Winston.
- Cronbach, L. J. (1988). Five Perspectives on the Validity Argument. In Wainer, H. & Braun, H. I. (Eds.), *Test Validity*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Cronbach, L. J. (1972). Validation of Educational Measures. In Bracht, G. H., Hopkins, K. D. & Stanley, J. C. (Eds.), *Perspectives in Educational and Psychological Measurement*. Englewood Cliffs, New Jersey: Prentice-Hall.

- Cronbach, L. J. (1971). Test Validation. In Thorndike, R. L. (Ed.), *Educational Measurement*. (Second Edition). Washington, D. C.: American Council on Education.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, Vol. 16, 297-334.
- Cronbach, L. J., Rajaratnam, N. & Gleser, G. C. (1963). Theory of Generalizability: A Liberalization of Reliability Theory. *The British Journal of Statistical Psychology*, Vol. 16, 137-163.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley and Sons.
- Demir, H. & Okan, T. (2009). Motivasyon Üzerinde Ulusal Kültür Etkisi. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 11(1), 121-142.
- Ebel, R. E. (1965). *Measuring Educational Achievement*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of Educational Measurement*. (Fifth Edition). Englewood Cliffs, New Jersey: Prentice Hall.
- Ellis, B. E. & Mead, A. D. (2002). Item Analysis: Theory and Practice Using Classical and Modern Test Theory. In Rogelberg, S. G. (Ed.), *Handbook of Research Methods in Industrial and Organizational Psychology*. Madlen, Massachusetts: Blackwell.
- Erkuş, A. (2003). *Psikometri Üzerine Yazılar*. (1. Basım). Ankara: Türk Psikologlar Derneği Yayınları, 24.
- Erkuş, A. (2000). Sık Kullanılan Bazı Psikolojik Ölçeklerin Güvenirliklerinin İrdelenmesi. *Türk Psikoloji Yazıları*, Cilt 3 (6), 3-17.
- Feldt, L. S. & Brennan, R. L. (1989). Reliability. In Linn, R. L. (Ed.), *Educational Measurement*. (Third Edition). New York: American Council on Education ve Macmillan.
- Ferguson, G. A. & Takane, Y. (1989). *Statistical Analysis in Psychology and Education*. (Sixth Edition). New York: McGraw-Hill.
- Gardner, H. (1999). Who Owns Intelligence? *Atlantic Monthly*. Vol. 283 (2), 67-76.
- Gazi Haber (2010). Türk Eğitim ve Biliminde Paradigma Değişikliği: Testler veya Ölçekler Güvenilir ve Geçerli Değildir. Nisan 2010, Sayı 104, 48-49.
- Glass, G. V. & Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology*. (Third Edition). Boston: Allyn and Bacon.
- Gronlund, N. E. (1965). *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and Evaluation in Teaching*. (Sixth Edition). New York: Macmillan.
- Guilford, J. P. (1954). *Psychometric Methods*. (Second Edition). New York: McGraw-Hill.
- Guthrie, A. C. (2000). *A Review of Coefficient Alpha and Some Basic Tenets of Classical Measurement Theory*. (ERIC Document Reproduction Service No. ED 438 307).
- Güneş, B. (2003). Paradigma Kavramı Işığında Bilimsel Devrimlerin Yapısı ve Bilim Savaşları: Cephelelerdeki Fizikçilerden Thomas S. Kuhn ve Alan D. Sokal. *Türk Eğitim Bilimleri Dergisi*, Cilt 1(1), 23-42.
- Henrysson, S. (1971). Gathering, Analyzing, and Using Data on Test Items. In Thorndike, R. L. (Ed.), *Educational Measurement*. (Second Edition). Washington, D.C.: American Council on Education.
- Henson, R. K. (2001). Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha. *Measurement and Evaluation in Counseling and Development*, Vol. 34, 177-189.
- Henson, R. K., Kogan, L. R. & Vacha-Haase, T. (2001). A Reliability Generalization Study of the Teacher Efficacy Scale and Related Instruments. *Educational and Psychological Measurement*, Vol. 61, 404-420.
- Hicks, C. R. (1973). *Fundamental Concepts in the Design of Experiments*. (Second Edition). New York: Holt, Rinehart and Winston.

- Hogan, T. P., Benjamin, A. & Brezinski, K. L. (2000). Reliability Methods: A Note on the Frequency of Use of Various Types. *Educational and Psychological Measurement*, Vol. 60, 523-531.
- Hopkins, C. D. & Antes, R. L. (1978). *Classroom Measurement and Evaluation*. Itasca, Illinois: F. E. Peacock.
- Hotaman, D. & Yüksel-Şahin, F. (2010). The Effect of Instructors' Enthusiasm on University Students' Level of Achievement. *Education and Science [Eğitim ve Bilim]*, Vol 35(155), 89-103.
- Hoyt, C. (1941). Test Reliability Estimated by Analysis of Variance. *Psychometrika*, Vol. 6, 153-160.
- Jones, L. V. (1971). The Nature of Measurement. In Thorndike, R. L. (Ed.), *Educational Measurement*. (Second Edition). Washington, D.C.: American Council on Education.
- Kan, A. (2006). Ölçme Araçlarında Bulunması Gereken Nitelikler. Atılğan, H. (Ed.), *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı.
- Kane, M. T. (1992). An Argument-Based Approach to Validity. *Psychological Bulletin*, Vol. 112(3), 527-535.
- Kane, M. T. (1990). *An Argument-based Approach to Validation*. ACT Research Report Series, 90-13. Iowa City, Iowa: ACT.
- Kartal, H. (2009). Öğretmen Adaylarının Uygulama Okullarındaki Zorbalıkla İlgili Değerlendirmeleri. *GÜ, Gazi Eğitim Fakültesi Dergisi*, Cilt 29(1), 141-172.
- Kartal, E. & Pekkanlı, İ. (2011). Yabancı Dil Öğretmen Adaylarının Anadil ve Yabancı Dilde İnternet Üzerinden Okuma Alanları ve Sıklıkları. *International Journal of Human Sciences*, Vol. 8(1), 1316-1326.
- Keppel, G. & Wickens, T. D. (2004). *Design and Analysis. A Researcher's Handbook*. (Fourth Edition). Upper Saddle River, New Jersey: Pearson.
- Kieffer, K. M. (1999). Why Generalizability Theory is Essential and Classical Test Theory is Often Inadequate. In Thompson, B. (Ed.), *Advances in Social Science Methodology*, Volume 5. Stamford, Connecticut: JAI
- Korkmaz, A. (2010). "Vahit Bademci'nin Paradigma Değişikliği Üzerine Bir Araştırma: "Testler Değil, Ölçümler Güvenilirdir" ". Yayımlanmamış Yüksek Lisans Tezi. Zonguldak: Zonguldak Karaelmas Üniversitesi, Sosyal Bilimler Enstitüsü.
- Kubiszyn, T. & Borich, G. (2007). *Educational Testing and Measurement: Classroom Application and Practice*. (Eighth Edition). USA: John Wiley and Sons.
- Kuder, G. F. & Richardson, M. W. (1937). The Theory of the Estimation of Test Reliability. *Psychometrika*, Vol. 2, 151-160.
- Kuhn, T. S. (1995). *Bilimsel Devrimlerin Yapısı*. (Çev.: Kuyaş, N.). (Dördüncü Baskı). İstanbul: Alan Yayıncılık.
- Li, H. & Wainer, H. (1998). *Toward a Coherent View of Reliability in Test Theory*. Research Report. Princeton, New Jersey: Educational Testing Service.
- Linn, R. L. & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. (Eighth Edition). Upper Saddle River, New Jersey: Merrill.
- Livingston, S. A. (1988). Reliability of Test Results. In Keeves, J. P. (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon.
- McCoach, D. B. (2002). A Validation Study of the School Attitude Assessment Survey. *Measurement and Evaluation in Counseling and Development*, Vol. 35, 66-77.
- McMillan, J. H. (2001). *Classroom Assessment. Principles and Practice for Effective Instruction*. (Second Edition). Boston: Allyn and Bacon.
- Mehrens, W. A. & Lehmann I. J. (1991). *Measurement and Evaluation in Education and Psychology*. (Fourth Edition). Fort Worth: Harcourt Brace College.
- Messick, S. (1995). Validity of Psychological Assessment. Validation of Inferences From Person's Responses and Performances as Scientific Inquiry into Score Meaning. *American Psychologist*, Vol. 50, 741-749.
- Messick, S. (1989). Validity. In Linn, R. L. (Ed.), *Educational Measurement*. (Third Edition). New York: American Council on Education & Macmillan.

- Mji, A. & Onwuegbuzie, A. J. (2004). Evidence of Score Reliability and Validity of the Statistical Anxiety Rating Scale Among Technikon Students in South Africa. *Measurement and Evaluation in Counseling and Development*, Vol. 36, 238-251.
- Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological Testing. Principles and Applications*. (Fifth Edition). Upper Saddle River, New Jersey: Prentice Hall.
- Nitko, A. J. (2001). *Educational Assessment of Students*. (Third Edition). Upper Saddle River, New Jersey: Merrill/Prentice-Hall.
- Norton, D. (2001). Giriş. "Yerleşik Düşünceler: Verip Veriştirmek". (Hazırlayan: Bouvet, J-F.) (Çev.: Atuk, E.) *Ispanaktaki Demir ve Diğer Yerleşik Düşünceler Üzerine*. İstanbul: YKY.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory*. (Third Edition). New York: McGraw-Hill.
- Oğuzkan, F. (1981). *Eğitim Terimleri Sözlüğü*. (Gözden Geçirilmiş ve Genişletilmiş İkinci Baskı). Ankara: Türk Dil Kurumu Yayınları.
- Osterlind, S. J. (1989). *Constructing Test Items*. Boston: Kluwer.
- Öncü, H. (1994). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Matser Basım.
- Özkan, S. & Sevil, Ü. (2007). Doğum Sonrası Fonksiyonel Durum Envanterinin Geçerlilik Güvenilirlik Çalışması. *TSK Koruyucu Hekimlik Bülteni*, Cilt 6(3), 199-208.
- Reynolds, C. R., Livingston, R. B. & Willson, V. (2009). *Measurement and Assessment in Education*. (Second Edition). Upper Saddle River, New Jersey: Pearson.
- Sayın, S. (2010). Bilimsel Araştırmalarda Yapılan İstatistiksel ve Yöntembilimsel Hatalar-II: Grafik, Tablo ve Gösterim Hataları. *Türk Eğitim Bilimleri Dergisi*, Cilt 8(1), 117-143.
- Sayın, S. (2008). Bilimsel Araştırmalarda Yapılan Bazı İstatistiksel ve Yöntembilimsel Hatalar-III: Güvenirlik Kestirimlerine Yönelik Hatalar. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, Sayı 15, 53-69.
- Serdar, Z. (2001). *Thomas Kuhn ve Bilim Savaşları*. (Çev.: Kılıç, E.). İstanbul: Everest.
- Sever, E. (2008). "Öğrenme Stilleri: İlköğretim 6-8. Sınıf Öğrencilerine Yönelik Bir Ölçek Geliştirme Çalışması". Yayımlanmamış Yüksek Lisans Tezi. Aydın: Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü.
- Shavelson, R. J., Webb, N. M. & Rowley, G. L. (1989). Generalizability Theory. *American Psychologist*, Vol. 44, 922-932.
- Shrout, P. E. (1998). Measurement Reliability and Agreement in Psychiatry. *Statistical Methods in Medical Research*, Vol. 7, 301-317.
- Spearman, C. (1913). Correlations of Sums or Differences. *British Journal of Psychology*, Vol. 5, 417-426.
- Spearman, C. (1910). Correlation Calculated from Faulty Data. *British Journal of Psychology*, Vol. 3, 271-295.
- Spearman, C. (1907). Demonstration of Formulae for True Measurement of Correlation. *The American Journal of Psychology*, Vol 15(2), 161-169.
- Spearman, C. (1904b). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, Vol 18(2), 201-292.
- Spearman, C. (1904a). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, Vol 15(1), 72-101.
- Spector, P. E. (1992). *Summated Rating Scale Construction. An Introduction*. Newbury Park: Sage.
- Stanley, J. C. (1971). Reliability. In Thorndike, R. L. (Ed.), *Educational Measurement*. (Second Edition). Washington, D.C.: American Council on Education.
- Strube, M. J. (2000). Reliability and Generalizability Theory. In Grimm, L. G. & Yarnold, P. R. (Eds.), *Reading and Understanding More Multivariate Statistics*. Washington, DC: American Psychological Association.
- Suen, H. K. (1990). *Principles of Test Theories*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Suen, H. K. & Lei, P.-W. (2007). Classical versus Generalizability Theory of Measurement. *Educational Measurement*, 4. <http://suen.ed.psu.edu/~hsuen/pubs/Gtheory.pdf>, en son 21 Eylül 2008'de alınmıştır.
- Tan, Ş. (2008). *Öğretimde Ölçme ve Değerlendirme. KPSS El Kitabı*. Ankara: Pegem Akademi.

- Tan, Ş. & Erdoğan, A. (2004). *Öğretimi Planlama ve Değerlendirme*. (Genişletilmiş 5. Baskı). Ankara: PegemA.
- Tavşancıl, E. (2005). *Tutumların Ölçülmesi ve SPSS ile Veri Analizi*. (2. Baskı). Ankara: Nobel Yayın Dağıtım.
- Thompson, B. (Ed.) (2003). *Score Reliability. Contemporary Thinking on Reliability Issues*. Thousand Oaks, California: Sage.
- Thompson, B. (1999). Five Methodology Errors in Educational Research: A Pantheon of Statistical Significance and other Faux Pas. In Thompson, B. (Ed.), *Advances in Social Science Methodology, Volume 5*. Stamford, Connecticut: JAI.
- Thompson, B. (1994a). Guidelines for Authors. *Educational and Psychological Measurement*, Vol. 54, 834-847.
- Thompson, B. (1994b). *It is Incorrect to Say "The Test Is Reliable": Bad Language Habits Can Contribute to Incorrect or Meaningless Research Conclusions*. (ERIC Document Reproduction Service No. ED 367 707).
- Thompson, B. ve Vacha-Haase, T. (2000). Psychometrics is Datametrics : The Test is Not Reliable. *Educational and Psychological Measurement*, Vol. 60, 174-195.
- Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin.
- Topdemir, H. G. (2002). Kuhn ve Bilimsel Devrimlerin Yapısı Üzerine Bir Değerlendirme. *Felsefe Dünyası*, Sayı 36, 45-62.
- Traub, R. E. (1994). *Reliability for the Social Sciences. Theory and Applications*. Thousand Oaks: Sage.
- Winer, B. J. (1971). *Statistical Principles in Experimental Design*. (Second Edition). New York: McGraw-Hill.
- Worthen, B. R., White, K. R., Fan, X. & Sudweeks, R. R. (1999). *Measurement and Assessment in Schools*. (Second Edition). New York: Longman.
- Yılmaz, E. & Sünbül, A. M. (2009). Üniversite Öğrencilerine Yönelik Girişimcilik Ölçeğinin Geliştirilmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Sayı 21, 195-203.
- Yurdabakan, İ. (2008). Eğitimde Kullanılan Ölçme Araçlarının Nitelikleri. Erkan, S. ve Gömleksiz, M. (Ed.), *Eğitimde Ölçme ve Değerlendirme*. Ankara: Nobel.