

**DEĞERBİÇİCİLER ARASI (INTERRATER) ÖLÇÜM GÜVENİRLİĞİNİN  
CRONBACH'IN ALFASI İLE KESTİRİLMESİ**Vahit BADEMCI<sup>1</sup>**ÖZET**

Güvenirlik ve geçerlik, testlerin ya da ölçme araçlarının veya değerbiçicilerin özellikleri *değildir*. Dolayısıyla, “test geçerlidir”, “ölçme aracı güvenilirlerdir”, “değerbiçicilerin güvenilirliği”, “ölçeğin güvenilirliği”, “bellilendirmenin geçerliği”, “sınavın güvenilirliği” ve benzeri ifadeler kullanmak *doğru değildir*.

Hesaplama kolaylığından dolayı, *Pearson çarpım moment korelasyon katsayısı* [r] kimi zaman değerbiçiciler arası ölçüm güvenirliliğinin bir göstergesi [indeksi] olarak rapor edilmektedir (Fleenor, Fleenor ve Grossnickle, 1996). Ancak, bu çalışmada, değerbiçiciler arası ölçüm güvenilirlik kestirimi için *Cronbach'ın alfası* kullanılmıştır.

**Anahtar Kelimeler:** Değerbiçiciler arası güvenilirlik, Cronbach'ın alfası, ölçüm güvenirliliği, güvenilirlik ve geçerlik

**ESTIMATION OF INTERRATER SCORE RELIABILITY BY THE CRONBACH'S  
ALPHA****ABSTRACT**

Reliability and validity are not the characteristics of the tests or measurement instruments or raters. Therefore, it is not correct to use statements as “the test is valid”, “the measurement instrument is reliable”, “the reliability of raters”, “the reliability of the scale”, “the validity of the assessment”, “the reliability of the examination” and similar statements.

Due to the simplicity of the calculation, the Pearson product moment correlation coefficient [r] is sometimes reported as an index of the interrater score reliability (Fleenor, Fleenor ve Grossnickle, 1996). However, in this study, the Cronbach's alpha has been used for the estimation of interrater score reliability.

**Keywords:** Interrater reliability, the Cronbach's alpha, score reliability, reliability and validity

**1. TÜRK EĞİTİM VE BİLİMİNDE 60 YIL SONRA BİLİMSEL DEVRİM:  
GÜVENİRLİK VE GEÇERLİK, TESTLERİN YA DA ÖLÇME ARAÇLARININ  
ÖZELLİKLERİ DEĞİLDİR**

Güvenirlik ve geçerlik, eğitim, psikoloji ve sosyal bilimler içerisinde sınıflanan diğer disiplinlerdeki ya da araştırmalardaki üst düzey kalitede bir ölçmenin kalbinde yatan iki kavramdır (Bademci, 2007; 2011b).

**Güvenirlik**, belirli bir evrene veya örnekleme uygulanmış bir test ya da ölçme aracından elde edilmiş ölçümlerin tutarlılığı veya tekrarlanabilirliğidir; **geçerlik** ise, belirli bir evrene veya örnekleme uygulanmış bir test ya da ölçme aracından elde edilmiş ölçümlerden

<sup>1</sup> Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Gölbaşı-Ankara, bademci@gazi.edu.tr

yapılmış belirli yorumların ve kullanımların uygunluğu ve yeterliğidir (Bademci, 2001a; 2007; 2011a).

Bademci (1999; 2001a; 2001b; 2002; 2004; 2005a; 2005b; 2005c; 2006a; 2006b; 2006c; 2007; 2008; 2010; 2011a; 2011b) tarafından yaklaşık 60 yıl sonra gerçekleştirilen Türk eğitim ve biliminde ölçme ve değerlendirme ve araştırma yöntembilimi alanlarındaki *paradigma değişikliği* ya da *bilimsel devrim* doğrultusunda, güvenilirlik ve geçerlikle ilgili yukarıda da ifade edilen bu *yeni* tanımlamalarda, testlerin veya ölçme araçlarının kendilerine *değil*, testlerden veya ölçme araçlarından elde edilen *ölçümlere* vurgu yapılmıştır.

Güvenirlik ve geçerlik, testlerin ya da ölçme araçlarının özellikleri *değildir* (Murphy ve Davidshofer, 2001; Barnes, Harp ve Jung, 2002; Worthen, White, Fan ve Sudweeks, 1999). Şüphesiz, güvenilirlik ve geçerlik, değerbiçicilerin veya gözlem yapanların ya da ölçümleyicilerin de özellikleri *değildir*. Böylelikle, güvenilirlik ve geçerlik bir testin ya da bir ölçme aracının veya bir değerbiçicinin değişmeyen, kalıcı özellikleriymiş gibi, “bir testin güvenilirliği”nden veya “bir ölçme aracının geçerliği”nden ya da “bir değerbiçicinin güvenilirliği”nden söz etmek ise, *uygun değildir, doğru değildir* (Bademci, 1999; 2001a; 2004; 2007; 2011a; 2011b; Ebel ve Frisbie, 1991; Kubiszyn ve Borich, 2007; Linn ve Gronlund, 1995).

Dolayısıyla, “test geçerlidir”, “test güvenilirdir”, “bu deneyin geçerliği”, “testin geçerliği”, “değerbiçicilerin güvenilirliği”, “ölçeğin güvenilirliği”, “bellilendirmenin geçerliği”, “ölçme aracının [ya da yönteminin] geçerliği”, “sınav geçerlidir”, “gözleyicilerin güvenilirliği”, “sınavın güvenilirliği”, “ölçme prosedürü geçerlidir”, “ölçümleyicilerin geçerliği” ve benzeri ifadeler kullanmak *doğru değildir*; bunların yerlerine, “test ölçümlerinin güvenilirliği”, “ölçümlerden yapılmış kullanım ve yorumun geçerliği”, “ABC ölçek ölçüm yorum geçerliği”, “ölçüm güvenilirliği”, “ABC testinden elde edilen ölçümlerden yapılmış bir yorumun geçerliği”, “değerbiçicilerin verdikleri [belirledikleri] ölçümlerin güvenilirliği”, “ABC testi ölçümlerinin test-tekrar test güvenilirliği” ve benzeri ifadeler kullanmak ise, *daha uygun ve doğrudur* (AERA, APA ve NCME, 1999; Bademci, 1999, 2001a; 2004; 2005a; 2007; 2008; 2011a; 2011b; Brookhart ve Nitko, 2008; Ebel ve Frisbie, 1991; Kieffer, 1999; Linn ve Gronlund, 1995; Linn ve Miller, 2005; McMillan, 2007; Nilsson, Schmidt ve Meek, 2002; Nitko ve Brookhart, 2011; Reynolds, Livingston ve Willson, 2010; Thompson, 2003).

[Ölçüm güvenilirliğiyle ilgili olası doğru ifadelerin kullanılmasına yönelik örnekler, ayrıca Bademci'nin (2004) bir başka çalışmasında da bulunmaktadır.]

Buraya kadar bahsedilen hususlarla ilgili olarak, Cronbach (1972), bundan tam 40 yıl önce, “test geçerliği” hakkında konuşmanın *mantksız* olduğunu, geçerlemenin (validation) ise, test yorumlayıcısının görevi olduğunu açıkça vurgulamıştır. Ebel (1965) ve Gronlund (1965) ise, bundan 47 yıl önce, “testin güvenilirliği”nden ziyade, “test ölçümlerinin güvenilirliği” hakkında konuşmanın çok daha uygun olduğunu altını çizmişlerdir.

## **2. GÜVENİLİR YA DA GÜVENİLMEZ OLAN, TESTLER VEYA ÖLÇEKLER YA DA ÖLÇME ARAÇLARI DEĞİL, ONLARDAN ELDE EDİLMİŞ OLAN ÖLÇÜMLERDİR**

Güvenirlik, testin kendisinin değil, elde edilmiş ölçümlerinin bir özelliğidir; o halde, *bir test ya da ölçme aracının kendisi ne güvenilir, ne de güvenilmezdir* (Bademci, 1999; 2001a;

2004; 2007; 2010; 2011a; Crocker ve Algina, 1986; Ebel ve Frisbie, 1991; Rouse, 2007; Rowley, 1976; Traub ve Rowley, 1991).

*Güvenilir ya da güvenilmez olan, testler veya ölçekler ya da ölçme araçları değil, onlardan elde edilmiş olan ölçümlerdir; bir başka söyleyişle, güvenilirlik özelliğine ölçümler sahiptir, testin veya ölçeğin ya da ölçme aracının veya değerbiçicinin kendisi değil* (Bademci, 1999; 2001a; 2004; 2007; 2008; Ebel ve Frisbie, 1991; Thompson, 2003; Traub ve Rowley, 1991). Kısaca, *testler değil, ölçümler güvenilirdir* (Bademci, 1999; Ebel ve Frisbie, 1991; Kieffer ve Reese, 2002; Vacha-Haase, 1998).

### **3. KLASİK GERÇEK ÖLÇÜM KURAMINDA ÖLÇÜM GÜVENİRLİĞİ KESTİRME YÖNTEMLERİ**

Test-tekrar test, almaşık (alternate) form, yarıya bölme (split-half), değerbiçiciler arası, Cronbach'ın (1951) alfası ( $\alpha$ ) [ya da alfa katsayısı veya Cronbach'ın alfası katsayısı], Kuder-Richardson 20 (KR-20; Kuder ve Richardson, 1937) ve Hoyt'un (1941) varyans analizi, ölçüm güvenirliliğinin kestiriminde kullanılan bazı yöntemlerdendir; değerbiçiciler arası, test-tekrar test, almaşık form, yarıya bölme ölçüm güvenirliliği kestirim yöntemlerinde, *Pearson çarpım moment korelasyon katsayısından yararlanılır* (Bademci, 2011b; Crocker ve Algina, 1986; Linn ve Gronlund, 2000; Mehrens ve Lehmann, 1991; Worthen, White, Fan ve Sudweeks, 1999).

### **4. DEĞERBİÇİCİLER ARASI ÖLÇÜM GÜVENİRLİK KESTİRİMİ İÇİN CRONBACH'IN ALFASI**

Hesaplama kolaylığından dolayı, Pearson çarpım moment korelasyon katsayısı kimi zaman değerbiçiciler arası ölçüm güvenirliliğinin bir göstergesi [indeksi] olarak rapor edilmektedir (Fleener, Fleener ve Grossnickle, 1996). Bu makalede ise, değerbiçiciler arası ölçüm güvenirlilik kestirimi için Pearson çarpım moment korelasyon katsayısı *değil*, Cronbach'ın alfası [katsayısı] kullanılacaktır (Bademci, 2001a; 2005a; 2010; Fleener, Fleener ve Grossnickle, 1996; Liao, Hunt ve Chen, 2010; Nitko, 2001).

Bilindiği üzere, Cronbach'ın alfası [katsayısı], ölçüm güvenirliliğinin kestiriminde kullanılan bir yöntemdir (Mehrens ve Lehmann, 1991; Linn ve Gronlund, 2000; Worthen, White, Fan ve Sudweeks, 1999). Ancak, hatırlanmalıdır ki, Cronbach'ın alfası [katsayısı] da bir *istatistiktir* ve de eğitim, sosyal ve sağlık bilimlerinde en yaygın kullanılmış olan bir ölçme istatistiğidir (Ferguson ve Takane, 1989; Liu ve Zumbo, 2007; Shelby, 2011); dolayısıyla, bu çalışmada değerbiçiciler arası ölçüm güvenirlilik kestirimi için Cronbach'ın alfası katsayısından yararlanılacaktır.

[Bir ölçüm güvenirlilik kestirim yöntemi olarak Cronbach'ın alfası ile ilgili de geniş ve ayrıntılı bilgiler, Bademci'nin (2001a; 2005a; 2006a; 2006b; 2006c; 2007; 2010; 2011b) ardışık bazı çalışmalarında bulunmaktadır.]

Aşağıda, Çizelge 1'de denencel veriler sunulmuştur; tabloda, 4 değerbiçicinin verdikleri [belirledikleri] ölçümler [değerbiçimler] bulunmaktadır ve değerbiçiciler arası ölçüm güvenirliliği, Cronbach'ın alfası ile kestirilecektir.

**Çizelge 1. DEĞERBİÇİCİLER ARASI ÖLÇÜM GÜVENİRLİĞİNİN CRONBACH'IN ALFASI İLE KESTİRİLMESİNE YÖNELİK DENENCEL VERİLER [Tablo, Nitko'dan (2001) Uyarlanmıştır.]**

Kişiler (i) (Öğrenciler)	Değerbiçiciler (j=4) [Öğretmenler/Jüri Üyeleri]				Toplam Ölçüm(y)
	1	2	3	4	
Azra	4	3	4	4	15
Büşra	2	5	5	5	17
Ceyda	3	5	5	3	16
Dilara	1	3	1	1	6
Esra	5	5	5	4	19
Fulya	4	3	4	4	15
$\sigma_j^2$	1.8055	1.0000	2.0000	1.5833	
$\sigma_y^2$					16.8888

$$\alpha = \frac{j}{j-1} \left(1 - \frac{\sum \sigma_j^2}{\sigma_y^2}\right)$$

$\alpha$  = Cronbach'ın alfası katsayısı

$j$  = değerbiçici sayısı

$$\alpha = \frac{4}{4-1} \left(1 - \frac{1.8055 + 1.0000 + 2.0000 + 1.5833}{16.8888}\right) = 0.8290$$

## 5. ÖLÇÜM GÜVENİRLİĞİ VE BELGELENDİRİLMESİ İLE İLGİLİ BAZI ÖNEMLİ HUSUSLAR

*Temel ya da genel araştırma amaçları için, ölçüm güvenilirlik katsayıları (örneğin, test-tekrar test, Cronbach'ın alfası vd.) en az .80 ve daha yukarısı olmalıdır* (Bademci, 2001a; 2007; 2010; 2011b).

Güvenirlilik analizi [sonuç veya sonuçları], yargı veya değerlendirme *olmaksızın* rapor edilmelidir; “iyi”, “kabul edilebilir”, “yüksek”, “orta” veya “zayıf” şeklinde veya benzeri hiçbir ifade **kullanılmamalıdır** (Bademci, 2010). Örneğin, “güvenirliliğin  $r=.70$ 'ten büyük olması, güvenirliliğin yüksek olduğu anlamına gelir” şeklinde bir ifadenin kullanılması

**doğru değildir** ve bu türden ifadeler bilimsel çalışmalarda ya da raporlarda **yer almamalıdır** (Bademci, 2010).

Bilimsel çalışmalarda, ölçüm güvenilirlik katsayıları için güven aralıkları kestirilmeli ve *mutlaka* rapor edilmelidir (Bademci, 2001a; 2002; 2007; 2010; test-tekrar test ölçüm güvenilirliği [katsayısı] için güven aralığı kestirilmesi ile ilgili, bkz., Bademci, 2005b).

Ölçüm güvenilirlik katsayıları, sınavı alanların evreni ya da örnekleminin özelliklerinden etkilenir; bir başka ifadeyle, ölçüm güvenilirliği, evren ya da örneklem bağımlıdır; dolayısıyla, belirlenmiş evren ya da örneklemin özellikleri (örneğin, örneklemin büyüklüğü) *mutlaka* rapor edilmelidir (Bademci, 2001a; 2002; 2005b; 2006c; 2007; 2010; 2011a)

Bilimsel araştırmalar ya da çalışmalarda, “güvenirlik doğuşturma” (reliability induction) ya da “güvenirlik doğuşması” **yapılmamalıdır** (Bademci, 2001a; 2005a; 2006c; 2010).

“**Güvenirlik**”, “**geçerlik**”, “**güvenirliği**”, “**geçerliği**”, güvenilirlik ve geçerlik terimlerinin **doğru yazılmış** durumlarıdır; “*reliability*”[“güvenirlik”] ve “*validity*”[“geçerlik”], *isimdir[ler]* ve *düzgün ve dikkatli yazılmalıdır* (Bademci, 2001a; 2005a; 2006c; 2010).

## 6. KAYNAKLAR

- AERA, APA ve NCME [American Educational Research Association, American Psychological Association ve National Council on Measurement in Education]. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Association.
- Bademci, V. (2011b). Kuder-Richardson 20, Cronbach’ın Alfa’sı, Hoyt’un Varyans Analizi, Genellenirlik Kuramı ve Ölçüm Güvenirliği Üzerine Bir Çalışma. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, Sayı 17, 173-193.
- Bademci, V. (2011a). Türk Eğitim ve Biliminde Bilimsel Devrim: Testler ya da Ölçme Araçları Güvenilir ve Geçerli Değildir. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, Sayı 16, 116-132.
- 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 2010. [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.
- Bademci, V. (2007). *Ölçme ve Araştırma Yöntembiliminde Paradigma Değişikliği: Testler Güvenilir Değildir*. Ankara: Yenyap Yayınları.
- 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 2006. [Konferansla ilgili haber için; **Gazi Haber**, Nisan 2006, Sayı 66, Sayfa 64.]
- 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.
- Bademci, V. (2006a). Güvenirliği Doğru Anlamak ve Bazı Klişeleri Yıkma: Bilinenlerin Aksine, Cronbach'ın Alfa Katsayısı, Negatif ve -1'den Küçük Olabilir. **İnönü Üniversitesi Eğitim Fakültesi Dergisi**, Cilt 7, Sayı 12, 3-26.
- 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, 46-75.
- 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.
- 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 2005.
- Bademci, V. (2004). "Testin Güvenirliği" veya "Test Güvenilirdir" Diye İfade Etmek Doğru Değildir. **Türk Eğitim Bilimleri Dergisi**, Cilt 2, 367-373.
- 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. (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. (2001a). **Düşünmenin Öğretilmesi ve Öğretimde Kullanılan Yöntemler-Teknikler**. Konferans. Düzenleyen: TÜRMOB. Bursa: Bursa SMMM Odası Konferans Salonu, 9 Kasım 2001.
- Bademci, V. (1999). Öğretmen Yetiştiren Programlar Nasıl Olmalı? **Türkiye'de Eğitim Fakülteleri ve Öğretmen Yetiştirme Paneli**. Düzenleyen: ESEF İşletme Araştırma Topluluğu. Ankara: G.Ü. Mesleki Eğitim Fakültesi Konferans Salonu, 21 Mayıs 1999.
- Barnes, L. L. B., Harp, D. ve Jung, W. S. (2002). Reliability Generalization of Scores on the Spielberger State-Trait Anxiety Inventory. **Educational and Psychological Measurement**, Vol. 62, 603-618.
- Brookhart, S. M. ve Nitko, A. J. (2008). **Assessment and Grading in Classrooms**. Upper Saddle River, New Jersey: Pearson.

- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Fort Worth: Holt, Rinehart and Winston.
- Cronbach, L. J. (1972). Validation of Educational Measures. Bracht, G. H., Hopkins, K. D. ve Stanley, J. C. (Eds.), *Perspectives in Educational and Psychological Measurement*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, Vol. 16, 297-334.
- Ebel, R. L. (1965). *Measuring Educational Achievement*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Ebel, R. L. ve Frisbie, D. A. (1991). *Essentials of Educational Measurement*. (Fifth Edition). Englewood Cliffs, New Jersey: Prentice Hall.
- Ferguson, G. A. ve Takane, Y. (1989). *Statistical Analysis in Psychology and Education*. (Sixth Edition). New York: McGraw-Hill.
- Fleener, J. W., Fleener, J. B. ve Grossnickle, W. F. (1996). Interrater Reliability and Agreement of Performance Ratings: A Methodological Comparison. *Journal of Business and Psychology*, Vol. 10, 367-380.
- Gronlund, N. E. (1965). *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Hoyt, C. (1941). Test Reliability Estimated by Analysis of Variance. *Psychometrika*, Vol. 6, 153-160.
- Kieffer, K. M. (1999). Why Generalizability Theory is Essential and Classical Test Theory is Often Inadequate. Thompson, B. (Ed.), *Advances in Social Science Methodology, Volume 5*. Stamford, Connecticut: JAI.
- Kieffer, K. M. ve Reese, R. J. (2002). A Reliability Generalization Study of the Geriatric Depression Scale. *Educational and Psychological Measurement*, Vol. 62, 969-994.
- Kubiszyn, T. ve Borich, G. (2007). *Educational Testing and Measurement: Classroom Application and Practice*. (Eighth Edition). USA: John Wiley and Sons.
- Kuder, G. F. ve Richardson, M. W. (1937). The Theory of the Estimation of Test Reliability. *Psychometrika*, Vol. 2, 151-160.
- Liao, S. C., Hunt, E. A. ve Chen W. (2010). Comparison between Inter-rater Reliability and Inter-rater Agreement in Performance Assessment. *Annals, Academy of Medicine, Singapore*, Vol. 39, 613-618.
- Linn, R. L. ve Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. (Eighth Edition). Upper Saddle River, New Jersey: Pearson.
- Linn, R. L. ve Gronlund, N. E. (1995). *Measurement and Assessment in Teaching*. (Seventh Edition). Upper Saddle River, New Jersey: Merrill/Prentice-Hall.
- Linn, R. L. ve Miller, M. D. (2005). *Measurement and Assessment in Teaching*. (Ninth Edition). Upper Saddle River, New Jersey: Merrill.

- Liu, Y. ve Zumbo, B. D. (2007). The Impact of Outliers on Cronbach's Coefficient Alpha Estimate of Reliability. Visual Analogue Scales. *Educational and Psychological Measurement*, Vol. 67, 620-634.
- McMillan, J. H. (2007). *Classroom Assessment. Principles and Practice for Effective Instruction*. (Fourth Edition). Boston: Allyn and Bacon.
- Mehrens, W. A. ve Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology*. (Fourth Edition). Fort Worth: Harcourt Brace.
- Murphy, K. R. ve Davidshofer, C. O. (2001). *Psychological Testing. Principles and Applications*. (Fifth Edition). Upper Saddle River, New Jersey: Prentice Hall.
- Nilsson, J. E., Schmidt, C. K. ve Meek, W. D. (2002). Reliability Generalization: An Examination of the Career Decision-Making Self-Efficacy Scale. *Educational and Psychological Measurement*, Vol. 62, 647-658.
- Nitko, A. J. (2001). *Educational Assessment of Students*. (Third Edition). Upper Saddle River, New Jersey: Merrill/Prentice-Hall.
- Nitko, A. J. ve Brookhart, S. M. (2011). *Educational Assessment of Students*. (Sixth Edition). Boston: Pearson.
- Reynolds, C. R., Livingston, R. B. ve Willson, V. (2010). *Measurement and Assessment in Education*. (Second Edition). Upper Saddle River, New Jersey: Pearson Education.
- Rouse, S. V. (2007). Using Reliability Generalization Methods to Explore Measurement Error: An Illustration Using the MMPI-2 PSY-5 Scales. *Journal of Personality Assessment*, Vol. 88(3), 264-275.
- Rowley, G. R. (1976). The Reliability of Observational Measures. *American Educational Research Journal*, Vol. 13, 51-59.
- Shelby, L. B. (2011). Beyond Cronbach's Alpha: Considering Confirmatory Factor Analysis and Segmentation. *Human Dimensions of Wildlife*, Vol. 16, 142-148.
- Thompson, B. (Ed.) (2003). *Score Reliability. Contemporary Thinking on Reliability Issues*. Thousand Oaks, California: Sage.
- Traub, R. R. ve Rowley, G. L. (1991). Understanding Reliability. *Educational Measurement: Issues and Practice*, Vol. 10(1), 37-45.
- Vacha-Haase, T. (1998). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. *Educational and Psychological Measurement*, Vol. 58, 6-20.
- Worthen, B. R., White, K. R., Fan, X. ve Sudweeks, R. R. (1999). *Measurement and Assessment in Schools*. (Second Edition). New York: Longman.