2008 Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi Sayı: 22, s.50-69

ARAŞTIRMALARDA ÖLÇME İLE İLGİLİ BAZI BÜYÜK HATALARI DÜZELTMEK VE EĞİTİMDE YENİDEN YAPILANMAYI SÜRDÜRMEK: GÜVENİRLİK, TESTLERİN BİR ÖZELLİĞİ DEĞİLDİR*

Vahit BADEMCİ[1]

ÖZET

Güvenirlik, çoğu kez yanlış anlaşılmış bir kavramdır. Güvenirlik, aracın kendisine değil, bir ölçme aracıyla elde edilmiş ölçümlere (veya sonuçlara) işaret eder. Güvenirlik testin kendisinin değil, eldeki veriler veya ölçümlerin bir özelliğidir. Böylelikle, testin güvenirliği veya aracın güvenirliği ya da test güvenilirdir diye ifade etmek doğru değildir. Nitekim testler değil, ölçümler güvenilirdir. Basit şekliyle, testler güvenilir değildir. Güvenirlik, yalnızca kullanılmış olan ölçme aracı tarafından değil, örneklem tarafından da etkilenmektedir. Güvenirlik örneklemin de bir fonksiyonudur. Çünkü güvenirlik, tasarlanmış evrenden seçilmiş bir örneklem üzerinde değerlendirilmektedir. Güvenirlik örneklemden örnekleme değişir ve bu suretle daha ayrışık [heterojen] örneklemler sıklıkla daha değişken ölçümlere ve bu durumda daha yüksek güvenirliğe yol açar. Bundan dolayı, aynı test veya ölçme aracı, daha ayrışık veya daha bağdaşık [homojen] gruplara uygulandığında, birbirini onamayan ölçümler güvenirliği ortaya çıkacaktır. O takdirde, ölçme güvenirliğinden veya test ölçümlerinin güvenirliğinden bahsetmek çok daha uygundur. Birkaç kelimeyle, bir test güvenilir veya güvenilmez değildir ve test ölçümleri veya eldeki veriler güvenilir veya güvenilmezdir.

Anahtar Kelimeler: Ölçüm (Score)** Güvenirliği, Güvenirlik, Yöntembilim Hataları

IN THE RESEARCHES, TO CORRECT SOME OF THE ENORMOUS ERRORS RELATED WITH MEASUREMENT AND TO MAINTAIN RESTRUCTURING IN THE EDUCATION: RELIABILITY IS NOT A CHARACTERISTIC OF TESTS

ABSTRACT

Reliability is often a misunderstood concept. Reliability refers to the scores (or results) obtained with an measurement instrument and not to the instrument itself. Reliability is a characteristic of scores or the data in hand, not of the test itself. Thus, it is incorrect to speak of "the reliability of the test" or "the reliability of the instrument" or " the test is reliable". Thus, the scores, not tests, are reliable. Simply, tests are not reliable. Reliability is not only ifluenced by a measurement instrument used, it is influenced by the sample as well as of the instrument. Reliability is also a function of sample. Because, reliability should be evaluated on selecting a sample from planned population. Reliability fluctuates from sample to sample, and so more heterogeneous samples often lead to more variable scores and thus to higher reliability. Therefore, the same test or the measurement instrument, when administered to more heterogeneous or to more homogeneous the groups, will yield with differing scores reliability. In that case, it is more appropriate to speak of the reliability of "test scores" or the "measurement". In a few words speaking, a test is not reliable or unreliable and the data in hand or the test scores are reliable or unreliable.

Key Words: Score Reliability, Reliability, Methodology Errors.

[1] Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Gölbaşı-Ankara, bademci@gazi.edu.tr

* Bu makalenin bir kısmı, Ankara'da, G.Ü. Gazi Eğitim Fakültesi'nce, 22-23-24 Eylül 2005 tarihleri arasında düzenlenen ***Eğitim Fakültelerinde Yeniden Yapılandırmanın Sonuçları ve Öğretmen Yetiştirme Sempozyumu***nda ***Bildiri*** olarak sunulmuştur.

** Ölçüm; (score)
M. Fuat Turgut, ölçme işlemleri sonunda elde edilen sayılara ölçüm denilmesini önermektedir (Bademci, 1999:7-8).
Ölçümleme; (scoring)
Bellilendirme; (assessment)

1. GİRİŞ

Eski bir Yunan mitine [hayali öyküsüne][***] göre, Thetis, bir oğlu olunca, oğlu Akhilleus'u [ya da diğer bir söyleyişle Aşil'i], silah işlemez kılmak, tehlikelerden korumak, ölümsüz kılmak için Ölüler Ülkesi'nin ırmağı Stiks'e daldırmıştır. Ancak, Thetis'in, oğlu Aşil'i topuğundan tutup suya batırması nedeniyle, Aşil'in topuğu büyülü suya değmemiş ve Aşil'in topuğu, vücudunun zayıf noktası olarak kalmıştır. Sonunda da, Aşil, bir savaşta, topuğuna saplanan bir ok yüzünden ölmüştür (Estin ve Laporte, 2003; Hamilton, 2003; Treays, 2000).

"Kaslar, kemiklere kiriş denen esnek olmayan sağlam şeritlerle bağlıdır. Bacağınızın alt kısmındaki kası ayağınızın arkasına bağlayan kiriş, egzersiz yaparken kolayca zedelenebilir. Bu kirişe bir Eski Yunan hikayesine dayanılarak Aşil kirişi adı verilmiştir. Güçlü bir kişinin zayıf noktasını belirtmek için de 'Aşil topuğu' deyimi kullanılır" (Treays, 2000:7).

Ölçme, sosyodavranışsal araştırmanın Aşil topuğudur (Pedhazur ve Schmelkin, 1991:2) yani, zayıf noktasıdır. Şüphesiz bu duruma, öncelikle lisansüstü programlar olmak üzere, yüksek lisans ve özellikle doktora programlarındaki ölçmeyle ilişkili konuların giderek azalmasının ve de ölçme konularıyla ilgili zayıf ve kalitesiz eğitim verilmesinin sebep olduğu söylenebilir.

Bu hususla ilgili olarak, Pedhazur ve Schmelkin (1991), özellikle doktora programlarında, araştırma deseni ve istatistiklere bir nebze de olsa gereksinim gösterilirken, ölçme üzerindeki vurgunun azaldığını, böylelikle de, pek çok öğrencinin, ölçülerin kullanılması ve geliştirilmesi için zorunlu olan özel yeterliklerin hiçbirini elde edemediğini, bunun sonucunda da, pek çok araştırmada kullanılmış olan ölçülerin özelliklerine hiç dikkat edilmediğini veya çok az dikkat edildiğini, ortaya çıkan bu durumun ise, beklenmedik bir olay olmadığını ileri sürmüştür (Henson, 2000b; Henson, 2001; Pedhazur ve Schmelkin, 1991; Thompson, 2001). Bu izlenim, Aiken ve arkadaşları (1990) tarafından yapılan bir çalışmada, doktora eğitim programları içindeki ölçmeyle ilişkili konuların esasen azaldığı şeklinde teyit edilmiş, bir başka ifadeyle doktora eğitim programları içindeki bu ölçme boşluğu, Aiken ve arkadaşlarınca da (1990) doğrulanmıştır (Aiken ve arkadaşları,1990; Henson, 2001; Thompson, 1999).

Eğitimin bu yetersizliği, yayımlanmış araştırmaların kalitesi üzerinde de kendisini göstermiştir. *American Educational Research Journal (AERJ)* ile ilgili yaptığı ve 1980 yılında yayımlanmış çalışmasında Wilson, AERJ makalelerinin yalnızca %37'since analiz edilmiş veriler için güvenirlik katsayılarının açıkça rapor edildiğini, diğer %18'inin daha önceki araştırmaya referans verme suretiyle güvenirliği dolaylı olarak rapor ettiğini, yayımlanmış araştırmanın yaklaşık yarısında ise, [mazeret kabul edilemez biçimde] güvenirliğin rapor edilmediğini ifade etmiştir (Bulunduğu yer, Thompson,1994a; Thompson, 1994b; Vacha-Haase,1998).

Meier ve Davis (1990) ise, *Journal of Counseling Psychology (JCP)* 'nin 1967, 1977 ve 1987 ciltleri üzerine yaptıkları çalışmalarında, üç JCP cildi içinde, tanımlanmış ölçeklerle ilgili olarak, 1967 cildi içindeki ölçeklerin %95'inin, 1977 cildi içindeki ölçeklerin %85'inin ve 1987 cildi içindeki ölçeklerin %60'ının psikometrik özelliklerinin [veriler güvenirlik kestirimlerine yöneliktir] raporlara eklenmediğini [ya da aktarılmadığını] ifade etmişlerdir (Meier ve Davis, 1990; Thompson, 1994a; Vacha-Haase, 1998). Bu çalışmalar ve benzerlerindeki zayıf güvenirlik rapor etme uygulamaları ve bunlarla ilgili ortaya konulmuş olan bulgular (Meier ve Davis 1990; Thompson, 1994b; Vacha-Haase,1998; Vacha-Haase, Kogan ve Thompson,2000; Whittington, 1998), bir başka söyleyişle ölçümlerin psikometrik özelliklerini rapor etmedeki bazı problemler, güvenirliğin ölçümlerin bir özelliği olduğunun karşıtı düşüncede olan ve güvenirliğin testin bir özelliği olduğuna inanan araştırmacılara değin izler olabilir (Whittington, 1998). Bu bulgular ise, güvenirliğin kesin görünüşünün iyi anlaşılamamış olabileceğini ifade etmektedir (Shields ve Caruso, 2004).

"Test ölçümlerinin güvenirliği" gibi uzun ama doğru bir ifadeyi, "testin güvenirliği" biçiminde kısaltarak konuşma tarzının (Thompson, 1999; Thompson ve Vacha-Haase, 2000), güvenirliğin doğası hakkında yanlış anlamalara yol açtığı ileri sürülmüştür (Baugh, 2002). Pedhazur ve Schmelkin'e (1991:82) göre, bir ölçünün güvenirliğine dair ifadeler, uygun değildir ve olanak dahilinde [uygun şart veya durum sağlandığında] yanlış yola sevk eder. Testin güvenirliği biçiminde kısaltarak ifade etme yolu doğal olarak şüpheli görünmemektedir,

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

ancak sonrasında, testin güvenirliği şeklindeki kısaltılan ifadeye, bilinçsizce, kelime kelimesine [kısaltılan] aslına uygun anlam yüklenir ve bu da doğru değildir (Thompson,1994a; Thompson, 1999).

Thompson (1992; 1994a; 1999; Thompson ve Snyder, 1998; Thompson ve Vacha-Haase 2000), "test ölçümlerinin güvenirliği" yerine, testin güvenirliği biçiminde kısaltarak konuşmayı [ve de yazmayı], "dikkatsiz (sloopy)" [sloppy; düzensiz-disiplinsiz-dikkatsiz-özensiz-şapsal anlamları da var, (Töreci, 2005)] konuşma olarak nitelendirmektedir. "Dikkatsiz" konuşma ise, bazen, "dikkatsiz" düşünmeye, "dikkatsiz" uygulamaya ve de daha fazla zararlı bir sonuca kılavuzluk etmektedir (Thompson, 1992; Thompson, 1994a; Vacha-Haase, 1998).

"Testin güvenirliği" biçiminde kısaltarak kullanma eğilimi için söylenmesi gereken özet ifade, "anlaşılır, fakat zararlıdır" (Thompson ve Vacha-Haase, 2000:178) şeklinde olacaktır. Kısaca söylemek gerekirse, güvenirlik, testlerin değil ölçümlerin bir özelliğidir (Caruso, 2000).

Thompson (1994a) ise, çok az araştırmacının, güvenirliğin eldeki veriler veya ölçümlerin bir özelliği olduğunu bilinçli kabul edip, buna göre davrandıklarını belirtmiştir. Ancak, pek çok insan da ölçüm güvenirliğini [hala] anlamamıştır (Mittag ve Thompson, 2000; Vacha-Haase, Kogan ve Thompson, 2000).

2. ÖLÇÜM GÜVENİRLİĞİNİN ÖZÜ

Her ne kadar test güvenirliğinin işevuruk bir tanımı başlığı altında da olsa, güvenirliğin testin bir özelliği olmadığını yorumlayan Ebel (1972), bu tartışma konusunun öncüleri arasında sayılabilir. Ebel'in bu yorumuna benzer bir vurgu, ölçme aracından ziyade ölçmelerin güvenirlik özelliğine sahip olduğu şeklinde, Guilford ve Fruchter'dan (1973) gelmiştir. Bu konuda başı çeken kişinin ise, Rowley (1976) olduğu söylenebilir. Güvenirliğin bir aracın (örneğin, test) değil ölçmenin bir özelliği olduğunu açıklamaya çalışan Rowley (1976: 53), bu konuda net bir ifade kullanmıştır; "…bir aracın kendisi ne güvenilirdir, ne de güvenilir değildir".

Güvenirlik, Aracın Kendisine Değil Bir Ölçme Aracı İle Elde Edilmiş Ölçümlere İşaret Eder.

Buraya kadar yapılan açıklamalara ilave edilebilecek aydınlatıcı ifadeler, Gronlund ve Linn'in l990 tarihli çalışmasında mevcuttur. Gronlund ve Linn, (1990: 78) güvenirliğin, aracın kendisine değil bir değerlendirme aracı ile elde edilmiş ölçümlere işaret ettiğine dikkat çekerek, aracın veya testin yerine, ölçmenin veya test ölçümlerinin güvenirliğinden bahsetmenin çok daha uygun olduğunu belirtmiştir. Bu görüşü destekleyici ve açımlayıcı tartışmalar, Bademci (2001a; 2001b; 2004; 2005a; 2005b; 2005c; 2005d), Ebel ve Frisbie (1991), Thompson (1994a; 2001), Thompson ve Vacha-Haase (2000) ve Vacha-Haase'de (1998) vardır.

[Güvenirliğin testin değil, testten elde edilen ölçümlerin bir özelliği olduğu şeklindeki düşünce tarzının ipuçları, bu hususta ki çok çok az sayıdaki çalışmadan ve [mümkün olduğunca] tarihsel bir akış dikkate alınarak verilmeye çalışılmaktadır. Test ölçümlerinin güvenirliği ve de güvenirliğin testin kendisinin değil, test ölçümlerinin [belirli] bir grubunun bir özelliği olduğunu vurgulama hususunda, Gronlund ve Linn (1990) ve Ebel ve Frisbie'nin (1991) derli toplu çalışmaları son derece önemli bir yer tutsa da, unutulmaması ve gözden kaçırılmaması gereken bir nokta, güvenirliğin eldeki veriler veya ölçümlerin bir özelliği olduğunu ısrarla vurgulayan -ve alanın en önde gelen süreli yayınlarından olan *Educational and Psychological Measurement* dergisinin 1995-2003 yılları arasında editörlüğünü de yapan- Thompson'un (1994a; 1994b), ölçme ve ölçmeyi yakından ilgilendiren değişik konularda, yurt dışındaki köklü reform hareketinin -*özellikle 1994 yılından bu yana*- gerçek başlatıcısı, öncüsü olduğu ve de düşünceleriyle, reform hareketi üzerinde de oldukça etkili bir bilim adamı durumunda bulunduğu söylenebilir, ileri sürülebilir].

Henson ve Thompson (2002), Gronlund ve Linn'in (1990) yukarıda da belirtilen görüşünün, American Educational Research Association, American Psychological Association ve National Council on Measurement in Education (AERA/APA/NCME) test etme standartlarının, Standart 2.l ve 2.2'sine yansımış olduğunu ifade etmiştir. Benzer düşünce APA Task Force on Statistical Inference (Wilkinson ve APA Task Force on Statistical Inference, l999) tarafından güvenirlik, sınavı alanların belirli bir evreni için bir test üzerindeki ölçümlerin bir özelliğidir şeklinde ifade edilmiştir.

Güvenirlik, Ölçümlerin Bir Özelliğidir.

Güvenirliğin ölçümlerin bir özelliği olduğuna dair belki de en çarpıcı açıklama Crocker ve Algina'dan (1986) gelmiştir. Crocker ve Algina (1986) güvenirlik katsayısını etkileyen faktörlerden grup bağdaşıklığını (homojenliğini) tartışmış ve denencel bir örnek vererek, güvenirliği, sınavı alanların belirli bir grubu için bir test üzerindeki ölçümlerin bir özelliği şeklinde ifade etmiştir. Bir başka söyleyişle güvenirlik, sınava giren belirli bir gruba uygulanmış bir testten elde edilmiş ölçümlerin bir özelliğidir. Yani güvenirlik, test sonuçlarının bir özelliğidir (Livingston, 1988). Livingston (1988), test sonuçlarının güvenirliğinin ise, testi alan öğrencilerin grubuna bağlı olacağına dikkat çekmiştir.

Grup Bağdaşıklığı Veya Ayrışıklığı, Güvenirliği Etkileyen Bir Faktördür.

Belirli bir grupta testi alan öğrencilerin veya kişilerin kendileri, ölçümlerin güvenirliğini etkilemektedir. Hal böyle iken, testin bir gruba uygulandığını dikkate almaksızın testin güvenirliğinden bahsetmek anlamsız olacaktır.

Güvenirlik, varyans tarafından yönlendirilmektedir; daha büyük ölçümler varyansı, daha büyük ölçümler güvenirliğine olanak tanır (Thompson, 1994a). Böylece daha ayrışık (heterojen) örneklemler sıklıkla daha çok değişken ölçümlere ve bu durumda daha yüksek güvenirliğe yol açar (Thompson, 1994a). Bu durumda aynı ölçme aracı veya test, daha bağdaşık (homojen) ya da daha ayrışık öğrencilerden oluşan gruplara uygulandığında, birbirlerini onamayan ölçümler güvenirliği ortaya çıkacaktır.

Bu durumun mantığı, Klasik Kuram (Suen, l990), Klasik Güvenirlik Modeli (Thorndike, l982), Klasik Test Kuramı (Pedhazur ve Schmelkin, 1991), Klasik Test Kuram Modeli (Lord ve Novick, l968), Klasik Gerçek Ölçüm Modeli (Crocker ve Algina,l986), Klasik Gerçek Ölçüm Kuramı (Allen ve Yen, l979) gibi adlandırılan ölçme kuramının bazı eşitliklerinden yararlanılarak ve bu kuramın derinliğine girilmeden açıklanabilir.

Güvenirlik katsayısı matematiksel olarak aşağıdaki şekilde tanımlanabilir (Allen ve Yen, l979).

$\rho_{XX'} = \sigma^2_T / \sigma^2_X$
$\rho_{XX'}$ = güvenirlik katsayısı (X ve X' paralel ölçmeler)
σ^2_T = gerçek ölçüm varyansı
σ^2_X = gözlenmiş ölçüm varyansı

Güvenirlik katsayısı, gerçek ölçüm varyansının gözlenmiş (toplam) ölçüm varyansına oranıdır (Allen ve Yen, l979; Nunnally ve Bernstein, l994).

Güvenirlik katsayısıyla ilgili yukarıda verilen formül, öteki biçimde de yazılabilir (Allen ve Yen, l979; Pedhazur ve Schmelkin, 199l).

$\rho_{XX'} = \sigma^2_T / \sigma^2x = \sigma^2x - \sigma^2_E / \sigma^2 = 1 - \sigma^2_E / \sigma^2x$
σ^2_E = hata ölçüm varyansı

$\rho_{XX'} = 1 - \sigma^2_E / \sigma^2x$ formülü, diğer şeyler eşit olmak üzere, daha ayrışık gruptan daha yüksek güvenirlik elde edileceğini açıklayıcı niteliktedir (Allen ve Yen, l979; Mehrens ve Lehmann, 1991). Aynı test, benzer büyüklükte biri bağdaşık diğeri ayrışık iki gruba uygulansın. Hata ölçüm varyansı eşit olma sayıltısı altında, ayrışık gruptaki gözlenmiş ölçüm varyansı, bağdaşık gruptaki gözlenmiş ölçüm varyansından daha büyük olacaktır. Çünkü ayrışık gruptaki ölçümler, bağdaşık gruptaki ölçümlere göre çok daha değişken olacaktır. Böylelikle ayrışık gruptaki gözlenmiş ölçüm varyansı büyüyeceğinden, güvenirlikte buna bağlı olarak artacaktır. Bu durumu, belki de en iyi Dawis (1987:486) açıklamıştır; "çünkü güvenirlik, aracın olduğu kadar, örneklemin de bir fonksiyonudur. [Zira,] güvenirlik, tasarlanmış hedef evrenden [alınmış] bir örneklem üzerinde değerlendirilmektedir", ancak Dawis'inde (1987:486) ifade ettiği gibi bu nokta, "bazen gözden kaçırılmıştır."

Ölçüm Güvenirliği, Örneklemden Örnekleme Değişir.

Örneklem özellikleri ölçüm güvenirliğini etkileyebilmekte (Henson, Kogan ve Vacha-Haase, 2001), bir testin veya ölçme aracının uygulandığı örneklemin bağdaşık ya da ayrışık olması, ölçüm güvenirliğinin azalmasına veya artmasına neden olmaktadır. Bir başka ifadeyle ölçüm güvenirliği, örneklemden örnekleme değişmektedir (Capraro ve Capraro, 2002). Aynı test, bağdaşık veya ayrışık örneklemlere uygulandığı zaman güvenirliğe ilişkin

farklı sonuçlar doğurabilecektir. Hal böyle iken "test güvenilirdir" ya da "testin güvenirliği" demek ve güvenirliği, testin veya aracın bir özelliği gibi ima veya ifade etmek uygun değildir, doğru değildir.

Ölçüm güvenirliğinin, örneklemden örnekleme değiştiğiyle ilgili, bir başka söyleyişle, aynı testin, bağdaşık veya ayrışık örneklemlere uygulandığı zaman ölçüm güvenirliğine ilişkin farklı sonuçlar doğurabileceğine ilişkin bazı yorumlar, Cronbach'ın (1951) alfa (α) katsayısından yararlanılarak da yapılabilir.

Test ölçümlerinin güvenirliğini kestirmenin yöntemlerinden birisi de, alfa katsayısı yöntemidir (Linn ve Miller, 2005; Mehrens ve Lehmann, 1991). Alfa katsayısı aşağıdaki formül yoluyla hesaplanmaktadır (Crocker ve Algina, 1986; Henson, 2000b; Reinhardt, 1996).

$\alpha = k / (k-1)*[1- (\Sigma\sigma^2_i / \sigma^2_T)]$
k = test üzerindeki madde sayısı
σ^2_i = i madde ölçüm varyansı [ya da bir madde üzerindeki bir grup bireyden elde edilen ölçümlerin varyansı]
$\Sigma\sigma^2_i$ = i madde ölçüm varyanslarının toplamı
σ^2_T = toplam test ölçümlerinin varyansı

Alfa katsayısı, "…madde varyansları toplamı, madde güçlüğü ve toplam test ölçüm varyansı tarafından etkilenmektedir" (Helms, 1999: 10). "Reinhardt (1996), …toplam test ölçüm varyansının alfa katsayısı üzerinde en büyük etkiye sahip olduğunu göstermiştir"(Bulunduğu yer, Helms, 1999: 10). Daha küçük toplam test ölçüm varyansı, daha küçük alfa katsayısına, daha büyük toplam test ölçüm varyansı ise, daha büyük alfa katsayısına vesile olmaktadır (Arnold, 1996; Helms, 1999). Zira klasik test kuram güvenirlik kestirimleri toplam test ölçüm varyansı tarafından (Capraro, Capraro ve Henson, 2001), toplam test ölçüm varyansı da, sınavı alan grubun ne derece bağdaşık ya da ayrışık olmasından çokça etkilenmektedir (Helms, 1999). Eğer bir test, bağdaşık [homojen] bir gruba verilirse, toplam test ölçümü içindeki değişkenlik azalacak, dolayısıyla alfa katsayısı küçülecek, aynı test daha ayrışık bir gruba verilirse toplam test ölçümü içindeki değişkenlik artacak, dolayısıyla alfa katsayısı da büyüyecektir (Arnold, 1996; Helms, 1999). Böylelikle aynı test farklı gruplara verildiğinde, farklı güvenirlik katsayıları elde edilmiş olacaktır. Bu durum bir sefer daha göstermiştir ki, testler güvenilir değildir ve de "testin güvenirliği" veya "test güvenilirdir" ifadelerini kullanmak da, söylemek de doğru değildir.

Bu durumda aynı ölçme aracı [veya test], daha ayrışık ya da daha bağdaşık öğrencilerden oluşan gruplara ya da örneklemlere uygulandığında, birbirlerini onamayan ölçümler güvenirliği ortaya çıkacaktır (Thompson, 1994a). Tüm bu bulgular, güvenirliğin, örnekleme, dolayısıyla örneklemden elde edilen ölçümlere bağlı olduğunu göstermektedir (Guthrie, 2000). Buradan, güvenirliğin testlerin değil, eldeki verilerin veya ölçümlerin bir özelliği olduğu (Thompson, 1999) ifade edilebilir. O halde güvenirliği, ölçümlerin değil de, testlerin veya ölçme araçlarının bir özelliği olarak kabul etmek, araştırmalarda bu "bilinçsiz paradigmatik inanç" (Cousin ve Henson, 2000:6) doğrultusunda hareket etmek ve "bir ölçme aracına işaret ettiği zaman kullanılan 'test güvenilirdir' veya 'testin güvenirliği' ifadelerini" (Guthrie, 2000) kullanmak, doğru değildir.

Cronbach'ın alfa katsayısıyla ve ölçüm güvenirliğini etkileyen bazı faktörlerle ilgili olarak Arnold (1996), Cousin ve Henson (2000), Dawson (1997), Henson (2000a), Henson (2000b) ve Reinhardt'ın (1996), çalışmaları, etkili çalışmalar olarak göze çarpmaktadır. [Alfa 0 ile 1 arasında değerler alır gibi bir mitin aksine, *Cronbach alfa katsayısının negatif, hatta –1'den daha küçük değerler alabileceği* (örneğin, α = –7.00), Türkiye'de ilk defa Bademci'nin (2001a; 2002; 2005a; 2005d) çalışmalarıyla gündeme getirilmiştir.]

3. ARAŞTIRMALARDA ÖLÇME İLE İLGİLİ BAZI BÜYÜK YÖNTEMTEMBİLİM (METHODOLOGY) HATALARI

3.1. Doğru Olmayan Biçimde Kısaltarak İfade Etme

"Test ölçümlerinin güvenirliği" gibi uzun ama doğru bir ifadenin "testin güvenirliği" gibi kısaltılarak kullanılmasının sakıncaları ve zararları Thompson (1994a; 1994b), Thompson ve Vacha-Haase (2000) ve Vacha-Haase'de (1998) tartışılmıştır. Kısaltılarak, hatalı şekilde "testin güvenirliği" biçiminde ifade etmenin ve iletişim kurmanın zararlı sonuçları olabileceği düşünülen hususlardan *en az* altısı, Baugh (2002), Capraro ve Capraro (2002), Henson ve Thompson (2002), Kieffer, Reese ve Thompson (2001), Onwuegbuzie ve Daniel (2000), Shields ve Caruso (2004), Thompson (1994a; 1994b), Thompson (2001), Thompson ve Snyder (1998),

Thompson ve Vacha-Haase (2000), Vacha-Haase (1998), Vacha-Haase, Kogan ve Thompson (2000), Vacha-Haase, Ness, Nilsson ve Reetz (1999), Vacha-Haase, Kogan, Tani ve Woodall (2001) ve Whittington'dan da (1998) yararlanılarak şu şekilde sıralanabilir; 1) güvenirliğin testlerin bir özelliği olduğuna ve dolayısıyla testlerin güvenilir olduğuna inanılarak, bir başkasına [ya da kendisine] ait olan ve daha önceden yayımlanmış veya yayımlanmamış çalışmalardaki ya da test elkitaplarındaki güvenirlik katsayılarının, hiç bir hesap yapılmaksızın kendi [mevcut] çalışmalarında aynen kullanılması, rapor edilmesi, 2) güvenirlik katsayılarının örneklemden örnekleme değiştiği unutularak, bir testten elde edilen ölçümler için hesaplanan güvenirlik katsayını, [eğer bir de tatmin edici düzeydeyse,] hatalı biçimde testin güvenirliği olarak kabul edip ve o testin veya ölçme aracının güvenilir olduğu ön sayıltısıyla, aynı araçla, sürekli biçimde ve uzun yıllar boyunca çeşitli çalışmalarda veri toplanılmasına çalışılması, 3) 'testler güvenilirdir' şeklindeki hatalı söylemin, yanlış algılamalar yaratması, böylelikle araştırmalarda ve öğretimde kullanılmak üzere her düzeyde öğrenciler için ölçme eğitimi ve eğitime duyulan gereksinimin gittikçe azalması, lisans, yüksek lisans ve doktora eğitim programlarındaki ölçme konularında çok büyük bir *ölçme boşluğu*nun meydana gelmesi; dolayısıyla, ölçme kuramları, güvenirliğin özü gibi konuların neredeyse tamamıyla göz ardı edilmesi, 4) pek çok insanın ölçüm güvenirliğini ve güvenirliğin özünü [hala] anlayamaması ya da sıklıkla yanlış anlamış olması 5)güvenirlik kestirimleriyle ilgili olarak, yaptıkları çalışmalara ve elde ettikleri verilere yönelik olarak güvenirlik katsayılarının araştırmacılarca rapor edilmemesi veya eksik rapor edilmesi ve 6) ölçüm güvenirliği üzerinde düşünmede ve yorumlamada başarısız olunmasıdır.

Bilim adamları ve araştırmacılar tarafından doğru olmayan biçimde kısaltarak kullanma şekline ilişkin örnekler, eğitim bilim çevrelerince alanında önemli olduğu vurgulanan ulusal süreli yayınlarından olan *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* [Yıl 2004, Sayı 26], *Türk Eğitim Bilimleri Dergisi* [Yıl 2004, Sayı 4] ve *Eğitim Araştırmaları Dergisinin* [Yıl 2005, Sayı 18], belirtilen [*en son*] sayılarındaki tüm makalelerin incelenmesi sonucunda Çizelge 1, 2 ve 3'de gösterilmiştir.[Adı geçen bu üç ulusal hakemli dergiden incelenmiş ve alıntı yapılmış makalelerin listesi, *EK A*'da sunulmuştur.]. Çizelge 4 ve 5'deki veriler için de, yine adı geçen bu üç ulusal düzeydeki hakemli [ya da inceleme kurullu] dergiden yararlanılmıştır. Bu üç dergiden biri olan *Eğitim Araştırmaları Dergisi*, uluslar arası düzeyde de makale kabul etmektedir. Bu üç derginin incelenmesi esnasında karşılaşılan birkaç nitel

Çizelge 1. Kısaltarak Hatalı İfade Etme Örnekleri (*Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* Yıl 2004, Sayı 26. –en son ve tüm sayı-)

"Ölçeğin Cronbach Alpha Güvenirlik Katsayısı 0,92 olarak hesaplanmıştır." (Altınok, 2004:3)

"Testin Cronbach alpha güvenirlik katsayısı 0.71 olarak bulunmuş..." (Bilgin ve Geban, 2004: 14)

"...anahtarla puanlama yönteminin güvenirliğini gösteren korelasyon katsayısının 0,45...olduğu görülmektedir." (Çetin ve Kelecioğlu, 2004: 24-25)

"The reliability of this scale was found as. 69" (Çetin, Ertepınar ve Geban, 2004: 29)

"The internal reliability of the survey was calculated by using Cronbach's Alpha formulae and found 0.95" (Demirci, 2004: 36)

"Bağımlılık eğilimi ölçeğinin güvenirlik katsayısı r= .68'dir." (Gürsoy, Aral, Ayhan ve Aydoğan, 2004: 64)

"Uzmanların güvenirliğini saptamak için... Birinci puanlamada uzman güvenirliği katsayısı r=0,71..." (Mutlu, Demirhan ve Şahin, 2004: 103-104)

"...çoktan seçmeli test maddelerinin iki kategorili (1,0) ve ağırlıklı (1,2,3,4) puanlama yöntemlerinin testin güvenirlik ve geçerliğine etkisi..." (Özdemir, 2004:117).

Çizelge 1'in Devamı

"Reliability analysis of the Instructional Management scale produced an alpha of .71." (Savran ve Çakıroğlu,2004:128)
"...test ve anket için Cronbach alpha güvenirlik katsayıları sırasıyla .74 ve .86 olarak hesaplanmıştır." (Sencar ve Eryılmaz, 2004:145)
"Testin geçerliği yüksektir ve tutarlığı (KR 21) 0,84 olarak bulunmuştur... Kullanılan kimya tutum ölçeğinin güvenirliği 0,88'dir." (Sepet, Yılmaz ve Morgil, 2004:151)
"The internal consistancy of the scale was determined to be .87 using Cronbach alpha." (Tuncer, Sungur, Tekkaya ve Ertepınar, 2004:169)
"The reliability (Cronbash's alpha) of the instrument was found to be 0,79." (Uzuntiryaki, Bilgin ve Geban,2004:184)
"...iki ayrı güvenirlik çalışması sonucunda ölçeğin Cronbach Alfa güvenirlik katsayısı "Aile Desteği" alt ölçeği için 0.88 ve 0.92 olarak bulunmuştur." (Yıldırım, 2004:223).

araştırma kapsam dışı bırakılmış olup, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* ve *Eğitim Araştırmaları Dergisinde* yabancı dilde yayımlanmış olan makalelerden yapılmış olan aktarımlar, yayımlanmış olduğu yabancı dildeki ifadesiyle aynen sunulmuştur.

Çizelge 2. Kısaltarak Hatalı İfade Etme Örnekleri (*Türk Eğitim Bilimleri Dergisi,* Yıl 2004, Sayı 4. –en son ve tüm sayı-)

"...geçerlik ve güvenirlik çalışması yapılan ölçeğin cronbach alfa güvenirlik katsayısı r= .81'dir." (Bozkurt Bulut, 2004:446)
"Laboratuvar şartları ile ilgili bölüm için 0,8260...laboratuvar uygulamalarındaki sınırlılıklar için 0,7770 olarak güvenirlik katsayısı ve anketin tamamı için coranbach-alfa katsayısı 0,7827 olarak bulunmuştur." (Uluçınar, Cansaran ve Karaca, 2004:467)

Çizelge 3. Kısaltarak Hatalı İfade Etme Örnekleri (*Eğitim Araştırmaları Dergisi,* Yıl 2005, Sayı 18. –en son ve tüm sayı-)

"Özgün ölçeğin test-tekrar test güvenirliği .74, faktörlerin güvenirlik katsayıları ise, .85 ile .63 arasında bulunmuştur... Ölçeğin madde analizine dayalı olarak hesaplanan Cronbach alfa iç-tutarlılık katsayıları birinci faktör için .83,... ölçeğin bütünü için ise .71 olarak bulunmuştur." (Deryakulu ve Büyüköztürk, 2005:59)
"Tutum ölçeğinin Cronbach-alfa değeri 0.90 olarak hesaplanmıştır." (Ekici, 2005:74)
"The analysis of the results showed a good and an acceptable level of reliability of the test (0.82)... The reliability of the test is 0.82." (Sharkova, 2005:142,146)

Çizelge 3'ün Devamı

"...araştırmacı tarafından geliştirilen anketin güvenilirliği Cronbach Alpha değeri olarak 0.7800 hesaplanmıştır." (Şahin, 2005:175)
"Anketin güvenilirliği ise Likert tipi sorular için hesaplanan Cronbach's Alpha katsayısı ile 0,784 olarak belirlenmiştir." (Şen ve Özgün-Koca, 2005:189)
"Ölçeğin geçerlik ve güvenirliği yeterli bulunduğundan..." (Üstüner, 2005:207)

3.2. Güvenirlik Katsayılarını Yanlış Yorumlama

Onwuegbuzie ve Daniel (2000), bazı araştırmacıların .70 olarak verilen [X testinden elde edilen] bir ölçümlerin güvenirlik katsayısını, aracın kendisi %70 güvenilir şeklinde ve doğru olmayan biçimde yorumladıklarını ifade etmişlerdir. Güvenirlik katsayısıyla ilgili buna benzer hatalı yorumlar yurt içindeki çalışmalarda da görülmektedir; örneğin, Büyüköztürk (2004:164 ve 2005:170) kitabında güvenirlik katsayısı ile ilgili yaptığı yorumunda, "Güvenirlik katsayısı .80 olan bir test için bireyler arası gözlenen test puanlarındaki farkların %80 oranında gerçek farkları, %20 oranında ise hatayı yansıttığı söylenebilir" şeklinde bir ifade kullanmıştır ve Büyüköztürk'ün (2004:164 ve 2005:170) güvenirlik katsayısı ile ilgili yaptığı bu yorumu yanlıştır. Çok daha doğru yorumlar, Bademci (2005a), Crocker ve Algina (1986) ve Onwuegbuzie ve Daniel'da (2000) vardır. Klasik test kuramında, güvenirlik katsayısı [ρ_{XX}'], matematiksel olarak, gerçek ölçüm varyansının gözlenmiş ölçüm varyansına oranı biçiminde tanımlanmıştır (Allen ve Yen,1979; Crocker ve Algina,1986; Helmstadter,1964; McDonald, 1999; Nunnally ve Bernstein,1994; Thompson ve Vacha-Haase,2000) ve aşağıdaki biçimde de ifade edilebilir;

$\rho_{XX}' = \sigma^2_T / \sigma^2x$
ρ_{XX}' = güvenirlik katsayısı (X ve X' paralel ölçmeler)
σ^2_T = gerçek ölçüm varyansı
σ^2x = gözlenmiş ölçüm varyansı.

Örneğin, X testi ölçümlerinin test-tekrar test güvenirliği .81 olsun. Bu aşamadan sonra, güvenirlik katsayısı ile ilgili yapılan yorumlar, uygun ve doğru olmalıdır. Birinci yorum, sınavı alan bu grup için, gözlenmiş ölçüm varyansının %81'i, gerçek ölçüm varyansına atfedilebilir [ya da dayandırılabilir] veya toplam ölçüm varyansının en azından %81'i, gerçek ölçüm varyansı nedeniyledir şeklinde yapılabilir. İkincisi, ikinci test üzerindeki gözlenmiş ölçüm varyansının ($.81^2$), veya [$\rho_{xx}'^{\,2}$ =] % 65'i, birinci test üzerindeki gözlenmiş ölçümlerin varyansından yordanmış olabilir biçiminde söylenebilir. Sonuncu yorum ise, sınavı alanlara dair, gerçek ölçümler ve gözlenmiş ölçümler arasındaki korelasyon [ρ_{XT}], $\sqrt{}$.81 veya .90'dır şeklinde yapılabilir (Bademci, 2005a; Crocker ve Algina,1986; Onwuegbuzie ve Daniel, 2000; Thompson ve Vacha-Haase, 2000). [Sonuncu yorum için gerekli bir not: Güvenirlik katsayısı, gözlenmiş ölçümler ve gerçek ölçümler arasındaki korelasyonun karesidir, $\rho_{XX}' = \sigma^2_T / \sigma^2x = \rho^2_{XT}$, (Lord ve Novick, 1968: 61; McDonald, 1999: 66)].

3.3. Güvenirlik Katsayılarının Rapor <u>Edilmemesi</u>

Vacha-Haase, Ness, Nilsson ve Reetz (1999) yaptıkları bir araştırmada, *Journal of Counseling Psychology (JCP), Psychology & Aging (P&A)* ve *Professional Psychology: Research and Practice (PP)* adlı üç derginin, 1990'dan, 1997'ye yayımlanmış makalelerini incelenmişler ve toplam 839 makalenin %36.4'ünde makale yazarlarınca güvenirliğin ifade edilmediğini, bu makalelerde ölçüm güvenirliğinden ziyade, hatalı biçimde test güvenirliği üzerinde odaklanıldığını belirtmişlerdir. *Araştırmalarda, güvenirlik çalışmaları için örneklem büyüklüğü, güvenirlik kestirim yöntemleri ve hesaplanılan ölçüm güvenirlik katsayıları* [*her çalışma için*] *mutlaka rapor edilmelidir*. Ölçüm güvenirlik katsayılarının rapor edilmemesi, araştırmalardaki ciddi bir yöntembilim hatasıdır ve incelenmiş olan ulusal üç hakemli dergide, bu hususla ilgili hatalı örnekler, Çizelge 4'de verilmiştir.

Çizelge 4. Güvenirlik Katsayılarının Rapor <u>Edilmemesi</u> Örnekleri (*Hacettepe* Üniversitesi *Eğitim Fakültesi Dergisi*, Yıl 2004, Sayı 26. –en son ve tüm sayı-; *Türk Eğitim Bilimleri Dergisi*, Yıl 2004, Sayı 4. –en son ve tüm sayı-; *Eğitim Araştırmaları Dergisi*, Yıl 2005, Sayı 18. –en son ve tüm sayı-)

"Data collection methods included mathematics pre-and posttests, follow-up interviews with all students after the mathematics posttest, computer-based clinical interviews at the end of the treatment with 5 students from each experimental group, and classroom and computer lab observations." (Özgün-Koca, 2004: 85)

"Veriler, 'Okullarda Bilişim Teknolojilerinin Yayılımı Anketi' (Mumcu,2004) aracılığıyla toplanmıştır. Verilerin çözümlenmesinde yüzde ve frekans kullanılmıştır (Kuşkaya Mumcu ve Koçak Usluel, 2004: 92)

"Öğrencilerin internet erişim olanakları ve kullanım amaçlarını belirlemek amacıyla, araştırmacılar tarafından bir anket geliştirilmiş ve bu veriler bu anket aracılığıyla toplanmıştır." (Orhan ve Akkoyunlu, 2004:110)

"Ölçme araçları otuzar kişilik iki gruba uygulanarak geçerlilik-güvenilirliği test edilmiştir." (Şimşek, 2004:505)

"Öğretmen adaylarının çeşitli matematiksel kavramlara yönelik problem çözme ve özellikle de problem kurma becerilerinin hangi durumda olduğunu belirlemek üzere, araştırmacılar tarafından açık uçlu tipte 5 sorudan oluşan bir test hazırlanmıştır." (Dede ve Yaman, 2005: 45)

"Ödevlerin değerlendirilmesinde kullanılan ölçme aracının geçerliği ilgili alanda üç öğretim üyesinin görüşleri alınarak sağlanmıştır. Güvenirlik çalışması ise, I. Katılımcı gruba ön çalışma niteliğinde uygulanan ölçme aracında belirlenen 30 kriterin tüm öğrenciler için aynı şekilde anlaşılan 24 kritere indirgenerek ölçeğe son hali verilmiş ve bu gruba ait veriler 24 kriter üzerinden puanlandırılmıştır" (Karamustafaoğlu ve Akdeniz, 2005:130)

3.4. Güvenirlik Katsayılarının Rapor Edilmesinde Muğlak İfadeler Kullanma

Sıklıkla, örnekler vermek gerekirse, "...X ölçme aracı Y (2000) tarafından geliştirilmiştir. X ölçme aracının Cronbach alfa güvenirlik katsayısı 0.86 bulunmuştur" ya da " X ölçeğinin Türkçe'ye çeviri ve uyarlaması ile geçerlik ve güvenirlik çalışması Y (2000) tarafından yapılmıştır. X ölçeğinin test-tekrar test güvenirlik katsayısı .85'dir" gibi ifadelerle güvenirlik katsayıları rapor edilmektedir; bu ve benzeri muğlak ifadelerle güvenirlik katsayılarını rapor etmek doğru değildir. Bu şekilde rapor edilen güvenirlik katsayıları, araçları geliştirenlerin çalışmalarına mı ait, yoksa önceki çalışma sonuçlarını rapor edenlerce, elde edilen yeni verilere yönelik olarak tekrar mı hesaplandı, bu ve benzeri hususlar belli değildir ve bu tür ifadeler yanıltıcıdır. Eğer ilgili çalışmaya ait veriler üzerinde konuşulduğuna dair herhangi bir ifade yoksa, "...bulunmuştur", "...saptanmıştır", "...açıklanmıştır" veya "bulundu" gibi yanıltıcı ifadelerin yerine, "...hesaplanmıştır" ya da "...kestirilmiştir" ifadelerinin kullanılması daha doğru olacaktır. Örneğin, " Bu çalışmada, X ölçme aracından elde edilen ölçümlerin iç tutarlılığı Cronbach α ile hesaplanmıştır ve α=.87'dir" gibi bir ifade veya benzeri ifadeler (bkz., Bademci, 2004:370) kullanılması daha yerinde olacaktır. Zaten, *ölçüm güvenirliğinin her seferinde* [*her yeni çalışmada*] *yeniden hesaplanması ve rapor edilmesi gerekliliğinden dolayı*, bu maddede bahsedilen [muğlak ifadelerle ilgili] sorun da tamamıyla ortadan kalkacaktır.

3.5. Güvenirlik Hakkında Yanlış Anlamaya Yol Açacak Hatalı İfadeler Kullanma

Türkçe literatürde pek çok değişik biçimde bulunmakta olan ve güvenirlik hakkında yanlış anlamaya yol açacak hatalı ifadeler kullanmayla ilgili bir örnek, Balcı'nın (2004:100) bir kitabında da görülmektedir. Güvenirlikle ilgili olarak, Balcı (2004:100) kitabında, "Güvenirlik bir ölçeğin tutarlılığını gösterir; onun her zaman aynı sonuçları vereceğini belirtir. Bir araç güvenilirse ölçmek istediği özellik(leri) tutarlı biçimde ölçer." şeklinde ifadeler kullanmıştır ve Balcı'nın (2004:100) güvenirlikle ilgili bu ifadeleri birkaç yönden yanlıştır: Öncelikle, Balcı'nın (2004:100) "bir araç güvenilirse..." gibi başlayan bir ifadeyle, güvenirliği, aracın bir özelliği gibi belirtmesi doğru değildir. Çünkü, [daha önce de açıklandığı gibi] *bir ölçme aracı [veya bir test] güvenilir veya güvenilmez değildir* (Crocker ve Algina,1986; Rowley, 1976) ve de *güvenilir veya güvenilmez olan ölçümlerdir* (Kieffer, 1999). Kısaca, güvenirlik, testin [ya da ölçme aracının] kendisinin değil, elde edilmiş ölçümlerinin bir özelliği (Lane, White ve Henson, 2002), ölçümlerin bir fonksiyonudur (Capraro, Capraro ve Henson, 2001). Balcı'nın (2004:100), "güvenirlik bir ölçeğin tutarlılığını gösterir; onun her zaman aynı sonuçları vereceğini belirtir" biçimindeki ifadesi de hatalıdır. Çünkü, "bir ölçeğin... *her zaman* aynı sonuçları" verebilmesi pratikte mümkün görünmemektedir. Zira, *aynı ölçek* [ya da *ölçme aracı*], [bırakılsın her zamanı] sadece bir bağdaşık ve de bir ayrışık gruba uygulandığında dahi, aynı sonuçlar yerine, farklı sonuçlar [ölçümler], dolayısıyla da değişik güvenirlik katsayıları elde edilebilecektir (bkz., Thompson, 1994a). Ayrıca, Balcı'nın (2004:100) "güvenirlik bir ölçeğin tutarlılığını gösterir; onun her zaman aynı sonuçları vereceğini belirtir" ifadesindeki, güvenirlikle ilgili, "güvenirlik bir ölçeğin... her zaman aynı sonuçları vereceğini belirtir" şeklindeki vurgusu da doğru değildir. Zira, güvenirlik, "bir ölçeğin... *her zaman aynı sonuçlar* vereceğini" de belirtmez. Çünkü güvenirlik, bir ölçeğin [ya da ölçme aracının] kendisinin değil, bir ölçekten [ya da ölçme aracından] elde edilen ölçümlerin bir özelliğidir ve de ölçeği alan [ya da ölçek ya da ölçme aracı uygulanan] grubun bağdaşık ya da ayrışık olmasından da çokça etkilenmektedir ya da en kısa ve öz biçimiyle ölçüm güvenirliği, örneklemden örnekleme değişmektedir (Capraro ve Capraro, 2002). Güvenirlik hakkında, Balcı'nın (2004:100), "güvenirlik bir ölçeğin tutarlılığını gösterir; onun her zaman aynı sonuçları vereceğini belirtir" şeklindeki ifadesi ve vurgusu, bir başka yönden yine hatalıdır. Çünkü, güvenirlik *bir ölçeğin* tutarlılığını göstermez ve de güvenirlik, en geniş anlamıyla, sınavı alan kişinin [/kişilerin] *ölçümleri*nde tutarlılıkların ve/veya tutarsızlıkların miktarını belirtmeyi kapsar (Brennan, 2001).

3.6. Daha Önce Yapılmış Çalışmalardaki Güvenirlik Katsayılarını Rapor Etme

Kendi örnekleminden elde ettiği veriler için, bir başkasının ölçümlerinden hesaplanmış olan bir güvenirlik katsayısını yorumlamak ve de rapor etmek, dikkatsizce yapılmış olan ve son derece hatalı bir uygulamadır. *Ölçüm güvenirliğini, örneklemdeki deneklerin ya da kişilerin kendileri etkilemektedir* (Arnold, 1996; Thompson, 1994a) ve testi [veya ölçme aracını] [daha] uygulamadan testin [veya ölçme aracının] güvenirliğinden bahsetmek veya bir başkasının hesapladığı bir teste [ya da ölçme aracına] ait ölçümlerin güvenirlik katsayısını hatalı bir kabul ile testin [ya da ölçme aracının] kendi özelliği gibi kabul edip ve yine yanlış bir biçimde ve bilinçsizce bir kabul ile ölçümlerin ve onlardan elde edilen güvenirlik katsayılarının değişmeyeceğini kabul ederek (Reinhardt, 1996; Vacha-Haase, Ness, Nilsson ve Reetz, 1999) uygulama yapmaksızın bir test hakkında 'test güvenilirdir' ya da 'güvenilir testler' şeklinde ifadeler kullanmak ve kabul etmek, Thompson'un (1994a:839) ifadesiyle bir "oxymoron" [oxymoron; yan yana kullanılması imkansız ve kullanıldığında da saçma olan iki kelime, (Turgut; 2002)] olmaktadır. Zira, güvenirlik, yalnızca testin kendisinin bir fonksiyonu değil (Reinhardt, 1996), örneklemin de bir fonksiyonu (Dawis, 1987; Henson,2000a), bir başka ifadeyle güvenirlik, en azından test ve testi alanların her ikisinin de bir fonksiyonudur (Arnold, 1996). O halde bir başkasının kendi örneklemine uyguladığı bir ölçme aracından elde edilen ölçümlere ait bir güvenirlik katsayısını, bir diğerinin [ya da aynı kişinin] o güvenirlik katsayısını, aynen kendi örnekleminde [yeni çalışmasında] kullanması, yani **daha önce bir testten elde edilen ölçümlere yönelik olarak hesaplanmış güvenirlik katsayısını, bir başkasının** [ya da aynı kişinin] **o güvenirlik katsayısını o testin bir özelliği gibi kabul ederek, aynen kendi** [yeni] **araştırmasında kullanması,** [Türkiye'de de sıklıkla yapılmış olan ve hala da yapılan] **ciddi bir ölçme ve de yöntembilim hatasıdır**. Zira iyice bilinmelidir ki, güvenirlik, bir testin değil, sınava giren belirli bir gruba uygulanmış o testten elde edilmiş ölçümlerin bir özelliğidir (Bademci, 2004) ve ölçüm güvenirliği de, örneklemden örnekleme değişir (Capraro ve Capraro, 2002).

Yurt dışındaki yayımlanmış çalışmalarda olduğu kadar (Meier ve Davis, 1990; Thompson, 1994a; Thompson ve Snyder,1998; Whittington, 1998), Çizelge 5'de de görüleceği üzere, yurt içindeki yayınlanmış çalışmalarda da, bir başkasının kendi örneklemine uyguladığı bir ölçme aracından elde edilen ölçümlere ait bir güvenirlik katsayısını, bir diğerinin [ya da aynı kişinin] aynen kendi örnekleminde [yeni çalışmasında] [hatalı biçimde] kullandığı, yani daha önce bir testten [ya da ölçme aracından] elde edilen ölçümlere yönelik olarak hesaplanmış güvenirlik katsayısını, bir başkasının o güvenirlik katsayısını o testin [ya da ölçme aracının] bir özelliği gibi kabul ederek, aynen kendi [yeni] araştırmasında [çalışmasında] [doğru olmayan biçimde] kullandığı görülmektedir.

Çizelge 5. Güvenirlik Katsayılarını Daha Önceden Yapılmış Çalışmalardan Rapor Etme Örnekleri (*Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, Yıl 2004, Sayı 26. –en son ve tüm sayı-; *Türk Eğitim Bilimleri Dergisi*, Yıl 2004, Sayı 4. –en son ve tüm sayı-)

"Flanders, Anderson ve Amidon tarafından geliştirilmiş olan 'Bağımlılık Eğilimi Ölçeği'nin Türk çocuklarına adaptasyonu ile geçerlilik ve güvenilirlik çalışması Uluğtekin (1976) tarafından yapılmıştır... Bağımlılık eğilimi ölçeğinin güvenirlilik katsayısı r=.68'dir." (Gürsoy, Aral, Bütün Ayhan ve Aydoğan, 2004: 64)

"Veriler 'Okullarda Bilişim Teknolojilerinin Yayılımı Anketi' (Mumcu, 2004) aracılığıyla toplanmıştır." (Kuşkaya Mumcu ve Koçak Usluel, 2004: 92). [*Güvenirlik katsayısı rapor edilmemiştir.*]

"Bu yeterlilik maddeleri Mahiroğlu (2004) tarafından bir ölçeğe dönüştürülmüştür. Bu ölçek dörtlü Likert tipi bir ölçektir ve güvenilirlik katsayısı 0.98'dir. Bu araştırmada veri toplama aracı olarak Mahiroğlu tarafından geliştirilen bu ölçek kullanılmıştır." (Seferoğlu, 2004:133)

"Öğrencilerin kimya dersine karşı tutumlarını ölçmek için Geban vd., (1995) tarafından geliştirilen 15 maddelik likert tipi ölçek kullanılmıştır. Kullanılan kimya tutum ölçeğinin güvenirliği 0,88'dir." (Sepet, Yılmaz ve Morgil, 2004:151)

"Çocukların bilişsel gelişimleri Bracken Kavram Ölçeği ve McCarthy Beceri Değerlendirme Ölçeği (MSCA) ile değerlendirilmiştir." (Üstün, Akman ve Etikan, 2004:205). [*Güvenirlik katsayısı rapor edilmemiştir.*]

"Araştırmada öğrencilerin ailelerinden algıladıkları destek düzeyini belirlemek amacıyla Yıldırım (1997) tarafından geliştirilen Algılanan Sosyal Destek Ölçeği (ASDÖ)'nin 'Aile Desteği' alt ölçeği kullanılmıştır... Yıldırım (1997, 1999) tarafından yapılan iki ayrı güvenirlik çalışması sonucunda ölçeğin Cronbach Alfa güvenirlik katsayısı 'Aile Desteği' alt ölçeği için 0.88 ve 0.92 olarak bulunmuştur." (Yıldırım, 2004:223)

"Çetinkanat (1997) tarafından geçerlik ve güvenirlik çalışması yapılan ölçeğin cronbach alfa güvenirlik katsayısı r = .81'dir." (Bozkurt Bulut, 2004: 446)

Daha önceden hesaplanmış bir ölçüm güvenirlik katsayısını, hatalı olarak, kendi araştırmalarında kullanmayla ilgili tipik bir örnek, Gürsoy, Aral, Bütün Ayhan ve Aydoğan'ın (2004) yayımlanmış bir makalelerinde görülebilir; Gürsoy, Aral, Bütün Ayhan ve Aydoğan (2004: 64) araştırmalarında, Flanders, Anderson ve Amidon tarafından geliştirilmiş ve Türk çocuklarına adaptasyonu ile 'geçerlilik' ve 'güvenilirlik' çalışması Uluğtekin (1976) tarafından yapılmış olan "Bağımlılık Eğilimi Ölçeği"ni kullanmış olduklarını ifade etmişler ve "Bağımlılık Eğilimi Ölçeği"nin güvenirlik katsayısı'nın r=.68 olduğunu da çalışmalarında rapor etmişlerdir. Gürsoy, Aral, Bütün Ayhan ve Aydoğan (2004: 64) tarafından atıfta bulunulan Uluğtekin'in (1976: 124)

"Çocuk Yetiştirme Açısından Anababa Çocuk İlişkileri. Anababa Davranışlarıyla Çocuğun Saldırganlık ve Bağımlılık Eğilimi Arasındaki İlişkilerin Araştırılması" başlıklı doktora tezi incelendiğinde ise, "Bağımlılık eğilimi ölçeğinin yazarları tarafından Hoyt'un varyans analizi tekniğine göre hesaplanan güvenirlik katsayısı r= .68'dir" şeklinde yine ve bir başkasının çalışmasına daha atıf yapıldığı görülmektedir; bu atıf ise, doktora tezini hazırlamış olan Uluğtekin (1976:124) tarafından yapılmış ve de "Flanders, Anderson, Amidon, Ön. Ver., s.583" şeklindedir. Uluğtekin (1976) tarafından yapılmış bu atıf doğrultusunda, Flanders, Anderson ve Amidon'un (1961:583) "*Measuring Dependence Proneness in the Classroom*" başlıklı makalesi incelendiğinde ise, "...the reliability coefficient is .68..." ifadesi, ilgili sayfada [sayfa 583] görülebilmektedir. Oradan oraya atıf yapılarak aktarıldığı görülen ve Flanders, Anderson ve Amidon (1961:583) tarafından 43 yıl önce hesaplanmış ölçüm güvenirlik katsayısını [.68], ciddi bir yöntem hatası yaptıkları da söylenebilerek, Gürsoy, Aral, Bütün Ayhan ve Aydoğan'ın (2004: 64) aynen ve sorgulamadan kendi çalışmalarında kullandıkları ve Flanders, Anderson ve Amidon (1961:583) tarafından 43 yıl önce kestirilmiş ölçüm güvenirlik katsayısıyla, hatalı biçimde, kendi çalışmalarında elde edilen ölçümlerle ilgili çeşitli [istatistiklerle] yorumlara gittikleri de ifade edilebilir.

4. SONUÇ

Nunnally (1982), bilimsel topluluk içinde ölçmenin oynadığı [kritik] rolü [mükemmel biçimde] teşhis etmiştir (Vacha-Haase, Ness, Nilsson ve Reetz, 1999). Nunnally'e (1982) göre, bilim tekrar edilir [edilebilir] deneylerle ilgilenmektedir. Yine, Nunnally'e (1982) göre, eğer deneylerden elde edilmiş veriler ölçmenin random hataları tarafından etkilenmişse, sonuçlar aynen tekrarlanmaz, böylelikle de bilim, ölçü araçlarının güvenirliği vasıtasıyla sınırlandırılmış olmaktadır. Güvenirliğin, "ölçme sonuçlarının tesadüfi [random] hatalardan ne derece arınık olduğu" (Turgut, 1993: 23) şeklinde de tanımlanabildiği dikkate alınırsa, Nunnally'nin (1982) bu görüşünün, güvenirliğin önemini (Vacha-Haase, Ness, Nilsson ve Reetz, 1999) ve tekrarlanmış ölçmelerde, elde edilen ölçümlerin ölçme hatalarından [olabildiğince] arınık olması gerektiğini çok iyi vurguladığı açıkça görülebilmektedir.

Klasik ölçme kuramı, esasen bir '*büyük örneklem*' kuramıdır (Nunnally ve Bernstein, 1994) ve güvenirlik, "test ölçümlerinin istendik tutarlılığı veya tekrarlanabilirliği" şeklinde tanımlanabilir (Crocker ve Algina, 1986:105). O halde, bilim tekrar edilir [edilebilir] deneylerle ilgileniyorsa, güvenirlik ise, "test ölçümlerinin istendik tutarlılığı veya tekrarlanabilirliği" şeklinde tanımlanabiliyorsa ve ölçümlerin güvenirliğinin örneklemden örnekleme değiştiği ifade ediliyor ve biliniyorsa, bir ölçme aracından elde edilmiş ölçümleri ve o ölçümlerden hesaplanmış güvenirlik katsayısını değişmez kabul edip "test güvenilirdir" demek veya "testin güvenirliği şudur" şeklinde ifade etmek veya daha önceden yapılmış araştırmalardaki güvenirlik katsayılarını kendi [yeni] çalışmalarındaki verilerde kullanmak ve de rapor etmek, doğru değildir. Çünkü "bir test güvenilir veya güvenilmez değildir" (Crocker ve Algina, 1986: 144), güvenilir veya güvenilmez olan ölçümlerdir (Kieffer,1999) ve güvenirlik, testlerin değil, eldeki veriler veya ölçümlerin bir özelliğidir (Thompson,1994a; Thompson, 1999). Bir başka söyleyişle, güvenirlik, testin kendisinin değil, elde edilmiş ölçümlerinin bir özelliği (Lane, White ve Henson, 2002), ölçümlerin bir fonksiyonudur (Capraro, Capraro ve Henson, 2001).

Güvenirlik, aracın kendisine değil, bir bellilendirme (assessment) [veya ölçme] aracı ile elde edilmiş ölçümlere işaret eder (Linn ve Gronlund, 2000). Böylelikle, bir ölçme aracına [testin kendisine] işaret ettiği zaman kullanılan "test güvenilirdir" veya "testin güvenirliği" ifadelerini (Guthrie,2000) kullanmak ise, doğru değildir, uygun değildir (Thompson, 1994b; Thompson ve Vacha-Haase, 2000). Bu bakış açısıyla, gözden kaçırılmaması gereken önemli bir nokta, *güvenilir ölçümler* ile *güvenilir testler* terimlerinin eşanlamlılıktan uzak olduğudur (Vacha-Haase, Kogan, Tani ve Woodall, 2001). Türkiye'deki kimi öğretim elemanlarının, ölçme ile bağlantılı bu ve benzeri birtakım terimleri birbirlerinden ayırt edemedikleri de görülmektedir (Bademci, 2005c).

Yurt dışındaki çalışmalarda da yaygın kullanılan "test güvenilirdir" veya "testin güvenirliği" ya da "aracın güvenirliği" gibi hatalı biçimde kısaltarak ifade etme biçimleri, Türkiye'de ve Türk eğitim ve bilim topluluğunda da 1940'lardan bu yana kullanılmaktadır ve Bademci (2001a; 2001b; 2002; 2004; 2005a; 2005b; 2005c) yaklaşık 60 yılı aşkın bir süredir süregelen bu doğru olmayan kullanım biçimine ve güvenirliğin hatalı yorumlanış ve uygulama şekillerine ve güvenirliği, testin ya da ölçme aracının bir özelliği gibi kabul eden düşünme tarzına karşı çıkmış; güvenirliğin, testlerin ya da kullanılan ölçme araçlarının değil, ilgili araçlar veya testlerden elde edilen ölçümlerin ya da ölçme sonuçlarının bir özelliği olduğunu vurgulamış ve doğru ifade biçiminin de,

testlerin [veya testin] güvenirliği biçiminde değil, test ölçümlerinin güvenirliği [veya ölçüm güvenirliği] şeklinde olması gerektiğini ifade etmiş ve de paradigma değişikliği gerekliliğini vurgulayarak, bir *yeni* paradigmayı [adayını] da yine bilimsel kanıtlarıyla Türk eğitim ve bilim topluluğunun gündemine taşımıştır.

"Testin güvenirliği" ya da "test güvenilirdir" veya "aracın güvenirliği" gibi kısaltarak veya aynı metin içinde, güvenirlikle ilgili olarak tutarsız biçimde hem "test ölçümlerinin güvenirliği", hem de "testin güvenirliği" ifadelerini birlikte kullanmak da, doğru değildir. Bilim adamlarınca yaygın ve neredeyse ortak kullanılan "test güvenilirdir" şeklindeki bir dil ise, doğru değildir. Ancak, ölçmenin doğruluğunu tanımlamada [bir ülkede, ölçme konusundaki] en iyi uzmanlar tarafından [yıllarca] kullanılan dil, uygun değilse veya hatalı biçimde de kısaltılarak kullanılmakta ise, ölçüm güvenirliği hakkında düşünenlerin ve de uygun dili kullananların sayısı da muhtemelen çok çok az olacaktır (Thompson,1994b).

5. YORUM

Güvenirlik, testin kendisinin değil, ölçümlerin bir fonksiyonudur (Capraro, Capraro ve Henson, 2001). Bir başka ifadeyle güvenirlik, en azından test ve testi alanların her ikisinin de bir fonksiyonudur (Arnold, 1996). Örneklem özellikleri ise, ölçümleri ve güvenirliği etkileyebilmektedir (Capraro, Capraro ve Henson, 2001). Klasik test kuram güvenirlik kestirimleri toplam test ölçüm varyansı tarafından (Capraro, Capraro ve Henson, 2001), toplam test ölçüm varyansı da, sınavı alan grubun ne derece bağdaşık ya da ayrışık olmasından çokça etkilenmektedir (Helms,1999). Buradan, güvenirliğin, testin kendisinin değil, örneklemin özelliklerinin bir fonksiyonu olduğu da söylenebilir (Capraro, Capraro ve Henson, 2001). Ölçüm güvenirliği ise, örneklemden örnekleme değişir (Buhi, 2005; Capraro ve Capraro, 2002). Örneğin, aynı ölçek [test veya ölçme aracı], 100 farklı örnekleme uygulansa, 100 farklı güvenirlik katsayısı ortaya çıkabilir (Buhi, 2005). Hal böyle iken, güvenirliği testin [veya ölçme aracının] bir özelliği gibi kabul etmek, aynı hatalı düşüncenin uzantısı olarak o test ya da ölçme aracından elde edilen ölçümleri ve hesaplanmış güvenirlik katsayısını [katsayılarını] da değişmez gibi kabullenmek, dolayısıyla aynı ölçme aracının kullanıldığı önceki çalışmalardaki hesaplanmış ve rapor edilmiş güvenirlik katsayılarını, hesaplama yapmaksızın kendi çalışmalarında aynen kullanmak, bir ölçme aracına işaret eden "testin güvenirliği" veya "test güvenilirdir" ya da "ölçme aracının güvenirliği" ya da "ölçeğin güvenirliği" ifadelerini kullanmak, doğru değildir.

Paradigmalar, günlük yaşam ve bilimsel çalışmalarda rehberlik etmektedir (Thompson, 1994b). Paradigmalar, *kuramlar değil*, düşünme tarzları veya araştırma için örnekler veya modellerdir (Gage, 1963) ya da bir paradigma, bir bilimsel topluluğun üyeleri tarafından paylaşılan inançlar, değerler, teknikler bütünü olarak da ifade edilebilir (Kuhn,1970). Bademci (2001a; 2001b; 2002; 2004; 2005a; 2005b; 2005c) güvenirliğin, testin bir özelliği olduğu şeklindeki "düşünme tarzı"nın ve bunun uzantısı olduğu söylenebilen kısaltarak "testin güvenirliği" ve benzeri ifade tarzlarının hatalı olduğunu bilimsel veri ve kanıtlarıyla ortaya koymaya çalışmış ve bir *yeni* paradigmayı [adayını] da yine bilimsel kanıtlarıyla [60 yılı aşkın bir süre sonra] Türk eğitim ve bilim topluluğuna sunmuştur. Şüphesiz, yeni bir paradigmanın [ya da adayının] başlangıçta çok az taraftarı olabilir (Kuhn, 1970). Yine şüphesiz, çalışmalarını yeni paradigmaya uydurmayı beceremeyen veya uydurmak istemeyen çok az sayıda kişi de olabilecek veya kalabilecektir. Ancak, bu satırların yazarı Türk eğitim ve bilim topluluğunda işlerinde yeterlik sahibi, yeniliğe ve yeni bilgiye ve bilimsel kanıtlarıyla desteklenmiş yeni bir düşünme tarzını almaya açık, her yaş ve akademik düzeyde bilim adamlarının var olduğuna inanmaktadır.

6. EK A: ALINTI YAPILMIŞ MAKALELERİN LİSTESİ (ÜÇ ULUSAL HAKEMLİ DERGİ)

(*Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* [Yıl 2004, Sayı 26. –en son sayı-], *Türk Eğitim Bilimleri Dergisi* [Yıl 2004, Sayı 4. –en son sayı-], *Eğitim Araştırmaları Dergisi*, [Yıl 2005, Sayı 18. –en son sayı-] isimli üç ulusal hakemli dergideki incelenmiş ilgili makaleler [tüm kaynaklar])

Altınok, H. (2004). Öğretmenlerin Fen Öğretimine Yönelik Tutumlarına İlişkin Öğrenci Algıları ve Öğrencilerin Fen Bilgisi Dersine Yönelik Tutum ve Güdüleri. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 1-8.

Bilgin, İ. ve Geban, Ö.(2004). İşbirlikli Öğrenme Yöntemi ve Cinsiyetin Sınıf Öğretmen Adaylarının Fen Bilgisi Dersine Karşı Tutumlarına, Fen Bilgisi Öğretimi I Dersindeki Başarılarına Etkisinin İncelenmesi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 9-18.

Bozkurt Bulut, N. (2004). İlköğretim Sınıf Öğretmenlerinin İletişim Becerilerine İlişkin Algılarının Çeşitli Değişkenler Açısından İncelenmesi. ***Türk Eğitim Bilimleri Dergisi***, Cilt 2 (4), 443-452.

Çetin, B. ve Kelecioğlu,H. (2004). Kompozisyon Tipi Sınavlarda Kompozisyonun Biçimsel Özelliklerinden Kestirilen Puanların Anahtarla ve Genel İzlenimle Elde Edilen Puanlarla İlişkisi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 19-26.

Çetin,G.,Ertepınar, H. ve Geban, Ö.(2004). The Effect of Conceptual Change Approach on Students' Ecology Achievement and Attitude Towards Biology. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 27-32.

Dede, Y. ve Yaman, S. (2005). Matematik Öğretmen Adaylarının Matematiksel Problem Kurma ve Problem Çözme Becerilerinin Belirlenmesi. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 41-56.

Demirci, N. (2004). Students' Attitudes Toward Introductory Physics Course. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 33-40.

Deryakulu, D. ve Büyüköztürk, Ş. (2005). Epistemolojik İnanç Ölçeğinin Faktör Yapısının Yeniden İncelenmesi: Cinsiyet ve Öğrenim Görülen Program Türüne Göre Epistemolojik İnançların Karşılaştırılması. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 57-70.

Ekici, G. (2005). Lise Öğrencilerinin Çevre Eğitimine Yönelik Tutumlarının İncelenmesi. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 71-83.

Gürsoy,F., Aral,N., Bütün Ayhan, A. ve Aydoğan, Y. (2004). Annesi Çalışan ve Çalışmayan Çocukların Bağımlılık Eğilimlerinin İncelenmesi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 62-71.

Karamustafaoğlu,O. ve Akdeniz, A.R. (2005). Özel Öğretim Yöntemleri Uygulamalarında Fizik Öğretmen Adaylarının Gerçekleştirdikleri Etkinliklerin Değerlendirilmesi. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 128-141.

Kuşkaya Mumcu, F. ve Koçak Usluel, Y. (2004). Mesleki ve Teknik Okul Öğretmenlerinin Bilgisayar Kullanımları ve Engeller. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 91-99.

Mutlu, Ş., Demirhan, G. ve Şahin,R. (2004). Hentbolde Sıçrayarak Atış Tekniğinin Öğretiminde Görsel Dönüt ve Materyal Kullanımının Erişiye Etkisi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 100-106.

Orhan, F. ve Akkoyunlu, B. (2004). İlköğretim Öğrencilerinin İnternet Kullanımları Üzerine Bir Çalışma. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 107-116.

Özdemir, D. (2004). Çoktan Seçmeli Testlerin Klasik Test Teorisi ve Örtük Özellikler Teorisine Göre Hesaplanan Psikometrik Özelliklerinin İki Kategorili ve Ağırlıklandırılmış Puanlanması Yönünden Karşılaştırılması. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 117-123.

Özgün-Koca, S.A. (2004). The Effects of Multiple Linked Representations on Students' Learning of Linear Relationships. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 82-90.

Savran, A. ve Çakıroğlu, J. (2004). Preservice Science Teachers' Orientations to Classroom Management. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 124-130.

Seferoğlu, S. (2004). Öğretmen Adaylarının Öğretmen Yeterlikleri Açısından Kendilerini Değerlendirmeleri, ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı26,131-140.

Sencar, S. ve Eryılmaz, A. (2004). Cinsiyetin Öğrencilerin Elektrik Konusunda Sahip Oldukları Kavram Yanılgıları Üzerindeki Etkisi ve Görülen Cinsiyet Farklılıklarının Nedenleri, ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 141-147.

Sepet, A., Yılmaz, A. ve Morgil, İ. (2004). Lise İkinci Sınıf Öğrencilerinin Kimyasal Denge Konusundaki Kavramları Anlama Seviyeleri ve Kavram Yanılgıları. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 148-154.

Sharkova, Z. (2005). Cube (Die) Rotation Along A Specified Route-Diagnostic Aspects. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 142-149.

Şahin, Ç. (2005). İlköğretim II. Kademesinde Matematik Dersinin Öğrenme-Öğretme Sürecinde Yapılan Etkinliklerin Öğretmen ve Öğrenci Açısından Değerlendirilmesi. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 171-185.

Şen, A.İ. ve Özgün-Koca, S.A. (2005), Orta Öğretim Öğrencilerinin Matematik ve Fen Derslerine Yönelik Olan Olumlu Tutumları ve Nedenleri. ***Eğitim Araştırmaları Dergisi***, Sayı 18, 186-201.

Şimşek, A. (2004). İlköğretim Okulu Sosyal Bilgiler Dersi Tarih Konularının Öğretiminde Hikaye Anlatım Yönteminin Etkililiği. ***Türk Eğitim Bilimleri Dergisi***, Cilt 2 (4), 495-509.

Tuncer,G., Sungur,S., Tekkaya, C. ve Ertepınar, H. (2004). Environmental Attitudes of the 6th Grade Students From Rural and Urban Areas: A Case Study for Ankara. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 167-175.

Uluçınar, Ş., Cansaran, A. ve Karaca, A. (2004). Fen Bilimleri Laboratuvar Uygulamalarının Değerlendirilmesi. ***Türk Eğitim Bilimleri Dergisi***, Cilt 2 (4), 465-475.

Uzuntiryaki, E., Bilgin, İ. ve Geban, Ö. (2004). The Relationship Between Gender Differences and Learning Style Preferences of the Pre-Service Teachers at Elementary Level. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 182-187.

Üstün, E., Akman, B. ve Etikan,İ. (2004). Farklı Sosyo-Ekonomik Düzeydeki Çocukların Bilişsel Gelişimlerinin Değerlendirilmesi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 205-210.

Üstüner, M. (2005). İlköğretim Okullarında Görev Yapmakta Olan Öğretmenlerin Öğrenci Doğasına İlişkin Görüşleri, ***Eğitim Araştırmaları Dergisi***, Sayı 18, 202-216.

Yıldırım, İ. (2004). Lise Öğrencilerinde Boyun Eğici Davranışların Yaygınlığı, ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 220-228.

7. KAYNAKLAR

Aiken, L.S. ve Arkadaşları (1990). Graduate Training in Statistics, Methodology, and Measurement in Psychology: A Survey of PhD Programs in North America, ***American Psychologist***, Vol. 45,721-734.

Allen, M. J. ve Yen, W. M. (1979). ***Introduction to Measurement Theory***, Monterey, California: Brooks/Cole

Arnold, M.E. (1996). ***Influences on and Limitations of Classical Test Theory Reliability Estimates***. (ERIC Document Reproduction Service No. ED 395 950)

Bademci, V. (2005a). Araştırmalarda Ölçme İle İlgili Bazı Büyük Hataları Düzeltmek ve Bir Reformu Başlatmak: Güvenirlik, Testlerin Bir Özelliği Değildir. ***Eğitim Fakültelerinde Yeniden Yapılandırmanın Sonuçları ve Öğretmen Yetiştirme Sempozyumunda Sunulan Bildiri***. 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.

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. (2005d). Güvenirliği Doğru Anlamak ve Bazı Klişeleri Yıkmak: Bilinenlerin Aksine, Cronbach'ın Alfa Katsayısı, Negatif ve –1'den Küçük Olabilir. Yayıma Hazırlanmış Makale.

Bademci, V. (2004). "Testin Güvenirliği" veya "Test Güvenilirdir" Diye İfade Etmek Doğru Değildir. ***Türk Eğitim Bilimleri Dergisi***, Cilt 2 (3), 367-372.

Bademci, V. (2002). *"Türkiye'deki Okullar Ne İşe Yarar? Türkiye'nin Anomi, Yabancılaşma, Ekonomik Büyüme, Demokratikleşme Sorunlarına Çözüm Önerisi."* Düzenleyen: ESEF Öğrenci Bilimsel Faal. Org. Kom. Ankara: G.Ü. Mesleki Eğitim Fakültesi Konferans Salonu, 30 Mayıs.

Bademci, V. (2001a). *"Düşünmenin Öğretilmesi ve Öğretimde Kullanılan Yöntemler-Teknikler."* Düzenleyen: TÜRMOB. Bursa: Bursa SMMM Odası Konferans Salonu, 9 Kasım.

Bademci, V. (2001b). *"Türkiye'deki Okullar Ne İşe Yarar?"* Düzenleyen: Çayyolu Türk Telekom Anadolu Teknik L. Ankara: Başkent Öğretmenevi Konferans Salonu, 9 Aralık.

Bademci, V. (1999). ***Hedefin Davranışlara Çevrilmesi, Davranışlardan Seçmeli Test Maddeleri Yazılması***. (Geliştirilmiş Üçüncü Baskı). Ankara: Gazi Kitabevi.

Balcı, A. (2004). ***Sosyal Bilimlerde Araştırma: Yöntem, Teknik ve İlkeler***. Dördüncü Baskı. Ankara: PegemA

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. (2001). An Essay on the History and Future of Reliability from the Perspective of Replications. ***Journal of Educational Measurement***, Vol. 38, 295-317.

Buhi, E.R. (2005). Reliability Reporting Practices in Rape Myth Research. ***Journal of School Health***, Vol. 75 (2), 63-66.

Büyüköztürk, Ş. (2004). ***Sosyal Bilimler İçin Veri Analizi El Kitabı***. (Dördüncü Baskı). Ankara: PegemA

Büyüköztürk, Ş. (2005). ***Sosyal Bilimler İçin Veri Analizi El Kitabı***. (Gözden Geçirilmiş 5. Baskı). Ankara: PegemA

Capraro, M. M., Capraro,R. M. ve Henson, R.K.. (2001). Measurement Error of Scores on the Mathematics Anxiety Rating Scale Across Studies. ***Educational and Psychological Measurement***, Vol. 61, 373-386.

Capraro, R. M. ve 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.

Caruso, J. C. (2000). Reliability Generalization of the Neo Personality Scales. ***Educational and Psychological* Measurement**, Vol.60, 236-254.

Cousin, S. L. ve Henson, R. K. (2000), ***What is Reliability Generalization, and Why is It Important***? (ERIC Document Reproduction Service No. ED 445 077).

Crocker, L. ve Algina, J. (1986). ***Introduction to Classical and Modern Test Theory***. Fort Worth: Holt, Rinehart and Winston.

Cronbach, L. J.(1951). Coefficient Alpha and the Internal Structure of Tests. ***Psychometrika***, Vol. 16, 297-334.

Dawis, R. V. (1987). Scale Construction, ***Journal of Counseling Psychology***, Vol. 34, 481-489.

Dawson, T. E. (1997). ***Basic Concepts in Classical Test Theory: Relating Variance Partitioning in Substantive Analyses to the Same Process in Measurement Analyses***. (ERIC Document Reproduction Service No. ED 406 443).

Ebel, R. L. (1972). ***Essentials of Educational Measurement***. (Second Edition). Englewood Cliffs, New Jersey: Prentice- Hall, Inc.

Ebel, R.L. ve Frisbie, D. A. (1991). ***Essentials of Educational Measurement***. (Fifth Edition). Englewood Cliffs, New Jersey: Prentice Hall.

Eğitim Araştırmaları Dergisi. (2005). Sayı 18. –en son ve tüm sayı-

Estin, C. ve Laporte, H. (2003). ***Yunan ve Roma Mitolojisi***. (On Dördüncü Basım). Ankara: Tübitak

Flanders, N. A., Anderson, J. P. ve Amidon, E. J. (1961). Measuring Dependence Proneness in the Classroom. ***Educational and Psychological Measurement***, Vol. 21, 575-587.

Gage, N. L. (1963). Paradigms for Research on Teaching. ***Handbook of Research on Teaching***. Ed. N.L. Gage. Chicago: Rand MçNally &Company

Gronlund, N. E. ve Linn, R. L. (1990). ***Measurement and Evaluation in Teaching***. Sixth Edition.New York: Macmillan.

Guilford, J. P. ve Fruchter, B. (1973). ***Fundamental Statistics in Psychology and Education***. (Fifth 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)

Gürsoy, F., Aral, N., Bütün Ayhan, A. ve Aydoğan, Y. (2004). Annesi Çalışan ve Çalışmayan Çocukların Bağımlılık Eğilimlerinin İncelenmesi. ***Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,*** Sayı 26, 62-71.

Hacettepe Üniversitesi Eğitim Fakültesi Dergisi.(Yıl 2004). Sayı 26. –en son ve tüm sayı-

Hamilton, E. (2003). ***Mitologya***. (On İkinci Basım). İstanbul: Varlık.

Helms, L. S. (1999). ***Basic Concepts in Classical Test Theory: Tests Aren't Reliable, the Nature of Alpha, And Reliability Generalization as Meta-Analytic Method***. (ERIC Document Reproduction Service No.ED 427 083)

Helmstadter, G. C. (1964). ***Principles of Psychological Measurement***. New York: Appleton-Century-Crofts.

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. (2000a). ***Sacrificing Reliability and Exalting Sampling Error at the Altar of Parsimony: Some Cautions Concerning Short-Form Test Development***. (ERIC Document Reproduction Service No. ED 447 211)

Henson, R. K. (2000b). ***A Primer on Coefficient Alpha***. (ERIC Document Reproduction Service No. ED 447 210)

Henson, R. K. ve Thompson, B. (2002). Characterizing Measurement Error in Scores Across Studies: Some Recommendations for Conducting " Reliability Generalization" Studies. ***Measurement and Evaluation in Counseling and Development***, Vol. 35, 113-126.

Henson, R. K., Kogan, L. R. ve Vacha-Haase, T. (2001). A Reliability Generalization Study of the Teacher Efficacy Scale and Related Instruments. ***Educational and Psychological Measurement***, Vol. 61, 404-420.

Kieffer, K. M. (1999). Why Reliability Theory is Essential and Classical Test Theory is Often Inadequate. ***Advances in Social Science Methodology****, Volume 5*. Ed. B. Thompson. Stamford, Connecticut: JAI.

Kieffer, K. M., Reese, R. J. ve Thompson, B.(2001). Statistical techniques Employed in AERJ and JCP Articles From 1988 to 1997: A Methodological Review. ***The Journal of Experimental Education***, Vol. 69(3), 280-309.

Kuhn, T. S. (1970). ***The Structure of Scientific Revolutions***. Second Edition, Enlarged. Chicago: The University of Chicago Press.

Lane, G. G., White, A. E. ve Henson, R. K. (2002). Expanding Reliability Generalization Methods with KR-21Estimates: An RG Study of Coopersmith Self-Esteem Inventory. ***Educational and Psychological Measurement***, Vol.62, 685-711.

Linn, R. L. ve Gronlund, N.E. (2000). ***Measurement and Assessment in Teaching***. Eighth Edition. Upper Saddle River, New Jersey: Merrill.

Linn, R. L. ve Miller, M.D. (2005). ***Measurement and Assessment in Teaching***. Ninth Edition. Upper Saddle River, New Jersey: Pearson.

Livingston, S. A. (1988). Reliability of Test Results. ***Educational Research, Methodology, and Measurement: An International Handbook***. Ed. John P.Keeves. Oxford: Pergamon.

Lord, F. M. ve Novick, M. R. (1968). ***Statistical Theories of Mental Test Scores***. Reading, Massachusetts: Addison-Wesley.

McDonald, R.P. (1999). ***Test Theory: A Unified Treatment***. Mahwah, New Jersey: Lawrence Erlbaum.

Mehrens, W. A. ve Lehmann, I. J. (1991). ***Measurement and Evaluation in Education and Psychology***. (Fourth Edition). Fort Worth: Harcourt Brace.

Meier, S. T. ve Davis, S. R. (1990). Trends in Reporting Psychometric Properties of Scales Used in Counseling Psychology Research. ***Journal of Counseling Psychology***, Vol. 37, 113-115.

Mittag, K. C. ve Thompson, B. (2000). A National Survey of AERA Members' Perceptions of Statistical Significance Tests and Other Statistical Issues. ***Educational Researcher***, May, 14-20.

Nunnally, J. C. ve Bernstein, I. H. (1994). ***Psychometric Theory***. (Third Edition). New York: McGraw-Hill.

Nunnally, J. C. (1982). Reliability of Measurement. ***Encylopedia of Educational Research***. (Fifth Edition). Ed. H.E.Mitzel. New York: The Free Press.

Onwuegbuzie, A. J. ve Daniel, L. G. (2000). ***Reliability Generalization: The Importance of Considering Sample Specificity, Confidence Interval, and Subgroup Differences***. (ERIC Document Reproduction Service No. ED 448 204).

Pedhazur, E. J. ve Schmelkin, L. P. (1991). ***Measurement, Design and Analysis: An Integrated Approach***.Hillsdale, New Jersey: Lawrence Erlbaum.

Reinhardt, B. (1996). Factors Affecting Coefficient Alpha: A Mini Monte Carlo Study. ***Advances in Social Science Methodology***, *Volume 4*, Ed. B. Thompson. Greenwich, Connecticut: JAI.

Rowley, G. R. (1976). The Reliability of Observational Measures. ***American Educational Research Journal***, Vol.13, 51-59.

Shields, A. L. ve Caruso, J. C. (2004). A Reliability Induction and Reliability Generalization Study of The Cage Questionnaire. ***Educational and Psychological Measurement***, Vol.64, 254-270.

Suen, H. K. (1990). ***Principles of Test Theories***. Hillsdale, New Jersey: Lawrence- Erlbaum.

Thompson, B. (2001). Significance, Effect Sizes, Stepwise Methods and Other Issues: Strong Arguments Move the Field. ***The Journal of Experimental Education***, *Vol*. 70, 80-93.

Thompson, B. (1999). Five Methodology Errors in Educational Research: A Pantheon of Statistical Significance and other Faux Pas. ***Advances in Social Science Methodology***, *Volume 5*. Ed. B. Thompson. Stamford, Connecticut: JAI.

Thompson, B. (1994a). Guidelines for Authors. ***Educational and Psychological Measurement***, Vol.54, 834-47.

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. (1992). Two and One-Half Decades of Leadership in Measurement and Evaluation. ***Journal of Counseling and Measurement***, Vol. 70,434-438.

Thompson, B. ve Vacha-Haase, T. (2000). Psychometrics is Datametrics: The Test is Not Reliable. ***Educational and Psychological Measurement***, Vol. 60, 174-195.

Thompson, B. ve Snyder, P.A. (1998). Statistical Significance and Reliability Analyses in Recent *Journal of Counseling & Development* Research Articles. ***Journal of Counseling and Development***, Vol. 76, 436-441.

Thorndike, R. L. (1982), ***Applied Psychometrics***, Boston: Houghton Mifflin.

Töreci, K. (2005). Yayın Etiği. http://www.endokrin.com (En son 25.10.2005'de ulaşılmıştır).

Treays, R. (2000). ***Kaslar ve Kemikler***. (Altıncı Basım). Ankara: Tübitak

Turgut, M. F. (1993). ***Eğitimde Ölçme ve Değerlendirme Metotları,*** (Dokuzuncu Baskı), Ankara, Saydam Matbaacılık.

Turgut, S. (2002). Akıl Defterimden Notlar. http://www.hürriyetim.com.tr (En son 25.10.2005'de ulaşılmıştır.).

Türk Eğitim Bilimleri Dergisi.(Yıl 2004). Sayı 4. –en son ve tüm sayı-

Uluğtekin, S. (1976).Çocuk Yetiştirme Açısından Anababa Çocuk İlişkileri. Anababa Davranışlarıyla Çocuğun Saldırganlık ve Bağımlılık Eğilimi Arasındaki İlişkilerin Araştırılması. Yayımlanmamış Doktora Tezi. Ankara: A.Ü.Eğitim Fakültesi.

Vacha-Haase, T., Kogan, L. R., Tani, C. R. ve Woodall, R. A. (2001). Reliability Generalization: Exploring Variation of Reliability Coefficients of MMPI Clinical Scales Scores. ***Educational and Psychological Measurement***, Vol. 61, 45-59.

Vacha-Haase, T., Kogan, L.R. ve Thompson, B. (2000). Sample Compositions and Variabilities in Published Studies versus Those in Test Manuals: Validity of Score Reliability Inductions. ***Educational and Psychological Measurement***, Vol. 60, 509-522.

Vacha-Haase, T., Ness, C., Nilsson,J. ve Reetz, D. (1999). Practices Regarding Reporting of Reliability Coefficients: A Review of Three Journals. ***The Journal of Experimental Education***, Vol. 67 (4), 335-341.

Vacha-Haase, T. (1998). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. ***Educational and Psychological Measurement***, Vol. 58, 6-20.

Whittington, D. (1998). How Well Do Researchers Report Their Measures? An Evaluation of Measurement In Published Educational Research. ***Educational and Psychological Measurement***, Vol. 58, 21-37.

Wilkinson, L. ve APA Task Force on Statistical Inference. (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. ***American Psychologist***, Vol. 54, 594-604.

Worthen, B. R., White, K. R., Fan, X ve Sudweeks, R. R. (1999), ***Measurement and Assessment in Schools***. (Second Edition). New York: Addison Wesley Longman.