

**HAKEMLERİN DEĞERLENDİRMELERİNDEKİ HATALAR ÜZERİNE:  
FİŞHER'İN Z DÖNÜŞÜMÜ VE GÜVENİRLİK ÇALIŞMALARI İÇİN ÖRNEKLEM  
BÜYÜKLÜĞÜ****Vahit BADEMCI<sup>1</sup>****ÖZET**

Bu makalenin amacı, *Türk Eğitim Bilimleri Dergisi (TEBD)* Editörleri tarafından belirlenmiş olan [kimlikleri gizlenmiş] iki hakemin değerlendirme raporlarındaki yargılarına yanıt vermedir. *TEBD* Editörleri tarafından atanmış olan hakemler, Bademci'nin "Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü" başlıklı makalesini değerlendirmişler ve *Türk Eğitim Bilimleri Dergisi* Editörlerine raporlarını da vermişlerdir. Özetle, hakemler, Bademci'nin makalesi hakkında "yayımlanması uygun değildir" yargısında bulunmuşlardır. Bu yazıda, hakemlerin değerlendirme raporlarındaki yargılarının tümü Bademci tarafından incelenmiş ve hakemlerin değerlendirme raporlarında bulunan yargılarındaki ciddi bilimsel hatalar veya şaşırtıcı yanlışlıklar, objektif [ya da bilimsel] ölçütlere dayalı olarak açıkça gösterilmiştir. Kısaca, hakemlerin değerlendirme raporlarındaki tüm yargıları hatalıdır

*Türk Eğitim Bilimleri Dergisi* Editörleri tarafından görevlendirilmiş hakemlerden birisi "test etme", "ölçme", "bellileştirme", "değerlendirme" gibi bazı kavramları hatalı biçimde kullanmaktadır. Yine, hakemlerden birisi Fisher'in z dönüşümünün kullanımı hususunda hatalı yargılarda bulunmuştur. Diğer hakem ise, test uzunluğu ve örneklem büyüklüğü ile ilgili hatalı bağlantılar kurmuş ve hatalı yargılarda bulunmuştur. Ayrıca bu çalışmada, iç ölçütlere ve dış ölçütlere göre bir bilimsel makalenin nasıl değerlendirileceği ve "özdeşsiz" ["özgün"] (Bloom ve diğerleri, 1956:162-184) kavramının neyi kapsadığı da açıklanmaya çalışılmıştır.

**Anahtar Sözcükler:** Hakemlerin hataları, Fisher'in z dönüşümü, örneklem büyüklüğü, güven aralığı, güvenirlilik

**ABSTRACT**

The purpose of this article is to respond to the judgements in the evaluation reports of two arbiters [identities concealed] appointed by the Editors of *The Journal of Turkish Educational Sciences (JTES)*. The arbiters nominated by the Editors of *JTES* have evaluated

Bademci's journal article with the heading "Tests Are Not Reliable : Sufficient Attention To judgements in the evaluation reports of the arbiters have been investigated by

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

Score Reliability And Sample Size For Reliability Studies” and have given their reports to the Editors of *The Journal of Turkish Educational Sciences*. In summary, the arbiters have judged about Bademci’s article as “unsuitable to be published”. In this paper, all the Bademci and the amazing biases or serious scientific errors found in the judgements of evaluation reports of the arbiters has been indicated clearly on the basis of objective [or scientific] criteria. In short, all the judgements in the evaluation reports of the arbiters were mistaken

One of the arbiters nominated by the Editors of *The Journal of Turkish Educational Sciences* has used some of the concepts like “testing”, “measurement”, “assessment”, “evaluation” in the wrong way. Again, one of the arbiters has given faulty judgements regarding the usage of the Fisher’s z transformation. The other arbiter has established incorrect connections related with the test length and sample size and has given faulty judgements. Furthermore, in this study how a scientific article can be evaluated according to the internal criteria and external criteria and what “*the unique*” (Bloom and others 1956:162-184) concept of the article comprises have been tried to be explained.

**Key Words:** The errors of the arbiters, the Fisher’s z transformation, sample size, confidence interval, reliability

## GİRİŞ

*Türk Eğitim Bilimleri Dergisi*’ne (*TEBD*) gönderilmiş olan “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleye yönelik olarak, iki adet isimleri gizlenmiş hakem değerlendirme raporu ve yine hakemlerin isimleri gizlenmiş iki adet *TEBD Makale Değerlendirme Formu*, Makalenin yazarı Bademci’ye *TEBD* Editörlüğü tarafından gönderilmiştir.\* [Hakemler hakkındaki edinilen tek bilgi bütünü, iki hakemden birisinin profesör unvanına, diğerinin ise, doçent unvanına sahip olduğu şeklindedir.] Bu sebeple, ilgili hakemlerin değerlendirmelerindeki hataları konu alan bu makalede, bilgisayar çıktısı olarak gönderilmiş olan rapor ve formu 1 nolu Hakem değerlendirme raporu ve formu, el yazısıyla gönderilmiş olan rapor ve formu da 2 nolu Hakem değerlendirme raporu ve formu olarak isimlendirilecektir. Her iki hakemin değerlendirme raporlarında da, bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, sübjektif ölçütler ve kişisel tercihler kullanılarak verilmiş yargılar görülmektedir ve bu açıdan, her iki hakemin de değerlendirme raporu hakkında hatalıdır yorumu yapılabilir; çünkü, bir bilimsel makalenin değerlendirilmesinde, objektif ölçü ve ölçütler kullanılması gerekli ve şarttır. Bademci’nin makalesine yönelik yazdıkları değerlendirme raporlarında, her iki hakemin de kişisel tercihlerinin dayanak olduğu yargılara ulaştıkları, yine her iki hakemin de sübjektif ölçü ve ölçütlere dayalı değerlendirme yaptıkları ve yargılarındaki mevcut bilimsel hataları, Bademci tarafından aşağıdaki satırlarda ortaya konulmaya çalışılmıştır.

## 1 NOLU HAKEMİN DEĞERLENDİRME RAPORUNDAKİ HATALI YARGILARINA KARŞIT YANITLAR

a) 1 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 5 nolu maddedeki yargısı: “Bir teste ait ölçümlerin güvenirligi konusunda karar verirken Fisher’in Z dönüştürümünden değil t ya da F dağılımından yararlanır. Fisher’in Z dönüştürümü bir korelasyon katsayısının anlamlılığında değil farklı korelasyonların homojenliğini sınamak ya da evren ortalamasını tahmin etmek amacıyla kullanılır”

**Bademci’nin Yanıtı:** Bademci’nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makalenin örneklem büyüklüğü ile ilgili hususunda, güvenirlilik katsayısı için güven aralığı saptanması işlemi ele alınmış, örnek olarak da, test-tekrar test güvenirlilik yöntemiyle kestirilmiş bir güvenirlilik katsayısı  $r = .90$  biçiminde verilmiş ve güven aralığı saptanması işlemi dahilinde de Fisher’in z dönüştürümü kullanılmıştır. Test-tekrar test güvenirlilik kestirim yönteminde [Makalede de belirtildiği üzere] Pearson moment çarpım korelasyon katsayısı ( $r$ ) kullanılmaktadır (Crocker ve Algina, 1986; Suen, 1990; Mehrens ve Lehmann, 1991). Test-tekrar test yöntemiyle kestirilmiş güvenirlilik katsayısının güven aralığını saptamak için de, dolayısıyla bir Pearson moment çarpım korelasyon katsayısının güven aralığı saptama işleminden yararlanılmaktadır (Charter, 1997; 1999). Pearson moment çarpım korelasyon katsayısının temelini teşkil eden sayıtlardan birisi, X ve Y olarak ifade edilmiş olan iki değişkenin, iki değişkenli normal dağılıma (bivariate normal distribution) (Edwards, 1976; Glass ve Hopkins, 1996; Sheskin, 2004) sahip olduğudur. Normal dağılım ise, simetriktir (Ferguson ve Takane, 1989; Popham ve Sirotnik, 1992). Örneklem korelasyon katsayısı ( $r$ ) vasıtasıyla kestirilmiş teorik bir evren korelasyon katsayısı ( $\rho$ ) düşünüldüğünde:  $\rho=0$  için, dağılım simetriktir ve de [Student]\*\* t dağılımı kullanılabilir, ancak,  $\rho \neq 0$  için, dağılım çarpıktır (Spiegel, 1961; 1988). Akhun (1986:29-30), evren korelasyon katsayısı sıfır ve N [örneklem büyüklüğü] büyük değilse,  $r$ ’nin örneklem dağılımının normal olmayacağını belirtmiş ve “ $r$  yüksek (0.80 ve daha büyük) ve N küçük ise,  $r$ ’nin örneklem dağılımı çarpıktır ve Pearson çarpım moment korelasyon katsayısının standart hatasını veren geleneksel formülden [ $\sigma_r = 1 - r^2 / \sqrt{N}$ ] hesaplanan SHr [korelasyon katsayısının standart hatası] yanıltıcı olacaktır” ifadesini kullanmıştır. Akhun (1986:30), bu durumun  $r$ ’nin düşük olması (0.20 veya daha düşük) durumunda da doğru olduğunu ifade ederek, *yüksek r değerlerinin örneklem dağılımının çarpık olmasının* [ki, Bademci’nin örneğinde  $r = .90$ ’dır],  $r$ ’nin -1.00 ile +1.00 arasında değişmesinden ileri geldiğini belirtmiştir. Korelasyon katsayısının örneklem dağılımının çarpıklığından meydana gelen güçlükler Fisher tarafından geliştirilen bir yöntemle çözülmüştür (Akhun, 1986:31). Dağılım çarpık olduğunda Fisher’in z dönüştürümü olarak bilinen bu yöntemle, bir  $r$  değeri, -yaklaşık-normal dağılım üzerine temellenmiş bir ölçeğe dönüştürülmektedir (Spiegel, 1961; Marascuilo ve Serlin, 1988; Sheskin, 2004). **Bir Pearson moment çarpım korelasyon katsayısı için güven aralığı saptanırken de Fisher’in z dönüştürümü kullanılmaktadır** (Spiegel, 1961; Guilford ve Fruchter, 1973; Glass ve Hopkins, 1996; Sheskin, 2004). Bademci, Makalesinde, test-tekrar test yöntemiyle kestirilmiş güvenirlilik katsayısının güven aralığını saptamak için güvenirlilik katsayısını örnek olarak  $r = .90$  olarak vermiş ve bir Pearson  $r$ ’sinin güven aralığını saptamak için kullanılan Fisher’in z dönüştürümünden de yararlanmıştır: **Bademci’nin yaptığı işlem doğrudur**.  $r = .90$  olarak verilmiş bir test tekrar

test güvenilirlik katsayısı için güven aralığı saptanması işleminde Fisher'in z dönüşümü kullanıldığı için Bademci'nin makalesine yönelik olarak, değerlendirme raporunda “bir teste ait ölçümlerin güvenilirliği konusunda karar verirken Fisher'in Z dönüştürümünden değil t ya da F dağılımından yararlanılır” şeklinde yargıda bulunan 1 nolu Hakemin bu yargısı, ciddi bir bilimsel hata olarak görünmektedir. Zira, t dağılımı da, normal dağılım gibi simetrik bir dağılımdır (Akhun,1982; Püskülcü ve İkiz,1989) ve yalnızca  $\rho=0$  için r'nin örneklem dağılımı simetrik ve t dağılımı kullanılabilir;  $\rho \neq 0$  için ise, dağılım çarpıktır. Örneklem dağılımının çarpık olduğu durumda, z dağılımı [bölünümü] veya t dağılımı doğrudan kullanılamaz (Ünver ve Gamgam, 1986).  $r=.90$  olarak verilen **bir test-tekrar test güvenilirlik katsayısının güven aralığının saptanması için**, Charter'ın (1997; 1999) ve de Bademci'nin yaptığı gibi **Fisher'in z dönüştürümünden yararlanılır**; 1 nolu Hakemin değerlendirme raporunun 5. maddesindeki Fisher'in z dönüştürümüyle ilgili yargısı ise, bilimsel yönden tam anlamıyla [ve de çok ciddi] hatalıdır.

Bademci'nin makalesinin örneklem büyüklüğü ile ilgili hususunda, güvenilirlik katsayısı için güven aralığı saptanması işlemi ele alınmış, örnek olarak da, test-tekrar test güvenilirlik yöntemiyle kestirilmiş bir güvenilirlik katsayısı  $r=.90$  biçiminde verilmiş ve güven aralığı saptanması işlemi dahilinde de Fisher'in z dönüşümü kullanılmıştır. Test-tekrar test güvenilirlik kestirim yönteminde Pearson moment çarpım korelasyon katsayısı kullanılmaktadır (Suen, 1990; Mehrens ve Lehmann,1991) ve test-tekrar test yöntemiyle kestirilmiş güvenilirlik katsayısının güven aralığının saptanılması için de, dolayısıyla bir Pearson moment çarpım korelasyon katsayısının güven aralığının saptanılması işleminden yararlanılmaktadır (Charter, 1997;1999). Bademci'de, Makalesinde, test-tekrar test güvenilirlik kestirim yönteminde Pearson moment çarpım korelasyon katsayısını kullanmış, test-tekrar test yöntemiyle kestirilmiş güvenilirlik katsayısının güven aralığının saptanılması için de Pearson moment çarpım korelasyon katsayısının güven aralığının saptanılması işleminden yararlanmış ve dolayısıyla bu işlem dahilinde Fisher'in z dönüşümünü kullanmıştır. 1 nolu Hakem ise, bu işlemlere yönelik olarak, “bir teste ait ölçümlerin güvenilirliği konusunda karar verirken Fisher'in Z dönüştürümünden değil t ya da F dağılımından yararlanılır” şeklinde bir yargıda bulunmuştur. Yapılan literatür taramaları sonucunda, test-tekrar test güvenilirlik kestirim yönteminde kullanılan Pearson moment çarpım korelasyon katsayısının güven aralığının saptanması işleminde, Bademci'nin de makalesinde faydalandığı gibi, Fisher'in z dönüşümünün kullanıldığı görülmektedir (Roscoe,1975; Popham ve Sirotnik,1992; Smithson,2000; Sheskin,2004); Test-tekrar test güvenilirlik kestirim yönteminde kullanılan Pearson moment çarpım korelasyon katsayısının güven aralığının saptanması işleminde, 1 nolu Hakemin yargısında belirttiği şekildeki gibi ‘t yada F dağılımından yararlanıldığına’ dair herhangi bir kaynağa , herhangi bir bilgiye ise, ulaşamamıştır. 1 nolu Hakemin yargısında belirttiği üzere, test-tekrar test güvenilirlik kestirim yönteminde kullanılan Pearson moment çarpım korelasyon katsayısının güven aralığının saptanması işleminde, t yada F dağılımından yararlanıldığına dair kaynakların ve bilgi bütünü, 1 nolu Hakemce bir makale halinde yazılarak Türk eğitim ve bilim dünyasına sunulması, kendilerinden beklenilmektedir.

1 nolu Hakemin değerlendirme raporunun 5. maddesindeki yargısının ikinci cümlesi ise, “Fisher'in Z dönüştürümü bir korelasyon katsayısının anlamlılığında değil farklı korelasyonların homojenliğini sınamak ya da evren ortalamasını tahmin etmek amacıyla kullanılır” şeklindedir. 1 nolu Hakemin Bademci'nin makalesiyle ilgili yazdığı değerlendirme raporunun 5. maddesinde bulunan bu ifade ve yargıya yönelik olarak, 1 nolu Hakemin yaptığı bir diğer bilimsel hatayı ortaya koymak amacıyla üzerinde durulmasında

fayda görülmektedir. 1 nolu Hakemin ‘Fisher’in Z dönüştürümünün bir korelasyon katsayısının anlamlılığında’ kullanılmayacağı biçimindeki yargısının tam tersi bir ifade, Akhun’un (1986) kitabında mevcuttur. Akhun’da (1986:29-33), *korelasyon katsayısının manidarlığı* başlığı altında ve Pearson çarpım-moment korelasyon katsayısı (r) ile ilgili olarak, “bir r değerinin manidarlığını [anlamlılığını] test ederken matematiksel olarak daha çok savunulabilen bir yöntem, korelasyon katsayısının özellikle çok yüksek veya çok düşük olması durumunda, r’nin Fisher’in z fonksiyonuna dönüştürülmesi ve z’nin standart hatasının ( $\sigma_z$ ) hesaplanmasıdır” şeklinde bir ifade bulunmaktadır. [Cümlelerin altı Bademci tarafından çizilmiştir]. Özellikle korelasyon katsayısının yüksek olması durumunda bir r değerinin manidarlığını [anlamlılığını] test etme ile ilgili Akhun’un (1986:29-33) ifadelerine benzer bir ifade, Garrett’da da (1953) vardır. 1 nolu Hakemin söylediğinin tam aksine, ‘bir korelasyon katsayısının anlamlılığı’ için de, Fisher’in z dönüşümünden yararlanılmaktadır. Yine, bir korelasyon katsayısının anlamlılığını [manidarlığını] test etme hususuyla ilgili olan ve Fisher’in z dönüşümünün de kullanıldığı bazı çalışmalar, manidarlık [anlamlılık] için tek bir korelasyon katsayısını test etme biçiminde Marascuilo ve Serlin’de (1988), bir korelasyon katsayısının z testi başlığı altında Kanji’de (1999) bulunmaktadır [ifadelerin altı Bademci tarafından çizilmiştir] ve bu noktada, 1 nolu Hakemin bu yargısı da, yine bilimsel olarak [ve de çok ciddi] hatalıdır.

1 nolu Hakem, Bademci’nin makalesinin 5. ve 6. sayfalarında yer alan, dört ayrı örneklem büyüklüğü (50,200,300,400) dikkate alınarak yapılmış olan ve Fisher’in z dönüşümünün de kullanıldığı test tekrar test güvenilirlik katsayısına yönelik güven aralıkları saptanması işlemleri için, yazdığı değerlendirme raporunun 6. maddesinde, aynen, “yazının 5.ve 6. sayfasındaki sayısal analizler, makale değil istatistik dersinde uygulama yapma amacına daha uygundur” ifadesini kullanmıştır. 1 nolu Hakemin bu yargısı, şık olmayan, bilimsel eleştiri ve akademik nezaket sınırlarını zorlayan bir ifadedir. 1 nolu Hakemin, Bademci’nin makalesine yönelik değerlendirme raporunun 5 nolu maddesindeki Fisher’in z dönüşümü ve istatistiği ilgilendiren bazı konularla ilgili, yani bahsettiği “sayısal analizler”le bağlantılı, bazı çok ciddi bilimsel eksiklikleri olduğu ve dolayısıyla bu hususlarla ilgili yargılarında da ciddi hatalar bulunduğu, yukarıda bilimsel kanıtlara dayalı biçimde açıkça ortaya konulmuştur; Pearson moment çarpım korelasyon katsayısı kullanılarak hesaplanmış bir test-tekrar test güvenilirlik katsayısı için güven aralığı saptanması işlemi dahilinde Fisher’in z dönüşümünün kullanıldığı işlemlerle, yani “sayısal analizler”le ilgili olarak, 1 nolu Hakemin, ciddi bilimsel eksikliklerinin olduğu yukarıda açıkça ortaya konulmasına rağmen, Bademci’nin makalesine yönelik “yazının 5.ve 6. sayfasındaki sayısal analizler, makale değil istatistik dersinde uygulama yapma amacına daha uygundur” biçiminde bir yargıda bulunması da, eğitim bilim ve 1 nolu Hakemin yargısının bilimsel güvenilirliği açısından son derece trajikomik bir durumdur, bilim adına da çok düşündürücüdür. [İki hakemden birinin unvanının profesör, diğerinin doçent olduğu düşünüldüğünde ise, eğitim, öğretim ve bilim yönünden işin vahametinin ne derece büyük olduğu görülebilir ve söylenebilir.]

Bademci’nin yazdığı Makaleye yönelik olarak, 1 nolu Hakem bir değerlendirme raporu yazmıştır ve yazılan bu değerlendirme raporunun 5. maddesinde, 1 nolu Hakemin iki cümle halinde, iki yargısı bulunmaktadır; yukarıda da bilimsel kanıtlarıyla açıkça ortaya konulduğu üzere, 1 nolu Hakemin bu iki yargısı da, bilimsel yönden ciddi biçimde hatalıdır.

**b) 1 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 7 nolu maddedeki yargısı:** “Yazıda İngilizce bazı kavramların Türkçe karşılıklarında hatalar bulunmaktadır. Örneğin, assessment değerlendirme olarak kullanılmaktadır. Homojen ise bağdaşık değil benzeşiktir.”

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makalede, Bademci, ‘assessment’ sözcüğüne karşılık olarak ‘bellilendirme’ sözcüğünü kullanmıştır. 1 nolu Hakem ise, yazıda İngilizce bazı kavramların Türkçe karşılıklarında hatalar bulunduğunu ifade ederek, örnek olarak da ‘assessment’ın ‘değerlendirme’ olarak kullanıldığını belirtmektedir. 1 nolu Hakemin değerlendirme raporundaki bu yargısı, bilimsel açıdan yine hatalı görünmektedir. ‘Assessment’, ‘measurement’, ‘evaluation’ sıklıkla birbirlerinin yerlerine kullanılmaktadır, ancak bu kavramlar ve kullanıldıkları yerler de birbirlerinden farklıdır (Keeves, 1988; Choppin,1990; Airasian,2000; Nitko, 2001). Bu kavramları birbirlerinden ayırt etmek ise, önemlidir. “Measurement”ın Türkçe karşılığı “ölçme”, “evaluation”ın ise, “değerlendirme” olarak yıllardır kullanılmaktadır. Bu sebeple, 1 nolu Hakemin ileri sürdüğü üzere, ‘assessment’ karşılığı olarak, ‘evaluation’ın karşılığı olan ‘değerlendirme’ sözcüğünün kullanılması, doğru değildir. Ertürk (1984:165), ‘assessment’ sözcüğüne Türkçe karşılık olarak “bellileme”, Bademci (2001:204) ise, “bellilendirme” sözcüğünü önermiş ve de kullanmıştır (Bademci, 2004). 1 nolu Hakem ise, ‘assessment’ın karşılığı olarak, ‘evaluation’ın karşılığı olan ‘değerlendirme’ sözcüğünü kullanmayı tercih etmekle ve kullanılmasını kabul etmekle, kavram kargaşası yaratılabilmesine ortak olmakta ve de kendisi hatalı kullanımda bulunmaktadır.

Yine, 1 nolu Hakem değerlendirme raporunun 7 nolu maddesinde, devamla, “homojen ise bağdaşık değil benzeşiktir” ifadesini kullanmıştır. ‘Homogeneous’ karşılığı olarak Bademci (2004) tarafından ‘bağdaşık’ kavramı tercih edilmekte ve kullanılmaktadır. Ayrıca, ‘homogeneous’ karşılığı olarak ‘bağdaşık’ sözcüğü, Demirel ve Ün (1987:238) ve Çoker’de de (1989:165) bulunmaktadır. Türkçe Sözlük de (1988:126) ise, ‘bağdaşık’ sözcüğünün tam karşısında “her yeri aynı özelliği gösteren, mütecanis, homojen” ifadeleri yer almaktadır ve 1 nolu Hakemin ‘homojen ise bağdaşık değil benzeşiktir’ görüşünün aksine, ‘homogeneous’ sözcüğünü, bağdaşık, benzeşikten daha fazla karşıladığı düşünülmekte ve bu yüzden de Bademci (2004) tarafından ‘bağdaşık’ sözcüğünün kullanılması yeğlenmektedir. ‘Homogeneous’ sözcüğünün karşılığı olarak ‘bağdaşık’ sözcüğünün Bademci tarafından kullanımı bilimsel bir hata değil, daha iyisi olduğu düşünülen bir tercihtir. Bilimsel hata, 1 nolu Hakemin yaptığı gibi, ‘evaluation’ karşılığı kullanılan ‘değerlendirme’ sözcüğünü, ‘assessment’ın karşılığı olarak da almak veya kabul etmek ve böylelikle iki farklı kavram olan ‘evaluation’ ile ‘assessment’ı birbirlerinin yerlerine koymak veya kullanmaktır. Yukarıda ortaya konulduğu üzere, 1 nolu Hakemin değerlendirme raporunun 7.maddesindeki yargısı ise, bilimsel yönden yine hatalıdır.

**c) 1 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 3 nolu maddedeki yargısı:** “Yazının 2. sayfasının başındaki ilk paragrafta test etme standartlarından 2.1 ve 2.2’ye atıfta bulunmaktadır. Bu standartlarda, yazıda öne sürülen görüş doğrultusunda herhangi bir ifade bulunmadığı gibi bunlar test etme ölçme standar[t]larıdır”

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makalede 1 nolu Hakemin bahsettiği ifade, Bademci'nin makalesinin 2. sayfasında da aynen belirtildiği üzere, Henson ve Thompson'a (2002) aittir ve de bu hususun 1 nolu Hakem tarafından gözden kaçırıldığı düşünülmektedir. Ayrıca, 1 nolu Hakem, değerlendirme raporunun 3 nolu maddesindeki yargısında “...bunlar test etme değil ölçme standartlarıdır” ifadesini kullanmıştır. Yazının orijinalinde bu ifade, “...testing standards” (Henson ve Thompson, 2002:113) şeklindedir ve Bademci tarafından da “test etme standartları” olarak çevrilmiştir; 1 nolu Hakem ise, “bunlar test etme değil ölçme standartlarıdır” şeklinde yargıda bulunmaktadır. ‘Assessment’, ‘measurement’, ‘test’, ‘testing’, ‘evaluation’ farklı kavramlardır (Airasian, 2000). “Measurement”ın Türkçe karşılığı, “ölçme”dir. “Testing” karşılığı olarak Bademci'nin kullandığı gibi “test etme”, dolayısıyla test etme standartları kullanılabilir veya içinde test kelimesi geçen ve test etmeyi çağrıştıran veya karşılayan bir başka ifade de kullanılabilir, ancak 1 nolu Hakemin kullandığı gibi “testing standards” karşılığı olarak “ölçme standartları” kesinlikle kullanılmamalıdır; zira, ‘testing’ ve ‘measurement’ kavramları birbirlerinden farklıdır ve 1 nolu Hakemin değerlendirme raporunun 3. maddesindeki bu yargısı da, yine hatalıdır. [1 nolu Hakemin diğer yargılarına karşıt yanıtlar, 2 nolu Hakemin yargılarına yönelik karşıt yanıtlar içerisinde sunulmaktadır.]

## 2 NOLU HAKEMİN DEĞERLENDİRME RAPORUNDAKİ HATALI YARGILARINA KARŞIT YANITLAR

2 nolu Hakemin Değerlendirme raporuna geçilmeden önce, objektif değerlendirme ve süreci üzerinde durulmasında, bu satırların yazarı [Bademci] tarafından fayda görülmektedir. Objektif bir değerlendirmenin nasıl yapılması gerektiği konusunda, Bloom taksonomisi olarak da bilinen bilişsel alan sınıflamasından (Bloom ve diğerleri,1956; Ertürk,1984; Turgut, 1993; Bademci,1999; Sadker ve Sadker, 1999; Sönmez, 2003) ve de değerlendirme ana basamağından faydalanılabilir. Değerlendirme basamağıyla ilgili olarak [lisans düzeyinde,Türkiye’de ve Türk eğitim sistemi içerisinde yıllarca kullanılan *Eğitimde Ölçme ve Değerlendirme Metotları* isimli kitabında] Turgut’un (1993) açıklamaları önemlidir:

“Burada değerlendirme terimi bilgilerin, bilgi edinme yöntemlerinin, fikirlerin, çözümlerin, teori ve yapıların ve benzeri bilişsel ürünlerin değeri hakkında bir amaç için yargıya varabilme anlamında kullanılmıştır. Değerlendirmede daima bir gözlem veya ölçme, bir de ölçüt (kıstas) kullanılır. Bir değer yargısına ulaşma, belli ölçütlerle objektif yapılabileceği gibi, sübjektif ölçütlerle, güvenirliliği olmayan gözlemlerle kişisel yargılara varılarak da yapılabilir. Öte yandan kişilerin tercihleri, sevdikleri veya sevmedikleri şeylerde kişisel değer yargılarına dayanır. **Eğitimde hedef alınan değerlendirme, bu kişisel tercihlerin dayandığı yargılara varabilme değil, objektif ölçü ve ölçütlerle değerlendirme yapabilme yeteneğidir**” (Turgut,1993:146). [Cümle, Bademci tarafından koyu yazdırılmıştır].

Şüphesiz, **bir bilimsel çalışmanın, bir bilimsel ürünün değerlendirilmesinde de objektif ölçü ve ölçütlerle değerlendirme yapılması gerekli ve şarttır**. Bir bilişsel ürün olan bir bilimsel makalenin de değerlendirilmesi, başlıca iki yolla yapılır. Bir makalenin kendi

içindeki ölçütlere göre doğru olup olmadığı, kendi içinde çelişkiye düşüp düşmediği aranır; “buna iç ölçütlerle değerlendirme denir (Turgut, 1993: 147).” Ancak, iç ölçütlere göre değerlendirme yapan bir kişi veya bir hakem, objektif ölçü ve ölçütler kullanarak yaptığı bir değerlendirme sonucunda, bir bilişsel ürünün veya makalenin kendi içindeki bir çelişkisini gördüyse, tutarsızlığını, yanlışlığını bulduysa, “bunları gerekçesiyle göstermek, tutarsızlıkları veya yanlışlıkları kanıtlayıp yazmak ( Sönmez, 2003)” zorundadır. Bazen de, bir bilişsel ürün [bir bilimsel makale], dıştaki bir ölçüte vurularak değerlendirilir, yani bir ürün, diğer bir ürün [makale] ya da ürünlerle de karşılaştırılabilir (Turgut, 1993; Sönmez, 2003). Buna da dış ölçütlerle değerlendirme denir. Dış ölçütlere göre değerlendirme yapan bir kişi veya hakem, “dış standartlara göre yargıda bulunarak bir eseri dahil olduğu alandaki bilinen en yüksek standartlarla karşılaştırır (Ertürk, 1984)” ya da “bir bilişsel ürünü diğer bir ürünle karşılaştırarak değer yargısına varır (Turgut,1993)”. Ancak, dış ölçütlere göre değerlendirme yapan bir kişi veya bir hakem de, objektif ölçü ve ölçütler kullanarak bir değerlendirme yaptıysa, yani bir bilişsel ürünü veya makaleyi, yayımlanmış veya bilinen en iyi bir başka ürünle karşılaştırıp bir yargıya vardıysa, “bunları da gerekçesiyle göstermek, kanıtlayıp yazmak veya [kullanılan dış ölçüte] uygunluğunu gerekçe göstererek yazmak (Sönmez, 2003)” zorundadır. Hem 1 nolu Hakemin, hem de 2 nolu Hakemin değerlendirme raporlarında, yaptıkları yargılara dayanak olması gereken “bilimsel kanıtlar” bulunmamaktadır. 1 nolu Hakemin değerlendirme raporunda kısmen ve kaynağı belli olmayan bazı bilimsel gerekçeler bulunmaktadır ve gerekçelerinin yanlışlığı da Bademci tarafından verilen bilimsel dayanaklı yanıtlarla açıkça ortaya konulmuş ve konulmaya da devam edilecektir. Bir bilimsel makaleye yönelik, objektif ölçü ve ölçütler kullanılarak iç ölçütlere veya dış ölçütlere göre değerlendirme yapılmasının ilkeleri göz önünde tutulduğunda ise, 2 nolu Hakemin yazdığı değerlendirme raporunun bilimsellikten çok uzakta olduğu ve hatta bilimsel dahi olmadığı söylenebilir. Bir bilimsel değerlendirme yapılmasının ölçüt ve ilkelerine uymadığı gözlemlenen 2 nolu Hakemin yazdığı değerlendirme raporundaki ifade ve yargılara, bu sebeple, bilimsel kanıt ve gerekçe göstererek cevap vermek çok zorlaşmaktadır, ancak bilimsel akıl yürütmeye ve çoğu zaman da bilimsel kanıt ve dayanak sunmayla, 2 nolu Hakemin bilimsellikten çok uzak, hatta bilimsel olarak yanıt verilmemesi gereken yargılarına, aşağıda, yine de yanıt verilmeye çalışılacaktır.

**a) 2 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 1 nolu maddedeki yargısı: “Çalışma özgün değildir”**

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makaleye, 2 nolu Hakemin yazdığı değerlendirme raporunun 1 nolu maddesindeki yargısı, sadece “çalışma özgün değildir” şeklindeki bir cümleden ibarettir ve başkaca bir şey de yoktur. Çalışmanın özgün olup olmadığı yargısını verebilmek için, yukarıda da açıklandığı gibi, Bademci'nin makalesinin, bir dış ölçütle, yani alanında iyi bilinen yayımlanmış bir makale veya yayımlanmış bir kitapla karşılaştırılması ve sonrasında da bunların bilimsel olarak kanıtlanıp, bilimsel gerekçeleriyle birlikte 2 nolu Hakem tarafından yazılmış olması gerekirdi; ancak, 2 nolu Hakemin, Bademci'nin makalesine yönelik olarak yazdığı değerlendirme raporunda, objektif ölçü ve ölçütler kullanarak, Bademci'nin makalesini, bilinen bir dış ölçütle karşılaştırdığına dair tek bir ifade, tek bir sözcük dahi yoktur. 2 nolu



Hakemin değerlendirme raporunun 1 nolu maddesindeki “çalışma özgün değildir” biçimindeki bilimsel dayanaklardan yoksun yargısının hatalı olduğu, bilimsel verilere ve kanıtlara dayalı şekilde aşağıdaki maddelerde ve de çeşitli yönlerden gösterilmeye çalışılmıştır;

**a/i) *Türk Eğitim Bilimleri Dergisi (TEBD)*** derleme yazıları ilkelerinde, derleme yazılarla ilgili özellikler açıkça belirtilmiştir:

“Yurt içinde ve dışında yaygın olarak üzerinde durulan ya da ilgi gören konularda en son literatürün kapsamlı bir şekilde sunulduğu derlemeler[.]Okuyucu kitlesinin yararlanabileceği konulardaki görgül çalışmaların derlenmesi veya gözden geçirilmesi[.] [bu] grupta yer alan yazılardaki esas, okuyucuya bir alanı değişik boyutları ile tanıtmak ve geniş bir perspektif sunabilmektir. Hedef, çoğunluğunu akademisyen, uygulayıcı ve öğrencilerin oluşturduğu okuyucu kitlesini kuramsal ve görgül literatürdeki farklı sorular, yaklaşımlar ve bulgularla tanıştırmak ve bu konuda güvenilir, zengin bir kaynak oluşturmaktır. Bu yazılarda yapılmış olan çalışmaların, yaklaşımların ve bulguların tekdüze bir şekilde sunulmasının önüne geçilmesi amaçlanmaktadır. Diğer bir deyişle, bu yazılarda iki boyutlu literatür özetlemeleri yerine, yeni sorular ve çalışmalara zemin hazırlayabilecek nitelikte akademik ürünlerin oluşturulması hedeflenmektedir" (*TEBD*, 2004).

Bademci'nin “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı yazısı okunduğunda da görüleceği üzere, *TEBD*'nin derleme yazılarla ilgili bahsedilen tüm ölçütlerine, Bademci'nin makalesi aynen uymaktadır. Hem güvenilirliğin, testin/ölçme aracının kendisinin değil elde edilmiş ölçümlerinin bir özelliği olduğu şeklinde vurgulanmakta olan noktayla ilgili, hem de bu konuyla doğrudan bağlantılı ve makalenin ana fikri olan ölçüm güvenilirliğine yeterli dikkat gösterildiğinde, gözden kaçırılmaması gereken örneklem özelliklerinden birisi olarak önem kazandığı ileri sürülen güvenilirlik çalışmaları için örneklem büyüklüğü ile ilgili en son literatür, Bademci'nin makalesinde derli toplu ve kapsamlı bir biçimde ortaya konulmaya ve sunulmaya çalışılmıştır. Bademci'nin makalesinin ana temasını oluşturan güvenilirlik çalışmaları için örneklem büyüklüğü ile ilgili olarak, hem en son literatürdeki görgül çalışma/çalışmalarla ilgili yayınlar, hem de bu hususla ilgili geçmişe yönelik literatürdeki bilgiler (Charter ve Feldt, 2002; Charter, 2001; Charter, 1999; Segall, 1994; Nunnally ve Bernstein, 1994; Kline,1986; Nunnally, 1982; Nunnally, 1978; Nunnally, 1967; Guilford,1954) derlenerek, Bademci'nin makalesinde sunulmuştur. Sadece *TEBD*'nin derleme yazılarla ilgili ilkeleri ölçüt alındığında dahi, Bademcinin makalesi için “yayımlanır” ana teması etrafında bir karar verilmesi gerekirdi, ancak hem 2 nolu Hakem, hem de 1 nolu Hakem tamamen subjektif ölçü ve ölçütlere dayandığı görülmekte olan kararlarıyla, makale hakkında “yayımlanması uygun değildir” yargılarını vermiştir. Her iki hakemin, Bademci'nin makalesi hakkında verdikleri “yayımlanması uygun değildir” kararı, sadece *TEBD*'nin derleme yazılarla ilgili ölçütü göz önüne alındığında dahi, hatalıdır ve de bilimsel dayanaktan da yoksundur.

**a/ii)** 2 nolu Hakemin “özgün” kavramına, bilimsel olarak yeterince doğru anlam yükleyemediği [ve dolayısıyla bilimsel olarak hatalı kullanımda da bulunduğu] düşünülmektedir. [Bu durum, 1 nolu Hakem için de geçerlidir.] *TEBD Makale*

*Değerlendirme Formunda* bulunan ‘puanlama’ ölçeği “0 zayıf... ve 5 çok iyi” aralıklarını (0,1,2,3,4 ve 5) kapsamaktadır. Bademci’nin makalesiyle ilgili olarak, 1 nolu Hakem, *TEBD Makale Değerlendirme Formunda* bulunan “özgün bir çalışma olarak alana katkısı” değerlendirme maddesinin karşısındaki puanlardan 1 puanına işaret koyarken, 2 nolu Hakem ise, 0 [sıfır] puanını işaretlemiştir. Özgünlüğün ne olduğuna dair bilgilenme için, yine Bloom taksonomisinin *Sentez* basamağına ve özellikle *5.10 özdeşsiz [özgün] bir iletişim muhtevası üretme [production of a unique communication]* alt basamağına [Bloom ve diğerleri (1956:162-184); Grobman (1970:235); Tekin (1982:184,203); Ertürk (1984:66); Özçelik, (1992:37); Turgut (1993:122,145); Demirel (1997:118,122); Sadker ve Sadker (1999); Sönmez (2003:47,73)] başvurulabilir. Sentez basamağı ve bilişsel özgün bir ürünle ilgili olarak Tekin (1982:203), “*sentezde bir özgünlük olmalıdır. Sentez gücü olan bir kişi, ilkin, değişik kaynaklardan uygun öğeleri toplayacak, sonra da bunları anlamlı bir bütün oluşturacak biçimde bir araya getirecektir. Böylece, ortaya çıkan ürün yeni ve özgün bir ürün olacaktır*” ifadesiyle ‘özgün’ bir ürünün, ‘özgün’ bir çalışmanın niteliğini, kolay anlaşılır ve net ifadelerle ortaya koymaktadır [Cümleler, Bademci tarafından italik ve koyu italik olarak yazdırılmıştır]. *Bir ana fikri yeterli delillerle savunmak veya fikirleri ve hükümleri çok iyi biçimde örgütleyerek yazmak* da (Ertürk, 1984; Turgut, 1993), *özgün bir iletişim meydana getirmenin örneklerindedir*. Bademci’nin makalesi okunduğunda da görüleceği üzere, burada ifade edilen özgün bir ürün veya iletişim oluşturma ile ilgili tüm gerekliliklerin, [örneğin güvenilirlik çalışmaları için örneklem büyüklüğünün en az 400 denek/kişiden oluşması ana fikrini yeterli delillerle savunması, Bademci’ye ait özgün bir çalışma ve karşılaştırmayla, bu ana fikri doğrulamaya çalışması ve tüm bunları örgütleyerek yazması gibi,] Bademci’nin makalesinde yerine getirilmiş olduğu açıkça görülecektir ve *Bademci’nin makalesi*, özellikle 2 nolu Hakemin iddia ettiğinin tam tersine, *özgündür*. Buradan, özellikle 2 nolu Hakemin ve de 1 nolu Hakemin *özgün* kavramına, bilimsel olarak doğru anlam yükleyemedikleri söylenebilir; Bademci’nin makalesine yönelik olarak 2 nolu Hakemin bilimsel dayanaktan yoksun [ve özgün kavramına bilimsel olarak doğru anlam yükleyemediği ileri sürülebilen] “çalışma özgündür” kararı, bu açıdan da yine hatalıdır.

**a/iii)** Bademci’nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makalenin ana teması, “güvenirlik çalışmaları için örneklem büyüklüğü, *en az* 400 kişi (denek) olmalıdır” biçimindedir. Bademci, Charter’ın (1999; 2001; Charter ve Feldt, 2002) görgül kanıt üzerine temellendirerek ileri sürdüğü örneklem büyüklüğünün en az 400 denekten oluşması gerektiği önerisi etrafında, bu konuyla ilgili yurt dışındaki kaynakları derlemiş, konuyu, özgün bir makale halinde de yazmış ve Türk eğitim bilim dünyasına da ilgili konu hakkında bilgi vermek istemiştir. 2 nolu Hakem yaptığı değerlendirmeye, Bademci’nin makalesi hakkında “yayımlanması uygun değildir” biçiminde bir yargıda bulunmuş ve bu yargısına dayanak olması gereken değerlendirme raporunda ise, sadece bir cümleden ibaret olan “çalışma özgündür” yargısını ve ifadesini kullanmıştır. “Çalışma özgündür” şeklindeki yargısına dayanak oluşturabilmesi için, *2 nolu Hakemin, Bademci’nin makalesinin ana teması olan güvenirlik çalışmaları için örneklem büyüklüğünün, en az 400 kişi (denek) olması hususunda, daha önce Türkçe yayımlanmış bilinen en yüksek standarttaki bir çalışmayı sunması, sunulan bu çalışmayla Bademci’nin yazdığı makaleyi karşılaştırması, Bademci’nin çalışmasında yararlandığı bilimsel yaklaşımların ve kullandığı istatistiksel işlemlerin aynısının [özdeşinin] daha önce yapıldığını bu karşılaştırma ile ortaya koyması, bunları belgelemesi ve ispat etmesi zorunludur; bu*

*zorunluluk, bilimsel kararların objektif ölçü ve ölçütlere dayandırılması gerekliliğinden doğmaktadır.* Bademci'nin çalışmasına o [sıfır] puan vererek özgün olmadığını ileri süren 2 nolu Hakem ise, bu yapılması gerekenlerin birini dahi, yapmamıştır. Artı, güvenilirlik çalışmaları için örneklem büyüklüğünün *en az* 400 denek olarak alınması önerisiyle ilgili, hem 1 nolu Hakemin "...bu tür iddialı sayılar öne sürmek uygun değildir" ifadesinden, hem 2 nolu Hakemin "niçin 400, niçin 600 değil, niçin 1000 değil" gibi Bademci'nin makalesine düştüğü notlarından ve ifadelerinden, "güvenirlik çalışmaları için *en az* 400 denek alınması" konusuyla ilk defa karşılaştıkları, dolayısıyla makalenin özgünlüğünü iki hakemin de aslında bilmeden kabul ettikleri düşünülmektedir; tüm bunlara rağmen, Bademci'nin makalesine yönelik olarak 2 nolu Hakemin verdiği "çalışma özgün değildir" kararı, yukarıda da açıkça ortaya konulduğu gibi, bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, sübjektif ölçütler kullanılarak verilmiş bir karardır ve 2 nolu Hakemin kararı, bu açıdan da yine hatalıdır.

**a/iv)** Charter (1999), çalışmasında kullandığı toplam 6.322 güvenilirlik katsayısı için örneklem büyüklüğüne de yönelik olarak çeşitli tablolar [örneğin, 6.322 güvenilirlik katsayısı için örneklem büyüklüğü yüzdeleri, '< 50', '≥50 ve >100'... ve '≥1000' gibi değişik grup aralıklarında gösterilmiş] ve şekiller kullanmış, bu şekillerde de farklı güvenilirlik katsayıları, değişik örneklem büyüklükleri ve güven aralığı genişliği arasındaki ilişkileri göstermiştir. Bademci ise, Makalesinde örneklem büyüklüğü olarak, Guilford (1954) ve Kline'in (1986) en az '200' veya daha fazla kişi-denek, Nunnally (1967;1978;1982), Segall (1994), Nunnally ve Bernstein'in (1994) en az '300' veya daha fazla kişi-denek ve Charter'ın (1999; 2001) görgül kanıt üzerine temellendirdiği en az '400' kişi-denek olması biçimindeki örneklem büyüklüklerini çalışmasına temel olarak almış ve çeşitli bilim adamlarınca ileri sürülen bu örneklem büyüklüklerini, Pearson moment çarpım korelasyon katsayısı için güven aralığı saptanması işleminden faydalanarak karşılaştırmıştır: Bademci'nin yaptığı bu karşılaştırma tamamıyla Bademci'ye ait olup, özgündür, ne yurt dışındaki, ne de yurt içindeki literatürde yoktur. 2 nolu Hakemin değerlendirme raporundaki "çalışma özgün değildir" ifade ve yargısı ise, tümüyle gerçeğe aykırıdır. 2 nolu Hakem, Bademci'nin makalesinde yapmış olduğu gibi, Bademci'nin makalesinde bahse konu olarak almış olduğu farklı bilim adamlarınca ileri sürülmüş olan güvenilirlik çalışmaları için örneklem büyüklüklerine [200, 300, 400 kişi/denek] yönelik bir karşılaştırmanın, daha önce aynısının [özdeşinin] yapıldığını belgelemek ve de bilimsel kanıtlarıyla ispat etmek zorundadır. Bademci'nin çalışmasının özgün olmadığını ileri süren 2 nolu Hakem ise, yine bunları yapmamış ve Bademci'nin makalesi hakkında, hiçbir bilimsel ölçü, ölçüt, kanıt kullanmadan "çalışma özgün değildir" yargısıyla birlikte, "yayımlanması uygun değildir" kararını vermiştir. *Türkiye Cumhuriyeti, sosyal bir hukuk Devletidir ve öğretim üyelerinin yayın yapma özgürlüğü anayasal güvence altındadır. Hiç bir bilimsel ölçü, ölçüt ve dayanak kullanılmadan, bilimsel kanıt, ispat veya gerekçe gösterilmeden, sadece bir "çalışma özgün değildir" denilerek, sonrasında da "yayımlanması uygun değildir" kararıyla birlikte, bir bilimsel çalışmanın yayımlanmasını açıkça engellemek, Türkiye Cumhuriyeti Devletinin değişik yasalarının kapsamı içine girmektedir* ve bu [bilimsel] makalenin de sınırları dışındadır. Yukarıdaki maddelerde ve çeşitli açılardan da ortaya konulduğu üzere, Bademci'nin "Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü" başlıklı bilimsel makalesine yönelik olarak 2 nolu Hakemin verdiği ve sadece bir cümleden ibaret olan "çalışma özgün değildir" kararı, bilimsel ve objektif kanıtlara ve gerekçelere

dayanmayan, güvenilirliği olmayan, sübjektif ölçütler ve kişisel tercihler kullanılarak verilmiş bir karardır ve 2 nolu Hakemin kararı, bu açıdan da yine hatalıdır.

[Buradaki tüm bu açıklamalar, Bademci'nin (2004) “ ‘Testin Güvenirliği’ Veya ‘Test Güvenilirdir’ Diye İfade Etmek Doğru Değildir” başlıklı ve *TEBD*'de yayımlanmış bulunan bir makalesinin hakem değerlendirmesi sürecinde yer almış olan [ve yine isimleri gizlenmiş] bir hakem için de tümüyle geçerlidir. Bademci'nin (2004) *TEBD*'de yayımlanmış olan makalesi hakkında [yine isimleri gizlenmiş] ikinci ve üçüncü hakem “yayımlanması uygundur” kararı verirken, “yayımlanması uygundur” kararı veren bu hakemlerden birisi de Bademci'nin (2004) makalesine yönelik olarak *TEBD Makale Değerlendirme Formu* üzerine “olağan üstü bir makale” şeklinde bir de not düşüp, Makalenin tümü için de 5 tam puan verirken, aynı Makale (Bademci, 2004) hakkında “yayımlanması uygun değildir” yargısı veren [yine ismi gizlenmiş] birinci hakem, hiçbir bilimsel ve objektif kanıt ve gerekçe göstermeden yazdığı değerlendirme raporunda “...bir makale olarak böyle bir dergide yayımlanacak özgünlük ve önemde değil. Bilgi vermenin genelde bilinenlerin ötesine geçilmemiş” biçiminde, yuvarlanmış, bilimsel güvenilirliği olmayan, sübjektif ölçüt ve kişisel tercihlerini kullanarak ve bilimsel eleştiri düzeyini ve akademik nezaket sınırlarını zorlayan bir ifadeyle, bir yargıda bulunmuştur. Bu çalışmadaki yanıtların ilgili büyük bir çoğunluğu, Bademci'nin (2004) *TEBD*'de yayımlanmış olan makalesi hakkında, Makale hakem değerlendirme sürecinde iken, tek bir objektif ölçü ve ölçüt ve gerekçe göstermeden “yayımlanması uygun değildir” yargısı veren o hakem için de, aynen geçerlidir.]

**b) 2 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 2 nolu maddedeki yargısı:** “Ölçümlerin güvenilirliği” Türk Bilim Dünyasına 2004'de ilk kez sunuldu [yazıdan bunun yazar tarafından yapıldığı anlaşılıyor] gibi gerçek dışı anlatımlar var. Oysa özgün (yurt dışı) kaynaklarda ne zaman tartışılmaya başlandıysa, Türkçe kitap, makale vb. eserlerde de eş zamanda vurgulanmaya başladı. “Testlerin güvenilirliği” denilirken dahi, burada, aslında “ölçümlerin, puanların güvenilirliğinden” bahsedildiği daima vurgulanır”

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makaleye, 2 nolu Hakemin yazdığı değerlendirme raporunun 2 nolu maddesindeki yargısı, hiçbir şekilde gerçeği yansıtmamaktadır. Bademci'nin makalesinde, 2 nolu Hakemin iddia ettiği şekilde, “ölçümlerin güvenilirliği” Türk Bilim Dünyasına 2004'de ilk kez sunuldu” gibi bir ifade, *kesinlikle yoktur*. 2 nolu Hakemin, yazı metni içinde de altını çizerek işaretlediği ifadenin yer aldığı cümlenin bulunduğu paragraf, Bademci'nin “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makalesinde aynen şu şekildedir;

“Güvenirliğin, “testlerin değil ölçümlerin bir özelliği” (Caruso, 2000) olduğu şeklinde vurgulanmakta olan bu nokta, Türkiye’de de ilk defa Bademci’nin (2004) çalışmasıyla eğitim ve bilim gündemine taşınmıştır. Türkiye’de de 1950’lerden [örneğin, ‘testin güvenliği –güvenirliği-’ ya da ‘güvenilir bir testtir’ (Weitzman ve McNamara,1953:145)] bu yana kullanılan, “testin güvenilirliği” veya “test güvenilirlidir” ya da “aracın güvenilirliği” ve benzeri

ifade biçimlerinin doğru olmadığını belirten Bademci (2004:367-372), *güvenirliğin, testin/ölçme aracının kendisinin değil elde edilmiş ölçümlerinin bir özelliği olduğunu* ve ifade edilmesi gerekenin de *ölçüm güvenilirliği* olduğunu vurgulamış, doğru ifadeler için de örnekler vermiştir” (Bademci, ‘Testler Güvenilir Değildir...’).

2 nolu Hakemin, Bademci’nin makalesine yönelik olarak “ ‘ölçümlerin güvenilirliği’ Türk Bilim Dünyasına 2004’de ilk kez sunuldu gibi gerçek dışı anlatımlar var” saptaması ve yargısı, yukarıdaki paragraftan da görüleceği üzere hiçbir biçimde doğru değildir. Paragrafın ilk cümlesinde de görüldüğü gibi Bademci’nin kullandığı ifade “güvenirliğin, “testlerin değil ölçümlerin bir özelliği” (Caruso, 2000) olduğu şeklinde vurgulanmakta olan bu nokta...” şeklinde başlamaktadır ve 2 nolu Hakemin kullandığı ifade olan “ölçümlerin güvenilirliği” ile Bademci’nin makalesinde yazılı olan “güvenirliğin, testlerin değil ölçümlerin bir özelliği olduğu” şeklinde başlayan ifadesi ve devamlı ilgili paragrafta açıkladığı husus, açıkça görülmektedir ki birbirlerinden farklıdır; 2 nolu Hakemin, 2 nolu maddedeki bu yargısı, tamamıyla gerçeğe aykırıdır. 2 nolu Hakem, “oysa özgün (yurt dışı) kaynaklarda ne zaman tartışılmaya başlandıysa, Türkçe kitap, makale vb. eserlerde de eş zamanda vurgulanmaya başladı. “Testlerin güvenilirliği” denilirken dahi, burada, aslında “ölçümlerin, puanların güvenilirliğinden” bahsedildiği daima vurgulanır” şeklindeki ifadesi için de, yine tek bir bilimsel ölçüt, dayanak, en başta da bu konuda özellikle kendisinin yazmış olduğu [varsa] kitap ya da makaleleri olmak üzere, tek bir kaynak, tek bir kanıt göstermemiş ve de bir bilimsel makale değerlendirmesinde olmaması gereken yuvarlanmış, bilimsel olarak içi boş ve desteksiz ifadeler kullanmıştır. 1 nolu Hakem de, bu hususla ilgili olarak, değerlendirme raporunun 2. maddesinde “testlerin güvenilirliği ifadesi, yazıda iddia edildiği gibi testin kendisine değil test puanlarına işaret etmektedir. Zaten güvenilirlik hesaplamaları da bu puanlara dayalı olarak yapıldığından, ‘testin güvenilirliği’ ifadesi, bir dil konusudur ve kolay anlatım için kullanılmaktadır” yargısında bulunmuştur. Bu konuda, her iki hakemin de yargılarının ve fikir yürütmelerinin yine doğru olmadığı açıkça söylenebilir; zira, **“bir ölçme aracına işaret ettiği zaman test güvenilirlidir veya testin güvenilirliği ifadesi...müşterek”** (Guthrie, 2000:4) **kullanılmaktadır** [İfade, Bademci tarafından koyu olarak yazdırılmıştır]. Yani, Guthrie’nin de (2004) ifade ettiği gibi, ‘test güvenilirdir’ veya ‘testin güvenilirliği’ ifadesi, hem 1 nolu hakemin, hem de 2 nolu hakemin söylediklerinin tam tersine, bir ölçme aracına [ya da testin kendisine] işaret etmektedir. Ayrıca, her iki hakemin de bir şekilde savundukları ‘testin güvenilirliği’ [veya ‘test güvenilirdir’] ifadelerinin kullanımına da, zaten Bademci (2004) karşı çıkmakta ve bilimsel kanıtlara dayalı biçimde de ‘testin güvenilirliği’ veya ‘test güvenilir’ şeklindeki ifade biçimlerinin doğru olmadığını ortaya koymaktadır. Türkçe literatürde, “test puanlarının [ölçümlerinin] güvenilirliği” ifadesinin kullanımı da, bu ifadenin kısaltılarak “testin güvenilirliği” biçiminde kullanımı da, yeni değildir ve 1940’lardan bu yana, 60 yılı aşkın bir süredir de, “testin güvenilirliği” şeklindeki kullanım biçimi sürmekte ve sürdürülmektedir; “test ölçümlerinin [puanlarının] güvenilirliği” şeklindeki uzun ama doğru bir ifadeyi, 1 nolu Hakemin de öne sürdüğü gibi “kolay anlatım için” “testin güvenilirliği” biçiminde kısaltarak kullanma eğilimine yönelik olarak söylenilebilecek bir çift söz, **“anlaşılır, fakat zararlıdır”** (Thompson ve Vacha-Haase, 2000:178) şeklinde olacaktır. “Test ölçümlerinin güvenilirliği” veya “X testi ölçümlerinin güvenilirliği” gibi bir ifadeyi, “testin güvenilirliği” şeklinde kısaltarak kullanmanın sıkıntıları ve zararları, Thompson (1992; 1994) ve Thompson ve Vacha-Haase’de (2000) tartışılmıştır ve bu konuya, burada girilmeyecektir. “Testin güvenilirliği” biçiminde ifade etme ve kullanma eğilimi, güvenilirliğin, testin de

[aracın da] bir özelliği olduğu şeklindeki düşünme tarzının da etkisiyle, Türkiye’de, hali hazırda sürmektedir. Bu düşünme tarzının uzantısı olan ifadelere, Türkçe yayımlanmış çok çeşitli çalışmalarda, “ölçme sonuçlarının veya araçlarının güvenilirliği”, “ölçme araçlarının güvenilirliği”, “testlerin güvenilirliği”, “güvenirlilik ölçme aracı veya aracın sonuçlarına ilişkin bir özelliktir”, “ölçeğin güvenilirliğinin”, ölçüm güvenilirliği başlığı altında “test formunun güvenilirliği”, “testin güvenilirlik katsayısı”, “bir araç güvenilirse”, “bir ölçme aracında; eş anlamlı olarak ölçme sonuçlarında...” gibi değişik biçimlerde ve de çok sık rastlanılmaktadır. Ayrıca, yine bu düşünme tarzının sonucunda, güvenilirlik konusuyla ilgili olarak, bazı çalışmalarda, ölçümlerin güvenilirliğiyle birlikte testlerin [veya ölçme araçlarının] güvenilirliklerinden ve güvenilirlik kestirimlerinden de bahsedilirken, bazı çalışmalarda ise, ölçme araçlarının [veya testlerin] güvenilirliklerinin kestirilmesiyle birlikte ölçümlerin [puanların] güvenilirliklerinden de bahsedilmektedir. Türkiye’de de halen devam eden bu düşünme tarzı, bu paradigma [paradigma ile ilgili; Gage,1963], ilk paradigma veya eski paradigma olarak nitelendirilebilir ve özetle, bu düşünme tarzında, güvenilirlik, hem aracı, hem de ölçümlerin [puanların] özelliği gibi düşünülmekte veya ifade edilmektedir. İkinci düşünme tarzında, ikinci paradigmadaki veya yeni paradigmadaki ise, [bilimsel kanıtlara dayalı net ifadeler, tutarlı bir dil kullanımı ve özgün bir bütün halinde,] güvenilirlik, testin/ölçme aracının kendisinin değil, elde edilmiş ölçümlerinin bir özelliği şeklinde ifade edilmiştir (Bademci, 2004). Bu düşünme tarzının ifadeleri ise, “X testi ölçümlerinin güvenilirliği...”, “X ölçme aracından elde edilen ölçümler için alfa güvenilirliği...”, “bu çalışmadaki ölçümlerin güvenilirliği iki yarı yöntemiyle kestirilmiştir...”, “eldeki mevcut veriler için iki yarı güvenilirliği...” gibi ifadeler olabilir. Güvenirlilik hakkında, kimi bilim adamları, doğru olmadığı Bademci (2004) tarafından ortaya konulmaya çalışılmış ilk düşünme tarzını, kimileri de Bademci’nin (2004:367-372) derleyip, *Türkiye’de de ilk defa açık ve net ifadelerle, tutarlı bir dil kullanarak, bilimsel kanıtlarıyla destekli ve özgün bir bütün halinde ortaya koyduğu* ikinci düşünme tarzını, yani yeni paradigmayı benimseyebilir. Bir başka çalışma veya çalışmaların konusunu oluşturacağından dolayı, burada, ayrıntısına girilmeden verilmiş olan iki düşünme tarzı arasında ayırım yapmanın gerekliliği önemlidir ve açıktır; konu, ne 1 nolu Hakemin bahsettiği gibi “bir dil konusu”, ne de 2 nolu Hakemin bahsettiği gibi “testlerin güvenilirliği denilirken dahi, burada, aslında...” konusudur, her iki hakemin de güvenilirlik hususunda kavram kargaşasına girdikleri düşünülmekte ve bu konuda hatalı yargılarda buldukları da görülmektedir.

Bu noktada, Bademci tarafından bir hususa daha dikkat çekilecektir; Bademci’nin makalesine yönelik olarak, 1 nolu Hakem *TEBD Makale Değerlendirme Formunda* bulunan ‘özgün bir çalışma olarak alana katkısı’ maddesi de dahil, 12 maddenin 11’ine ve makalenin tümüne 1 [bir] puan işareti koymak ve değerlendirme raporunun yukarıda da belirtilen 2 nolu maddesindeki yargısını belirtmek suretiyle, 2 nolu Hakem de *TEBD Makale Değerlendirme Formunda* bulunan ‘özgün bir çalışma olarak alana katkısı’ maddesine de 0 [sıfır] puan vermek ve değerlendirme raporunun yukarıda da aynen yazılmış bulunan 1 nolu ve 2 nolu maddelerindeki yargılarını beyan etmek ve Makale üzerine çeşitli notlar düşmek suretiyle suretiyle, Bademci’nin makalesinde sunduğu konuların yeni olmadığını, ‘özgün olmadığını’, başta kendileri üzere de, Bademci’nin üzerinde durduğu konuların Türk eğitim ve bilim çevrelerince bilindiğini [öncelikle kendilerinin yayımlanmış çalışmaları olmak üzere, yayımlanmış tek bir Türkçe kaynak dahi gösterememelerine rağmen] ima etmişler ya da vurgulamışlardır. Bademci’nin, hem yayımlanmış çalışmasında (Bademci, 2004), hem de “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlilik Çalışmaları İçin Örneklem Büyüklüğü”

isimli makalesinde, *tutarlı bir dil kullanımıyla*, güvenilirliğin testlerin değil ölçümlerin bir özelliği olduğu şeklinde konu aldığı husus, her iki hakemin de bahsettiği şekilde basit bir ifade biçimi değil, Bademci'nin (2004) makalesi yayımlandıktan sonra bazı araştırma görevlileri ve öğretim üyelerinin de doğru yorumladıkları gibi, bir paradigma değişikliği başlangıcı, bir yeni düşünme tarzı değişikliğinin başlangıcı olan bir husustur; bu yeni düşünme tarzı kapsamı dahilinde, güvenilirliğin, testlerin değil ölçümlerin bir özelliği olduğu ve ölçüm güvenilirliğine yeterli dikkat ve özen gösterilmesi vurgusu, tekrarlamak gerekirse, *tutarlı bir dil kullanımıyla birlikte, bilimsel kanıtlara dayalı ve özgün bir bütün halinde* Türkiye'de ilk defa Bademci'nin (2004) çalışmasıyla gündeme getirilmiştir. Bu hususun bilindiğini iddia veya ima eden her iki hakemin bu yargılarının da yine bilimsel dayanağı olmadığı, bir başka yönden de yine burada gösterilmeye çalışılacaktır; Bademci'nin makalesiyle doğrudan ilgili olan ve Bademci'nin ulaşabildiği yayımlanmış kaynaklara göre, güvenilirliğin, testlerin değil verilerin [ölçümlerin] bir özelliği olduğuna dair, yurt dışında, bu hususta, ciddi anlamda tartışmayı başlattığı düşünülen ilk çalışmanın üzerinden, yayım yılı ve içinde bulunulan bu yıl da dahil 15 yıl (Thompson, 1991), ölçüm güvenilirliği söyleminin, alanının önde gelen ve en etkili yayımlarından biri olan *Educational and Psychological Measurement* Dergisinin editöryel politikalarından biri olacağına belirtilmesi üzerinden 12 yıl (Thompson,1994), test-tekrar test gibi bazı güvenilirlik katsayıları için güven aralığı saptanmasına ilişkin makalenin yayımlanmasının üzerinden 9 yıl (Charter,1997), bazı güvenilirlik katsayıları için, örneklem genişliğinin bir fonksiyonu gibi güvenilirlik katsayısı için güven aralığı genişliğinden yararlandığı ve görgül kanıt üzerine dayalı biçimde güvenilirlik çalışmaları için örneklem büyüklüğünün *en az* 400 [denek-kişi] olması gerektiği ifade edilen çalışmanın üzerinden 7 yıl (Charter, 1999) geçmekte olup veya geçmiştir. Güvenirlik çalışmaları için güven aralığı saptanması ve örneklemin bazı özellikleri [örneğin, büyüklüğü] gibi hususlar, yurt dışında, güvenilirliğin ölçümlerin bir özelliği olduğu şeklindeki düşünme tarzının etkisiyle birlikte üzerlerinde yeniden yoğunlaşılana, birbirleriyle ilişkili konulardan sadece ikisi durumundadır. Yaklaşık 15 yıl önce yurt dışında ciddi tartışma başlatılan “güvenirliğin, ölçümlerin bir özelliği olduğu” ana temalı düşünme tarzıyla ilgili hususun Türkiye’de de bilindiğini ve de Türkçe kitap, makale vb. eserlerde vurgulandığını, “oysa özgün (yurt dışı) kaynaklarda ne zaman tartışılmaya başlandıysa, Türkçe kitap, makale vb. eserlerde de eş zamanda vurgulanmaya başladı” biçiminde ifade eden 2 nolu Hakem, ayrıca, Bademci'nin *tutarlı bir dil kullanarak ve özgün bir bütün halinde* sunduğu makalesindeki “güvenirliğin, “testlerin değil ölçümlerin bir özelliği” (Caruso, 2000) olduğu şeklinde vurgulanmakta olan bu nokta, Türkiye’de de ilk defa Bademci'nin (2004) çalışmasıyla eğitim ve bilim gündemine taşınmıştır” şeklinde başlayan ifadesinin yanına, “çok iddialı bir görüş, sanırım araştırmacı literatürü izlemiyor” diye de not düşmüş ve de Bademci'nin (2004) öne sürdüğü hususla ilgili literatürü, kendisinin izlediğini ve bildiğini de [*öncelikle kendisinin yayımlanmış çalışmaları olmak üzere, yayımlanmış tek bir Türkçe kaynak dahi gösterememesine ve ortaya koyamamasına rağmen*] ima etmiştir. [*Ki, 2 nolu Hakem, Bademci'nin (2004:367-372) derleyip, Türkiye’de de ilk defa açık ve net ifadelerle, tutarlı bir dil kullanarak, bilimsel kanıtlarıyla destekli ve özgün bir bütün halinde ortaya koyduğu -ve yeni paradigma değişikliği başlangıcı olarak sunduğu- çalışmasının, Türkçe literatürde – kitap,makale vb.-, daha önce ayısının [özdeşinin] yayımlandığını da belgelemek ve de bilimsel kanıtlarıyla ispat etmek zorundadır. Ancak, 2 nolu Hakem, bu yapılması gerekenlerin birini dahi, yine yapmamıştır.*] Literatürü, kendisinin izlediğini ve bildiğini ima eden aynı 2 nolu Hakem, değil 15 yıl, yayımlanmasının üzerinden 7 yıl geçmekte olan (Charter, 1999) ilgili hususta, yani güvenilirlik çalışmaları için örneklem büyüklüğüyle ilgili

konuda, Bademci'nin makalesinin üzerine ve değerlendirme raporuna “niçin 400, niçin 600 değil de 400, niçin 1000 değil, bunun sınırı nedir, n=600 olsa ne olur” gibi ifade ve yargılarını da, rahatlıkla yazabilmiştir. Şüphesiz 2 nolu Hakemin sorduğu bu sorular, literatürü takip ettiğini ima eden birisinin ifadeleri değildir; zira, literatürü izleyen birisi, Bademci'nin makalesinde defalarca tekrar edilen “*en az 400*” kavramının ne anlama geldiğini iyi bilir, ayrıca bu tür soruları da yöneltmezdi: Çünkü, 2 nolu Hakemin sorduğu soruların cevapları da, yayımlanmış ilk makalenin üzerinden 7 yıl geçmekte olan bir süredir, literatürdeki bazı çalışmalarda (Charter, 1999, Charter 2001, Charter ve Feldt, 2002, Charter, 2003) zaten mevcuttur. Ayrıca, Bademci'nin güvenilirliğin ölçümlerin bir özelliği olduğu biçimindeki düşünme tarzıyla ilgili çalıştığı ve Makalelerinde sunmaya başladığı konuların tamamı da, *birbirlerine bağlı veya yakın ilişki içerisinde*; bu yeni düşünme tarzının çalışılan bir ögesini literatürden takip ettiğini söyleyen birisinin, diğer ögeler hakkında da en azından yeterli düzeyde bilgi sahibi veya kulak aşinalığı olacağı da bir gerçektir. Bu ifadeler, Bademci'nin makalesine yönelik olarak, “testlerin güvenilirliği” ifadesi, yazıda iddia edildiği gibi kendisine değil test puanlarına işaret etmektedir” şeklindeki [hatalı olduğu yukarıda ortaya konulmuş olan] yargısıyla birlikte, hem *TEBD Makale Değerlendirme Formunda* bulunan ‘özgün bir çalışma olarak alana katkısı’ maddesine, hem de Bademci'nin makalesinin tümüne 1 puan işareti koyan, ancak, güvenilirliğin ölçümlerin bir özelliği olduğu konusuyla bağlantılı, yayımlanmasının üzerinden 9 yıl geçen bir çalışmadaki test-tekrar test güvenilirlik katsayısı için güven aralığı saptanması işleminde de kullanılmış bulunulan Fisher'in z dönüşümünden (Charter,1997) yararlanılması hususunu doğru bulmayan, fakat bu yargısının da hatalı olduğu, Bademci tarafından yukarıdaki satırlarda bilimsel kanıtlara dayalı olarak açıkça ispat edilen, 1 nolu Hakem için de geçerlidir. Buradan, her iki hakemin de bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, sübjektif ölçütler ve kişisel tercihler kullanarak bir değerlendirme yaptıkları, yine rahatlıkla söylenebilir.

**c) 2 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 3 nolu maddedeki yargısı:** “Yazarın sıkıcı bir anlatımı var “ölçümlerin güvenilirliği” ifadesi, bir sayfada 12 kez vurgulanıyor, Sanırım okuyucu kitlesinin özellikleri farklı algılanıyor”

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makaleye yönelik olarak, 2 nolu Hakemin yazdığı değerlendirme raporunun 3 nolu maddesindeki yargısında [ve makalenin bütününe düştüğü “daha önce hiç söylenmemiş!” gibi bazı notlarda da] ilk göze çarpan husus, 2 nolu Hakemin kullanmış olduğu bilimsel eleştiri düzeyini ve akademik nezaket sınırlarını zorlayan, sübjektif ölçüt dayanıklı ifadelerdir. Bademci'nin makalesine yönelik olarak, 2 nolu Hakemin “yazarın sıkıcı bir anlatımı var” kişisel tercih dayanaklı yargısını, Bademci'nin makalesiyle bağdaştırmak mümkün değildir; zira, Bademcinin makalesi, bir edebi metin değil, güvenilirlik çalışmaları için örneklem büyüklüğü ana temasıyla ilgili olarak yazılmış bir bilimsel makedir; bir bilimsel makalenin ise, kişisel tercihler ölçüt alınarak değerlendirilmemesi gerektiği, daha önce de ifade edilmişti. 2 nolu Hakem, Bademci'nin makalesine yönelik olarak “sanırım okuyucu kitlesinin özellikleri farklı algılanıyor” şeklinde bir ifade kullanmıştır; 2 nolu Hakemin kullandığı bu ifadesinden, 2 nolu Hakemin, Bademci'nin makalesini ve bahsedilen sayfayı okuyarak, Bademci'nin “okuyucu kitlesinin özelliklerini farklı



algıladığını” iddia edebilecek ve Bademci’nin “algılaması” hakkında karar verebilecek özel bir yeteneğe sahip olduğu akla gelmektedir. 2 nolu Hakemin değerlendirme raporunda bulunan 3 nolu maddesinde [altına notta düşerek] bahsettiği sayfa ise, Bademci’nin yazdığı makalenin 2. sayfasıdır. 2 nolu Hakemin aynı sayfada 12 kez vurgulandığını [2 nolu Hakem 12 kez hususıyla ilgili “tekrar-tekrar” notu da düşmüş] ileri sürdüğü ve tırnak içine alarak yazdığı “ölçümlerin güvenilirliği” ifadesi, Bademci’nin yazdığı makalenin 2. sayfasında, yani **aynı sayfada, sadece 1 (bir) yerde** “...ölçümlerin güvenilirliği şeklinde” sunulmuştur; bir yerde ise, “ölçüm güvenilirliği söylemi” ve bir başka yerde de Gronlund ve Linn’den (1990) yapılmış alıntı içerisinde “...test ölçümlerinin güvenilirliğinden” biçiminde, hepsi, toplam 3 kez kullanılmıştır. 2 nolu Hakem ise, “ ‘ölçümlerin güvenilirliği’ ifadesi bir sayfada 12 kez vurgulanıyor” ifadesiyle açıkça gerçeğe aykırı ve doğru olmayan bir beyanda bulunmaktadır; 2 nolu Hakemin *gerçeğe aykırı ve doğru olmayan söylemle* sunduğu bu yargısı da, yine hatalıdır.

**d) 2 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 4 nolu maddedeki yargısı:** “Uygun örneklem büyüklüğü için, test uzunluğu vb. diğer ölçütlerin de tartışılması gerek, bu konuda çok sayıda kaynağa ulaşılabilir”

**Bademci’nin Yanıtı:** Bademci’nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makalenin 4. sayfasına, 2 nolu Hakem, yukarıda yazdığı yargıyla ilişkili olarak “örneklem büyüklüğü aynı zamanda araçta bulunan maddelerin sayısına bağlı de değişir [test uzunluğu]” şeklinde bir de not düşmüştür. 2 nolu Hakemin bu ifadelerinin ve yargılarının ciddi bilimsel hataları üzerine, burada çok genişçe durulmasına gerek bile duyulmamaktadır. Kısaca, güvenirlilik konusuyla bağlantılı olarak ifade edilirse, *test uzunluğu, grup bağıdaşıklığı, zaman sınırı, maddelerin güçlüğü gibi faktörler, güvenirlilik katsayılarını etkileyen faktörlerdir* (Crocker ve Algina, 1986; Mehrens ve Lehmann, 1986; Lyman, 1991; Mehrens ve Lehmann, 1991; Thorndike, Cunningham, Thorndike ve Hagen, 1991) ve *test uzunluğu da, test maddeleri ile ilgilidir* ve de *test uzunluğunun, maddelerin sayısının doğrusal bir fonksiyonu* olduğu varsayılmıştır (Magnusson, 1967). Tekin’de (1982:67), ‘testin uzunluğu’yla ilgili olarak, “*eğer güvenirligi etkileyen diğer etkenlerin tümü kontrol edilmişse* ve teste sonradan katılan sorular öncekiler gibi aynı davranışlarla ilgiliyse, soru sayısı artırılarak [bir] testin [ölçümlerinin] güvenirligi artırılabilir” ifadesi mevcuttur [Cümle, Bademci tarafından italik yazdırılmıştır]. *Bir ölçme aracında [bir testte] bulunan maddelerin sayısına bağlı olarak değişen ise, test ölçümlerinin [puanlarının] güvenirligidir*. 2 nolu Hakem ise, örneklem büyüklüğünün aynı zamanda araçta bulunan maddelerin sayısına [test uzunluğuna] bağlı da değiştiğini ifade ederek, ciddi bir bilimsel hata yapmıştır. Görüldüğü üzere, literatürde, test uzunluğu, grup bağıdaşıklığı gibi faktörler, güvenirligi veya güvenirlilik katsayısını etkileyen faktörler başlığı altında aktarılmakta veya tartışılmaktadır. 2 nolu Hakem, hem “örneklem büyüklüğü aynı zamanda araçta bulunan maddelerin sayısına bağlı de değişir [test uzunluğu]”, hem de “uygun örneklem büyüklüğü için, test uzunluğu vb. diğer ölçütlerin de tartışılması gerek, bu konuda çok sayıda kaynağa ulaşılabilir” şeklinde ifade ve yargılarda bulunmuştur; literatürde güvenirlilik çalışmaları için ‘uygun örneklem büyüklüğü’ saptanırken, test uzunluğu vb. diğer ölçütlerin tartışıldığı herhangi bir kaynağa rastlanılmamıştır. “Bu konuda çok sayıda kaynağa ulaşılabilir” ifadesini kullanan 2 nolu Hakemden, hem

güvenirlik çalışmalarında “uygun örneklem büyüklüğü için, test uzunluğu vb. diğer ölçütleri de” kendisinin “tartıştığı”, hem de bu konunun tartışıldığı “çok sayıda ki kaynağı”, bir makale halinde toparlayarak yazması ve bunu da Türk eğitim ve bilim dünyasına sunması, bilhassa kendilerinden beklenilmektedir.

Ayrıca, 2 nolu Hakemin konu aldığı biçimiyle test uzunluğu, test geliştirme [ya da test oluşturma] süreci olarak da adlandırılan ve çeşitli biçimlerde meydana getirilebilen ardışık adımlar (Ross,1954; Tinkelman,1971; Yıldırım,1973; Crocker ve Algina,1986) bütününde, ihtiyaç duyulduğunda gündeme getirilen bir husustur. Test geliştirme sürecindeki ardışık adımlar; *i)* test ölçümlerinin hangi amaç/amaçlar için kullanılabileceğinin belirlenmesi, *ii)* ölçülecek niteliklerin [ya da psikolojik yapıların] tespit edilmesi, *iii)* maddelerin yazılması ve gözden geçirilmesi, *iv)* testin denemeye hazır hale getirilmesi *v)* testin deneme uygulamasında yararlanılacak olan örneklem belirlenmesi, *vi)* belirlenmiş olan deneme örneklemini üzerinde testin deneme uygulamasının yapılması ve *vii)* testin deneme uygulaması yapıldıktan sonra uygulamadan elde edilen sonuçların analizi [test ölçümleri analizi, madde ölçümleri analizi] ve testin son halinin meydana getirilmesi biçiminde, ana çizgileriyle özetlenebilir ve de bu adımların sayısı, azaltılabilir veya çoğaltılabilir. Bir testin deneme uygulamasından elde edilen sonuçlar, değişik amaçlar için test ölçümleri veya madde ölçümleri düzeylerinde veya her iki düzeyde de analiz edilebilir, test ölçümlerinin güvenilirliğinin veya test ölçümlerinin güvenilirliğiyle bağlantılı olarak testteki madde sayısının [testin uzunluğunun] yeterli olup olmadığı gibi değişik konularda karar verebilmek için de bu analiz sonuçlarından yararlanır. Kısaca bir testi uzatmak ya da bir teste kaç madde daha katılacağı kararını verebilmek için, testin deneme uygulaması yapıldıktan sonra elde edilen sonuçlarının analizinin yapılması ve bunlardan yararlanılarak test ölçümlerinin güvenilirliğinin kestirilmesi gerekmektedir. Yani, bir testin uzatılmasına karar verilebilmesi için, ilgili testin önceki uzunluğundaki test ölçümlerinin güvenilirliğinin bilinmesi gereklidir. Yukarıdaki adımlardan da görüleceği üzere, örneklem büyüklüğünün belirlenmesi testin deneme uygulamasından önce olan **besinci adımda**, test uzunluğuyla ilgili bir hususa karar verilebilmesi ise, testin deneme uygulaması yapıldıktan sonra olan ve test ölçümlerinin güvenilirliğinin de kestirildiği **vedinci adımda** gerçekleşmektedir. Bir başka ifadeyle, önce [deneme] örneklem belirlenmekte, sonra belirlenmiş olan [deneme] örneklem üzerinde testin deneme uygulaması yapılmakta, daha sonra testin deneme uygulamasından elde edilen sonuçların analizi yapılmakta ve de bu basamakta test ölçümlerinin güvenilirliğiyle bağlantılı olarak, testteki madde sayısının [testin uzunluğunun] yeterli olup olmadığı gibi konularda karar verilebilmektedir; 2 nolu Hakemin ise, genel geçer bu sırayı birbirine karıştırmış olduğu rahatlıkla ifade edilebilir. Bu açıklamalardan da görüleceği üzere, 2 nolu Hakemin, “uygun örneklem büyüklüğü için, test uzunluğu vb. diğer ölçütlerin de tartışılması gerek” ve “örneklem büyüklüğü aynı zamanda araçta bulunan maddelerin sayısına bağlı de değişir [test uzunluğu]” ifadeleri, 2 nolu Hakemin bağlantısını kurduğu güvenilirlik çalışmaları için örneklem büyüklüğü ile test uzunluğu konularıyla ilgili de, yine çok ciddi bilimsel hatalarının bulunduğunu gösterir işaretlerdir.

2 nolu Hakemin, Bademci'nin makalesindeki güvenilirlik çalışmaları için örneklem büyüklüğü ile ilgili hususa da, bilimsel yönden yeterince hakim olmadığını düşünmektedir; zira, *1)* güvenilirlik çalışmaları için örneklem büyüklüğü kestirimindeki bir yaklaşım, SABİT BİR SAYI kullanmadır (Streiner ve Norman, 1995) ve Bademci, Makalesinde bu yaklaşımın kullanıldığını da, açıkça yazmıştır. 2 nolu Hakemin, Bademci'nin makalesine yönelik olarak bulunduğu yargıda, “araçta bulunan maddelerin

sayısına [test uzunluğuna] bağlı da örneklem büyüklüğünün değiştiğini” söylemesi ve “uygun örneklem büyüklüğü için, test uzunluğu vb. diğer ölçütlerin de tartışılması” gerektiğini ifade etmesi ve iddiada bulunması, bırakılsın bilimsel hatayı, açıkça mantık hatalarıyla yüklü [ve bilim adına da talihsiz ve vahim] ifadeler olarak görülmektedir. Çünkü, tekrarlamak gerekirse, *Bademci'nin makalesinde güvenilirlik çalışmaları için örneklem büyüklüğü kestiriminde kullandığı yaklaşım, sabit bir sayı kullanma yaklaşımıdır ve Bademci'de bu yaklaşımdan yararlanarak makalesinde sabit sayılar kullanmıştır.* Bademci'nin makalesinde örneklem büyüklüğü kestiriminde **sabit bir sayı** yaklaşımı kullanıldığı ve örneklem büyüklüğü “SABİT bir sayı” olarak alınmış olduğu halde, 2 nolu Hakemin iddia ettiği şekilde örneklem büyüklüğünün aynı zamanda araçta bulunan maddelerin sayısına bağlı da nasıl değiştiğini [ya da değişeceğini] açıklaması, yani zaten **sabit [değişmeyen, hep aynı kalan] bir sayı** olarak alınmış bulunan bir örneklem büyüklüğünün, araçta bulunan maddelerin sayısına da bağlı nasıl değişebileceği sorusunun cevabını 2 nolu Hakemin vermesi ve bunu da Türk eğitim ve bilim dünyasına yazılı biçimde sunması, yine 2 nolu Hakemin kendisinden beklenilmektedir. 2) Bademci'nin makalesinde, örnek olarak Pearson moment çarpım korelasyon formülü kullanılarak hesaplanmış bir test-tekrar test güvenilirliği  $r=.90$  ve örneklem büyüklükleri de 50, 200, 300, 400 [kişi/denek] olarak belirlenmiş olup,  $r=.90$  olarak verilmiş olan test-tekrar test güvenilirlik katsayısı, 4 farklı örneklem büyüklüğü ile ilgili yapılan işlemler için de aynı değer olarak alınmakta, bir başka ifadeyle  $r=.90$  olarak verilmiş olan test-tekrar test güvenilirliği sabit tutulmakta ve sadece örneklem büyüklükleri değiştirilmekte ve de örneklem büyüdükçe, daha duyar bir evren güvenilirlik kestiriminin yapılıp yapılamayacağı ortaya konulmaya çalışılmaktadır; 2 nolu Hakemin ifade ettiği test uzunluğunun, yapılan işlemlerle ve de Bademci'nin makalesinde çalıştığı hususla doğrudan bir ilgisi yoktur ve 3) değerlendirme raporunun 4 nolu maddesindeki yargısında 2 nolu Hakem, “...bu konuda çok sayıda kaynağa ulaşılabilir” ifadesini kullanmıştır, ifade doğrudur, ancak adres yanlıştır; yukarıda yapılan tüm bu bilimsel açıklamalardan sonra, Bademci'nin makalesinde çalıştığı hususlarla ilgili çok sayıda kaynağa 2 nolu Hakemin ulaşması gerektiği de, Bademci tarafından 2 nolu Hakemin kendisine tavsiye edilmektedir; 2 nolu Hakemin değerlendirme raporundaki 4 nolu maddede bulunan yargısı da, bilimsel yönden ciddi biçimde yine hatalıdır.

**e) 2 nolu Hakemin, “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleyle ilgili yazdığı değerlendirme raporunda bulunan 5 nolu maddedeki yargısı:** “Örneklem büyüklüğü için ölçütler için daha net gerekçeler olmalı, kesim noktaları tartışılmalıdır.  $n=400$  doyurucu denilirken sadece görelî bir saptama var.  $n= 600$  olsa ne olur ? Nerede durulmalı?”

**Bademci'nin Yanıtı:** Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makaleye, 2 nolu Hakemin yazdığı değerlendirme raporunun 5 nolu maddesindeki yargısı da [4 nolu maddedeki yargısıyla paralellik gösterir şekilde, yine] anlamlı değildir. Zira, örneklem büyüklüğünün en az 400 denekten oluşması gerektiği önerisi ve ifadesinin Charter'a (1999;2001) ait olduğu ve Charter'ın (1999;2001) çalışmalarının ise, “görgül kanıt üzerine temellendiği”, Bademci'nin makalesi içinde, dört-beş farklı yerde belirtilmiştir. Görgül kanıt üzerine temellenmiş çalışmalarında güvenilirlik çalışmaları için örneklem büyüklüğünün en az 400 denekten oluşmasına yönelik tartışmalar ve gerekçeler,

önerinin sahibi, Charter'ın (1999;2001) çalışmalarında vardır, okunduğunda da görülebilir. Bademci ise, çalışmasının güvenilirlik çalışmaları için örneklem büyüklüğüyle ilgili işlemlerinde, Guilford (1954) ve Kline'in (1986) en az '200' veya daha fazla kişi-denek, Nunnally (1967;1978;1982), Segal (1994), Nunnally ve Bernstein'in (1994) en az '300' veya daha fazla kişi-denek ve Charter'ın (1999;2001) görgül kanıt üzerine temellendirdiği en az- '400' kişi-denek olması biçiminde önerdiği örneklem büyüklüğü sayılarını ele almış ve bu sayıları, güvenilirlik katsayısı için güven aralığı saptanması işleminden faydalanarak, incelemiştir; inceleme sonucunda da Bademci, Makalesinde, “*en az '400' veya daha büyük kişiden-denekten oluşan örneklem üzerine temellenmiş güvenilirlik kestiriminin, daha duyar, daha isabetli olduğu bu çalışmada ortaya konulmuştur*” şeklinde yorumunu yapmış, böylelikle Charter'ın (1999;2001) ileri sürdüğü düşüncesinin isabetli olduğunu [özgün bir çalışmayla ve] bir başka yönden doğrulamıştır. Bademci'nin makalesine, örneklem büyüklüğüyle ilgili olarak, 2 nolu Hakemin “niçin 600 değil, niçin 1000 değil, niçin 400, n=600 olsa ne olur” gibi yazdığı ifadelerinden ve düştüğü notlarından, örneklem büyüklüğünün en az 400 olması fikrini kabul etmemesine rağmen, bu fikrin temel dayanağı olan Charter'ın (1999;2001) yayımlanmış çalışmalarını da incelemediği düşünülmektedir [Bu durum, “...bu tür iddialı sayılar öne sürmek doğru değildir” şeklinde yargıda bulunan 1 nolu Hakem için de geçerlidir]. Çünkü, 2 nolu Hakemin, güvenilirlik çalışmaları için örneklem büyüklüğü ile ilgili olarak Bademci'nin makalesine düştüğü “niçin 400” gibi sorularının cevapları, Charter'ın (1999;2001) çalışmalarında mevcuttur; 2 nolu Hakem, özellikle de Charter'ın (1999) tarihli çalışmasının tablolarında ve şekillerinde, sorularının cevaplarını bulabilir ve görebilirdi. Ayrıca, 2 nolu Hakem, “n=400 doyurucu denilirken sadece görel bir saptama var. n= 600 olsa ne olur? Nerede durulmalı?” gibi ifadeler kullanmıştır. 2 nolu Hakemin ısrarla kullandığı “n=400” ifadesi, Bademci'nin makalesinde, bir karşılaştırmanın gereği olarak *sadece* istatistiksel işlem yapılırken kullanılmıştır; onun dışında Bademci'nin makalesi içinde, birisi ara başlıkta ve altı çizili olmak üzere, değişik sayfalarda, toplam 7 sefer, örneklem büyüklüğünün “en az 400” olması ifadesi yazılmış ve de belirtilmiştir. “**En az 400**” ifadesi, **bu sayının bir başlangıç** olduğunu ve “güvenirlik çalışmalarına yönelik örneklem büyüklüğü için, 400 ve 400'den daha büyük sayılar kullanılabileceğine işaret eder”; bu sözel ifade, “ $n \geq 400$ ” şeklinde sembollerle de gösterilebilir. Charter'ın (1999;2001), güvenilirliğin daha duyar kestirimi için örneklem büyüklüğünün en az 400 denek olması gerektiğini görgül kanıt üzerine temellendirerek ileri sürdüğü çalışmalarından sonra da, yayımlanmış çalışmalarıyla güvenilirlik konusundaki önde gelen isimlerden biri olan Feldt (örneğin, Feldt ve Brennan, 1989) ile yaptığı ortak çalışmasında da Charter'ın bu önerisi kabul görerek tekrarlanmış (Charter ve Feldt, 2002) ve de Charter'ın bu önerisi, -Bademci'nin makalesi *TEBD*'ne gönderildikten sonra da, Bademci tarafından hala devam edilen literatür taramaları sonucunda ulaşılan-2003 yılında *The Journal of General Psychology*'de yayımlanan bir makalesiyle de literatürdeki yerini pekiştirmiştir (Charter, 2003). Yurt dışındaki literatürde tüm bu gelişmeler sürerken, güvenilirlik çalışmaları için örneklem büyüklüğüyle ilgili en son gelişmelerle birlikte bu konudaki farklı önerileri çeşitli çalışmalardan derleyerek, özgün bir çalışma ve karşılaştırmayla Türk bilim dünyasına sunmaya çalışan Bademci'nin makalesine yönelik olarak, 1 nolu Hakem “...güvenirlik çalışmalarında 400 denek ya da cevaplayıcıdan ölçüm alınması gerektiği iddia edilmiştir. Araştırmalardaki denek sayısının kaç olacağını bazı durumlarda araştırmacılar belirleyemeyebilir. Dolayısıyla, bu tür iddialı sayılar öne sürmek doğru değildir” biçimindeki bir yargıyla, 2 nolu Hakem de “niçin 600 değil, niçin 1000 değil” biçimindeki ifade, notlar ve yargularla, “yayımlanması uygun değildir” biçimindeki bir karara imza atmışlar ve Bademci'nin makalesinin yayımlanmasını

engellemişlerdir; tüm bu aktarılanlardan da açıkça görüleceği üzere, bu hususta, her iki hakemin de bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, sübjektif ölçütler ve kişisel tercihler kullanarak bir değerlendirme yaptıkları rahatlıkla söylenebilir. Tüm bu gelişmelerden, hem 1 nolu, hem 2 nolu Hakemin literatürü yeterince takip etmedikleri, takip etmedikleri gibi Bademci'nin makalesinin içinde ve kaynakçasında konuyla ilgili verilmiş olan temel makalelere veya çalışmalara ulaşmak için de çaba sarfetmedikleri [buna rağmen Bademci'nin makalesi hakkında karar verdikleri], ileri sürülebilir. Ayrıca da, Bademci'nin en son literatürü de içeren “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makalesi hakkında “yayımlanması uygun değildir” kararını veren, hem 1 nolu, hem 2 nolu Hakem için de, öncelikle, Bademci'nin bu çalışmada kullandığı bazı kaynakların tarihlerine göre de en az 50 yıllık konular olan ve de Bademci'nin çalışmasında Pearson r'sinin güven aralığının saptanmasında faydalanılan biçimiyle Fisher'in z dönüşümü, Fisher'in z dönüşümünün kullanımı ve Bademci'nin makalesiyle doğrudan ilgili olmayan ve güvenilirliği etkileyen faktörlerden biri olan ve de 2 nolu Hakemin değerlendirme raporuna taşıdığı test uzunluğu (Gulliksen,1950; Garrett,1953; Spiegel,1961; Downie ve Heath,1965; Magnusson, 1967; Guilford ve Fruchter,1973; Roscoe,1975; Crocker ve Algina, 1986; Popham ve Sirotnik,1992; Glass ve Hopkins,1996; Smithson,2000; Sheskin,2004) konularında, bu çalışmada bilimsel kanıtlarıyla ortaya konulmuş olan [en az 50 yıllık konulardaki] bilimsel eksikliklerini tamamlamaları, ondan sonra Bademci'nin en son literatürü de kullandığı Makalesi hakkında karar vermeleri gerektiği de ifade edilebilir. 2 nolu Hakemin değerlendirme raporundaki 5 nolu maddesinde bulunan yargısı da, yine hatalıdır.

## SONUÇ

Bademci'nin yazdığı “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makaleye yönelik olarak, *Türk Eğitim Bilimleri Dergisi (TEBD)* Editörlerince belirlenmiş olan her iki hakemin de değerlendirme raporlarındaki yargıları teker teker bu çalışmada ele alınmıştır; [birisi profesör, diğeri ise doçent unvanlı] her iki hakemin de bazı konular da kavram kargaşasına girdikleri ve bazı kavramlara farklı anlamlar yükledikleri, her iki hakemin de eleştirilerindeki ve yargılarındaki mevcut çok ciddi bilimsel hataları ve yine her iki hakemin de Bademci'nin makalesinin tümüne yönelik olarak, bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, sübjektif ölçütler ve kişisel tercihler kullanarak bir değerlendirme yaptıkları ve hatalı yargılarda buldukları, bilimsel veri ve kanıtlara dayalı biçimde ve gerekçeleriyle Bademci tarafından bu çalışmada ortaya konulmuştur; kısaca, [birisi profesör, diğeri ise doçent unvanlı] her iki hakemin değerlendirme raporlarındaki yargılarının tümü de hatalıdır.

## YORUM

Gardner (1995), bir makalesine, “on yıllık bir sessizlik, bazen iyi bir fikirdir” cümlesiyle başlamıştır; bu cümle, bu satırların yazarı Bademci için de geçerlidir. 1988 yılından bu yana kalite yönetim felsefesinin eğitim sistemine yerleştirilmesi üzerine çalışan Bademci, 1992 yılında Milli Prodüktivite Merkezi'nde Anahtar Verimlilik Grubunu kurmuş ve eğitimde toplam kalite yönetimi düşüncesinin yayılması için çalışmalar başlatmıştır; Bademci, eğitimde kalite geliştirme çalışmaları dahilinde, 1989 yılında öğretimin ölçülmesinde yeni

eğilimler, 1990 yılında esnek düşünme, eleştirici, problem çözücü, yaratıcı düşünme becerileri ve çok boyutlu zeka, 1991 yılında program geliştirmede ve öğretmede yeni yaklaşımlar ve özellikle disiplinler arası eğitim programı, 1992 yılında karakter eğitimi, 1993 yılında beyin temelli öğrenme konularında araştırmalar ve çalışmalar yapmış, 1995 yılından itibaren bu konuları deneysel araştırmalar yapmak amacıyla çeşitli kademelerdeki okullara taşımış ve bu konular üzerinde Türk eğitim camiasında ilgi uyanmasını sağlamıştır (Kutay, 2004).

Eğitimde toplam kalite yönetimi, çok boyutlu zeka, beyin temelli öğrenme, yapılandırıcılık, bütünleştirilmiş [integrated] program geliştirme uygulamalarının başarı veya başarısızlığının arkasında yatan etkenlerden belki de başta geleninin, ‘ölçme ve değerlendirme’ çalışmalarının bu hususlara yönelik dönüştürülebilir başarı ya da başarısızlığına bağlı olduğunu tespit eden Bademci, 1994’ten bu yana, bu bahsedilen konuları hem tek tek, hem de birbirleriyle bağlantılı ve etkileşimli biçimde ve ‘ölçme ve değerlendirme’ hususunu hiç gözden kaçırmadan, derinlemesine ve kuramsal boyutlarıyla çalışmaktadır; bu 10 yıllık yoğun araştırma ve çalışma sürecinin ilk ürünü de, bir paradigma değişikliği başlangıcı biçiminde, *TEBD*’de yayımlanan bir makalesiyle (Bademci,2004) Türk eğitim ve bilim topluluğuna sunmuştur. Bu sürecin ardışık ürünlerinden ikincisi olan “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makale de, yaklaşık 3 yıllık araştırma ve yazım sürecinin sonucunda, büyük bir emeğin ürünüdür. Bademci’nin 3 yıllık araştırma ve yazım sürecinin bir ürünü olan adı geçen çalışmasının ve de daha sonraki bilimsel çalışmalarının, hiçbir bilimsel kanıt ve bilimsel olarak doğru ve tutarlı gerekçeler gösterilmeden, kişisel tercihlere ve sübjektif ölçütlere dayalı biçimde verilen yargılarla, bir şekilde yayımlanmasının engellenmesine, hem bilimin, hem de yasaların çerçevesinde izin ve imkan verilmeyeceği, burada açıkça ifade edilebilir.

Bademci’nin “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı makalesine yönelik olarak, *TEBD*’nce belirlenmiş olan [biris profesör, diğeri ise doçent unvanlı] her iki hakemin verdiği kararların ve ulaştıkları yargıların tümünün, bilimsel ve objektif kanıtlara ve gerekçelere dayanmadığı, güvenilirliği olmadığı, sübjektif ölçütler ve kişisel tercihlere dayalı ve bilimsel yönden de hatalı olduğu, bilimsel veri ve kanıtlara dayalı biçimde ve bilimsel gerekçeleriyle, Bademci tarafından bu çalışmada ortaya konulmuştur. Şüphesiz, görevini olanca itina ve en iyi biçimde yürütmeye çalışan ve de yerine getiren bilimsel dergi hakemleri de Türk eğitim topluluğunda vardır. Ancak, son zamanlarda, *hem yayımlanmış bilimsel makalelerden, hem de -Bademci’nin makalesine/makalelerine yönelik de olduğu gibi- sübjektif ölçü ve ölçütlere dayalı biçimde “yayımlanması uygun değildir” yargısı verilen raporlardan da görüleceği üzere, bilerek ya da bilmeyerek olduğu düşünülen hakem değerlendirmelerindeki [bazı] hataların çok ciddi biçimde artış göstermekte olduğu ifade edilebilir. Bu artıştaki belki de büyük sorumluluk, yeterince hakim olmadıkları konularda karar vermekten geri duramayan hakemler de [yani üniversite öğretim üyelerinin de] olduğu kadar, hakem seçimlerindeki isabetlilik derecesi de dahil pek çok hususu kapsayan biçimde de bilimsel dergilerin editör ve yardımcılarına düşmektedir. Pek çok bilimsel dergi editör ve yardımcıları, gelen makaleleri hakemlere, hakem değerlendirme raporlarını da makale sahibine yönlendiren makale yönlendiriciliğinden daha fazla sorumluluk almak ve dergilerinde etkili olmak zorundadır; çünkü, bilimsel dergi editör ve yardımcılarının makale yönlendiriciliğine dönüşen hatalı davranış biçimlerinin, “hakem*

doğrudur, gerisi yanlış” gibi bilim dışı ve Türk eğitim bilimi adına tehlikeli bir zihniyetin yerleşmesine ve de hakem diktasının oluşabilmesine olanak sağlayabileceği düşünülmektedir. Bu nokta önemlidir, zira, böylesine tehlikeli bir zihniyetin yerleşmesinin, Türk eğitim biliminin gelişmesine de, yeni ve farklı yaklaşımların Türk eğitim bilimine kazandırılması çabalarına da, genç bilim insanlarının veya araştırma görevlilerinin literatürdeki yeni konuları Türk eğitiminin gündemine getirme çabalarına da, ağır darbeler vuracağı açıktır. Bu doğrultuda, Bademci'nin bu çalışması, [eğitim bilimle ilgili] bilimsel dergilerde ciddi ve tehlikeli bir biçimde yerleşmeye başlayan “hakem doğrudur, gerisi yanlış” zihniyetine karşılık, bilimsel doğrultuda, bilimsel temeller üzerine oturtulmaya çalışılmış ve Türk eğitim biliminin daha iyiye gitmesi adına verilmiş bir tepkidir. Bademci'nin bu çalışmasının, [eğitim bilimle ilgili ulusal hakemli dergilerde] bir ilki başlatarak, Türk eğitim biliminin çağdaş yönde ilerlemesi, demokratikleşmesi, bilimsel tutumun yerleşmesi, bir bilimsel makale veya çalışmanın objektif ölçü ve ölçütler kullanılarak ne şekilde değerlendirilmesi veya eleştirel düşünmenin geliştirilmesi gibi bazı konulara [da] katkı sağlayabileceği düşünülmektedir.

\* Bademci'nin yazmış olduğu “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makale, 2005 yılı Nisan ayı içerisinde *Türk Eğitim Bilimleri Dergisi*'ne ulaştırılmıştır. Bademci'nin adı geçen bu makalesine yönelik olarak, iki adet [kimlikleri gizlenmiş] *Türk Eğitim Bilimleri Dergisi Makale Değerlendirme Formu* ile yine iki adet [kimlikleri gizlenmiş] hakem raporuyla birlikte, *Türk Eğitim Bilimleri Dergisi (TEBD)* Editörlüğünün 17 Haziran 2005 tarihli [üst] yazısı, yine aynı gün, yani 17 Haziran 2005'de, *TEBD* Editörlüğünce Bademci'ye gönderilmiştir. *TEBD* Editörlüğünün yazısına ve her iki hakemin bilimsel ve objektif kanıtlara ve gerekçelere dayanmayan, güvenilirliği olmayan, subjektif ölçütler ve kişisel tercihlere dayalı olduğu görülen ve de bilimsel yönden de hatalı olduğu da tespit edilen raporlarına karşılık ve bir bilimsel cevap olarak, bilimsel veri ve kanıtlara ve bilimsel gerekçelere dayalı biçimde “Hakemlerin Değerlendirmelerindeki Hatalar Üzerine: Fisher'in Z Dönüşümü Ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” başlıklı bu makale, Bademci tarafından 20 Haziran-10 Ağustos 2005 tarihleri arasında yazılmıştır. Ekleriyle birlikte toplam 63 sayfadan oluşan bir dilekçe ile, Bademci'nin daha önce yazıp göndermiş olduğu “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makale [bilimsel içeriğine hiç dokunulmadan, aynen] ve de “Hakemlerin Değerlendirmelerindeki Hatalar Üzerine: Fisher'in Z Dönüşümü Ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” isimli makale, [yani iki makale] 22 Ağustos 2005 tarihinde [kayıtlı olarak], *Türk Eğitim Bilimleri Dergisi*'ne teslim edilmiştir. 4 sayfalık dilekçede, özetle; “Testler Güvenilir Değildir: Ölçüm Güvenirliğine Yeterli Dikkat ve Güvenirlik Çalışmaları İçin Örneklem Büyüklüğü” makalesiyle ilgili olarak *TEBD* Editörlüğünce atanmış iki hakemin, bu makaleye yönelik olarak yazdıkları değerlendirme raporlarındaki yargılarının tümünün hatalı olduğu, bilimsel kanıtlara ve gerekçelere dayanmayan ve bir bilimsel çalışmanın değerlendirilmesinde olmaması gereken subjektif ölçütler ve kişisel tercihler kullanılarak ve de yanlış değerlendirme yaptıkları, her iki hakemin de yaptıkları yargılarında ve

değerlendirme raporlarında ciddi bilimsel hatalarının olduğu, *TEBD* Editörlüğünün hakem seçimlerinde ve makale hakkındaki karar verme sürecinde daha dikkatli olmaları gerektiği, hakem sürecini tamamlamış bulunan “Testler Güvenilir Değildir...” başlıklı makaleyle birlikte, iki hakemin değerlendirme raporlarındaki bilimsel hataları ve subjektif yargıları bilimsel veri ve kanıtlarıyla ortaya koyan “Hakemlerin Değerlendirmelerindeki Hatalar Üzerine..” başlıklı makalenin, her ikisinin de, [artık] *TEBD* Editörlüğünce incelenerek, aynı sayıda ve peş peşe bir sıra takip edecek şekilde, baskıya hazırlanan ilk sayıda, iki ayrı makale olarak yayınlanmasının talebi ve gerekliliği ve de bu gerekliliğin nedenleri belirtilmiştir. Yine olayları ve gelişmeleri kısaca özetlemek gerekirse; bir bilim adamı [Bademci] tarafından bir makale yazılmış, bir bilimsel dergiye [*TEBD*] gönderilmiş, bu dergi, Makaleyi incelemek ve yayınlanma veya yayınlanmamasına karar vermeleri amacıyla birisi profesör, diğeri doçent unvanlı iki hakeme göndermiş, bu iki hakem Makalenin yayınlanmaması yönünde karar vermişler ve bu yöndeki raporlarını dergiye göndermişler, dergi, bir üst yazıyla hakemlerin verdikleri “yayınlanmama” şeklindeki bu karara bir şekilde uyduğunu belirterek, iki hakemin raporlarını da Makale sahibine [Bademci’ye] göndermiş, Makale sahibi de her iki hakemin değerlendirme raporlarındaki bilimsel hataları ve subjektif ölçü ve kişisel tercihlere dayalı yargılarını, bilimsel veri ve kanıtlara dayalı biçimde ortaya koyarak, ikinci bir Makale yazmış, ilk Makalenin de, ikinci Makalenin de dergi editörlerince incelerken baskıya hazırlanan ilk ve aynı sayıda yayımlanması amacıyla ve de bir dilekçeyle tekrar dergiye göndermiştir. Demokratik, çağdaş ve bilimsel tutum, Bademci tarafından yazılmış olan ilk Makalenin de, hakemlere yanıt olarak yazılmış ve hakemlerin değerlendirme raporlarındaki hatalarını bilimsel veri ve kanıtlarıyla ortaya koyan ikinci Makalenin de birlikte basılarak, Türk eğitim ve bilim topluluğuna sunulmasını ve eğer bir bilimsel tartışma varsa, dergi sayfalarında sürdürülmesinin sağlanmasını gerektirirdi; bu gereklilik, açıktır ki, bilimsel gelişme ve tartışmaların yerinin bilimsel dergilerin ya da yayınların sayfalarında veya ortamında olmasından veya yapılmasından ileri gelmektedir. Ancak *TEBD* Editörlüğü ilginç [ve de ilginç] bir karara imza atmıştır; Makalelerin baskıya hazırlanan ilk sayıda yayınlanılması ya da yayınlanılması amacıyla sıraya konulması yerine, *TEBD* Editörlüğü, Makaleler için tekrar hakem sürecinin başlatıldığını ve [yeniden başka] hakeme gönderildiğini [Makalelerin sahibi Bademci’ye] ifade etmiştir. Şüphesiz *TEBD* Editörlüğünün bu davranış biçiminin, çağdaş, demokratik ve objektif bilim yapma tutumuna aykırı olduğu düşünülmektedir; *TEBD* Editörlüğünün bu davranış biçiminin, bilimin ilerlemesi ve gelişmesinin can damarı olan ve olması gereken bilimsel tartışmaları, bilimsel dergi sayfalarından ve bilimsel ortamlardan ziyade, Türkiye’deki üniversitelerde gittikçe ağırlığını ve çirkinliğini artıran ve Türk eğitim ve bilimine de bir saç teli kadar dahi katkısı bulunmayan ve Türk eğitim biliminin kalite çitasını da çok aşağılara çeken kapı arkalarındaki dedikodu çarkının içine doğru sürüklenmesine katkıda bulunabilecek nitelikte olduğu söylenebilir. *TEBD* Editörlüğünün bu hatalı olduğu düşünülen davranış şekline karşı gelişen bir tepki ve de bilimsel etik ve sorumluluğun bir yansıması olarak, *TEBD*



Editörlüğüne iletilmesi amacıyla, bu satırların yazarı Bademci tarafından bir dilekçe yazılarak, bu iki makalenin *TEBD*'nde yayınlanılmasından vazgeçildiği ifade edilmiş ve makaleler Bademci tarafından geriye çekilmiştir. Adı geçen iki makale, Türk eğitim biliminde demokratik ve objektif bilim yapma ve tutumun gelişmesine katkı sağlayabileceği düşüncesiyle, on yılı aşkın süredir yayın yapan *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi (GÜESEFD)* adlı ulusal hakemli dergiye, Kasım 2005 tarihinde Bademci tarafından gönderilmiş ve makaleler tekrar *GÜESEFD* adlı dergide [hakemlere gönderilmesi amacıyla] yeniden hakem değerlendirme sürecine girmiştir.

*Hakemli dergilerde bir ilk olabileceği düşüncesiyle ve bilimsel bir örnek teşkil etmesi amaçlandığından dolayı da, sadece ve sadece bu makaleye özgü olarak, hakemlerin isimlerinin ilgili yasalar uyarınca Bademci tarafından TEBD Editörlüğü'nden istenmesi ve de ilgili kişilerin [hakemlerin] açık unvan ve isimlerine yönelik olarak [bilimsel] yanıt verilmesi yoluna [bu makalede] gidilmemiştir. Ayrıca, ulusal hakemli dergilerde, genel geçer davranış biçimi olan ve makale yazar veya yazarlarının isimlerinin kapatılarak hakemlere gönderilmesi ve hakemlerin değerlendirme raporlarındaki isimlerinin kapatılarak da makale sahiplerine tekrar gönderilmesi tutum ve davranışının, bilimin demokratik ve objektif ilerlemesi ve gelişmesi yönüne, olumsuz katkı yapacağı da düşünülmektedir.*

\*\* Metin içindeki [...] arasındaki ifadeler, Vahit Bademci tarafından eklenmiştir.

## KAYNAKLAR

- Airasian, P.W. (2000). *Assessment in the Classroom: A Concise Approach*.(Second Edition). Boston:McGraw-Hill.
- Akhun, İ.(1986).*İstatistiklerin Manidarlığı ve Örneklem*.(Geliştirilmiş İkinci Baskı). Ankara.
- Akhun, İ. (1982). *Hipotez Testi ile İlgili Bir Araştırma*. Ankara: A.Ü.Eğitim Fakültesi Yayınları.
- Bademci, V. (2004). “Testin Güvenirliği” veya “Test Güvenilirdir” Diye İfade Etmek Doğru Değildir.*Türk Eğitim Bilimleri Dergisi*, Cilt 2, 367-372.
- Bademci, V. (2001). *Türkiye'deki Okullar Ne İşe Yarar?* (Geliştirilmiş İkinci Baskı). Ankara: Alp.
- Bademci, V. (1999). *Hedefin Davranışlara Çevrilmesi, Davranışlardan Seçmeli Test Maddeleri Yazılması*. (Geliştirilmiş Üçüncü Baskı). Ankara: Gazi.

- Bloom, B.S. (Ed.) ve diğeri (1956). *Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook 1: Cognitive Domain*. New York: David McKay.
- Caruso, J.C. (2000). Reliability Generalization of the Neo Personality Scales. *Educational and Psychological Measurement*, Vol.60, 235-254.
- Charter, R.A. (2003). Study Samples Are Too Small to Produce Sufficiently Precise Reliability Coefficients. *The Journal of General Psychology*, Vol. 130, 117-129.
- Charter, R.A. (2001). Damn the Precision, Full Speed Ahead With the Clinical Interpretation. *Journal of Clinical and Experimental Neuropsychology*, Vol. 23, 692-694.
- Charter, R.A. (1999). Sample Size Requirements for Precise Estimates of Reliability, Generalizability, and Validity Coefficients. *Journal of Clinical and Experimental Neuropsychology*, Vol. 21, 559-566.
- Charter, R.A. (1997). Confidence Interval Procedures for Retest, Alternate-Form, Validity, and Alpha Coefficients. *Perceptual and Motor Skills*, Vol.84, 1488-1490.
- Charter, R.A. ve Feldt, L.S. (2002). The Importance of Reliability as It Relates to True Score Confidence Intervals. *Measurement and Evaluation in Counseling and Development*, Vol.35, 104-112.
- Choppin, B.H. (1990). Evaluation, Assessment, and Measurement. *The International Encyclopedia of Educational Evaluation*. Ed.H.J.Walberg ve G.D.Haertel. Oxford: Pergamon.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Fort Worth: Holt, Rinehart and Winston.
- Çoker, D. (1989). *Bilimsel ve Uygulamalı Terimler Sözlüğü*. Ankara: Yetkin.
- Demirel, Ö. (1997). *Kuramdan Uygulamaya Eğitimde Program Geliştirme*. Ankara: USEM.
- Demirel, Ö. ve Ün, K. (1987). *Eğitim Terimleri*. Ankara: Şafak Matbaası.
- Downie, N.M. ve Heath, R.W. (1965). *Basic Statistical Methods*. (Second Edition). New York :Harper and Row.
- Edwards, A. L. (1976). *An Introduction to Linear Regression and Correlation*. San Francisco: W.H. Freeman and Company.
- Ertürk, S. (1984). *Eğitimde "Program" Geliştirme*. (Beşinci Baskı). Ankara: Meteksan.

- Feldt, L.S. ve Brennan, R.L. (1989). Reliability. *Educational Measurement*. (Third Edition). Ed. R.L.Linn. New York: American Council on Education ve Macmillan.
- Ferguson, G.A. ve Takane, Y. (1989). *Statistical Analysis in Psychology and Education*. (Sixth Edition). New York: McGraw-Hill.
- Gage, N.L. (1963). Paradigm for Research on Teaching. *Handbook of Research on Teaching*. Ed.N.L.Cage. Chicago: Rand MçNally.
- Gardner, H. (1995). Reflections on Multiple Intelligences: Myths and Messages. *Phi Delta Kappan*, Vol.77, (3), 200-209
- Garrett, H.E. (1953). *Statistics in Psychology and Education*. (Fourth Edition). New York: Longmans,Gren and Co.
- Glass, G.V. ve Hopkins, K.D. (1996). *Statistical Methods in Education and Psychology*. (Third Edition). Boston: Allyn and Bacon.
- Grobman, H. (1970). *Developmental Curriculum Projects: Decision Points and Processes*. New York: F.E.Peacock
- Gronlund, N. E. ve Linn, R. L. (1990). *Measurement and Evaluation in Teaching*. ( Sixth Edition). New York: Macmillan.
- Guilford, J. P. (1954). *Psychometric Methods*. (Second Edition). New York: McGraw-Hill.
- Guilford, J.P. ve Fruchter, B. (1973). *Fundamental Statistics in Psychology and Education*. (Fifth Edition). New York: McGraw-Hill.
- Gulliksen,H.(1950). *Theory of mental Tests*. New York: John Wiley and Sons.
- 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)
- 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.
- Kanji, G.K. (1999). *100 Statistical Tests*. (New Edition). London: Sage.
- Keeves, J. P. (1988). Introduction: The Improvement of Measurement for Educational Research.*Educational Research,Methodology,and Measurement: An International Measurement*. Ed.J.P.Keeves. Oxford: Pergamon.
- Kline, P. (1986). *A Handbook of Test Construction: Introduction to Psychometric Design*. New York: Methuen.

- Kutay, B. (2004).Eğitimde Müzik,Matematik Kadar Önemli.*Eğitimbilim*.Yıl 6,Sayı 67,34-38.
- Lyman, H. W. (1991). *Test Scores and What They Mean*. (Fifth Edition). Boston: Allyn and Bacon.
- Magnusson, D. (1967). *Test Theory*. Massachusetts: Addison-Wesley.
- Marascuilo, L.A. ve Serlin, R.C. (1988). *Statistical Methods for Social and Behavioral Sciences*. New York: W.H.Freeman and Company.
- Mehrens, W.A. ve Lehmann, I.J. (1991). *Measurement and Evaluation in Education and Psychology*. Fort Worth: Harcourt Brace.
- Mehrens, W.A. ve Lehmann, I.J. (1986). *Using Standardized Tests in Education*. (Fourth Edition). New York: Longman.
- Nitko, A.J. (2001). *Educational Assessment of Students*. (Third Edition). New Jersey: Prentice-Hall.
- Nunnally, J. C. (1982). Reliability of Measurement. *Encyclopedia of Educational Research*. (Fifth Edition). (Ed. H.E.Mitzel). New York: The Free Press.
- Nunnally, J. C. (1978). *Psychometric Theory*. (Second Edition). New York: McGraw-Hill.
- Nunnally, J. C. (1967). *Psychometric Theory*. New York: McGraw-Hill.
- Nunnally, J.C. ve Bernstein, I.H. (1994). *Psychometric Theory*. (Third Edition). New York: McGraw-Hill.
- Özçelik, D.A. (1992). *Eğitim Programları ve Öğretim (Genel Öğretim Yöntemi)*. (Üçüncü Baskı). Ankara: ÖSYM Yayınları.
- Popham, W.J. ve Sirotnik, K.A. (1992). *Understanding Statistics in Education*. Itasca, Illinois: F.E. Peacock.
- Püskülcü, H. ve İkiz, F. (1989). *İstatistiğe Giriş*. (Üçüncü Baskı). İzmir: Bilgehan Basımevi.
- Roscoe, J.T. (1975). *Fundamental Research Statistics for the Behavioral Sciences*. Second Edition. New York: Holt,Rinehart and Winston.
- Ross,C.C. (1954). *Measurement in Today's Schools*. Third Edition. Rev. J.C.Stanley. Englewood Cliffs,N.J.: Prentice-Hall.
- Sadker, M. ve Sadker, D. (1999). Questioning Skills. *Classroom Teaching Skills*. (Sixth Edition). Ed. J. M. Cooper.Boston: Houghton Mifflin.

- Segall, D.O.(1994).The Reliability of Linearly Equated Tests.*Psychometrika*,Vol.59,361-375.
- Sheskin, D.J. (2004). *Handbook of Parametric and Nonparametric Statistical Procedures*. (Third Edition). Boca Raton: Chapman and Hall/Crc
- Smithson, M. (2000). *Statistics with Confidence*. London: Sage.
- Sönmez, V. (2003). *Program Geliştirmede Öğretmen Elkitabı*. (Onuncu Baskı). Ankara: Anı.
- Spiegel, M.R. (1988). *Schaum's Outline of Theory and Problems of Statistics*. (Second Edition). New York: McGraw-Hill.
- Spiegel, M.R. (1961). *Schaum's Outline of Theory and Problems of Statistics*. New York: Schaum Publishing Company.
- Streiner, D.L. and Norman, G.R. (1995). *Health Measurement Scales: A Practical Guide to Their Development and Use*. (Second Edition). Oxford University Press.
- Suen, H.K. (1990). *Principles of Test Theories*. New Jersey: Lawrence Erlbaum.
- Tekin, H. (1982). *Eğitimde Ölçme ve Değerlendirme*. (Üçüncü Baskı). Ankara: Daily News Web Ofset.
- Thompson, B. (1994). Guidelines for Authors.*Educational and Psychological Measurement*, Vol. 54, 837-847.
- Thompson, B. (1992). Two and One-Half Decades of Leadership in Measurement and Evaluation. *Journal of Counseling and Development*, Vol.70, 434-438.
- Thompson, B. (1991). Review of Generalizability Theory: A Primer by Richard J. Shavelson and Noreen M. Webb. *Educational and Psychological Measurement*, Vol. 51, 1069-1075.
- Thompson, B. ve Vacha-Haase, T. (2000). Psychometrics is Datametrics : The Test is Not Reliable. *Educational and Psychological Measurement*, Vol. 60, 174-195.
- Thorndike, R.M., Cunningham,G.K., Thorndike,R.L. ve Hagen, E.P. (1991). *Measurement and Evaluation in Psychology and Education*. (Fifth Edition). New York: Macmillan.
- Tinkelman, S.N. (1971). Planning the Objective Test. *Educational Measurement*. Second Edition. Ed. R.L. Thorndike. Washington,D.C.: American Council on Education
- Turgut, M. F. (1993). *Eğitimde Ölçme ve Değerlendirme Metotları*. (Dokuzuncu Baskı).

Ankara: Saydam Matbaacılık.

*Türk Eğitim Bilimleri Dergisi (TEBD)*. (2004).

*Türkçe Sözlük*. (1988). Ankara:Türk Dil Kurumu Yayınları.

Ünver, Ö ve Gangam, H. (1986). *Uygulamalı İstatistik Yöntemler*. Ankara.

Weitzman, E. ve Mc.Namara, W. J. (1953). *Sınıfta Test Nasıl Yapılır*. (Çeviren:V. D. Pars). İstanbul: Milli Eğitim Basımevi.

Yıldırım, C. (1973). *Eğitimde Ölçme ve Değerlendirme*. Genişletilmiş İkinci Baskı. İstanbul: Milli Eğitim Basımevi.