

Ölçme ve Öğretmen

Robert L. EBEL*

Çev. : Dr. Nizamettin KOÇ**

Bir başarı testinin etkililiği, büyük ölçüde onun yapımının altında yatan ilkelerin akla uygunluğuna dayanır. Bu makalede Robert L. Ebel'in, eğitimde başarının geçerli olarak ölçülebilmesine temel teşkil eden ilkeler olarak belirlediği 10 ilke sunulmaktadır. Bu ilkelerden her biri hakkındaki kısa bir tartışma, söz konusu olan ilkenin önemini ve Sınıf Başarı Testleri'nin yapımı ve kullanımıyla ilişkisini göstermeyi amaçlamaktadır.

Eğitimde başarının ölçümüne ilişkin olarak bu makalede sunulan ilkeler, Sınıf Başarı Testleri Uygulamaları'nın geliştirilmesinde çalışmakta olan çok sayıda kişinin araştırma ve deneyimlerine dayanmaktadır. Ayrıca burada tartışılan belirli ilkeler, eğitimde, başarı testlerinin hazırlanması ve kullanılmasının düşünüldüğü zamanlarda en çok ortaya çıkan soru ve sorunlarla ilişkileri temeline dayalı olarak seçilmişlerdir. İlkelerden bazılarının eleştiriye açık olabilmelerine karşın, yine de herbirini destekler nitelikte bir tartışma ortaya konulabilir.

1. Eğitimde başarının ölçülmesi, etkili öğretim için gereklidir. Öğrenme, insan yaşamının doğal ve kaçınılmaz bir sonucudur. Okullarda hiçbir özel önlem alınmazsa ya da öğrenmeyi kolaylaştırmak için hiçbir çaba harcanmazsa bile, bir parça öğrenme olacaktır. Fakat, okuma, bilimin anlaşılması ya da yazınsal yapıtlardan zevk alınması gibi karmaşık davranışların tam öğrenimi özel güdülenme, rehberlik ve yardım gerektirir. Çabalar, belirli amaçlara ulaşılmasına yönlendirilmelidir. Eğitim sürecinde yer alan öğrenciler, öğretmenler ve diğer ilgililer, amaçların ne dereceye kadar başarılı olduğunu bilmek zorundadırlar. Eğitimde başarının ölçülmesi bu etkinliklere katkıda bulunabilir.

Zaman zaman, okulların testler (dolayısıyla ölçme) olmadan da işleyebileceği ya da testlerin yasaklanması durumunda daha iyi işlev

*EBEL, Robert L., «Measurement and the Teacher», Educational And Psychological Measurement — Contributions To Theory And Practice (Ed., David A. Payne And Robert F. McMorris), Blaisdell Publishing Company, Waltham, Massachusetts, 1967, p. 20-27.

**Ankara Üniversitesi Eğitim Fakültesi, Eğitimde Psikolojik Hizmetler Bölümü.

görebilecekleri ileri sürülür. Ancak, özel hiçbir amaçları olmayan ya da ne kadar öğreniliyor olduğuna önem vermeyen öğretmen ve öğrenciler tarafından eğitimin etkili olarak yürütülebileceği savları ya çok seyrek ya da hiç yoktur. Eğer testler yasa dışı kılınırsa, eğitimde başarıya değerlendirmenin diğer bazı araçlarının bunların yerine kullanılmak zorunluluğu ortaya çıkacaktır.

2. Bir eğitimsel test, öğretmenin, öğrenci başarısına ilişkin gözlemlerini kolaylaştıran, genişleten ve özleştiren bir araçtan başka bir şey değildir. Günlük yaşamamızda çoğumuz kendimizi, oldukça sık olarak, başkaları hakkında yargılarda bulunurken görürüz. Örneğin; Aday A, Oylama aday B'den daha mı değerlidir? Doktor C, D'den daha iyi bir doktor mudur? İşçi E, hizmetleri karşılığı maaş artırımını ya da ödüllendirilmeyi hak etmiş midir? Sınıf temsilinde başrolü oynamaya öğrenci M yerine öğrenci L' mi seçilmelidir? Öğrenci F'ye zayıf bir not verilmeli midir?

Bu tür yargılarda bulunmakla yükümlü olanlar, çoğunlukla bunu oldukça yetersiz kanıtlara dayanarak yapmak zorunda olduklarını hissederek. Kararın dayanması gereken nitelikler açıkça tanımlanmamış olabilir. Çeşitli adayların edimleri (performansları) yeteri genişlikte ya da karşılaştırılabilir koşullar altında gözlenmemiş olabilir. Bu durumda karar verici, kaydedilmiş veriler yerine, yanılması olası ve söylen-ti kanıtlarla tamamlanmış belleğine dayanmak zorunda kalabilir

Öğrencilerinin başarılarını değerlendirme girişiminde bulunurken öğretmenler bunlara benzer sorunlarla karşılaşılır. Bu sorunları çözme çabalarının bir sonucu olarak testler geliştirilmiştir. Sözlü ve objektif sınavlar, öğretmen için daha geniş bir öğrenci davranışı örneklemine daha dikkatli ve kontrol edilmiş koşullar altında gözlemeyi kolaylaştıran araçlardır.

Öğrenci başarılarının gözlenmesinde testlerin, yeterlilik ve kontrol açılarından sağladıkları üstünlükler, gözlenen davranışın doğallığından bazı şeylerin yitirilmesini göze almayı gerektirir. Öğrencinin tipik davranışını, özellikle davranışın; ilgiler, tutumlar, değerler ya da duygusal tepkilere bağlı yönlerini ölçen testlerde, test durumunun yüzeyselliği elde edilen ölçümü ciddi olarak çarpıtabilir. Fakat bu sorun, öğrencinin ne kadar bildiğini ve bilgisiyle ne yapabilme yeteneğinde olduğunu ölçmeye çalışan testlerden çok daha az ciddidir. Ölçümün yeterliliği ve doğruluğu açısından kazanılanlar, içinde öğrenci davranışının gözlemlendiği durumun yüzeyselliğinden dolayı kaybedilenlere çok daha ağır basar.

3. Eğitimin önemli her ürünü (sonucu) ölçülebilir. Bir eğitim ürününün önemli olabilmesi için bir fark yapması zorunludur. Belli bir ürünün daha fazlasına sahip olan bir kişinin davranışı o ürünün daha azına sahip olan kişinin davranışından gözlenebilir bir farklılık göstermek zorundadır. Eğitimin çok fazla kişisel bir sonucu düşünülebilir, o kadar kişisel ki kişinin ne söylediğini, ne yaptığını ya da zamanını nasıl harcadığını hiçbir şekilde etkilemez. Fakat bu kadar iyi gizlenmiş bir başarının önemli olduğunu tartışabilmek için nedenler bulmak zordur.

Eğer bir başarı kişinin ne yapabildiği üzerinde bir farklılık yapıyorsa aynı zamanda ölçülebilir de. Çünkü; en temel ölçme tipleri, kişi ya da nesne X'in tanımlanmış bir niteliğe kişi ya da nesne Y'den daha fazlasına sahip olduğunun doğrulanabilir bir gözleminin yapılması olasılığından fazla bir şey değildir.

Önemli her eğitimsel ürünün ölçülebilir olduğunu söylemek, halihazırda doyurucu ölçme yöntemlerinin var olduğunu söylemek değildir. Kesinlikle, her eğitimsel ürünün kağıt-kalem testleri aracılığı ile ölçülebilir olduğunu söylemek değildir. Fakat, bazı önemli eğitimsel ürünlerin ölçülebilmek için çok karmaşık ya da soyut olduğu yolundaki savı reddetmek demektir. Önemlilik ve ölçülebilirlik mantıksal olarak birbirinden ayrılamaz.

4. En önemli eğitim başarısı yararlı bilginin (kullanılabilir bilginin) kazanılmasıdır. Eğer eğitimsel bir ürünün önemi öğretmen ve öğrencilerin zamanlarının çoğunu ne yapmaya harcadıkları temeline dayalı olarak değerlendirilebilirse yararlı bilgiye hakimiyetin kazanılmasının oldukça önemli bir ürün olduğu açıktır. Ya da eğer diğer amaçlara — kendini gerçekleştirme, insan ilişkileri, ekonomik yeterlilik, medeni sorumluluk amaçları — nasıl ulaşılabileceği sorulursa yine açıktır ki yararlı bilgiye hâkimiyet tek temel araçtır.

Bir kişinin bir problem hakkında ne kadar etkili olarak düşünebileceği geniş ölçüde onun probleme ilişkin bilgiye ne kadar etkili olarak hakim olabildiğine bağlıdır. Bilgiye hakimiyet başarıyı, mutluluğu ya da haklılığı garanti etmez. Fakat bir okulun bu amaçlara götürmede bilgiye hakimiyetin yarısı kadar başarı olasılığı bulunan başka bir şeyi geliştirme girişiminde bulunabileceğini düşünmek zordur.

Eğer biz öğrencilere bilgiye hakimiyeti kazandırırsak, düşünme yeteneklerini geliştirirsek onları zihinsel olarak özgür ve bağımsız yaparız. Bu onların mevcut durumu muhafaza etmek için çok çalışacakları ve bizim inançlarımızın hepsini benimseyecekleri, bütün değerle-

rimizi kabul edecekleri konusunda bize güvence vermez. Bununla beraber onları özgürlüğün en önemli olduğu alanlarda özgür kişiler yapar. Öğrencilerin davranışını, akla uygun bir biçimde, kendini yöneltme temelinden, yani bilgiye hakimiyetin sağladığı bir temelden başka bir temele dayalı olarak değiştirmeyi ya da kontrol etmeyi amaçlayan bir eğitim programı konusunda da uyanık olmalıyız.

5. Yazılı testler öğrencinin yararlı bilgiye hakimiyetini ölçmeye oldukça uygundur. Bütün bilgiler öneriler içinde ifade edilebilirler. Öneriler yanlış ya da doğru olarak yargılanabilen ifadelerdir. Bilginler araştırmacılar, incelemeciler — bilgi anbarımıza birşeyler eklemeyi düşünen herkes — zamanlarının çoğunu öneriler geliştirmeye ve bunları doğrulamaya harcarlar.

Her doğru — yanlış ya da çoktan seçmeli test maddesinde bir ya da bir çok öneri gizlidir. Kolay testler de bir öğrencinin bilgiye hakimiyetini gerektirir.

Bir öğrencinin bilgiye hakimiyetini test etmeyi amaçlayan her soruda bazı yenilik elemanları gereklidir. Öğrencinin ezbere ya da sözel çağrışıma dayalı olarak başarılı tepkide bulunmasına izin vermemelidir. Önceden zihinde cevabı hazır kalıplaşmış sorular sorulmamalıdır.

6. Sınıf öğretmeni eğitimsel başarıyı ölçmede kullanılan testlerin çoğunu sınıfta hazırlamalıdır. Bilginin çeşitli alanlarında eğitimsel yetenek ya da başarıyı ölçmede kullanılan birçok yayınlanmış test sınıf kullanımı için hazır durumdadır. Fakat özel bir çalışma (iş) biriminin ya da özel bir öğretim döneminin amaçlarında başarıyı ölçmeye özellikle uygun sadece birkaç test vardır. Bazen ders kitabı yayıncıları test soruları kitapçıklarını ders kitabı ile birlikte sunarlar. Bu gibi test sorularının genellikle düşük kalitede acele yazılmış, gözden geçirilmemiş, ön test edilmemiş ve ezberlemeye dayalı olarak doğru tepkileri kendiliğinden verebilmelerine karşın faydalı olabilirler.

Önceden iyi bir biçimde hazırlanmış testler el altında bulunsalar da genellikle öğretmen yapısı testler lehine bir tartışma yine de ortaya konulabilir. Bunun temel nedeni test geliştirme sürecinin öğretmenin amaçlarını tanımasına yardımcı olmasıdır. Bu süreç öğretmen yapısı bir testin dış bir teste nazaran öğretmenin konusu ile ilişkili olmasının bir sonucu olabilir. Diğer yandan, öğretmenin dersi ve konusu ile ilgili olarak test geliştirmesi, eğitimsel başarıyı ölçme sürecini olması gerektiği gibi tüm öğretim sürecinin bütünleyici bir parçası yapabilir.

7. Başarıyı etkili olarak ölçmek için sınıf öğretmeni; a) test edilecek hüner ya da bilgide, b) uygulamalı test etme sanatında, usta olmak zorundadır. Ne eğitimde ölçme konusunda açılan kurslar ne de

«öğretmen yapısı test geliştirme» konusunda yazılmış kitap ya da makaleler zayıf öğretmenlerin iyi testler yapabilmesini sağlar. Bir öğretmenin öğretmeye çalıştığı bilgiye hakimiyeti, bu içeriğe ilişkin genel yanlış kavramlaştırmaları anlayışı, yeni sorular ve problemler üretme-deki yeteneği ve bunları açık-seçik ifade yeteneği; onun test yapma-daki başarısında kesin öneme sahiptir. Ne yazık ki, iyi öğretmenin ve test etmenin bu ön istemlerinden bir ya da daha fazlasından yoksun bazı kişilerin öğretmenlik sertifikasına sahip oldukları da doğrudur.

Bu arada, test yapma işinin de bazı kurnazlıkları vardır. Eğitimde ölçme dersi ya da sınıf başarı testi uygulaması alanında yazılmış bir kitap ya da makale bunları öğretebilir. Böyle bir ders ayrıca bir öğretmenin eğitimsel başarıyı test etme hakkındaki bazı yaygın yanlış kavramlaştırmalara güvenini — yapıcı ve sağlam olarak — sarsmaya hizmet edebilir. Bir öğrencinin yüksek zihinsel süreçlerinin gelişimini ölçmede yalnızca açık-uçlu testler (esse tipi testler) yararlıdır; bir test planı, öğrencinin ne oranda ne bilmek zorunda olduğunu bildiğini göstermelidir, puanlamadaki yanlışlıklar test puanlarındaki hataların başlıca kaynağıdır gibi inançlar bu yanlış kavramlaştırmalar arasındadır.

8. Bir öğretmen yapısı testin niteliği içindeki işlevlerin içerikle ilgililiğine, öğretimin bütün yönlerinin örneklenmesinin temsil ediciliğine ve verdiği puanların güvenilirliğine bağlıdır. Eğer bir test sorusu öğrencinin sınıf dışında ve sonraki yaşamında karşılaşmayı bekleyebileceği türden problemler sunarsa ve başarısının ölçülmekte olduğu ders gerçekten ona böyle problemlerle uğraşmayı öğretmeye çalışmışsa soru içerikle ilişkilidir. Eğer test soruları önemleri oranında dersin geliştirmeyi amaçladığı başarının bütün yönlerini içerirse temsil eder nitelikte örneklem almış olur. Eğer öğrencilerin testten aldıkları puanlar bağımsız eşdeğer bir başka testten aldıkları puanlarla uyşurlarsa, bu testin verdiği puanların güvenilir olduğu söylenebilir.

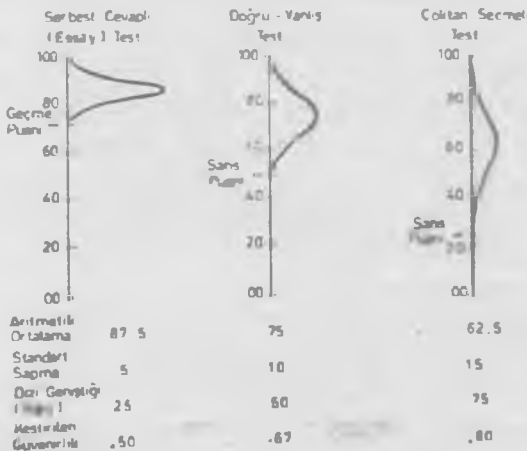
İçerikle ilişki, temsil edicilik ve güvenilirlik bir derece sorunudurlar. Test güvenilirliği tahminlerinin hesaplanmasında kullanılan işlem ve formüller iyi geliştirilmiştir ve eğitimde ölçme üzerine yazılmış kitapların pek çoğunda betimlenmiştir. Temsil edebilirlik ve içerikle ilişki tahminleri daha öznel ve daha az nicelikseldir. Bununla beraber, bu söylenenler, içerikle ilişki ve temsil edebilirliğin güvenilirlikten daha az önemli olduğu biçiminde bir yanlış anlamaya yol açmamalıdır. Bir test her birinin daha fazlasına sahip olduğu ölçüde iyidir. İçerikle ilişgisiz ve temsil edebilirliği olmayan ama oldukça güvenilir bir test yapmak olası iken bu üç özelliğin herhangi birinden ötekiler lehine fedakârlıkta bulunmak çok seyrek olarak gerekebilir. Böyle bir fedakârlıkta bulunmak ise hiçbir zaman arzu edilen bir durum değildir.

Hem açık uçlu hem de kapalı uçlu kısa cevap maddelerinden oluşan test formları sınananlara içerikle ilişkili sorular sunmada kullanılabilirler. Alışlageldiği gibi, bir test sorusunun yeniliği ölçüsünde, yani öğrencinin o soru ile önceden karşılaşmış olma ya da ona uygun bir cevap öğretilmiş olma olasılığının düşüklüğü ölçüsünde içerikle ilişkisi yüksektir. Daha büyük sayıda sorunun içeriğe katılmış olması yönünden bazan kısa cevap maddelerinden oluşan (objektif) bir testte soruların temsil edebilir örneklemini almak, aynı şeyi açık uçlu — serbest cevaplı sorulardan oluşan testlerde yapmaktan daha kolaydır. Aynı nedenle ve ayrıca puanlamadaki daha yüksek tekdüzelik yüzünden kısa cevap maddelerinden oluşan testler serbest cevaplı testlerden biraz daha güvenilir puanlar vermek eğilimindedirler.

9. Belirli bir en yüksek olası puanı amaçlayan test: en gelen puanların değişkenliği ölçüsünde bu puanların beklenen güvenilirlikleri yüksek olacaktır. Güvenirlik bazan ölçme hatalarına atfedilemeyen test puanları arasındaki toplam değişebilirlik oranı olarak tanımlanır. Ölçme hatalarının büyüklüğü testin doğasına bağlıdır — yani, testteki soruların (maddelerin) sayı ve çeşidine bağlıdır. — Böylece belirli bir test için puanların toplam değişebilirliklerindeki herhangi bir artış, ölçme hatalarından gelmeyen oranı ve testin güvenilirliğini artırmak eğilimindedir.

Şekil 1, üç test için kuramsal puan dağılımlarını göstermektedir. Açık uçlu serbest cevaplı test herbiri 10 puan değerinde 10 sorudan oluşur ve böyle bir testte 75'i geçer puan olarak gören bir öğretmen tarafından puanlanır. Doğru-yanlış test, doğru olarak cevaplandırıldığı da herbiri 1 puan alan, yanlış cevaplar için çıkarma yapılmayun 100 sorudan oluşur. Çoktan seçmeli test ayrıca herbiri dört seçeneikli cevap seçimini içeren 100 sorudan oluşur. Bu da sadece doğru cevapların sayısı ile puanlanır ve şans başarısı için düzeltmesi yapılmaz.

Şekil 1 Üç test için kuramsal puan dağılımları

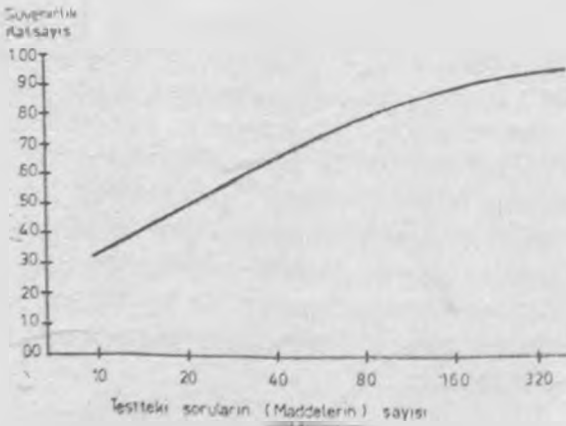


Şekil 1'in tabanındaki verilerde ortalama puan (aritmetik ortalama)'da testler arasındaki farklılıklara, değişebilirliğe (standart kaymayı), etkili ranj'a ve tahmin edilen güvenilirliğe dikkat ediniz. Belli varsayımlar temeline dayalı hesaplamalardan elde edilen bu kuramsal verilerden çoğu, öğretmenlerin bu tip testleri kullanarak başardıkları sonuçları belki de anlaşılabilir ölçüde temsil etme yeteneğindedirler.

100'ün çoktan seçmeli test maddesi kullanarak güvenilirliği, 90'ın üzerinde puanlar elde etmek olasıdır. Fakat bunu yapmak oldukça güçtür ve sınıf öğretmenleri hazırladıkları testlerde bunu çok seyrek yaparlar. Güvenirlikleri 100 maddelik çoktan seçmeli testinkine eşit olabilecek 100 puanlı serbest cevaplı testler ve 100 maddelik doğru-yanlış testlerle uğraşmak mümkündür. Fakat tekrar belirtmek gerekir ki bunu yapmak kolay değildir ve sınıf öğretmenleri bunu yapmada çok seyrek olarak başarılıdırlar.

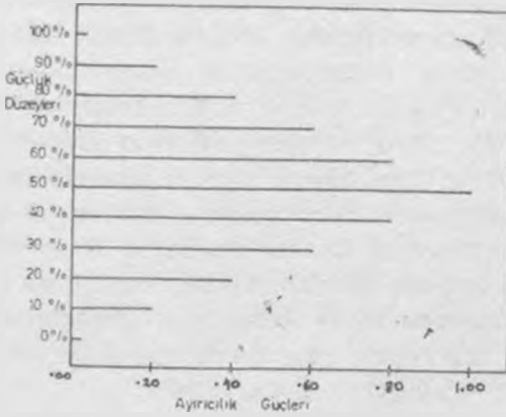
10. Testin güvenilirliği, içerdiği soruların (bağımsız olarak puanlanabilen) artırılmasıyla ve bu arada yüksek ve düşük başarılı öğrenciler arasında ayırım yapmak için bireysel soruların güçlüklerinin duyurulmasıyla artırılabilir. Şekil 2 bir testte bağımsız olarak puanlanabilen soruların sayısını artırmanın sonucu olarak test güvenilirliğinde beklenebilecek artışları göstermektedir. Güvenirlik katsayısı .33 olan 10 soruluk bir testin uzunluğunu iki katına çıkarmak güvenilirliği .50'ye yükseltir. Uzunluğu tekrar iki katına çıkarma güvenilirliği .67 yapar ve soru sayısındaki artışlar devam ettikçe güvenilirliğin artışı da devam eder. Bu hesaplamalar, soru sayısı artırılmış bir testin güvenilirliğini kestirme (prediction) için kullanılan ve Spearman-Brown tarafından geliştirilen formüle dayanmaktadır. Bu formül bütün durumlarda haklı gösterilemeyecek varsayımları gerektirirse de kestirmeleri (prediotions) çoğunlukla ve oldukça doğrudur.

Şekil 2 Test güvenliğinin test uzunluğu ile ilişkisi



Şekil 3 bir sorunun en yüksek ayırıcılık gücünün (discriminating power) aynı sorunun güçlük düzeyi (level of difficulty) ile nasıl ilişkili olduğunu göstermektedir. Bu ayırıcılık göstergeleri (discrimination indices) iyi öğrenciler ile zayıf öğrencilerden gelen doğru cevap oranları arasındaki farkları yansıtır. İyi öğrenciler toplam test puanları test edilen öğrencilerin üst % 27'si arasına düşenlerdir. Zayıf öğrenciler ise puanları alt % 27'yi oluşturanlardır. Güçlük düzeyi .50 olan (yani, test edilen grubun % 50'si tarafından doğru cevaplanan) bir sorunun ayırıcılık göstergesi her zaman 1.00 olmak zorunda değildir. Ayırıcılık gücü 0 (sıfır) ve hatta negatif de olabilir. Fakat orta güçlükteki sorular kendi ayırıcılık güçlerinde daha yüksek tavanlara sahiptirler. Daha önemli olanı ise bunlar çok kolay ya da çok zor sorulardan daha yüksek bir ayırıcılık gücüne sahip olabilmekle kalmazlar, bu gerçekten böyledir. Herkesin doğru olarak cevaplayabildiği bir sorunun ayırıcılık gücü hiç yoktur. Böyle bir soru testin güvenilirliğine hiçbir şey eklemeyiz.

Şekil 3 Farklı güçlük düzeylerinde sorulara elde edilebilecek (Ulaşılabilecek) en yüksek ayırıcılık güçleri



Özet olarak, bu makalede belirtilen ve tartışılan 10 ilke eğitimde ölçme konusunda sınıf öğretmenlerinin bilmek gereksiniminde oldukları önemli şeylerin sadece bir örneğini temsil eder. Bu ilkeler ve herbirinin burada sunulan kısa tartışması sınıf başarı ölçümlerindeki bazı yaygın uygulamalara gölge düşürmeye ya da öğretmen yapısı testlerin iyileştirilebileceği bazı yollar ileri sürülmesine hizmet edebilir. Ayrıca bu ilkeler sınıf başarı testlerinin geliştirilmesi için söylenmesi gerekenlerin tümünü söylemek, yapılması gerekenlerin tümünü yapmak anlam ve niyetini taşımamaktadır. Ancak, bizim içten inancımız odur ki, hazırladığı sınıf başarı testleri ile bu ilkelerin anlaşıldığını yansıtan bir öğretmen, öğrenci başarısını ölçme gibi orta bir işten daha iyisini yapmış olacaktır.