

## Geniş Ölçekli Başarı Testlerinin Eğitimindeki Yeri ve Önemi

### The Role and Importance of Large-Scale Achievement

Mehtap ÇAKAN

Abant İzzet Baysal Üniversitesi

#### Öz

Bu çalışmada, geniş ölçekli başarı testlerinin eğitim sistemimizde nasıl kullanılabileceği ve ne tür faydalar sağlayabileceği üzerinde durulmaktadır. Çalışmada ilk olarak, geniş ölçekli başarı testlerinin başta Amerika Birleşik Devletleri olmak üzere yurtdışında nasıl kullanıldığı örnekler verilerek incelenmiştir. Daha sonra ülkemizde bu alanda yapılan çalışmalar gözden geçirilmiştir. Son olarak yurtdışı ve yurtiçi çalışmalar dikkate alınarak geniş ölçekli başarı testlerinin eğitim sistemimize getireceği katkılar; öğrenci, öğretmen, öğretim programları, öğretim strateji ve materyalleri, okul yöneticileri ve veliler boyutlarıyla incelenmiştir.

*Anahtar Kelimeler:* Geniş ölçekli başarı testleri, öğrenci başarıları, eyalet testleri.

#### Abstract

In this study, different ways of using large-scale achievement tests and approaches for benefiting from these tests in our educational system are discussed. First, applications of large-scale achievement tests in other countries, especially in USA, were investigated and related examples were presented. Then, studies of the use of large-scale testing that have been conducted in Turkey were reviewed. Finally, possible contributions that large-scale achievement tests can make to Turkey's educational system are discussed. Suggestions are made in terms of various aspects of education; students, teachers, educational curriculums, teaching strategies and materials, school managers, and parents.

*Key Words:* Large-scale achievement tests, student achievements, state testing

#### Giriş

Eğitim sistemimizin sağlıklı bir yapıya kavuşturulabilmesi için öğretim kurumlarının işleyişlerinin ve bu süreçte meydana gelen aksaklıkların bilimsel anlamda saptanması ve giderilmesi gerekmektedir. Öğrenci başarılarının yanı sıra uygulanmakta olan öğretim programlarının, öğretim yapmakla sorumlu personelin izlenmesi ve saptanan aksaklıkların giderilmesi, üzerinde durulması gereken önemli noktalar. Öğrencilerin akademik başarıları, pek çok ülkede ve eğitim camiasında, eğitimin etkililiğinin en önemli göstergesi olarak kabul edilmektedir. Amaç, öğrencilerin gerekli bilgi ve

beceriyi kazanmaları ve istenen düzeye ulaşmalarıdır. Eğitim sistemi ve öğelerine yönelik kararların çoğu, öğrencilere uygulanan çeşitli başarı testlerinden elde edilen bulgulara dayalı olarak verilmektedir. Öğrencilerin akademik başarılarının, eğitimin etkililiği olarak algılanması, eğitimin -yukarıda değinilen boyutları dahil- değişik boyutlarına yönelik kararlar verilirken ya da plan yapılırken öğrenci başarılarının tespit edilmesini gerekli kılmaktadır. Bu amaçla uygulanan testlerden elde edilen bilgiler, ihtiyaç alanlarının tespitinde ve hangi alanların geliştirilmesi gerektiğini öğrenmede kullanılabilir. Her yıl öğretim kurumlarına yapılan harcamalar dikkate alındığında, buradaki öğrencilerin istenen düzeyde başarı gösterip göstermediğinin saptanması ve başarısızlık durumu var ise bu başarısızlığı giderici önlemlerin alınması özel bir önem taşımaktadır.

Yrd. Doç. Dr. Mehtap Çakan, Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü 14280, Bolu.

Bu çalışmanın amacı, geniş ölçekli başarı testlerinin ilk ve ortaöğretim kurumlarındaki öğrencilerin çeşitli ders alanlarındaki başarılarını saptamada nasıl kullanılacağı başta olmak üzere, genel olarak, eğitim sistemimizde nasıl kullanılabileceğini incelemek ve böylesi bir yaklaşımın eğitim sistemimize katkılarını tartışmaktır. Çalışmada geniş ölçekli başarı testlerinden elde edilen sonuçların öğrenci, öğretmen, okul yöneticisi ve öğretim programlarına yönelik ne tür dönütler sağlayabileceği de ele alınmıştır. Ayrıca bu çalışmada geniş ölçekli başarı testlerinin başta Amerika Birleşik Devletleri olmak üzere gelişmiş ülkelerde nasıl uygulandığı ve sonuçlarından nasıl yararlanıldığı incelenerek bu uygulamaların ülkemize ne şekilde adapte edilip yararlanılabileceği konusu da tartışılmıştır. Kısacası, araştırmada böylesi bir yaklaşımın eğitim sistemimize katkıları tartışılacaktır. Geniş ölçekli başarı testleri, genel olarak, farklı sınıf düzeyi ve/veya ders alanlarında saptanmış olan bilgi ve becerileri içeren ve birden fazla alt testten ya da boyuttan oluşan başarı testlerini ifade etmektedir. Bu şekilde geliştirilen başarı testleri daha sonra geniş öğrenci kitlelerine uygulanmaktadır. Böylesi testlerde hazırlanan çok sayıda bilgi ve becerinin geniş öğrenci kitlesine uygulanması önemlidir. Geniş ölçekli başarı testleri, başta ABD olmak üzere gelişmiş pek çok ülkede eğitime yönelik karar verme ve planlama aşamalarında sıkça kullanılmaktadır. Böylesi bir testin geliştirilmesinde ilk aşama, bu testlerle ölçülecek olan bilgi ve becerilerin saptanmasıdır. ABD’de yaklaşık olarak bütün eyaletlerde öğrencilerin öğrenmeleri gereken bilgi ve beceriler, ilk ve ortaöğretim seviyesindeki tüm dersler için hazırlanmış ya da hazırlanma aşamasındadır (Reoben, 1997). Bu eyaletlerde testler öğrenciler için daha uygun hedeflerin saptanması, uygun öğretim programı geliştirme ve öğretmen yetiştirme gibi konularda bilgi, dönüt sağlama ve dolayısıyla reform yapmak için bir araç olarak görülmektedir (Reoben, 1997).

Milli Eğitim Bakanlığı’nın hazırladığı öğretim programları ve bunlarda her ders ve sınıf düzeyi için belirlenen hedef davranışlar dikkate alınarak hazırlanacak olan standart başarı testleri, belirlenen okullar genelinde uygulanıp sonuçlar öğrenci başarılarının izlenmesinde ve değerlendirilmesinde kullanılabileceği gibi, yukarıda değinildiği üzere bu sonuçlardan uygulanmakta olan programların, öğretimin ve öğretmenlerin

değerlendirilmesi amacıyla da yararlanılabilir. Böylesi bir uygulamadan elde edilen verilerin eğitim sisteminin tümüne genellenebilirliğinin sağlanabilmesi için bu testlerin Türkiye’ye genellenebilir bir grup üzerinde uygulanması gerekmektedir. Her bir coğrafi bölge için hazırlanmış testler, o bölgeyi temsil edecek sayı ve nitelikteki ilk ve ortaöğretim kurumlarında uygulanmalıdır. Ayrıca, bu testler illeri ve bölgeleri temsil eden okullara uygulanacağından, okullar ve bölgeler arasındaki başarı farklılıkları saptanacak, başarısızlığın giderilmesi için gerekli önlemler alınabilecektir.

#### *Geniş Ölçekli Başarı Testlerinin Yurtdışında ki Kullanımı*

Pek çok ülkede (özellikle ABD’ de olmak üzere İngiltere, Yeni Zelanda ve Avustralya’da) ulusal eğitim hedeflerini saptayıp geliştirme ve okullarca bu hedeflere ulaşma düzeylerini ölçme işlemlerini tek bir merkezden kontrol etme yönünde bir eğilim dikkat çekmektedir. Eğitim hedeflerinin geliştirilmesi ve ölçülmesi işlemlerindeki kontrol, ortak bir merkeze verilmekle birlikte, okula ayrılan bütçenin harcanması, kontrolü ve çalışanlarına yönelik kararların alınması, okul yönetimlerinin sorumluluğuna bırakılmıştır (Townsend, 1996). Türkiye, özellikle son yıllarda yaptığı bazı çalışmalarla bu ülkeler grubuna katılmış görünmektedir. Bu yaklaşımı benimseyen ülkelerde izlenen ortak strateji, saptanan ulusal eğitim hedeflerinin (öğrencilerin 21. yüzyılın gerektirdiği bilgi ve becerilerle donanık olması vb.) gerçekleştirilmesine dönük ülke genelinde ilk ve ortaöğretim kurumlarında ana hatlarıyla ortak özellik gösteren öğretim programlarının uygulanması ve ortak bir merkez tarafından bu hedeflere ulaşma düzeyinin geniş ölçekli testlerle saptanmasıdır. Uygulanan öğretim programlarının ana hatlarıyla ortak özellik göstermesi, test sonuçlarının çeşitli okulların karşılaştırılmasında kullanılabilmesi için bir zorunluluktur.

Avustralya’da ülke genelinde uygulanan standart test programı (Australasian Cooperative Testing Program - ACAP) tanılayıcı (diagnostic) bir yaklaşım kullanarak öğretmenlerin, öğrencilerin başarılarını öğrenmeyi ve geliştirmeyi hedeflemektedir. Testin kullanılmasındaki bir başka hedef ise, halkı öğrenci başarıları hakkında haberdar etmektir (Dunn, 1989).

ABD’de bütün eyaletlerde ülke genelinde uygulanmak üzere geliştirilen “National Assessment of Educational Progress (NAEP)” adıyla standart bir başarı



testi kullanılmaktadır. Bu test, eyaletlerin her bir ders alanı için ve diğer bazı kriterler açısından karşılaştırılmasında ve yıllar genelinde eyaletlerin başarılarının izlenmesinde kullanılmaktadır. Bütün ülkede uygulanan bu teste ek olarak eyaletlerin tümünde o eyalete mahsus standart eyalet-test programları (state-wide standardized testing programs) uygulanmaktadır (O'Neil, 1992). Eyaletler genelinde uygulanan geniş ölçekli standart başarı testlerinden oluşan bu programların uygulanması 1970' li yıllarda başlamıştır (Barry, Nielsen, Glasnapp, Poggio ve Sundbye, 1997; Barton ve Coley, 1994). Bu programların başlatılmasındaki ana hedef, eğitimde reform yapmak ve buna bağlı olarak okulların verimliliğini -school effectiveness- arttırmaktır (Strong ve Sexton, 1996; Barton ve Coley, 1994).

ABD'de yaklaşık olarak bütün eyaletlerde ilk ve ortaöğretim seviyesindeki öğrencilerin ilgili bütün ders alanları için öğrenmeleri gereken bilgi ve beceriler hazırlanmış ya da hazırlanma aşamasındadır (Roeben, 1997). Standart eyalet testleriyle bu bilgi ve becerilere öğrencilerin ne ölçüde sahip oldukları test edilmektedir. Bu testler öğrenciler için daha uygun hedeflerin saptanması, uygun öğretim programı geliştirme ve öğretmen yetiştirme gibi konularda bilgi ve dönüt sağlama ve buna bağlı olarak eğitimde reformlar yapma amaçlı kullanılmaktadırlar (Roeben, 1997). İlk aşamada, eyaletler tarafından okulların ulaşması gereken hedefler belirlenmekte ve bunları ölçen standart testler geliştirilmektedir. Daha sonra bunu, geliştirilen testlerin eyalet genelinde tüm öğrencilere ya da eyaleti temsil edici okullara uygulanması izlenmektedir. Uygulanan testlerin yaklaşık % 70'i çoktan seçmeli ve %12'si yazılı yoklama formatında olmaktadır (Barry ve diğerleri, 1997). Son yıllarda çoktan seçmeli testlerden performansa dayalı testlere (performance-based assessment) doğru yönelme başlamıştır (Madaus, 1994). Ayrıca son yıllarda ölçüt -dayanaklı geniş ölçekli testlere (criterion-referenced state testing) olan ilgi de artmış görünmektedir. Bu ilginin artmasındaki ana neden, planlamada ve karar verme aşamalarında norm-dayanaklı testlerin sağladığından daha fazla tanılayıcı (diagnostic) bilgiye olan ihtiyacın artmasıdır (Ascher, 1975). Bu da ölçüt-dayanaklı testler ile sağlanabilmektedir.

Testleri kullanım amaçları ve izledikleri işlemler açısından ABD'deki eyaletler, yüksek oranda benzerlikler göstermekle beraber, belli farklılıklar da dikkat

çekmektedir. Bu farklılıklardan birisi eyaletlerin test sonuçlarından yararlanmada izledikleri yaklaşımlardır. Bunlardan Maryland Eyaleti'nde, eğer okulların uygulanan test sonucuna göre önceden belirlenen standartlara ulaşmadığı saptanırsa, bu okullar performans düzeylerini artırmak amacıyla bir program hazırlayıp uygulamakla yükümlü bırakılmaktadır. (Yen ve Ferrara, 1997). Sürekli olarak başarısızlık gösteren okullar ise diledikleri takdirde "yardıma muhtaç okullar (Challenge Schools)" adı verilen okullar grubuna alınmaktadır. Bu sınıflandırmaya giren okullara eyaletin belirlediği eğitim standartlarına ulaşmada ve öğrenci devamsızlığı, okul terkleri gibi konularda eyaletten uzman ve para yardımı verilmektedir (Yen ve Ferrara, 1997). Maryland Eyaleti'nde uygulanan test programı, bireysel öğrenci başarısından ziyade okulların performanslarını değerlendiren bir test programı olma özelliği göstermektedir. Bu programda matematik, okuma, yazma, sosyal bilimler, fen bilimleri gibi alanların her biri için okullar "yeterli", "yetersiz", ve "mükemmel" olmak üzere üç aralıkta derecelendirilmektedirler. Test sonuçları ayrıca okulların yıllara göre izlenmesinde ve okulları iyileştirici programların geliştirilmesinde de kullanılmaktadır (Yen ve Ferrara, 1997).

Delaware Eyaleti'nde test sonuçları öğrenci başarılarının tespit edilmesi ve her bir öğrencinin başarı gelişiminin ailelerine rapor edilmesi, okulların ve illerin her ders alanı ve sınıf düzeyi için başarılarının hesaplanması, okulların gelişim programlarını hazırlamaya yönelik dönüt sağlama, öğretim programının ve öğretme tekniklerinin geliştirilmesinde kullanılmaktadır (Delaware Educational Assessment Program: 1984 Statewide Test Results, Summary Reports, 1984).

Wisconsin Eyaleti, devlet ve özel okullara gönüllülük esasına göre uygulanan eyalet testi (Wisconsin Pupil Assessment Program) sonuçlarını bu eyaletin her bir ders ve sınıf düzeyindeki başarısını ülke genelinin başarısı ile karşılaştırmada ve öğrencilerin yıllara göre değişik alanlardaki başarısını takip etmede kullanmaktadır (Local Option Testing Manual-1981. Wisconsin Pupil Assessment Program, 1981).

Diğer bazı eyaletlerin aksine Indiana Eyalet testi, 3., 6., 8. ve 10. sınıflara zorunluluk ilkesine göre uygulanmaktadır. Bu testler öğrencilerin, okulların ve eyaletlerin karşılaştırılmasında kullanılmaktadır (Costello, 1999). Ayrıca, öğrenciler buradan ve diğer bazı standart

testlerden sağladıkları başarı düzeyine göre okul ve eyalet bazında karşılaştırılmaktadır. Testi, okullarda zorunlu olarak uygulayan eyaletlerden birisi de Utah Eyaleti'dir. Bu eyalette geliştirilen norm dayanaklı eyalet testi, değişik ders alanları için 5., 8. ve 11. sınıflarda zorunlu olarak uygulanmaktadır (Nelson, 1994). Bu eyalette de test sonuçları eyalet başarısını ülke geneliyle karşılaştırmak amacıyla kullanılmaktadır.

North Carolina Eyaleti'nde, eyalet testi eğitimsel gelişmeleri değerlendirme ve özel okulların başarı durumlarını saptama amaçlı da kullanılmaktadır (Wilson,1993). Öğrenci başarıları diğer eyaletlerin ortalamalarıyla karşılaştırılmaktadır. Ayrıca testler, öğrenci başarılarının yıllara göre dağılımın eğilimini saptamada da kullanılmaktadır (State of the state: Educational performance in North Carolina, 1992).

Georgia Eyaleti'nde, eyalet testleri benzer şekilde 4., 8. ve 11. sınıfların değişik alanlardaki başarılarını tespit için kullanılmaktadır (Hutcheson ve Schabacker, 1974). Ayrıca, artan sayıda eyalet, eyalet testlerini, okulları, öğrencileri ve eyaletleri karşılaştırmak amaçlı kullanma yanında, özellikle devlet okulu öğrencilerinin bu testlerden aldıkları puanları, bu öğrencilerin ortaöğretimden mezun olmaları için bir zorunluluk olarak da kullanılmaktadırlar (Natriello ve Pallas, 1999). Elli eyaletten 22' sinde lise diploması almak için eyalet testinden geçme zorunluluğu aranmaktadır.

Liseden mezun olabilmek için eyalet testini geçme zorunluluğu getiren eyaletlerden birisi de North Carolina Eyaleti'dir. Burada eyalet testi liseden mezun olma için bir ölçüt olarak kullanılmaktadır. Bu test ayrıca eğitimsel gelişmeleri değerlendirme ve özel okulların başarı durumlarını saptama amaçlı olarak da kullanılmaktadır (Wilson,1993).

Eyalet testlerini lise bitirme ölçütü olarak kullanan eyaletlerden birisi de Louisiana Eyaleti'dir. Burada uygulanan eyalet testi öğrenci, öğretmen, okul başarılarını izlemek ve eğitim sürecindeki aksaklıkları tespit etmek amacı yanında liseden mezun olmak için bir ölçüt olarak da kullanılmaktadır (Louisiana Department of Education, 1995-96).

Özetle, bazı farklılıklar göstermekle beraber hemen hemen tüm eyaletler standart başarı testlerini eğitimsel kararlar verme ve planlar geliştirme amacıyla kullanılmaktadırlar.

### *Türkiye'de Geniş Ölçekli Başarı Testlerinin Kullanımı*

Türkiye'de geniş ölçekli başarı testlerine yönelik çalışmalar 1990'lı yıllarda başlamış ve 1990'lı yılların sonlarına doğru ortaya somut örnekler çıkarmıştır. Bu alandaki ilk önemli çalışmalar, Milli Eğitim Bakanlığı'na bağlı bir birim olan ve 1992'de kurulan Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı (EARGED) tarafından gerçekleştirilmiştir. Bu birim Türkiye genelinde Türkçe, sosyal bilgiler, fen bilgisi ve matematik alanlarında 5., 6., 8., 9. ve 11. sınıf düzeylerinde durum tespit sınavları yapmaktadır (MEB, 1999). Bu testler, Türkiye genelini temsil edici bir örneklem üzerinde uygulanmaktadır. Örnekleme yedi coğrafi bölgenin her birinden seçilmiş illerdeki Müfredat Laboratuvar Okulları (MLO) ve MLO dışı okullar (MLD) oluşturmaktadır. Bu testlerin kullanım amacı EARGED tarafından şöyle ifade edilmiştir: "Türkiye genelinde farklı sınıf düzeylerinde ve farklı konu alanlarında öğrencilerin neler bildiğini, hangi zihinsel becerilere sahip olduğunu, eksikliklerinin neler olduğunu tespit etmek ve bu bilgileri bakanlığın ilgili birimlerine iletmek" (MEB, 1997:1).

Bu testler her bir konu alanı ve sınıf düzeyi için farklı beceri düzeylerinde (bilgi, kavrama, problem çözme vb.) hazırlanmış sorulardan oluşmaktadır. EARGED tarafından yapılmış olan bu testlere ilişkin konu alanları, uygulandıkları sınıf düzeyleri ve diğer bazı bilgiler şöyledir: Fen bilgisi alanında 1994, 1995, 1996, 1997, 1999 yıllarında 7 bölgedeki MLO okullarının 5., 6., 8., 9. ve 11. sınıflarına durum tespit sınavları uygulanmıştır (1994 ve 1999 yılı uygulamasında bazı bölgeler kapsamıştır ve 1999 yılı uygulamasında MLO dışı okullar da kapsama alınmıştır). Matematik alanında 1996 ve 1997 yıllarında yine 7 bölgedeki MLO okullarının 5., 8., ve 11. sınıflarına testler uygulanmıştır. 1999 yılında MLO dışı okullar da kapsama alınarak 5., 6., 8., 9. ve 11 sınıflara matematik seviye tespit sınavı yapılmıştır. Türkçe alanında 1997 yılında , yine tüm bölgelerdeki MLO okullarının 5. ve 8. sınıflarına ve 1998 ve 1999 yıllarında da tüm bölgelerdeki MLO ve bazı bölgelerdeki MLO dışı okulların 5., 6., 8. ve 9 sınıflarına testler uygulanmıştır. Sosyal bilgiler alanında ise 1997 yılında tüm bölgelerdeki MLO okullarının ve 1998 ve 1999 yıllarında MLO ve MLO dışı okulların 5., 6., 8., 9. ve 11. sınıflarına durum tespit testleri uygulanmıştır (MEB, 1999).



EARGED testleri çoktan-seçmeli ve açık-uçlu maddelerden oluşmaktadır. Testlerde yalnız bilgi düzeyi değil, daha üst düzey becerileri ölçen sorular da (kavrama, uygulama vb.) yer almaktadır. Testlerin geçerliğini ve güvenilirliğini sağlamada özen gösterilmektedir. Geçerlik ve güvenilirlik; test uygulamalarının standart olması, konu alanlarının yeterince temsil edilmesi, konu alanı uzmanları, branş öğretmenleri ve teknik elemanların (ölçme ve değerlendirme uzmanları, program geliştirme uzmanları vb.) çalışmaya katılması ile sağlanmaktadır. Testlerin geçerlik ve güvenilirliğini sağlamaya yönelik çalışmalarda gelişmiş ülkelerde uygulanan test koşulları ve standartlar sağlanmaya çalışılmaktadır. Testler asıl gruba uygulanmadan önce seçilen örneklem gruplar üzerinde bu testlerin pilot uygulamaları yapılmaktadır. Pilot uygulamaların ardından testlere ait güvenilirlik çalışmaları yapılmaktadır. Bu çalışmalarda 1997, 1998 ve 1999 yıllarında yapılan pilot uygulama sonucu Türkçe testlerine ait güvenilirlik için Alpha değerleri en düşük 0.66 ve en yüksek 0.91 olarak bulunmuştur. Matematik testleri için 1997 ve 1999'da yapılan pilot uygulamalarda Alpha değerleri en düşük 0.68 ve en yüksek 0.90 olarak gözlenmiştir. Fakat 5. sınıfların 1997'deki pilot uygulamalarındaki güvenilirliği 0.54' ün altında gözlenmiştir. Sosyal bilimler testlerinin 1999 yılı pilot uygulamasında bu testin değişik formları (form-1 ve form-2) için Alpha değerleri en düşük 0.66 ve en yüksek 0.82 olarak gözlenmiştir. Benzer şekilde 1999 yılı pilot uygulamasında fen bilimleri testlerinin değişik formlarına ait Alpha değerleri en düşük 0.61 ve en yüksek 0.89 olarak gözlenmiştir. Testlerin diğer yıllardaki pilot çalışmalarına ve nihai uygulamalarına ait güvenilirlik çalışmalarına ulaşılammıştır.

EARGED' in uyguladığı başarı testleri sırasında öğrenci anketleri de uygulanmaktadır. Bu anketlerin uygulanmasındaki amaç, öğrencilerin sosyo-ekonomik düzeylerini, ders çalışma alışkanlıklarını, farklı derslere karşı tutumlarını ortaya çıkarmak ve bunlarla test sonuçları arasındaki ilişkileri incelemektir. Ayrıca, bu değişkenler açısından öğrencilerin başarıları arasında fark olup olmadığı da araştırılmaktadır (MEB, 1998). Testler, her bir konu alanı ve sınıf düzeyi için farklı bölgelerin, illerin ve hatta okulların başarı bakımından karşılaştırılmasına olanak vermektedir. Böylesi bir karşılaştırma; bölgeler, iller ve okullar genelinde ihtiyaç tespiti yapmak, eksikleri belirlemek için de önemli bir

fırsat yaratmaktadır. EARGED tarafından da belirtildiği gibi bu bilgiler öğrenci merkezli öğretim faaliyetlerini geliştirmede, öğretim programları ve materyallerini planlama ve geliştirme çalışmalarında uygulanmalıdır. Yoksa okul, öğretmen ya da öğrencileri başarı sırasına koymak ya da seçmek amaçlı kullanılmamalıdır (MEB, 1998).

EARGED'in uyguladığı testler dışında, bazı il milli eğitim müdürlükleri kendi illerindeki ilk ve ortaöğretim okullarının gelişimini takip etmek ve eksiklerini tespit etmek için geniş ölçekli başarı testleri uygulamaya başlamıştır. Bu testler o ilin tümünde ya da ili temsil edici bir örneklem üzerinde uygulanmaktadır. Bu testlerin öğrenci, veli, okul ve MEB il müdürlüklerinin gerekli birimlerine bilgi sağlamak için kullanılması amaçlanmaktadır. Bu tür çalışmaları yürüten iller arasında Bolu, İzmir, vb. iller bulunmaktadır. Bu testlerde yer alan sorular farklı konu alanı ve sınıf düzeyi için branş öğretmenleri, zümre öğretmenleri, ölçme ve değerlendirme uzmanları ve program geliştirme uzmanlarından oluşan bir heyet tarafından geliştirilmektedir.

Son yıllarda EARGED, yukarıda belirtilen çalışmalarına ek olarak, geniş ölçekli başarı testlerinin kullanımını gerektiren bazı uluslararası çalışmalara da katılmıştır. Bu çalışmaların bir parçası olarak EARGED, 1998'de merkezi Hollanda'da bulunan Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'na (International Association for the Evaluation of Educational Achievement, IEA) üye olmuştur. EARGED bu kuruluşun yürüttüğü "Üçüncü Uluslararası Matematik ve Fen Bilgisi Araştırmasının Tekrarı Projesi"ne (Third International Mathematics and Science Study Repeat, TIMSS-R) katılmıştır. TIMSS-R 'ın amacı, üye 13 ülkenin 13 yaş grubu (8. sınıflar) öğrencilerinin matematik ve fen alanlarındaki başarılarını uluslararası düzeyde karşılaştırmaktır (MEB, 2000a).

Ayrıca EARGED, aynı kuruluşun Uluslararası Okuma Becerisini Geliştirme Projesi (PIRLS)'ne de katılmıştır. PIRLS ile üye ülkelerdeki 9 yaş grubu (4. sınıf) öğrencilerinin okuma becerileri ve bu ülkelerde okuma becerilerini geliştirmek için kullanılan öğretim metotları uluslararası bir boyutta ve karşılaştırmalı olarak değerlendirilmektedir. Bu değerlendirme sonucu üye ülkelere öğrencilerin okuma becerilerini geliştirmede

kullandıkları kaynakları ve yapacakları çalışmaları nerelere yönlendirmeleri gerektiğine dair bilgiler de verilmektedir (MEB, 2000b).

Bu çalışmalardan da görüldüğü gibi son yıllarda Türkiye, hızla geniş ölçekli testlerin kullanımına yönelmiştir. Başka bir deyişle eğitim sisteminde gerekli olan kalite kontrolünü yapmaya ve bu amaçla eksiklikleri gidermeye doğru bir çaba harcanmaktadır. Bu çalışmaların eğitim sistemimize gerekli katkıyı sağlayabilmesi için bu alanda gelişmiş ülkelerin bu konuyla ilgili yaptıkları çalışmalar incelenmeli, bu ülkelerde geniş ölçekli testlerin eğitim sistemlerine katkısının nasıl sağlandığı araştırılmalı ve uygulanabilir hususlar ülkemiz eğitimine uyarlanmalıdır.

Diğer önemli bir nokta ise bu çalışmaların sonuçları dikkatlice incelenip yorumlanmalı ve bunlara bağlı olarak sonuçlar uygulamaya geçirilmelidir. Eğer bu sonuçlar doğrultusunda eğitim sisteminizde gerekli reformlara gidilmezse yapılacak çalışmalar, verilecek emekler, boşa gitmiş olacaktır.

#### *Geniş Ölçekli Başarı Testlerinin Kullanım Alanları*

Geniş ölçekli başarı testlerinin sonuçları, eğitimin farklı öğelerine yönelik dönütler sağlayabilmektedir. Testler sonucunda belli bir alanda kazanılmamış ya da eksik kazanılmış olan davranışların nasıl daha iyi öğretilebileceği araştırılmalı; içerik, öğretim stratejileri ya da materyalleri açısından değerlendirilmelidir. Eğer öğrencilerin tümü belli davranışlarda düşük performans göstermiş ise bu hedeflerin düzeyleri ayarlanabilir ya da duruma göre o hedefle ilgili öğretim stratejileri, materyalleri, ders kitapları dahil olmak üzere, gözden geçirilebilir.

Bu bölümde diğer ülkelerde ve Türkiye’de yapılan çalışmaların ışığında, ülke genelini temsil gücü olan (ya da testlerden hangi gruba genelleme yapılacaksa o grubu temsil gücü olan) bir örnekleme uygulanan geniş ölçekli başarı testi sonuçlarından nasıl yararlanılabileceği üzerinde durulmuştur. Bu testlerden şu amaçlarla yararlanılabilir:

1) *Öğrenci:* Testteki maddeler öğrencilerin sahip olmaları gereken bilgileri ve becerileri ölçmeye yönelik olduğundan, öğrenci, düzeyine uygun hazırlanmış bir raporla kendisiyle ilgili sonuçlardan haberdar edilebilir. Öğrenciler için okullarda hazırlanacak öğrenci dosyalarında öğrenciye ait kişisel bilgiler, testlerin her bir alt dalından aldığı puanlar ve toplam puanlar

saklanabilir. Öğrencinin her bir ders alanı için zayıf olduğu ya da geliştirmesi gereken noktalar davranış boyutunda tespit edilebilir ve öğrenciye anlayabileceği şekilde açıklanabilir. Ayrıca, öğrenci başarılarının yıllara göre gelişimi ortaya çıkarılıp nasıl bir gelişme gösterdikleri de saptanabilir. Her bir ders için yıllara göre bir ilişkinin olup olmadığı saptanabilir. Böylesi bir yaklaşım her yıl test sonuçlarına göre alınan önlemlerin ya da yapılan reformların ne derecede başarıya ulaştığını takip etmede kullanılabilir.

Öğrenci başarıları ihtiyaca göre bölgeler, iller, hatta okullar bazında birbirleriyle karşılaştırmada, başarı farklılıklarını ortaya çıkarmada kullanılabilir. Değişik öğrenci gruplarının (örneğin, o il’e bağlı ilçelerdeki öğrenciler) birbiriyle karşılaştırılması, ihtiyaç alanlarının tespiti için gerekli görülmektedir. Buna bağlı olarak başarı düzeyleri düşük okullarda gerekli önlemler alınabilir.

Öğrenciler değerlendirilirken daha önceden belirlenmiş bir ölçüte dayalı olarak (ölçüt-dayanaklı) il, bölge, ya da Türkiye normlarına göre değerlendirilebilir. Bu, testin ne amaçla kullanılacağına göre değişebilir.

Öğrencilerin tümünü test etmek kuşkusuz ki pahalı bir olaydır. Bu nedenle genelleme yapılmak istenilen grubu (il, bölge, Türkiye vb.) temsil edici bir örneklem yeterli olacaktır.

2) *Öğretmen:* Öğretmenler test sonuçlarına dayalı olarak bilgilendirilebilir. Öğretmen test sonuçlarına dayalı olarak izlediği öğretim stratejilerinde ve kullandığı ölçme-değerlendirme yaklaşımlarında gerekli değişikliklere gidebilir. Amerika’da yapılan bir çalışmada (McMillan, Myran ve Workman, 1999) görüşme yapılan 722 öğretmenden %80’i eyalet genelinde yapılan geniş ölçekli başarı testlerinin sınıfta kullandıkları ölçme-değerlendirme yaklaşımlarında ve özellikle de öğretim stratejilerini belirlemede etkili olduğunu vurgulamışlardır. Öğretmenler sınav sonuçlarına dayalı olarak kullandıkları öğretim-öğrenme stratejilerinde ve ölçme ve değerlendirme alışkanlıklarında gerekli değişikliklere gitmektedirler.

Ayrıca, testler sonucunda elde edilen bilgiler, öğretmen yetiştirme sistemine yönelik eksiklikleri gidermek ya da reformları yapmak amaçlı da kullanılabilir. Test sonuçları, öğretmen eğitiminde dikkat edilmesi gereken bazı noktalar hakkında ipuçları verebilir. Tespit edilen aksaklıklar, örneğin öğrencilerin güçlük çektikleri



alanlar dikkate alınarak öğretmen yetiştirmede (eğitim fakültelerinin öğretmen yetiştirmeye yönelik programları) kullanılabilir. Bunlar belli bir alanda yetiştirilecek öğretmenlerin hangi öğretim stratejilerine, materyallerine ağırlık vermeleri gerektiği konularında olabilir.

Ayrıca test sonuçlarına göre gerekirse öğretmenler hizmetiçi eğitime alınabilir ya da böyle bir eğitimde nelere dikkat edilmesi gerektiğine karar verilebilir. Testlerin geniş ölçekli olması, Türkiye genelinde ortak sesliliği sağlar.

3) *Öğretim programı:* Testler, uygulanan öğretim programındaki gelişmeleri yorumlamak, programı geliştirmek, iyileştirmek ve değerlendirmek için kullanılabilir. Farklı iller ya da okulların karşılaştırılabilmesine olanak sağlamak için okullarda uygulanan öğretim programlarının ana hatlarıyla (hedef ve davranış boyutlarıyla) aynı olması gerekir. Hangi ders kitabı, öğretim yaklaşımı, materyal vb. kullanılacağı, o öğrenci grubunun özelliğine, özel ihtiyaçlarına ve öğretmene göre değişebilir. Kısacası her program, o okulun özel ihtiyaçlarını karşılayacak esnekliğe de sahip olmalıdır. Her sınıf ve konu düzeyinde öğrencilerin neler bilmesi, hangi bilgi ve becerilere sahip olması gerektiği, programlarda belirtilmelidir.

Test sonuçları, uygulanan bu programlara yönelik kararlar almada kullanılabilir. Programların aksayan yönleri, hangi hedeflerin gerçekleştirilip gerçekleştirilemediği ve bunların olası nedenlerine yönelik ipuçları sağlanabilir. Örneğin hedef-davranış boyutunda mı, içerik boyutunda mı, öğretim-etkinlikleri boyutunda mı aksaklıklar var? Bu öğeler gözden geçirilebilir. Eğer öğrencilerin tümü belli davranışlarda düşük performans göstermişse bu hedeflerin düzeylerinin yeniden gözden geçirilmesi ya da söz konusu duruma göre o hedefle ilgili öğretim stratejileri, materyalleri ya da ders kitaplarının gözden geçirilmesi gerekebilir.

Program geliştirme uzmanları ya da programlardan sorumlu kişiler, gerekli düzenlemeleri ya da değişiklikleri yapmak amacıyla test sonuçlarından haberdar edilmelidir.

4) *Öğretim stratejileri ve materyalleri:* Yukarıda da değinildiği gibi, testler, öğretim yöntem ve stratejilerinin, materyallerini geliştirmek, iyileştirmek için kullanılabilir. Test sonuçlarından, hangi hedeflerin kazanılmadığını ve bunları kazandırmak için öğretim stratejilerinin ve materyallerinin etkililiğine ilişkin ne tür önlemler

alınacağı kararlaştırılabilir. Yenilikler kullanılacak strateji seçiminden ders kitabı ve derste kullanılacak diğer öğretim materyallerinin seçimine kadar değişik boyutlarda olabilir.

5) *Veliler:* Test sonuçları, velileri, çocuklarının akademik başarılarından, gelişimlerinden haberdar etmek için kullanılabilir. Veliler, düzeyine uygun dille hazırlanmış bir raporla öğrencilerinin her bir ders alanındaki başarısından ve gelişiminden haberdar edilebilir. Böylesi bir yaklaşım, velinin çocuğunun eğitiminde daha aktif rol oynamasını sağlayabilir.

### Sonuç

Yukarıda da belirtildiği gibi, geniş ölçekli başarı testleri, eğitim-öğretim sürecinin farklı boyutlarında var olan aksaklıkların belirlenmesinde ve giderilmesinde etkili olarak kullanılabilir. Fakat bu, büyük ölçüde, test sonuçlarının doğru yorumlanmasına ve kullanılmasına bağlıdır. Bunu sağlamak için test sonuçlarının farklı boyutlardan bakılarak analiz edilmesi ve etkili kullanılması gerekmektedir. Test sonuçlarının analizinde farklı tekniklerin ve yaklaşımların kullanılması, elde bulunan verilerden yüksek oranda yararlanılmasını ve elde edilen sonuçların doğruluk derecesini artırmaktadır.

Eğer geniş ölçekli başarı testleri öğrenci, öğretmen ve okullar hakkında önemli kararlar vermek amacıyla kullanılacaksa, özel bir itinaya gösterilmesi gerekmektedir. Klein, Hamilton, McCaffrey ve Stecher (2000), geniş ölçekli başarı testlerinin öğrenci, öğretmen ve okul hakkında önemli kararlar vermede, öğrenci başarısını gösteren tek ölçüt olarak kullanılmasının zorluğu üzerinde durmuştur. Elbette öğrenci, öğretmen ve okullara yönelik kararlar vermede yalnızca geniş ölçekli test sonuçlarının kullanılması güçlük teşkil etmektedir. Fakat diğer bilgilerle de birleştirildiğinde, dikkatli ve doğru kullanıldığında önemli bir bilgi kaynağı olduğu da bir gerçektir. Unutulmamalıdır ki test sonuçları doğru yorumlandıkları ölçüde faydalıdır.

Üzerinde durulması gereken diğer önemli nokta ise, sürekli yeni veri toplamaktan ziyade, toplanmış olan verilerin etkili şekilde ve farklı boyutlarda analiz edilip kullanılmasıdır. Böylesi bir yaklaşım daha etkili ve daha ekonomik olacaktır. Bir o kadar önemli nokta ise, elde edilen sonuçların uygulamaya geçirilmesidir. Testlerin sonuçları yeni uygulamalara yansıtılmadığı takdirde, geniş ölçekli testlerin uygulanması zaman, para ve insan gücü israfından öteye geçmeyecektir.

## Kaynakça

- Ascher, G. (1975). *Some nonparametric approaches to the use of criterion-referenced statewide test results in the evaluation of local district educational programs*. (ERIC Document Reproduction Service No. ED106369)
- Barton, P. E. & Coley, R. J. (1994). *Testing in America's schools: Policy information report*. Princeton, N. J. : Educational Testing Service, Policy Information Center. (ERIC Document Reproduction Service No. ED366 616)
- Barry, A. L. , Nielsen, D. C., Glasnapp, D. R., Poggio, J. P. & Sundbye, N. (1997). Large scale performance assessment in writing: Effects of student and teacher choice variables. *Contemporary Education*, 69 (1), 20-26.
- Costello, R. W. (1999). *Using standardized testing results to establish standarts for the improvement of student academic performance in Indiana*. (ERIC Document Reproduction Service No. ED431798)
- Delaware Educational Assessment Program. (1984). *1984 statewide test results, summary report*. (ERIC Document Reproduction Service No. ED306261)
- Dunn, S. S. (1989). *Public accountability in Australian education: A discussing paper*. Occasional Paper No. 11. (ERIC Document Reproduction Service No. ED321354)
- Hutcheson, L. D. & Schabacker, W. H. (1974). *Utilization guide for Georgia Statewide Testing Program test scores*. (ERIC Document Reproduction Service No. ED109166)
- Klein, S. P., Hamilton, L. S., McCaffrey, D. F. & Stecher, B. M. (2000). *What do test scores in Texas tell us? Issue paper*. (ERIC Document Reproduction Service No. ED447219)
- Local Option Testing Manual (1981). *Wisconsin Pupil Assessment Program*. (ERIC Document Reproduction Service No. ED207988)
- Louisiana Department of Education, Office of Research and Development, Bureau of Pupil Accountability. (1995-96). *Louisiana Educational Assessment Program, 1995-96 Annual Report Bulletin 1867*.
- Madaus, G. F. (1994). A technological and historical consideration of equity issue: Associated with proposals to change the nation's testing policy. *Harvard Educational Review*, 64 (1), 76-91.
- McMillan, J. H., Myran, S. & Workman, D. (1999). *The impact of mandated statewide testing on teachers' classroom assessment and instruction practices*. (ERIC Document Reproduction Service No. ED431041)
- MEB/EARGED. (1995). *Bülten (Özel Sayı)*. Ankara: Milli Eğitim Basımevi.
- MEB/EARGED. (2000a). *Bülten (7)*. Ankara: Milli Eğitim Basımevi.
- MEB/EARGED. (2000b). *Bülten (9)*. Ankara : Milli Eğitim Basımevi.
- MEB/EARGED. (1997). *EARGED, Durum tespit raporu: Fen bilgisi ve matematik*. Ankara.
- MEB/EARGED. (1998). *EARGED, Durum tespit raporu: Türkçe ve sosyal bilimler*. Ankara.
- MEB/EARGED. (1999). *EARGED, Nihai test uygulamaları dökümü*. Ankara.
- Natriello, G. & Pallas, A. M. (1999). *The development and impact of high stakes testing*. (ERIC Document Reproduction Service No. ED443871)
- Nelson, D. E. (1994). *Accountability reports for all districts and schools. The Utah Statewide Testing Program—1993. Appendix to the Annual Reports of the State Superintendent of public Instruction*. (ERIC Document Reproduction Service No. ED 373077)
- North Carolina Assessment Program (1993). *State of the state: Educational performance in North Carolina, 1992*. (ERIC Document Reproduction Service No. ED357461)
- O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49 (8), 14-19.
- Roeben, E. D. (1997). The technical and practical challenges in developing innovative assessment approaches for use in statewide assessment programs. *Contemporary Education*, 69 (1), 6- 10.
- Strong, S. & Sexton, L. C. (1996). Kentucky performance assessment of reading: Valid? *Contemporary Education*, 67 (2), 102- 106.
- Townsend, T. (1996). *School effectiveness and restructuring schools: What does the research tell us?* (ERIC Document Reproduction Service No. ED398625)
- Wilson, H. E. (1993). Ethics in testing. *School Management Advisor*, 21.
- Yen, W. M.& Ferrara, S. (1997). The Maryland School Performance Assessment Program: Performance sssssment with psychometric quality suitable for high stakes usage. *Educational and Psychological Measurement*, 57, 60-84.

Geliş	11 Eylül 2001
İnceleme	20 Eylül 2001
Kabul	1 Nisan 2002