International Online Journal of Educational Sciences, 2014, 6 (3), 790-804

iojes
www.iojes.net

International Online Journal of Educational Sciences

ISSN: 1309-2707

# Measurement and Evaluation Methods for Academic Progress: An Example from the U.S.A.

**Mehmet Ali Öztürk**[1]

[1]*Bahçeşehir University, Istanbul, Turkey*

ARTICLE INFO

*Article History*:
Received 25.05.2014
Received in revised form 15.06.2014
Accepted 30.06.2014
Available online 02.12.2014

ABSTRACT

This article introduces the concept of Adequate Yearly Progress (AYP), which has been used as the measure of academic progress within the context of the No Child Left Behind (NCLB) Act in the United States. After providing brief information about administering tests and reporting test results within the NCLB framework, the methods of measurement and evaluation used to determine AYP were explained. Then, methods used to develop content and achievement standards for exams were covered. The ways in which the National Assessment of Educational Progress (NAEP) has been used as a means of quality control for these standards were also presented. Finally, lessons to be learned from AYP experience were discussed.

© 2014 IOJES. All rights reserved

Keywords:
Measurement of academic progress, educational reform, adequate yearly progress, no child left behind, AYP

## Extended Abstract

This article is an introduction to Adequate Yearly Progress (AYP), which has been used as the measure of academic progress within the framework of the U.S. No Child Left Behind (NCLB) Act. NCLB can be seen as an effort to establish a quality control system for core courses in elementary and secondary education. After introducing AYP, lessons are extracted from the AYP experience.

AYP is related to the following goal in NCLB: "All students will achieve high academic standards, by attaining proficiency or better, in reading and mathematics by the 2013-2014 school year" (Yell & Drasgow, 2009, p. 10). To determine progress, states have been required to annually administer a reading and a mathematics test to all students in grades 3 through 8, starting with the 2005-2006 school year. They have also been required to test every student in each of these subject areas at least once during 10$^{th}$ through 12$^{th}$ grades (U.S. Department of Education, n.d.b). NCLB has defined eight student subgroups and required that test results be reported separately for each subgroup. It has also required that all students in all subgroups reach proficiency by 2014.

Three models have been used to determine AYP: 1) Based on the achievement levels reached by student subgroups at a single point in time 2) Based on the academic progress from one year to the next of the same student subgroup at the same grade level 3) Based on growth models. In the context of AYP, growth models can be used to create the achievement projection of individual students, taking into account each student's achievement level at the starting point and achievement rate in the past. These projections can then be used

[1] Corresponding author's address: Bahçeşehir University, Faculty of Educational Sciences, Çırağan Cad. No: 4 Beşiktaş Istanbul Turkey
Telephone: + 90 212 381 0303
Fax: + 90 212 381 0300
e-mail: mehmet.ozturk@bahcesehir.edu.tr
DOI: http://dx.doi.org/10.15345/iojes.2014.03.021

© 2014International Online Journal of Educational Sciences (IOJES) is a publication of Educational Researches and Publications Association (ERPA)

to determine whether a student exhibits the achievement level expected of him/her at different points in time. Students can be tracked individually in growth models.

The most important reason why states wanted to implement growth models for AYP purposes was that schools which would be considered failing under the first two models would be counted successful if a growth model were used to determine AYP. Another reason for the implementation of growth models has been the desire to prevent the usage of resources (money, teachers' time, etc.) disproportionately for the students who are just below the proficiency level. Preventing teachers' efforts to transfer from low-achieving schools to high-achieving ones to avoid sanctions for not realizing AYP has been perceived as another benefit of implementing growth models (Welner, 2008). Finally, measuring every student's academic progress from one year to the next on a scale common to these years has been deemed to be a healthier approach than defining and measuring proficiency at every grade level independently of the definition and measurement of proficiency at other grade levels (Welner, 2008).

Another topic covered in this article is how content and achievement standards for tests are developed. A major criticism toward NCLB has been that the requirement of 100 % proficiency might force states to develop low academic standards, since those which develop high standards would run the risk of facing sanctions in case they could not reach these standards (Darling-Hammond, 2007).Standards for tests are developed in three stages. First is to develop content standards; second is to develop or purchase tests that can measure these standards; and third is to decide upon the cutoff scores that would mark academic achievement / proficiency levels (Rotherham, 2006). Two common methods at the third stage that can be used with the first two AYP models are the modified Angoff and the bookmarking methods (Rotherham, 2006). It is important to know that these methods are applied on a particular test and that the cutoff values determined for performance levels are meaningful for that particular test. Therefore, a comparison between the cutoff values for two different tests will be meaningless. It is obvious that the difficulty levels of questions on different tests will be different, and this difference will lead to different cutoff values for these tests (Rotherham, 2006).

The most common method used with growth models, however, has been vertically-scaled assessments that allow to measure performance at different grade levels on the same scale. A commonly used method for vertically-scaled assessments in growth models is a different version of the bookmarking method, which is based on Item Response Theory (IRT). Developing a common scale for different grade levels in such vertical scaling is possible through one of the following three methods. First is to include common items in the tests given to students at different grade levels. Second is to ask students at the same grade level questions at the difficulty level of their own grade level as well as questions at the difficulty level of a lower and / or higher grade level. Third is to give a common scaling test to students at different grade levels that allows linking these grade levels in addition to tests at students' own grade levels (Patz, 2007).

The first limitation of vertical scaling is the assumption of unidimensionality of the measured knowledge / skill. This assumption brings about the limitation of using vertical scaling only on types of knowledge / skill that is on the same continuum from one grade level to the next. The second limitation is the fact that the linking between grade levels weakens statistically as grade levels go further apart and the resulting comparisons between grade levels become meaningless (McCall, 2007; Patz, 2007). In sum, use of the growth models to determine AYP as well as vertical scaling in growth models to measure progress are sophisticated applications.

# Akademik İlerlemeyi Ölçme ve Değerlendirme Yöntemleri: Amerika Birleşik Devletleri'nden Bir Örnek

**Mehmet Ali Öztürk**[1]

[1]*Bahçeşehir Üniversitesi, İstanbul, Türkiye*

MAKALEBİLGİ

*Makale Tarihçesi*:
Alındı25.05.2014
Düzeltilmiş hali alındı15.06.2014
Kabul edildi30.06.2014
Çevrimiçi yayınlandı02.12.2014

ÖZET

Bu çalışmada, Amerika Birleşik Devletleri'nin No Child Left Behind (NCLB – Hiçbir Çocuk Geride Kalmasın) kanununun uygulanmasında merkezi bir rol üstlenen Adequate Yearly Progress (AYP – Yeterli Yıllık İlerleme) kavramı açıklanmıştır. Akademik ilerlemenin ölçütü olarak kullanılan AYP, Türkiye'deki eğitimcilerin sağlam bir fikir edinebilecekleri şekilde ve güvenilir kaynaklara dayanılarak sunulmuştur. NCLB çerçevesinde öğrencilere sınav uygulanması ve sınav sonuçlarının raporlanması ile ilgili özet bilgiler verilmiş ve AYP kararlarında kullanılan ölçme ve değerlendirme yöntemleri açıklanmıştır. Sınavların içerik ve başarı düzeyi standartlarının belirlenme yöntemleri tartışılmış ve ülke çapında uygulanan National Assessment of Educational Progress (NAEP – Eğitimde İlerlemenin Ulusal Değerlendirmesi) sınavlarının bu standartların kalite kontrolleri için nasıl kullanıldığı üzerinde durulmuştur. Son olarak, AYP deneyiminden Türkiye için çıkarılabilecek dersler sunulmuştur.

© 2014 IOJES. Tüm hakları saklıdır

Anahtar Kelimeler:
Akademik ilerleme ölçümü, eğitim reformu, yeterli yıllık ilerleme, hiçbir çocuk geride kalmasın,AYP

## Giriş

Günümüz şartları, gelişmiş ülkeler de dahil olmak üzere, birçok ülkeyi, eğitim sistemlerinde reform arayışı içine itmiştir. Böyle bir arayışa zorlayan nedenler arasında, öğrenciye kazandırılması gereken bilginin çok hızlı artması ve değişmesi; ülkelerin, bir köy haline gelen dünya için vatandaşlarını sosyal, kültürel ve mesleki açıdan başarılı bir şekilde yetiştirme isteği; her geçen gün daha fazla yatırım isteyen eğitim alanında, etkin bir kalite kontrol sisteminin yerleştirilme zorunluluğu ve eğitimde eşitliğin geliştirilmesi arzusu sayılabilir.

Türkiye'de de eğitimde reform tartışmaları uzun süredir devam etmektedir. Böyle bir arayış sürecinde, her ülkenin, kendine özgü şartları gözönünde bulundurması doğal olmakla birlikte, benzer problemler için başka ülkelerin uygulamalarını incelemesi ve onların deneyimlerinden de yararlanması sağlıklı bir yaklaşım olacaktır. Bu reform deneyimlerinden bir tanesi, Amerika Birleşik Devletleri'nin No Child Left Behind (Hiçbir Çocuk Geride Kalmasın) Kanunu ve takip eden uygulamalardır. Aralık 2001 tarihli No Child Left Behind (NCLB), 1000 sayfadan fazla hacmiyle, ABD tarihinin, ülke ilk ve orta öğretimiyle ilgili günümüze kadarki en kapsamlı federal kanunudur (U.S. Department of Education, n.d.a; Yell & Drasgow, 2009). Yürürlüğe girdiği günden bu yana da, Amerikan eğitim sisteminde çok sayıda düzenleme ve uygulamalara kaynaklık etmiştir.

NCLB, kısaca, ABD ilk ve orta öğretiminin temel derslerinde kalite standartları belirleme, belli bir takvime bağlı olarak bir kalite kontrol sistemi yerleştirme, sonrasında da bu sistemin takipçisi olma niyet ve gayretidir. Kanunun geniş kapsamına paralel olarak, bu kanun etrafındaki uygulamalar için ciddi bir bütçe ayrılmış, başarısızlığa da önemli yaptırımlar getirilmiştir. NCLB, getirdiği yaptırımlar ve öğretmen kalitesine yönelik düzenlemeleri gibi birçok değişik açıdan irdelenebilir. Bu çalışmada, NCLB'nin ölçme ve

[1] Sorumlu yazarın adresi: Bahçeşehir Üniversitesi, Eğitim Bilimleri Fakültesi, Çırağan Cad. No: 4 Beşiktaş İstanbul Türkiye
Telefon: + 90 212 381 0303
Faks: + 90 212 381 0300
e-posta: mehmet.ozturk@bahcesehir.edu.tr
DOI: http://dx.doi.org/10.15345/iojes.2014.03.021

değerlendirmeyle ilgili merkezi bir kavramı olan Adequate Yearly Progress (Yeterli Yıllık İlerleme) ele alınacaktır.

Bu çalışmanın iki amacı bulunmaktadır. Birincisi, Türkiye'deki eğitimci ve akademisyenlere, NCLB çerçevesinde akademik ilerlemenin ölçütü olarak kullanılan Adequate Yearly Progress (AYP) kavramını ve ilgili uygulamaları bir makale hacminin elverdiği kadar tanıtmaktır. İkincisi ise, eğitim sisteminde önemli reformlara giden Türkiye için AYP örneğinden alınabilecek dersleri çıkarmaktır. AYP, zamanın yalnızca bir noktasındaki akademik başarının ölçülmesiyle değil, zaman içinde akademik ilerlemenin ölçülmesiyle ilgili bir kavramdır. Akademik ilerlemenin başarılı bir şekilde ölçülmesi hedefi ise, yalnızca Amerikan eğitim sistemine özgü değildir. Türkiye'de yaygın olarak veya yalnızca akademik çevrelerde tartışılan birçok başlık da, bu hedefle doğrudan ilgilidir. Bu başlıklardan en önemlileri arasında, eğitim sisteminin nabzını tutmak amacıyla öğrencilere her yıl temel derslerde verilmesi düşünülen sınavlar, performans ölçme amacıyla öğretmenlerin her yıl eğitiminden sorumlu oldukları öğrencilere ne kadar akademik ilerleme kaydettirdiklerinin belirlenmesi sayılabilir. Akademik ilerlemenin ölçülmesinin temel teşkil edeceği bu başlıklar etrafında, olası birçok yeni uygulama düşünülebilir. Öğretmenler için bir hesap verebilirlik sisteminin kurulması, öğrencilerine gereken ilerlemeyi kaydettiremeyen okulların belirlenip gereken maddi desteğin, personel desteğinin sağlanması ve öğrencilerin akademik ilerlemelerinin demografik özelliklere göre farklılık gösterip göstermediğinin belirlenip gereken tedbirlerin alınması, bu uygulamalara yalnızca birkaç örnektir. Bu olası uygulamaların, başka yönleriyle yerinde uygulamalar olup olmayacağından bağımsız olarak ifade edilebilecek bir husus, akademik ilerlemenin sağlıklı bir şekilde ölçülmesinin, bu uygulamalar için hayati öneme sahip olmasıdır. Özetle, AYP, uzak bir ülkenin Türkiye'yi ilgilendirmeyen bir kavramı değil; aksine, günümüz Türk eğitim sisteminde uygulanması düşünülen birçok yeniliğe doğrudan ışık tutacak bir kavramdır.

Bu çalışmanın sınırlarını belirlemek de yararlı olacaktır. Birincisi, çalışma, NCLB ve AYP'nin Türkiye'yi ilgilendiren yönlerini ele almakta; federal hükümet ile eyaletler arası ilişkiler gibi ilgisiz yönlerinden ancak gerektiği kadar söz etmektedir. İkincisi, çalışma, yalnızca ABD'deki bir uygulama örneğini vererek, Türkiye'de herhangi bir eğitim reformu için en iyi örneğin ABD olduğu mesajını hedeflememektedir. Üçüncüsü, Türkiye'deki akademisyenler, Türk eğitim sistemi ve ölçme değerlendirme uygulamaları etrafında değerli birçok düşünce ve teklif dile getirmişlerdir. Ancak, bu düşünce ve tekliflerin, AYP uygulamaları çerçevesinde ele alınıp değerlendirilmesi, bu çalışmanın sınırları dışındadır.

**NCLB Çerçevesinde Uygulanan Sınavlar ve AYP'nin Merkezi Konumu**

AYP, NCLB kanununun ilk halinde genel ifadelerle sözedilen hedeflerden biri olan şu hedefle ilgilidir: 2013–2014 akademik yılına kadar tüm öğrenciler, okuma ve matematik alanlarında yüksek akademik standartlara [yeterlilik düzeyinde veya daha yüksek (proficiencyor better)] ulaşmış olacaklar (Yell & Drasgow, 2009). Bu standartlara ulaşma yönünde ilerleme kaydedilip kaydedilmediğini belirlemek için, eyaletlere, 2005–2006 okul yılından başlayarak, 3. sınıftan 8. sınıfa kadar tüm öğrencilere okuma ve matematik alanlarında her yıl bir sınav uygulama zorunluluğu getirilmiştir. Buna ek olarak kanun, her öğrenciye 10. sınıfla 12. sınıf arasında aynı iki alanda en az bir kere sınav uygulanmasını zorunlu kılmıştır (U.S. Department of Education, n.d.b).

NCLB, öğrencileri sekiz kategoride gruplamış; sınav sonuçlarının bu kategorilerin her biri için ayrı ayrı rapor edilmesini ve her bir kategorideki öğrencilerin yüzde yüzünün 2013–2014 okul yılı sonuna kadar yeterlilik düzeyine ulaştırılmasını zorunlu kılmıştır. Bu sekiz kategori şunlardır: 1) Amerikan yerlisi (American Indian / Native Alaskan), 2) Asyalı (Asian / Pacific Islander), 3) İspanyol kökenli (Hispanic), 4) Afrika kökenli Amerikalı (African American), 5) beyaz ırka mensup (White), 6) İngilizcede sınırlı yeterliliğe sahip (limited English proficient), 7) özel eğitim kategorisinde (special education), 8) düşük gelirli aileden geldiği için okulda verilen öğle yemeğini bedava veya indirimli yeme hakkını kazanan (free/reduced price lunch) (U.S. Department of Education, 2003). Eğer bir okulda herhangi bir kategoride yeterli sayıda (genelde 20'den 50'ye kadar değişen sayılar) öğrenci varsa, okul, o kategoriye ait sınav sonuçlarını, öğrencilerinin tümü için hazırladığı rapordan ayrı olarak rapor etmek zorundadır (Erpenbach, Forte-Fast, & Potts, 2003). Bunun yanında, sınavlara yeterli sayıda öğrencisi olan her kategorinin en az yüzde 95'inin katılımı sağlanmak zorundadır (U.S. Department of Education, n.d.b).

NCLB, eğitim sisteminde elde etmek istediği sonuçlara ulaşmak için yaklaşık 12 yıllık (2002-2014) bir takvim belirlemiştir. Bununla birlikte, 12 yıllık süreç boyunca, hedefte belirlenen çıktılara yönelik düzenlemelerden sorumlu birimlerden (eyaletlerden, okullardan, vb.) hedefe ne kadar yaklaşıldığının ve bu ilerlemenin tatmin edici olup olmadığının yıllık raporunu istemiştir. İşte tam da bu noktada, bu çalışmanın merkezini tutan AYP kavramı işin içine girmektedir. İlerlemenin kıstası olarak neyin ölçüleceği, ilerlemenin tatmin edici olup olmadığının ve dolayısıyla yaptırım uygulanıp uygulanmayacağının nasıl belirleneceği sorusu bu bağlamda önem kazanmaktadır. Her ne kadar bu çalışmada, çok yer tutacağından dolayı yaptırımlar konusuna girilmese de, eyaletler tarafından belirlenen ilerleme hedeflerine bir veya birkaç yıl içinde ulaşılmadığı durumlarda, okullara ve okul bölgelerine (school district) her başarısız yıl sonunda giderek artan ciddi yaptırımlar getirildiğini söylemek, AYP kavramının NCLB'deki merkezi yerini ifade etmek adına önemlidir (Yell & Drasgow, 2009).

**AYP'yi Belirlemek İçin Kullanılan Modeller**

**Zamanın belli bir noktasında öğrenci kategorilerinde ulaşılan başarı durumuna göre AYP'yi belirleme (grup durum modeli).** AYP'nin kanıtlanmasının en basit yöntemi, her bir kategoride yeterliliğe ulaşmış öğrenci yüzdesinin her yıl artarak son yıl olan 2014'te yüzde 100'e ulaşmasıdır. Örneğin, başlangıç noktası (baseline) olan 2001–2002 akademik yılı sonunda bir okulun öğrencilerinin yüzde 30'unun yeterliliğe ulaştığı varsayılacak olursa, 2014 yılına kadarki 12 yılda bu değerin yüzde 70 oranında artması gerektiği görülecektir. Bu yüzde 70, oniki yıla eşit bölündüğünde ise, her yıla yaklaşık yüzde 6'lık bir artış düşmektedir. Dolayısıyla, AYP'yi kanıtlamak için, okul öğrencilerinin yüzde 6'lık artışlarla 2003 yılında yüzde 36 oranında, 2004 yılında yüzde 42 oranında yeterliliğe ulaşması gerekecektir. Başlangıç noktasında yeterliliğe ulaşmış öğrenci oranları ve yüzde 100'e ulaşmak için gereken kazanımların alacağı zaman ve gayret, matematik ve okuma için ve çeşitli sınıf düzeyleri için farklı olabileceğinden, eyaletler, bu oranları bu iki ders için ve çeşitli sınıf düzeyleri için farklı olarak belirleyebilirler. Ancak, yıllar için belirlenen bu yüzdelik değer hedeflerinin, tüm öğrenci kategorileri için aynı olma zorunluluğu vardır. AYP, zamanın belli bir noktasında öğrenci kategorilerinde ulaşılan başarı durumuna göre belirlendiğinden dolayı bu modele, grup durum modeli (group status model) denilmektedir (Riddle, 2008).

**Birbirini takip eden iki yıl arasında aynı sınıf düzeyindeki aynı öğrenci kategorisinin akademik ilerlemesine göre AYP'yi belirleme (ardışık grupların ilerlemesine dayalı model).** NCLB, tüm öğrenci kategorilerinin hedeflenen yüzdelik değerlere ulaşamadığı durumlarda, AYP'yi gerçekleştirmiş sayılmanın istisnai bir yöntemini sunmuştur. Yukarıda açıklanan grup durum modelinin şartlarına esneklik getiren bu yönteme göre yerine getirilmesi gereken şartlar şunlardır: 1) Okul öğrencilerinin tümü tek bir grup olarak ele alındığında, bu grubun AYP'yi gerçekleştirmiş olması 2) AYP'yi gerçekleştirememiş öğrenci kategorilerinin her birinde, yeterliliğe ulaşmış öğrenci yüzdesinin, bir önceki yıla göre en az yüzde 10 oranında artmış olması 3) AYP'yi gerçekleştirememiş öğrenci kategorilerinin her birinin, eyaletin belirlediği en az bir akademik kıstas açısından (okula devam oranı gibi) daha, yine eyalet tarafından önceden tanımlanmış olan bir ilerleme kaydetmiş olması. Bu alternatif şartları yerine getirmeye NCLB literatüründe güvenli liman (safe harbor) adı verilmiştir (U.S. Department of Education, 2005). İkinci şarta şöyle bir örnek verilebilir: Bir okuldaki İspanyol kökenli dördüncü sınıf öğrencileri, okuldaki tüm öğrenci grubu ve her bir öğrenci kategorisi için 2008 yılı hedefi olarak belirlenen yüzde 60 yeterlilik düzeyine ulaşamamış; yüzde 55 yeterlilik düzeyinde kalmışlardır. Bununla beraber, 2007 yılında dördüncü sınıfta olan İspanyol kökenli öğrencilerin yüzde 25'i yeterliliğe ulaşmışken, 2008 yılında dördüncü sınıfta olan İspanyol kökenli yeterliliğe ulaşmış öğrenciler, yüzde 10'dan fazla bir oranda (yüzde 30) artarak yüzde 55'e çıkmıştır. Böyle bir durumda okul, güvenli liman şartlarının ikincisini yerine getirmiş sayılmaktadır. Birbirini takip eden iki yıl arasında aynı sınıf düzeyindeki aynı öğrenci kategorisinin akademik ilerlemesi kıstas olarak kullanıldığından dolayı bu model, ardışık grupların ilerlemesine (successive group improvement) dayalı model olarak adlandırılmaktadır (Riddle, 2008). Grup durum modeli ile bu model arasındaki fark, bu modelde ilerlemeye veya değişime bakılmasıdır.

**Büyüme Modeli Kullanarak AYP'yi Belirleme**

NCLB'nin orijinal halinde, AYP'nin belirlenmesi için yukarıda açıklanan iki modelin kullanılabileceği ifade edilmiştir. Kanunun ilk halinde bulunmadığı halde sonradan kabul edilen büyüme modelinin (growth model) ölçme ile ilgili teknik yönlerini açıklayabilmek için AYP bağlamından bağımsız olarak bir giriş

yapmak doğru olacaktır. Büyüme modelleri, eğitim alanında en yaygın kullanıldıkları şekliyle, öğrencilerin zaman içindeki değişimlerinin yanında, bu değişim sürecindeki bireysel farklılıklarının da incelenebildiği istatistiksel modellerdir. Zaman içindeki değişim, modelin ayrılmaz bir parçası olduğu için bu modeller, ancak uzunlamasına (longitudinal) veri üzerinde uygulanabilir. Uzunlamasına veri şartı da, zamanın en azından iki ayrı noktasında aynı değişken(ler) hakkında aynı öğrencilerden veri toplanmasını gerektirmektedir. Bu istatistiksel modeller, herbir öğrenciden gelen veride, hem zamanla ilintili (time-variant) değişkenlerin (örneğin, yaş, sınıf düzeyi) hem de zamanla ilintili olmayan (time-invariant) değişkenlerin (örneğin, cinsiyet, sosyoekonomik durum) etkisinin olduğu kabul edilerek kurulmaktadır. Örneğin, büyüme modeli kullanılarak, bir grup öğrencinin başarı düzeyinin bir akademik yıldan diğerine değişip değişmediği ve değişim varsa, bunun cinsiyete göre farklılaşıp farklılaşmadığı araştırılabilir. Bu çalışmada açıklanamayacak kadar geniş ve karmaşık bir istatistiksel yöntem olan büyüme modelleri ile ilgili birçok kaynak bulmak mümkündür (örneğin, Singer & Willett, 2003).

AYP bağlamında ise, büyüme modelleri, her öğrencinin başlangıç noktasındaki bireysel başarı düzeyi ve geçmişteki başarı oranı hesaba katılarak bireysel başarı projeksiyonunun yapılması ve bu projeksiyona göre zamanın farklı noktalarında kendisinden beklenilen başarıyı gösterip göstermediğinin belirlenmesi için kullanılabilir. Büyüme modellerinde öğrenciler, bireysel olarak takip edilebilir. Ancak, NCLB bağlamında başarısızlığın tanımlandığı en düşük birim öğrenci kategorileri olduğundan dolayı, öğrencilerin başarı durumları, kanunda adı geçen her bir öğrenci kategorisi için toplu olarak (aggregate) rapor edilmektedir. Böylesi bir model, bireysel veya grup büyüme modeli (individual/cohort growth model) olarak adlandırılmaktadır (Riddle, 2008).

NCLB'nin orijinal halinde söz edilmeyen bu model, daha sonra federal düzeydeki eğitim bakanlığı (U.S. Department of Education) tarafından Kasım 2005'te en fazla on eyaleti içine alacak bir pilot çalışma başlatılarak kullanılmış; Aralık 2007'de ise tüm eyaletlere büyüme modeli kullanma teklifi yolu açılmıştır. Bu uygulama, eyaletlere, aşağıdaki şartları sağlamak koşuluyla, AYP'yi belirlemek için büyüme modeli kullanma imkanı tanımıştır:

1. 2014 yılı için belirlenen yeterlilik hedeflerinden sapma olmayacak.

   Bu şart, AYP amaçlı kullanılması düşünülen büyüme modellerinde, öğrencilerin başlangıç noktasındaki bireysel başarı durumlarının ve geçmişteki bireysel ilerleme hızlarının hesaba katılabilmesiyle beraber, başarı projeksiyonunun son noktası olan 2014 yılı başarı düzeyinin en azından yeterlilik çizgisi olarak tanımlanması zorunluluğunu getirmektedir. Bu bağlam dışında kullanılan büyüme modellerinde ise bu projeksiyon, öğrencinin geçmişteki ilerleme hızına göre yapılır ama ilerleme grafiğinin (trajectory) son noktasının belli bir değer olma zorunluluğu yoktur. Tam tersine, böylesi durumlarda büyüme modelleri, öğrencinin zamanın hangi noktasında hangi başarı düzeyinde olacağını tahmin etmek için kullanılır.

2. Öğrenci kategorileri arasındaki başarı farklılıkları giderek azalacak.

   Bu şart, başarı farklılıklarını giderme konusundaki gayretlerin yalnızca 2014'e yakın yıllarda değil, her yıl gösterilmesi gerektiğini ifade etmektedir.

3. Yıllık başarı hedefleri belirlenirken, öğrenci kategorilerinin veya okulların demografik özellikleri gözönünde bulundurulmayacak. Bu özellikler, yıllık hedef belirlemede ayrımcılığa neden olmayacak.

   Bu şart, yalnızca NCLB süreci sonunda ve çıktılar olarak değil, süreç boyunca ve beklenti olarak bile öğrenciler arasında demografik faktörlere bağlı olarak ayrımcılık yapılmaması gerektiğini ifade etmektedir. Büyüme modellerinin katma değer (value-added) modelleri olarak adlandırılan birtakım versiyonlarında, demografik değişkenlerin ve/veya geçmiş başarı düzeyinin istatistiksel olarak kontrol edildiğini, hatta böylesi bir uygulamanın öğretmenlerin öğrencilere sağladığı katma değeri ölçmede çok daha doğru bulunduğunu ifade etmek yararlı olacaktır (Riddle, 2008).

4. Öğrenci değerlendirme sistemi, yıldan yıla (bir sınıf düzeyinden diğerine) karşılaştırılabilir sonuçlar üretecek.

> Bu şartın yerine getirilmesinde en yaygın kullanılan yöntem, farklı sınıf düzeylerindeki performansın aynı ölçekte ölçülebilmesini mümkün kılan dikey ölçeklendirilmiş değerlendirmelerdir (vertically-scaled assessments). Bu yöntemle, bir öğrenci için iki sınıf düzeyi arasındaki başarı farkı hesaplanabilir. Bu yöntem, çalışmanın ilerleyen bölümlerinde açıklanacaktır.

5. Her öğrencinin akademik gelişimi, eyalet çapında bir veri sistemi bünyesinde bireysel olarak takip edilecek.

> Bu şart, bir eyaletteki her bir öğrenciye bir kimlik numarası vermeyi gerektirir ve okul değiştiren öğrenciler gözönünde bulundurulduğunda takip için zorunlu bir durumdur.

6. Başvuran eyaletlerin, 3. sınıftan 8. sınıfa kadar tüm sınıf düzeyleri için federal düzeydeki eğitim bakanlığıtarafından onaylanmış yıllık değerlendirmeleri mevcut olacak. Aynı zamanda, bu değerlendirmeler, büyüme modelinin AYP amaçlı kullanılmaya başlanmasından en geç bir yıl öncesinden itibaren uygulanmış olacak.

> Bu şart, değişimi hesaplamak için gerekli olan başlangıç noktası (baseline) verisinin en az bir yıl öncesine ait olmasını gerektirmektedir (Riddle, 2008).

Eyaletlerin, büyüme modellerini AYP amaçlı kullanma isteklerinin en önemli nedeni, ilk iki model kullanıldığında AYP'yi gerçekleştirmiş sayılamayacak birçok okulun, büyüme modeli kullanıldığı zaman AYP'yi gerçekleştirmiş bir konuma gelebilmeleri olmuştur. Örneğin, 2007–2008 akademik yılında, AYP kararları için büyüme modellerini deneyen pilot çalışma kapsamındaki dokuz eyalette, kanunun orijinalinde sunulan iki modelle AYP'yi gerçekleştirmiş sayılamayan okullardan 1246'sı, büyüme modeli kullanıldığında AYP'yi gerçekleştirmiş sayılabilmişlerdir (U.S. Department of Education, 2011a). AYP'yi gerçekleştirme oranlarının büyüme modelleri yoluyla yükselmesinin ancak ilk yıllar için geçerli olacağı; üç modelde de 2014 yılı itibarıyla her öğrencinin yeterliliğe ulaştığının kanıtlanması gerektiği için, bu yükselmenin 2014'e yaklaştıkça giderek azalacağı tespiti de yapılmıştır (Riddle, 2008). Bir başka ifadeyle, bir grafikte y ekseni üzerinde farklı noktalarda başlayıp grafikte aynı (x,y) noktasına varan çizgiler arasındaki uzaklık, ortak hedef noktasına (x,y) yaklaştıkça azalacaktır.

Büyüme modellerine rağbetin bir nedeni de, kaynakların (maddi kaynaklar, öğretmenlerin mesaileri, vb.) yeterlilik çizgisinin hemen altındaki öğrenciler için orantısız bir biçimde kullanılmasının önüne geçme isteği olmuştur. Büyüme modeli dışındaki iki modelde, yeterlilik çizgisinin üstünde bilgi sahibi olan öğrencilere kaynak ayrılması, AYP açısından bir yarar sağlamamaktadır. Bu iki modelde, yeterlilik çizgisinin çok altındaki öğrencilerin ise, ancak çok büyük gayretlerle yeterlilik çizgisine yükseltilebilecekleri için, AYP açısından emek verilmeye değmez olarak görülmeleri olasılığı vardır. Bu durumda AYP'yi gerçekleştirme açısından en ekonomik yaklaşım, kaynak kullanımının, yeterlilik çizgisinin hemen altında olup az bir ek gayretle çizginin üstüne çekilebilecek öğrencilere yoğunlaştırılması olacaktır (Welner, 2008). Büyüme modellerinde ise, her öğrencinin bireysel gelişimi önemli olduğu için her öğrencinin kendi içinde gelişim göstermesi AYP için gereklidir. Ohio ve Tennessee gibi birtakım eyaletlerin büyüme modellerinde, halen yeterlilik çizgisinin üstünde olduğu halde, başarı projeksiyonunda bir veya birkaç yılda bu çizginin altına düşmesi beklenen öğrenciler, AYP'yi gerçekleştirememiş olarak hesaba katılmaktadırlar (Center for Public Education, n.d.). Bu yönüyle büyüme modellerinin, yeterlilik çizgisinin zaten üstünde olan ve dolayısıyla ihmal edilme riski yaşayan üstün zekalı öğrencilerin eğitimi için bir cankurtaran olduğu söylenebilir.

AYP'yi gerçekleştirememenin yaptırımlarından kurtulmak için öğretmenlerin, başarısı düşük okullardan yüksek olanlara geçme gayretinin engellenmesi, büyüme modeli kullanmanın başka bir faydası olarak görülmüştür (Welner, 2008). Her öğrencinin bir akademik yıldan diğerine gelişiminin bu yıllar için ortak bir ölçek kullanımıyla hesaplanmasının, her sınıf düzeyindeki yeterliliğin, diğer sınıf düzeylerindeki yeterlilik tanımlarından bağımsız olarak belirlenmesine göre daha sağlıklı bir ölçme yöntemi olduğu da ifade edilmiştir (Welner, 2008). Özetle, ne kadar karmaşık bir yöntem olursa olsun, büyüme modelleri birçok eyalet tarafından rağbet görmüş ve AYP'yi belirlemek için kullanılmıştır. Bu yöntemin karmaşıklığını anlamak için eyaletlerin federal düzeydeki eğitim bakanlığına sundukları büyüme model tekliflerine bir göz

atmak yeterlidir (U.S. Department of Education, n.d.c). Uygulamaya konduktan sonra ise eyaletleri ve okulları ne kadar detaylı bir veri toplama, veri düzenleme ve hesaplama iş yüküyle karşı karşıya bıraktığını anlamak için Missouri eyaletinin hazırladığı AYP kılavuzunu incelemek yararlı olacaktır (Missouri Department of Elementary and Secondary Education, 2011).

2009 yılında kamuoyuna duyurulan ve NCLB yaptırımlarına esneklik getiren Race to the Top (Zirveye Doğru Yarış) uygulamaları da, öğrenci izleme, veriye dayalı karar verme ve hesap verebilirlik sisteminin bir bileşeni olarak büyüme modellerinin kullanımının devamını öngörmektedir (Lohman, 2010).2008 yılında ülke yönetimindeki politik parti, başkan (president) ve eğitim bakanı değişmiş olmasına ve NCLB'ye12 yıllık süreç içerisinde birçok konuda ciddi eleştiriler yöneltilmiş olmasına rağmenbüyüme modellerinin kullanımından vazgeçme şimdilik sözkonusu değildir.

**Sınavların İçerik Standartlarının ve Başarı Düzeyi Standartlarının Belirlenmesi**

Tüm kategorilerde öğrencilerin yüzde 100'ünün yeterliliğe ulaşması hedefine yönelik bir eleştiri, özel eğitim ve İngilizce'de sınırlı yeterliliğe sahipöğrenci kategorilerinde, bunun teknik olarak mümkün olmadığıdır (Darling-Hammond, 2007). Bu öğrenciler, öğrenme ile ilgili özel gereksinimleri olduğu ve dolayısıyla öğretimin genel formatında tipik öğrenciler kadar başarılı olamadıkları için bu kategorilere yerleştirilmekte; bu gereksinimler ve başarısızlık bittiği durumda bu kategorilerden çıkarılmaktadırlar. Dolayısıyla, bu iki kategori, tipik öğrenciden beklenen başarıyı gösteremeyen öğrencilerden oluşacağı için yüzde 100 yeterliliğe ulaşmaları hiçbir zaman mümkün olmayacaktır (Darling-Hammond, 2007).

Başka bir konu ise, uygulanan sınavların içerik standartlarının ve başarı düzey standartlarının nasıl belirlendiğidir. Hemen ifade etmek gerekir ki, içerikle ve yeterliliğin / başarının belirlenmesi ile ilgili akademik standartlar, federal düzeyde kararlaştırılmamış; her eyalete, federal düzeydeki eğitim bakanlığındanonay almak şartıyla, kendi standartlarını belirleme özgürlüğü tanınmıştır (U.S. Department of Education, n.d.d). Amerika Birleşik Devletleri'nde eğitimle ilgili karar yetkisi federal hükümete değil, eyaletler ve yerel otoritelere verilmiş olup bu husus ülkenin anayasasında da yer almaktadır (U.S. Department of Education, n.d.e). Bu konudaki başlıca eleştiri, öğrencilerin yüzde 100'ünü yeterlilik (proficiency) düzeyine çıkarma zorunluluğunun, eyaletleri, yaptırımlarla karşılaşmamak amacıyla, yeterlilik için düşük akademik standartlar belirlemeye zorlayacağıdır (Darling-Hammond, 2007).

Eyaletlerin standart belirleme süreçlerinin üç aşamalı olduğu söylenebilir. Birincisi, içerik standartlarının belirlenmesi; ikincisi, bu içerik standartlarını ölçebilecek sınavların geliştirilmesi veya satın alınması; üçüncüsü ise bu sınavlarda yeterlilik / başarı düzeylerinin alt sınırlarını oluşturacak test puanlarının belirlenmesidir (Rotherham, 2006). İlk aşama olan içerik standartlarının belirlenmesi, alan uzmanlarından oluşan bir heyetin görüşüyle olup öznel bir süreçtir. Heyet, bu süreçte, kendi standartlarını, uluslararası ve/veya piyasanın talep ettiği standartlarla karşılaştırarak belirlemeye (benchmarking) çalışsa bile, tüm uzmanların aynı şekilde kullanacağı ve/veya verilen bir kararın yanlışlığını kanıtlayabilecek ölçütler yoktur. Belirlenen içerik standartlarına göre sınav geliştirilmesi ya da ölçme firmalarının sunduğu sınav seçeneklerinden birinin seçilmesi olan ikinci aşama da, uzman görüşüne dayanan öznel bir süreçtir. Bu ikinci aşamada, satın alınacak sınavın ücreti, sınav sonrası sonuçların firma tarafından raporlanması hizmetinin kalitesi gibi ölçütler de kararda etkili olmaktadır.Üçüncü aşama olan yeterlilik / başarı düzeyi standartlarını belirleme sürecinde de öznellik bulunmaktadır. Fakat AYP'yi belirlemek için büyüme modeli kullanıldığında sözkonusu olan dikey ölçeklendirme işlemi, önemli bir nicel ölçme değerlendirme başlığı olduğu için üçüncü aşama burada biraz detaylı ele alınacaktır. Rotherham'ın da (2006) açıkladığı gibi, yeterlilik / başarı düzeyi standartlarını belirlemek için kullanılan çok yöntem olmakla birlikte birkaçı yaygın olarak kullanılmaktadır. Bunlardan, AYP'yi belirlemede kullanılan ilk iki model için uygulanabilecek ikisi ise, değiştirilmiş Angoff (modified Angoff) yöntemi ve yer işaretleme (bookmarking) yöntemidir. Değiştirilmiş Angoff yönteminde uzmanlardan oluşan bir hakem heyetinin her bir üyesi, sözkonusu sınavı önce kendisine uygular. Sonra üyeler kendi aralarında tartışarak bu sınav için başarı düzey [örneğin, temel (basic), yeterli (proficient) ve ileri (advanced)] tanımları üzerinde mümkün olduğu kadar görüş birliğine varmaya çalışırlar. Daha sonra, her üye herbir soruyu inceleyerek her bir başarı düzeyinin alt çizgisinin hemen üstünde olan öğrencilerin yüzde kaçının o soruyu doğru yanıtlayabileceğini tahmin eder. Her bir üye için bu yüzdeliklerin ortalaması alınarak her bir başarı düzeyine ulaşılması için doğru yanıtlanması gereken toplam soru sayısı, dolayısıyla her bir düzeyin alt puanı hesaplanır. Bireysel ve diğer heyet üyelerinden

bağımsız olarak gerçekleştirilen bu aşamadan sonra heyet üyeleri, birbirlerinin tahminlerini de öğrenerek birkaç oturumda başarı düzeylerinin alt puanları hakkında uzlaşmaya ve tek bir alt puanda buluşmaya çalışırlar. Yer işaretleme yönteminde ise, uzmanlardan oluşan bir hakem heyeti sınavı inceler ve belirlenecek başarı düzeyleri hakkında tartışırlar. Daha sonra, sınavdaki sorular, sınavın daha önce uygulandığı öğrencilerin her bir sorudaki performansına göre en kolaydan en zora doğru sıralanır. Her bir hakem bu sıralanmış soru listesinde başarı düzeylerini ayırt etmek için nereden çizgi çekilmesi gerektiğini, yani belli bir başarı düzeyine ulaşmış olmak için listede hangi soruya kadar doğru yanıt verilmesi gerektiğini belirler. Birkaç döngüde gerçekleştirilen bu işlemin sonunda hakemlerin belirledikleri soru sıra sayılarının ortanca değeri alınarak her bir başarı düzeyi için minimum doğru yanıt sayısı, dolayısıyla alt puan belirlenir. Bu iki yöntemle ilgili dikkat edilmesi gereken bir konu, bu yöntemlerin belirli bir test üzerinde uygulandığı ve belirlenen alt puanların yalnızca belirlenen test için anlamlı olduğudur. Dolayısıyla, bir test için belirlenen alt puanın başka bir test için belirlenen alt puanla karşılaştırılması anlamsız olacaktır. Her bir sınavın içerdiği soruların zorluk derecelerinin farklı olacağı ve bu zorluk derecelerinin bu yöntemler kullanılarak alt puanların belirlenmesinde etkili olacağı açıktır (Rotherham, 2006).

Bu iki yöntem, AYP'yi belirlemede kullanılan ilk iki modelde kullanılabilmekle birlikte üçüncü model olan büyüme modelinde kullanılamamaktadır. AYP'yi belirlemede büyüme modeli kullanan bir eyaletin değerlendirme sistemi, yıldan yıla (bir sınıf düzeyinden diğerine) karşılaştırılabilir sonuçlar üretmek zorundadır. Bu şartın yerine getirilmesinde en yaygın kullanılan yöntem ise, farklı sınıf düzeylerindeki performansın aynı ölçekte ölçülebilmesini mümkün kılan dikey ölçeklendirilmiş değerlendirmelerdir (vertically-scaled assessments).

Büyüme modellerinde dikey ölçeklendirilmiş değerlendirmeler için yaygın olarak kullanılan yöntem ise yer işaretlemenin farklı bir versiyonudur. Bu yöntem, madde tepki kuramını(item response theory - IRT) kullanmaktadır. Bu kuramın uygulamalarında, bir sınavda sorulan her bir soru, bir veya birkaç parametre [örneğin, zorluk (difficulty), ayrıştırma (discrimination) ve tahmin ederek doğru yanıtlanabilme (pseudo-chance)] açısından etiketlenmektedir. Etiketlenmiş sorulardan oluşan bir soru bankasından bir sınavın uygulandığı herhangi bir bireyin bilgi düzeyi, sınavda sorulan diğer soruların veya aynı sınavı alan diğer bireylerin özelliklerinden bağımsız olarak (group-independent; test-independent) belirlenmektedir. Belli bir bilgi düzeyindeki bir bireyin etiketlenmiş olan bir soruyu doğru yanıtlama olasılığı veya etiketlenmiş bir soru için farklı doğru yanıtlanma olasılık değerlerine karşılık gelen bilgi düzeyleri hesaplanabilmektedir. Yer işaretlemenin bu farklı versiyonunda, IRT kullanılarak en kolaydan en zora doğru sıralanmış sorular üzerinde hakemler doğru yanıtlanma için en düşük olasılık değerini belirlerler ve sıralanmış sorular listesinde o noktada yer işaretleme yapılır. Örneğin, hakemler yeterlilik düzeyine ulaşmış sayılabilmenin şartını, sıralanmış soruların ilk üçünün en az yüzde seksen olasılıkla doğru yanıtlanması olarak belirlemişlerse, bu olasılığa karşılık gelen bilgi düzeyi hesaplanır ve yeterliliğin alt sınırı olarak belirlenir. Aynı bilgiyi ölçen ortak bir ölçek uzantısında etiketlenmiş sorular üzerinde bir sonraki sınıf düzeyinin yeterlilik alt sınırını da belirlemek böylelikle mümkün olmaktadır. Örneğin, üçüncü sınıf okuma bilgisinin yeterlilik alt düzeyi sıralanmış soruların ilk üçünün en az yüzde seksen olasılıkla doğru yanıtlanmasına karşılık gelen bilgi düzeyi ise, dördüncü sınıfınki ilk altı sorunun en az yüzde seksen olasılıkla doğru yanıtlanmasına karşılık gelen bilgi düzeyi olarak belirlenebilir (Patz, 2007).

Böylesi bir dikey ölçeklendirmede, farklı sınıf düzeyleri için ortak bir ölçek geliştirme, üç yöntemden biriyle mümkün olmaktadır. Birincisi, farklı sınıf düzeylerindeki öğrencilere uygulanan sınavlara ortak sorular konulabilir (common items). İkincisi, aynı sınıf düzeyindeki bir öğrenci grubuna hem kendi sınıf düzeyi hem de bir alt ve/veya bir üst sınıf düzeyini de temsil eden zorlukta sorular sorulabilir (common examinees). Üçüncüsü, her sınıf düzeyine kendi seviyelerinde sınav uygulanmakla beraber farklı sınıf düzeylerini birbirleriyle ilişkilendirmeye (linking) olanak sağlayacak ortak bir ölçeklendirme sınavı ek olarak tüm öğrencilere verilebilir (common scaling test) (Patz, 2007). Dikey ölçeklendirme, bazı eyaletlerin NCLB amaçlı geliştirdikleri sınavlarda ve ETS (Educational Testing Service), McGraw Hill ve Harcourt gibi büyük ölçekli ölçme firmalarının NCLB'den bağımsız geliştirdikleri ve ABD'de yaygın olarak kullanılan Metropolitan Achievement Test ve TerraNova gibi sınavlarda kullanılmaktadır (McCall, 2007; Patz, 2007).

Dikey ölçeklendirme ile ilgili bir sınırlılık, IRT kullanılarak dikey ölçeklendirilen sınavların tek boyutlu bilgi ölçtüğü (unidimensionality) varsayımıdır ki, herhangi bir alanda öğrenci başarısının birçok faktörden etkilendiği düşünülecek olursa oldukça sınırlayıcı bir varsayımdır. Aynı varsayım, bu ölçeklendirmenin,

ancak bir sınıftan diğerine aynı uzantıda devam eden bilgi veya beceri türleri üzerinde uygulanabileceği sınırlılığını getirmektedir. Dolayısıyla, bu tür bir uygulamaya, ilkokul düzeyindeki okuma / dilbilgisi dersi iyi bir adayken, birbirinden bağımsız farklı konulardan oluşan tarih dersi, iyi bir aday değildir. İkinci sınırlılık ise, sınıf düzeyleri birbirinden uzaklaştıkça hem ilişkilendirmeninistatistiksel olarak zayıflaması ve güvenilirliğini kaybetmesi, hem de sınıf düzeyleri arasında karşılaştırma sırasında yapılan yorumların anlamsızlaşmasıdır. Örneğin, dördüncü sınıfla beşinci sınıf öğrencileri arasında dikey ölçeklendirme temelli bir karşılaştırma sağlıklı ve anlamlı olurken, dördüncü sınıfla sekizinci sınıf öğrencileri arasında böyle bir karşılaştırma anlamlı veya istatistiksel olarak sağlıklı olmamaktadır. Bunun nedeni ise, ilişkilendirmeiçin çok önemli olan, hem içerik standartları hem de farklı başarı düzeylerindeki öğrencilerin dağılımı açısından, karşılaştırılan sınıf düzeyleri arasındaki örtüşmenin yetersiz kalmasıdır (McCall, 2007; Patz, 2007).

AYP'yi belirlemek için büyüme modelleri kullanıldığında dikey ölçeklendirmenin bir zorunluluk olmadığı, birkaç eyaletin büyüme modeli kullanmasına rağmen dikey ölçeklendirme yapmadığı literatürde ifade edilmekle birlikte, böylesi uygulamalarda bile dikey ölçeklendirmenin değerli ek bilgi sunacağı dile getirilmektedir (O'Malley, 2008; Patz, 2007). Bunun yanında, büyük ölçekli ölçme firmaları, ABD'nin birçok eyaletinde yaygın olarak kullanılan sınavları için dikey ölçek geliştirmeyi tercih etmişlerdir. Bu nedenle, dikey ölçek kullanmayan büyüme modelleri istisnai uygulamalar olarak görülmüş ve bu çalışmada değinilmemiştir.

AYP'yi belirlemek için büyüme modellerinin ve bu modelleri kullanırken ilerlemeyi ölçmek için dikey ölçeklerin kullanımı, oldukça karmaşık uygulamalar olarak karşımıza çıkmaktadır. Bu uygulamalarla ilgili tüm sorunların da tamamıyla çözüldüğü söylenemez. Bununla birlikte, bu yöntemlerin kullanımındaki sorunların giderilmesine ve alternatif yaklaşımların geliştirilmesine yönelik araştırmalar, NCLB ile ivme kazanmış göreceli olarak yakın tarihe ait araştırmalardır (Yen, 2009). Konunun uzmanları, eğitim politikalarıyla ilgili kararların nesnel yöntemlerle alınması gerektiğinin altını çizerken bu yöntemlerde kullanılan istatistiksel uygulamalarla ilgili araştırmaların sürdürülmesini doğru bulmaktadırlar. Uzmanlar, herhangi bir hesap verebilirlik sistemine ait ölçme değerlendirme bileşeni ile ilgili birkaç önemli tavsiyede bulunmaktadırlar. Bunlardan biri, kullanılan istatistiksel yöntemler açısından temkinli olunmasıdır. Örneğin, yerine getirilmesi zorvarsayımları olan istatistiksel yöntemlerin tercih edilmemesinin ve çok sayıda insanın yaşamını olumsuz etkileyecek kadar yüksek oranlı ölçümsel hatalarla karar verilmemesinin altı çizilmektedir. Bir başka tavsiye de, hem aktif olarak kullanılan hem de yeni geliştirilecek yöntemlerin sağlamasının yapılabileceği uygulamaların geliştirilmesi ve başarı/başarısızlığın belirlenmesinde birden fazla bilgi kaynağının ve ölçütün kullanılmasıdır. Son olarak, ölçme değerlendirme sonuçlarının, bunlardan etkilenecek bireylere yeteri kadar açıklanması ve doğru yorumunun sağlanması tavsiye edilmektedir (U.S. Department of Education, 2009; Yen, 2009).

**Standartlar İçin Bir Kalite Kontrol Mekanizması Olarak NAEP**

İçerik ve yeterlilik / başarı düzeyi standartları federal düzeyde belirlenmemekle birlikte, her eyaletin kendi belirlediği standartlarla ilgili bir kalite kontrol mekanizması görevi yapabilecek, aynı zamanda akademik standartlar ve başarı açısından eyaletler arası karşılaştırmalara olanak sağlayacak bir uygulama olarak tüm eyaletlerin NAEP (National Assessment of Educational Progress – Eğitimde İlerlemenin Ulusal Değerlendirmesi) sınavlarına katılmaları zorunlu kılınmıştır. NAEP, 1969 yılında federal düzeydeki eğitim bakanlığıtarafından başlatılan ve periyodik olarak çeşitli alanlarda (okuma, matematik, coğrafya, tarih, vb.) uygulanan sınavlar yoluyla ülkenin eğitim sisteminin durum ve ilerlemesini değerlendiren bir sistemdir. Ulusun Karnesi (Nation's Report Card) olarak da adlandırılan bu sistem, ülke ilk ve orta öğretiminin nabzını tutmada önemli rol oynamakta ve değerlendirme sonuçları hem ulusal hem de yerel medya tarafından yayınlanmaktadır. Ülke çapındaki değerlendirmelerde tüm ülkeyi temsil eden bir örneklem, eyalet bazındaki değerlendirmelerde ise her eyaletten o eyaleti temsil eden bir örneklem kullanılmaktadır (National Center for Education Statistics, n.d.a).

NCLB'nin NAEP'le ilgili getirdiği zorunluluk, eyaletlerin 2002–2003 akademik yılından itibaren, okuma ve matematik alanlarında, en azından iki yılda bir NAEP çatısı altında dördüncü ve sekizinci sınıf öğrencilerine uygulanacak olan eyaletler bazındaki değerlendirmelere katılmaları olmuştur (National Center for Education Statistics, n.d.b). Eyaletler ve yerel otoriteler, eyaleti temsil eden bir örneklem oluşturabilmek için NAEP'in istediği öğrenci listelerini NAEP'e bildirmek, sonrasında da örnekleme giren öğrencilere

sınavların uygulanmasını sağlamakla sorumludurlar (National Center for Education Statistics, n.d.c). Bununla birlikte, NAEP sonuçları, eyaletlerin NCLB amaçlı uyguladıkları sınavların yerine geçmemenin yanında eyaletlere veya yerel otoritelere herhangi bir yaptırıma da neden olmamaktadır (National Center for Education Statistics, n.d.b). Dolayısıyla, bu değerlendirmeler, eyaletlerin NCLB amaçlı uyguladıkları ölçme değerlendirme sistemine bir müdahale değil; federal hükümetin kendi başına yürüttüğü ve masraflarını da kendisinin karşıladığı bir nabız tutma ve kalite kontrol mekanizması olarak görülebilir.

NAEP'in içerik ve yeterlilik / başarı düzeyi standartları, NCLB'den bağımsız olarak geliştirilen standartlardır. Bu nedenle, eyaletlerin NAEP değerlendirmelerine katılımları, federal hükümete eyaletlerin NCLB uygulamaları ile ilgili üç tür bilgi sağlama imkanına sahiptir. Birincisi, federal hükümet; her bir eyalette, o eyaletin NCLB amaçlı geliştirdiği akademik standartlara göre yeterliliğe ulaşan öğrenci oranıyla, NAEP standartlarına göre yeterliliğe ulaşan öğrenci oranını karşılaştırarak, o eyaletin gerçekten kabul edilebilir düzeyde standartlar geliştirip geliştirmediğiyle ilgili bir fikre sahip olabilir. İkincisi, her bir eyaletin kendi standartlarına göre ulaştığı yeterlilik oranıyla NAEP standartlarına göre ulaştığı oran arasındaki farka bakarak ve bu fark bazında eyaletleri karşılaştırarak, hangi eyaletlerin yüksek hangi eyaletlerin düşük standartlar geliştirdiğiyle ilgili çıkarımlar yapabilir. Üçüncüsü ise, eyaletlerin NCLB amaçlı kullandıkları ölçme değerlendirme sistemlerinden bağımsız bir şekilde, eyaletlerin başarı düzeyleri hakkında bilgi sahibi olabilir.

National Center for Education Statistics (2011) tarafından NAEP değerlendirmeleriyle eyaletlerin NCLB amaçlı değerlendirmelerinin karşılaştırıldığı örnek bir çalışmada, 2009 yılı eyalet standartlarının NAEP standartlarının çok altında kaldığı gözlemlenmiştir. Örneğin, dördüncü sınıf okuma standartlarında, hiçbir eyaletin yeterlilik (proficiency) alt sınırını NAEP kadar yüksek belirlemediği, hatta 50 eyaletten 35'inin yeterlilik alt sınırını NAEP'te yeterlilik başarı düzeyinin bir altındaki düzey olan temel (basic) alt sınırından bile daha aşağıda belirlediği bulunmuştur. NAEP'in ve her bir eyaletin sınavlarındaki sorular birbirinin aynı olmadığı gibi bu sınavların uygulandığı öğrenci grupları da farklıdır. Bu nedenle, bu karşılaştırmalar aynı ölçek üzerinde birebir karşılaştırmalar olmayıp yukarıdaki paragrafta kısaca açıklanan karşılaştırma yöntemlerine bağlı çıkarımlar yoluyla yapılmakta ve bu çıkarımların doğruluğu farklı yöntemlerle sınanmaktadır. NAEP ve eyaletlerin standartlarını birbirleriyle karşılaştırılabilir hale getirmek için kullanılan yöntemlere ilişkin detaylı bilgi, ilgili teknik rapordan edinilebilir (National Center for Education Statistics, 2010).

**AYP ve Standartlarla İlgili Son Durum**

Bu çalışma, NCLB bağlamında kullanılan AYP kavramı ile sınırlı olduğu için, 2009 yılında başlatılan ve NCLB'yi kaldırmayıp onunla birlikte yürürlükte olan Race to the Top (Zirveye Doğru Yarış) uygulamaları açıklanmayacak; kısa bir değerlendirmeye imkan tanıyacak kadar sözedilecektir. Race to the Top (RTTT) programı, ABD ekonomisini canlandırmak için federal hükümetin onayladığı ek bütçenin eğitime düşen kısmıyla yürütülmektedir. Hedefleri, NCLB hedefleriyle büyük oranda örtüşmesine rağmen yaklaşımları farklıdır. NCLB, eğitimde hedeflediği iyileşmeyi sağlamak için federal hükümetin eyaletlere NCLB öncesinden beri vermekte olduğu para için zorunluluklar getirmekte; RTTT ise benzer hedefler için ek bütçeden parasal teşvikler sunmaktadır (Lohman, 2010). RTTT, 2008 yılı sonunda ülke yönetimindeki politik partinin, başkanın ve eğitim bakanının değişmesinden hemen sonra başlatılan bir uygulamadır. Yeni eğitim bakanı, eski yönetimin yürürlüğe soktuğu bir kanun olan NCLB ile ilgili görüşlerini çok güzel özetleyen bir açıklamasında, 2011 yılı için AYP'yi gerçekleştiremeyen okul oranının % 82'ye kadar çıkabileceğinin öngörüldüğünü ifade etmekte; bu öngörüyü takiben, okulları 'başarısız' olarak etiketlemekten vazgeçilmesi ve NCLB'de revizyona gidilmesi gerektiğini savunmaktadır (U.S. Department of Education, 2011b). RTTT teşvik fonuna başvuracak eyaletlerden beklentilerin, standartlar ve kullanılacak değerlendirme sisteminin özellikleri açısından NCLB'den düşük olmadığı, hatta bazı başlıklarda daha yüksek olduğu söylenebilir (Lohman, 2010). Kısacası, RTTT, eyaletlere farklı bir formatta ikinci bir şans tanıyan bir uygulama olarak tanımlanabilir. Bu kısa karşılaştırmadan AYP ile ilgili bulunulabilecek çıkarım ise, akademik başarının kullanılabilir sonuçlar üretecek, nesnel, geçerli, güvenilir ve mümkün olduğu kadar detaylı bir şekilde ölçülmesi hedefinden bir sapma yaşanmamış olması; bunun için gerekli yöntemsel araştırmaları yapma ve bu tür bir ölçmenin üreteceği veriyi kullanacak kapsamlı değerlendirme sistemlerini kurma niyetinden vazgeçilmemiş olmasıdır.

**AYP Uygulamalarından Türkiye İçin Çıkarılabilecek Dersler**

Çalışmanın buraya kadarki kısmında bir eğitim reform hareketinin yalnızca bir yönü olan ölçme değerlendirme alanındaki tek bir kavrama odaklanılmıştır. Bununla birlikte, eğitim alanındaki hemen her uygulamanın birçok yönü olduğu bir gerçektir. Dolayısıyla, çalışmanın ilk kısmından bu bölümde çıkarılan derslerin ve bu dersleri takiben yapılan tavsiyelerin, ilgili konuların diğer yönleriyle de ele alınması ve değerlendirilmesi gerektiği açıktır. AYP işlenirken birçok detayın ve özellikle NCLB süreci sırasında yaşanan sayısız tartışma ve gelişmenin zorunlu olarak çalışma dışı bırakıldığını da ifade etmek yararlı olacaktır. Farklı bir ülkedeki bir uygulamadan Türkiye için dersler çıkarırken kısıtlı hacme sahip bir çalışmada, bu tür sınırlılıkların olması kaçınılmazdır. Çıkarılabilecek derslerin yalnızca konu başlıklarına bile dikkat çekilse, çalışmanın bu bölümü başarıya ulaşmış sayılacaktır.

Çıkarılabilecek birinci ders, yeni herhangi bir uygulama için bir takvim belirlenmesi ve uygulamanın doğru, yanlış ve eksik yönleri hakkında süreç sırasında ve sonunda yeterli bilgi ve geribildirim elde edildikten sonra karar verilmesi gerektiğidir. NCLB, ulaşmak istediği hedefler için 12 yıllık (2002-2014) bir takvim belirlemiş ve bu takvimi işletmiştir. Süreç sırasında uygulamanın başarı düzeyini ölçmek için bir ölçütün olması gerektiğine en güzel örnek, bu çalışmanın merkezini tutan AYP kavramıdır. NCLB sürecinde, sağlam bir ölçme değerlendirme sistemi kurma gayretinin yanısıra, tüm paydaşlardan geribildirim alındığı gibi, uzmanlardan oluşan birçok heyete raporlar hazırlatılmış ve ara değerlendirmeler yaptırılmıştır. Bu raporlar, veriye dayalı raporlar olmanın yanında, uygulama ile ilgili sorunları, sorunların kaynaklarını ve bu sorunların olası çözümlerini içeren kapsamlı ve apolitik raporlardır. Uygulama yoluyla değiştirilmek istenen değişken(ler)in ölçümünün en sağlıklı şekilde nasıl yapılacağının belirlenmesi ve gereken ölçme değerlendirme sisteminin kurulması, çoğunlukla yüksek miktarda kaynak harcanan bu tür deneyimlerden en yüksek düzeyde yararlanmak adına önem taşımaktadır. Böyle bir yaklaşım, süreçten etkilenen insanlara saygının ifadesi olarak onları bilgilendirmeyi ve sürecin bilinçli yönetildiği mesajını vermeyi de mümkün kılacaktır.

Çıkarılabilecek ikinci ders, NCLB'nin hedefinde olan ve AYP kavramının uygulandığı derslerle ilgilidir. Standartlar geliştirilip kalite açısından her zaman kontrol edilmek istenen derslerin, matematik ve okumayla sınırlandırılmasının üç ana nedeni olduğu söylenebilir. Birinci neden, bu iki dersin başka alanlarda gerçekleşmesi istenen birçok farklı öğrenmeye temel oluşturacak bilgi ve becerileri kazandıracak dersler olmasıdır. Diğer derslerde ve hatta yaşam boyu öğrenmede okuma becerisinin ve temel matematik bilgisinin önemi yadsınamaz. İkincisi, bu derslerin öğrenme hedeflerinin zaman içerisinde kolay kolay değişmeyecek olmasıdır (öğretim yöntemleri değişse bile). Okuma, içeriği yalnızca bir araç olarak kullanan, dolayısıyla zaman içerisinde derste kullanılan içerik değişse bile kazandırılmak istenen becerinin değişmeyeceği bir derstir. Matematik ise evrensel kurallardan oluşan ve doğruları değişmeyecek, dolayısıyla da yoruma açık olmayan bir disiplindir. Bu yönleriyle bu iki ders, sosyal bilimler alanının yoruma açık derslerinden ve hatta yeni buluş, keşif veya kuramlarla içeriği değişebilecek olan fen bilgisi gibi derslerden farklıdır. Üçüncüsü ise, tarihsel olarak ABD'deki halk ve akademisyenlerin önemli bir kısmının eğitim sistemlerindeki müfredat genişliğini (breadth of curriculum) bir zenginlik olarak görmeleridir. Bu nedenle, ABD'de de uluslararası karşılaştırmalarda en başarılı çıkan ülkeler gibi temel akademik derslerden oluşan dar bir müfredat takip edilmesi gerektiği düşüncesine karşı çıkanlar olmuş; ABD'li öğrencilerin diğer birçok ülkedeki yaşıtlarına göre daha genel bir eğitim aldığı ve akademik olmayan daha çok ilgi alanları ve müfredat dışı (extracurricular) etkinlikleri olduğu ifade edilmiştir (Berliner & Biddle, 1995). Hesap verebilirlik sisteminin, birçok ders için değil de, bu iki temel ders için kurulmak istenmesinin bir nedeni de budur. Bu iki dersin, NCLB'nin başlangıcında hesaba katılmayan ama AYP'yi belirlemek için büyüme modellerinde dikey ölçeklendirilmiş sınavların kullanılmaya başlamasıyla önem kazanan dördüncü bir özelliği de, her birinde kazandırılmak istenen bilgi veya becerinin farklı sınıf düzeylerinde bile olsa aynı uzantıda ölçülebilmesidir. Başka bir ifadeyle, her bir dersin bilgi veya becerisinin birden fazla sınıf düzeyini kapsayacak şekilde tek boyutlu olmasıdır. Bu özellik, derste kazandırılmak istenen bilgi ve becerilerin kümülatif olmasıyla, yani bir sonraki aşamada kazandırılacak bilgi ve becerinin, bir önceki aşamayla bağlantılı olması ve onun üzerine bina edilmesiyle ilgilidir. Sonuç olarak, Türkiye'de benzer bir kalite kontrol ve/veya hesap verebilirlik sistemi kurulmak istendiğinde, Türkiye için de geçerli bu nedenlerle, en azından ilk planda sistemin bu iki dersle sınırlandırılması en doğrusu olacaktır.

NCLB sürecindeki problemlerden bir tanesi de, sınav uygulama ve ölçme değerlendirme sisteminin işletilebilmesi konusunda yaşanan para ve teknik eleman azlığıdır. Eyaletlerin çoğu, hem NCLB öncesi bu kadar kapsamlı bir sisteme sahip olmadıkları hem de böylesi istatistiksel olarak karmaşık ve emek isteyen bir işin üstesinden gelebilecek kadar teknik eleman bulamadıkları için, ülkedeki büyük ölçekli ölçme firmalarından hizmet satın alma yoluna gitmişlerdir. Örneğin, tüm eyaletlerin NCLB amaçlı ölçme değerlendirmeye yalnızca 2005–2006 okul yılında harcadıkları paranın, 700–750 milyon dolar gibi yüksek bir miktar olduğu tahmin edilmektedir. Teknik eleman konusunda ise yalnızca eyaletlerin değil, ölçme firmalarının da sıkıntı çektiği söylenebilir. Üniversitelerin ölçme değerlendirme ve istatistik alanlarında sunduğu programların ve her yıl bu programlardan mezun olanların sayısının çok az olduğu dile getirilmiştir. Yapılması gereken ölçme değerlendirme ve raporlamanın çok emek gerektirmesi, bu işler için tanınan sürenin azlığı ve teknik eleman eksikliği birleşince sistemde eksiklikler ve yanlışlıklar da görülmüştür (Center on Education Policy, 2007; Toch, 2006). Bu paragrafta aktarılan sorunlar, ölçme değerlendirme alanındaki üniversite programları ve yetişmiş eleman sayısının azlığı açısından Türkiye'nin yalnız olmadığını, gelişmiş ülkelerde de aynı sorunların yaşandığını göstermesi yönüyle güzel bir teselli olabilir. Başka bir açıdan ise, gerçekten işe yarayacak bir kalite kontrol ve/veya hesap verebilirlik sisteminde sağlam bir ölçme değerlendirme yapılanması gerekeceği, bunun için de üniversitelerde kaliteli eğitim veren ve yeterli sayıda eleman yetiştiren ölçme değerlendirme ve istatistik programlarının bulunmasının bir zorunluluk olduğu sonucu çıkarılabilir. Özellikle, bir hesap verebilirlik sisteminde, NCLB'de olduğu gibi bazılarına haklar tanınacağı (çocuğunun gittiği okulun başarısızlığı durumunda velilere tanınan okul seçme gibi haklar), bazılarına ise yaptırımlar uygulanacağı (uzun süre başarısızlık durumunda okul ve öğretmenlerin görevden alınması gibi yaptırımlar) akıldan çıkarılmamalıdır. Böyle bir durumda, ölçme değerlendirme sisteminin her bileşeninin (sınav soruları, puanlama, karar mekanizmaları, vb.) savunulabilir olması gerekecektir. Bu da, ancak sağlam bir sistem ve kadro ile mümkün olacaktır.

NCLB'nin başarısız bir reform görüntüsü vermesinin nedeni, kurulmak istenen ölçme değerlendirme sistemi değildir. 2014 hedeflerine ulaşılamayacağının veriyle de desteklenerek öngörülmesinin belki de en büyük nedeni, hem süreç için hem de süreç sonu için konulan hedeflerin gerçekçi olmayışıdır. Süreç sonu için konulan tüm okullardaki tüm öğrencilerin başarılı olması hedefi, okullardan yüzde yüz verim beklemek demektir ki, eğitim gibi malzemesi insan olan bir etkinlik alanı için, hem de milyonlarca öğrenci sözkonusu olunca, gerçekçi değildir.

Süreç boyunca kullanılacak AYP tanımları için federal hükümetin getirdiği şartlardan en önemlisinin, öğrencilerden beklenen akademik gelişmenin hiçbir şekilde öğrencilerin veya okulların demografik özelliklerine göre belirlenmemesi; beklenen gelişmenin, ancak öğrencilerin geçmiş bireysel akademik düzeylerine göre belirlenebilmesi olduğu söylenebilir. Geçmiş akademik düzeyin, öğrencinin içinde bulunduğu demografik şartları belli bir derecede yansıttığı söylenebilse de, geçmiş akademik düzeyin hesaba katılmasıyla, demografik özelliklerin de hesaba katılmış olacağını iddia etmek mümkün değildir. Bu nedenle, Türkiye'de kurulacak olası bir sistemde, öğrencinin akademik başarısını etkileyen demografik değişkenlerin hesaba katıldığı katma değer (value-added) modellerinin kullanılması daha doğru bir tercih olacaktır. Bu modellerde hesaba katılacak değişkenlerin listesi, ancak tartışmaya açıldıktan ve ölçme değerlendirme sisteminin yaptırımlarından etkileneceklerin onayları alındıktan sonra belirlenmelidir. Öğrenci düzeyinde ekonomik durum, okul düzeyinde ise ortalama sınıf mevcudu, bu listeye girmeye aday olabilecek değişkenlerdendir. Katma değer modellerinin, istatistiksel olarak işi daha da karmaşıklaştıracağı doğrudur. Bununla birlikte, veri analizine kadarki aşamalarda, istatistiksel sınırlamaları gözetmeksizin en doğru kararı vermeye çalışmak; veri analizinde ise, sağlıklı olarak kullanılabilecek yöntemlerden istenilen hedefleri en iyi şekilde gerçekleştirebilecek olanları seçmek, en doğru yaklaşım olacaktır. Aksi halde, amaç, araca feda edilmiş olacaktır. Özet olarak, süreç için ve süreç sonu için gerçekçi hedeflerin konulması, reformun başarı olasılığını artıracaktır.

Yüksek ama gerçekçi hedefler konulmadığında, iki durumdan biriyle karşılaşma olasılığı vardır. Birinci olası durum, hedef olarak tanımlanamayacak kadar düşük düzeyde çıktılar belirlenip bunlara ulaşıldığında, reformun istenileni yaptığı yanlış sonucuna varıp aldanmadır. İkinci olası durum ise, ulaşılamayacak kadar yüksek, dolayısıyla gerçekçi olmayan hedefler konulduğunda, reform süreci boyunca durmadan ödün verip hedef küçültme; böylelikle hedefleri, zaten ulaşılmış başarı düzeylerine çekme ve başarıya ulaşıldığı yanlış sonucuna varıp aldanmadır. Kurulacak sağlam bir ölçme değerlendirme sisteminin yararlı olması da,

hedeflerin doğru bir şekilde belirlenmesiyle yakından ilgilidir. Sonuçta, böyle bir sistem, kendisiyle ölçülmek istenen hedeflere anlam ve değer kazandıran bir mekanizma değil, bu hedefleri ölçüp istenen veriyi sunan bir araç olacaktır.

**Kaynakça**


Berliner, D.C., & Biddle, B.J. (1995). *The manufactured crisis*. New York: Longman.

Center for Public Education. (n.d.). *NCLB pilot state growth model summaries.* Retrieved May19, 2014, from http://www.centerforpubliceducation.org/Main-Menu/Policies/Measuring-student-growth-At-a-glance/NCLB-Pilot-State-Growth-Model-Summaries.html

Center on Education Policy. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington, DC: Author.

Darling-Hammond, L. (2007, May 21). Evaluating No Child Left Behind. *The Nation*, p. 11, 13, 14, 16, 18.

Erpenbach, W.J., Forte-Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB.* Washington, DC: Council of Chief State School Officers.

Lohman, J. (2010). *Comparing No Child Left Behind and Race to the Top.* Retrieved May19, 2014, from http://www.cga.ct.gov/2010/rpt/2010-R-0235.htm

McCall, M. (2007, December). *Vertical scaling and the development of skills*. Paper presented at the WERA/OSPI State Assessment Conference, SeaTac, WA.

Missouri Department of Elementary and Secondary Education. (2011). *Understanding your Adequate Yearly Progress (AYP).*(ERIC Document Reproduction Service No. ED 541 038). Retrieved May19, 2014, from http://files.eric.ed.gov/fulltext/ED541038.pdf

National Center for Education Statistics. (2010). *Mapping state proficiency standards onto NAEP scales: 2005-2007.* Retrieved May19, 2014, from http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf

National Center for Education Statistics. (2011). *Mapping state proficiency standards onto the NAEP scales: Variation and change in state standards for reading and mathematics, 2005-2009.* Retrieved May19, 2014, from http://nces.ed.gov/nationsreportcard/pdf/studies/2011458.pdf

National Center for Education Statistics. (n.d.a). *National Assessment of Educational Progress (NAEP): Frequently asked questions.* Retrieved May19, 2014, from http://nces.ed.gov/nationsreportcard/faq.asp

National Center for Education Statistics. (n.d.b). *National Assessment of Educational Progress (NAEP): Important aspects of No Child Left Behind relevant to NAEP.* Retrieved May19, 2014, from http://nces.ed.gov/nationsreportcard/nclb.asp

National Center for Education Statistics. (n.d.c). *National Assessment of Educational Progress (NAEP): How states join.* Retrieved May19, 2014, from http://nces.ed.gov/nationsreportcard/about/statejoin.asp

O'Malley, K. (2008). *Review of student growth models used by states*. Retrieved May19, 2014, from http://www.pearsonassessments.com/NR/rdonlyres/BBBD784B-B0BC-4B0B-ACB2-7069F1239BB3/0/StudentGrowthModels_0608.pdf

Patz, R.J. (2007). *Vertical scaling in standards-based educational assessment and accountability systems.* Washington, DC: Council of Chief State School Officers.

Riddle, W.C. (2008). *Adequate Yearly Progress (AYP): Growth models under the No Child Left Behind Act* (Congressional Research Service Report RL33032). Washington, DC: Library of Congress, Congressional Research Service.

Rotherham, A.J. (2006). *Making the cut: How states set passing scores on standardized tests*. Washington, DC: Education Sector.

Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195152968.001.0001

Toch, T. (2006). Turmoil in the testing industry. *Educational Leadership, 64*(3), 53-57.

U.S. Department of Education. (2003). *Report cards, Title I, Part A: Non-regulatory guidance.* Retrieved May19, 2014, from http://www2.ed.gov/programs/titleiparta/reportcardsguidance.doc

U.S. Department of Education. (2005). *No Child Left Behind: A road map for state implementation.* Retrieved May19, 2014, from http://www2.ed.gov/admins/lead/account/roadmap/roadmap.pdf

U.S. Department of Education. (2009). *Evaluation of the National Assessment of Educational Progress: Final report.* Retrieved May19, 2014, from http://www2.ed.gov/rschstat/eval/other/naep/naep.pdf

U.S. Department of Education. (2011a). *Final report on the evaluation of the growth model pilot project.* Retrieved May19, 2014, from http://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp-final.pdf

U.S. Department of Education. (2011b). *Duncan says 82 percent of America's schools could "fail" under NCLB this year.* Retrieved May19, 2014, from http://www.ed.gov/news/press-releases/duncan-says-82-percent-americas-schools-could-fail-under-nclb-year

U.S. Department of Education. (n.d.a). *Fact sheet on the major provisions of the Conference Report to H.R. 1, the No Child Left Behind Act.* Retrieved May19, 2014, from http://www2.ed.gov/nclb/overview/intro/factsheet.html

U.S. Department of Education. (n.d.b). *Testing: Frequently asked questions.* Retrieved May19, 2014, from http://www2.ed.gov/nclb/accountability/ayp/testing-faq.html

U.S. Department of Education. (n.d.c). *Growth models.* Retrieved May19, 2014, from http://www2.ed.gov/admins/lead/account/growthmodel/index.html

U.S. Department of Education. (n.d.d). *The standards and assessments peer review: Content standards.* Retrieved May19, 2014, from http://www2.ed.gov/admins/lead/account/peerreview/requirements_pg3.html#content

U.S. Department of Education. (n.d.e). *Policy overview.* Retrieved May19, 2014, from http://www2.ed.gov/policy/landing.jhtml

Welner, K.G. (2008). The overselling of growth modeling. *School Administrator, 65*(6), 6.

Yell, M.L. & Drasgow, E. (2009). *No Child Left Behind: A guide for professionals* (2nded.). Boston: Pearson.

Yen, W.M. (2009). *Measuring student growth with large-scale assessments in an education accountability system.* Retrieved May19, 2014, from http://www.k12center.org/rsc/pdf/YenReactorSession1.pdf