

Are NCLB's Measures, Incentives, and Improvement Strategies the Right Ones for the Nation's Low-Performing High Schools?

Robert Balfanz, Nettie Legters, Thomas C. West,
and Lisa M. Weber
Johns Hopkins University

This article examines the extent to which adequate yearly progress (AYP) is a valid and reliable indicator of improvement in low-performing high schools. For a random subsample of 202 high schools, the authors investigate the school characteristics and the federal and state policy contexts that influence their AYP status. Logistic regression models reveal that the strongest predictors of AYP status in low-performing high schools are the number of student subgroups for which schools are accountable and their No Child Left Behind improvement status. Analysis of state report card data further paints a confusing landscape in which improving low-performing high schools are sanctioned whereas similar schools showing less improvement are not.

KEYWORDS: high schools, adequate yearly progress, graduation rates, No Child Left Behind

Just more than 2,000, or about 12%, of regular and vocational high schools in the United States produce nearly half of the nation's dropouts. In these high schools, a freshman cohort shrinks 40% or more by senior year, and a majority of students fail to graduate in the standard number of years, if at all. Predominately poor and minority students attend these schools. More than 600 of these high schools educate only minority students. Nearly half of the nation's African American and 40% of Latino students attend one of these high schools where graduation is little more than a 50–50 proposition (Balfanz & Legters, 2004).

The No Child Left Behind Act of 2001 (NCLB) is designed to identify such schools—schools that are consistently failing to serve poor and minority students—and to instigate school-based and systemic remedies so that all students are provided with access to a high-quality, standards-based education. At the high school level, the intent of NCLB is to identify high schools

where students are not achieving proficient levels of academic skills and/or graduating with a regular high school diploma in the standard number of years. If schools fail to make adequate yearly progress (AYP) toward proficiency and graduation goals within the framework established by their states, NCLB requires schools and districts to take action to improve the schools and provide students with access to enhanced or alternative educational options.

AYP toward academic and graduation goals is the central measure of success or failure for high schools under NCLB. Accordingly, achieving AYP has become the dominant goal of high school reform in low-performing high schools. High schools that make AYP are viewed as on track and succeeding. Those that do not are viewed as struggling, in need of additional or different reforms, or outright failures that need new principals and staff or even that need to be closed. What is unclear is if these perceptions are accurate. Are low-performing high schools that are making AYP improving more than those that are not? Are the school improvement actions being taken or not taken based on AYP results justified by low-performing high schools' actual levels of performance and rates of improvement?

Emerging literature on NCLB points to weaknesses in the law, both in concept and implementation, that call into question its ability to effectively identify and stimulate improvement in low-performing schools, especially at the high school level (see Commission on No Child Left Behind, 2007; Forum on Educational Accountability, 2007; also see *Harvard Education Review*, vol. 76, no. 4). Though a recent study reports overall gains in student

ROBERT BALFANZ is a research scientist at the Center for Social Organization of Schools, Johns Hopkins University, 3003 N. Charles St., Suite 200, Baltimore, MD 21218; e-mail: rbalfanz@csos.jhu.edu. His research interests focus on secondary school reform, understanding the causes and consequences of dropping out of school, and developing tools and strategies to increase the nation's high school graduation rate.

NETTIE LEGTERS is a research scientist at the Center for Social Organization of Schools, Johns Hopkins University, 3003 N. Charles St., Suite 200, Baltimore, MD 21218; e-mail: nlegters@csos.jhu.edu. Her research interests focus on school organization, teachers' work, and implementation, impact, and scale-up of secondary school reform.

THOMAS C. WEST is a research associate at the Center for Social Organization of Schools, Johns Hopkins University, 3003 N. Charles St., Suite 200, Baltimore, MD 21218; e-mail: twest@csos.jhu.edu. His research interests focus on secondary school reform, standardization of educational outcome measures, and understanding how race, class, and gender intersect to shape the experiences of students.

LISA M. WEBER is a former research assistant at the Center for Social Organization of Schools, Johns Hopkins University, and is a research associate with Mathew Greenwald & Associates, 4201 Connecticut Ave. NW, Suite 620, Washington, DC 20008; e-mail: lisa.weber@greenwaldresearch.com. Her research interests focus on health and health care, aging and the life course, and the impact of workplace and government policy on each.

achievement and a narrowing of achievement gaps since NCLB was enacted in 2002, more states showed declines in reading and math achievement at the high school level than at the elementary and middle school levels. Study authors also note that achievement gains cannot be directly attributed to NCLB and may result from a combination of increased learning and other factors such as teaching to the test, more lenient tests, scoring or data analyses, and changes in the student populations tested (Center on Education Policy, 2007).

One recent study of NCLB in six states hones in on how inconsistent implementation among states makes it difficult to discern whether schools achieve AYP because they actually improve or because improvement targets and proficiency standards are more easily met in their state. In a case study of California schools, the researchers also find that the NCLB subgroup accountability requirements concentrate sanctions in socially and economically disadvantaged schools in spite of little evidence that the schools are actually failing to improve student achievement over time (Sunderman, Kim, & Orfield, 2005). Other case study and anecdotal evidence focuses on the NCLB graduation rate accountability requirements and finds that, contrary to the intentions of the law, the lack of standards and subgroup accountability requirements around graduation rates has created loopholes and perverse incentives for schools to make AYP by pushing out or holding back students likely to score low on academic proficiency tests (Darling-Hammond, 2006; Losen, 2005; Orfield, Losen, Wald, & Swanson, 2004).

In this article, we extend the research base by using quantitative analysis and case examples to examine the drivers of AYP status in more than 200 low-performing high schools across 34 states. We examine the extent to which AYP, as it is currently constructed at the high school level, is a reliable indicator of improvement in low-performing high schools and if, as designed, it is serving as an effective tool for high school reform. We then discuss our findings with an eye on providing guidance for policy makers to improve NCLB as the law moves toward reauthorization in the coming months.

Research Design

Questions and Conceptual Framework

Is NCLB achieving its goals of accurately identifying and stimulating improvement in low-performing high schools? Is AYP, as the central accountability measure under NCLB, a valid and reliable indicator of improvement in low-performing high schools? Is it serving as an effective tool for high school reform?

This study addresses these questions by exploring three hypothetical sources of influence that could determine whether a low-performing high school makes AYP (Figure 1). One hypothesis is that NCLB is working as intended (i.e., it quantifies school performance and improvement in a valid and reliable manner and propels schools to improve through a series of effective

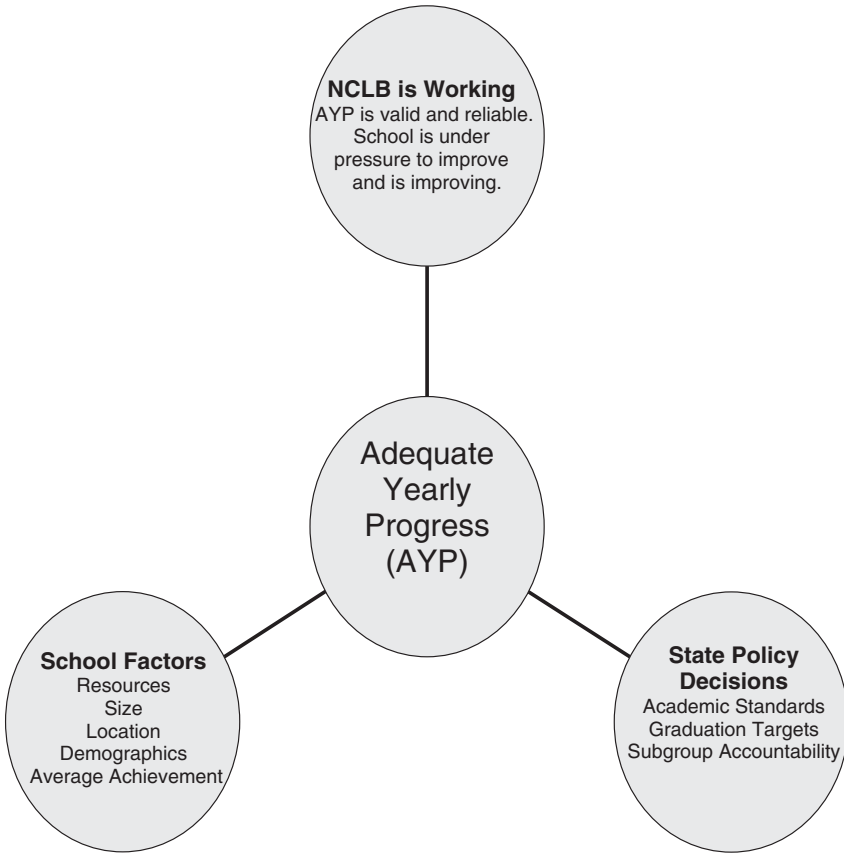


Figure 1. Hypothesized determinants of whether a low-performing high school makes adequate yearly progress (AYP). NCLB = No Child Left Behind.

supports and sanctions). States, however, have considerable latitude in setting academic standards and improvement rates, establishing graduation rate goals and improvement rates, and establishing baselines for the number of students needed to activate subgroup accountability. All of these state-level policy decisions influence how difficult it is for a school to achieve AYP. Finally, high schools themselves considerably differ in their resources, size, location, and student body composition, all of which likely influence whether they will meet the standards or be able to make the improvements required to make AYP.

The intent of NCLB is that it will galvanize schools to perform better. AYP is a valid and reliable measure of progress toward this goal if, in the main, schools that make AYP do so because they have become more effective. If the reason why some high schools make AYP and others do not is more

a function of the school's characteristics—its size, resources, or students—then, as constructed, AYP becomes less valid as a measure of school improvement. It may provide important policy information (e.g., small schools or integrated schools do better), but it would be indicating primarily that some schools are better positioned than others to improve. If differences in how NCLB is being implemented across states turn out to be a dominant factor in determining which high schools make AYP, then AYP becomes neither a valid nor a reliable indicator of school improvement.

Simply put, if the degree to which high schools make AYP is determined by how hard or easy a state has made it to achieve in a given year, then we cannot say that a school that has made AYP in one state is improving more than a school that has not made AYP in another state or that a school that has not made AYP is not improving. With only a weak relationship between AYP and improvement, it also is possible that, as constructed at the high school level, AYP is functioning counter to the intent of NCLB. Instead of focusing schools' attention on key improvement variables, the quest for AYP could focus the reform efforts and attentions of low-performing high schools either too narrowly or even in the wrong direction.

Research Plan

To investigate these questions and hypotheses, we use quantitative data to identify systematic ways in which low-performing high schools that make AYP differ from low-performing high schools that do not make AYP. Descriptive and multivariate analyses shed light on the relationship and relative influence of federal, state, and school factors on a school's AYP status. We then conduct more detailed case studies of a subset of the low-performing high schools in our sample to more fully address the extent to which it is possible to say that low-performing high schools that make AYP are performing better or improving faster than are low-performing high schools that are not making AYP.

Method

Identifying Low-Performing High Schools

As described above, we focus our study on the extent to which NCLB effectively identifies and stimulates improvement in low-performing high schools. Doing so required a measure of "low performing" independent of NCLB designations obtained through AYP. We also could not use graduation or dropout rates, given the high degree of variability within and among states in measuring these outcomes and the large amount of missing data related to these outcomes in national data. Hence, to identify low-performing high schools, we used a measure of schools' promoting power.

As the ratio of 12th graders to 9th graders 3 years prior, promoting power estimates the success with which a high school achieves on-time promotion

of a cohort of students from their freshmen to their senior year. Based on widely available enrollment data, promoting power provides a good estimate of a school's dropout rate and a school's capacity to keep students on track to graduation (Balfanz & Legters, 2004, 2005). We identified schools as low performing if they displayed chronically weak promoting power (i.e., if they promote 60% or fewer freshmen to senior status on time averaged across three successive cohorts of students). This enabled us to identify a set of schools that were clearly low performing at the outset of NCLB and to investigate the extent to which they have been identified and improved by the NCLB accountability framework.

Data and Measures

This study draws from three data sources. The Common Core of Data (CCD), the U.S. Department of Education's census of all schools in the United States, collected by the National Center for Educational Statistics, provided all student enrollment data and other school data (type, demographics, number of teachers, free and reduced lunch participation, and Title I eligibility) used in the analyses (see Table A1 for detail on variables and measures). CCD enrollment data also were used to construct the measure of promoting power that identifies high schools as low performing. Promoting power is defined as the number of seniors enrolled in Year X divided by the number of freshmen enrolled in Year X-3 (or Year X-2 in a 10th to 12th grade senior high school).

The study also draws from state and district report cards for the 2003–2004 and 2004–2005 school years for the 202 high schools in the sample. These data provided information about states' AYP requirements and about schools' academic performance, graduation rates, AYP status, whether AYP was met through safe harbor provisions, subgroup breakdowns for AYP, and NCLB improvement status (e.g., in improvement, corrective action, restructured or restructuring).¹ Because these data are not yet centralized, we constructed statistical profiles for each sample school through a thorough scan of data published on official state and district Web sites (see Table A2 for further detail).

A third source of data is the National Assessment of Educational Progress (NAEP) state assessment of eighth graders in math and English collected and profiled by the U.S. Department of Education's National Center for Education Statistics. These data enabled us to evaluate the rigor of state AYP exams by comparing states' performance on the state exams to their performance on the NAEP exams. This comparison is measured by taking the difference between the percentage proficient or basic on the NAEP and the percentage proficient on the NCLB or AYP exam in each state (see Table A3).

Sample

To create the base population of 2,030 low-performing high schools, we filtered the 2003 CCD on high schools (defined as having at least 10th to 12th

grades), school type (regular or vocational), enrollment (100 students or more), and a 3-year promoting power average of 60% or less (averaged across the classes of 2000–2001, 2001–2002, and 2002–2003).

We then drew a 10% random sample of the 2,030 high schools, resulting in a total of 202 high schools for our analysis sample.² The 202 high schools in the sample reflect the main characteristics of the nation's 2,030 high schools with weak promoting power. They are predominately located in cities in the Northeast, industrial Midwest, and West and throughout the South and Southwest. Included are high schools from the nation's largest urban school districts and rural, single-high school districts. The sample includes high schools from 34 states, with slightly more than half of the high schools located in 11 Southern states (110 of 202). Texas has the most high schools in the sample (22), followed by Florida (19), Georgia (15), New York (14), and California (14). Eleven high schools are in New York City, the district with the most high schools in the sample. Overall, the schools in the 10% random sample of schools are marginally more rural and less suburban than the base population of 2,030 low-performing high schools and serve a slightly higher proportion of students from poverty backgrounds (as measured by free and reduced-price lunch eligibility). These differences are small, however, and the two groups do not significantly differ on other salient variables (type, size, percentage minority, resources). Statistical comparisons between the base population and the analysis sample can be found in Table A4.³

Method

We conducted a thorough scan of state report card data to construct a database with AYP information on each of the 202 schools in our sample and merged that with our CCD data. We used basic cross-tabulation and correlation analysis to determine how many low-performing high schools in the sample did and did not make AYP and to explore features of schools or state policies that appeared to influence that outcome. We then used logistic regression to determine the relative influence of school and state variables on making AYP. Finally, deeper analysis of state report card data and state and district Web sites enabled us to construct comparative case examples of selected schools to assess AYP as an indicator and stimulus of school improvement in low-performing high schools.

Findings

Which Low-Performing High Schools Made AYP in 2005?

In 2005 the majority of low-performing high schools in our sample did not make AYP. Success in making AYP also considerably varied by state, irrespective of achievement gains. Of the low-performing high schools, 41% made AYP in 2004–2005, and 59% did not make AYP. This was nearly

Table 1
**Poverty Levels, Resources, Enrollment, and
 Locale of Low-Performing High Schools that Did and
 Did Not Make Adequate Yearly Progress (AYP)**

School Characteristic	Made AYP	Did Not Make AYP
# of schools (<i>n</i>)	82	120**
% free lunch	53	57
% school Title I	47	47
Average pupil–teacher ratio	15.85	17.18*
Average total enrollment	901	1,426**
% urban	30	52**
% rural	43	21**

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003, and data on AYP status were taken from state report cards.

Note. $N = 202$.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed (indicates means and percentages are significantly different than schools that made AYP).

identical to the prior year (2003–2004), when 39% of the low-performing high schools made AYP.

Descriptive analysis revealed variations across states in the rates at which low-performing high schools made AYP. In six states (Colorado, Louisiana, Massachusetts, Mississippi, Virginia, and Wisconsin), 75% or more of the low-performing high schools in the sample made AYP in 2004–2005. By contrast, in 18 states, one fourth or fewer of the low-performing high schools made AYP. Moreover, there is a strong regional bias. Half of the low-performing high schools in the South made AYP, compared to about one third of the low-performing high schools in the North and West.

The clustering of low-performing high schools that made AYP within a subset of states and one region could indicate that either school characteristics or successful high school reform efforts vary by state and region. Clustering could also mean that state policy differences in NCLB implementation are playing a role in determining which low-performing high schools make AYP. One indication of this is that differences in achievement gains do not seem to be the driving factor in why some low-performing high schools are making AYP and others are not. Of low-performing high schools that reported gains in the percentage of their students achieving proficiency in mathematics or English (or had their proficiency levels stay the same), 44% made AYP. Yet nearly the same rate (38%) of the low-performing high schools that reported *declines* in either mathematics or English proficiency levels also made AYP.

Table 2
Comparison of Adequate Yearly Progress (AYP) Status for Types of Schools Based on Size, Percentage Minority, and Percentage Free or Reduced Lunch (FRL)

Type of School	# Meeting AYP	# Failing to Meet AYP	Total
Type I	11	43	54
More than 1,200 students			
More than 50% minority			
More than 40% FRL			
Type II	12	6	18
Less than 1,200 students			
Less than 50% minority			
Less than 40% FRL			
Type III	59	71	130
Neither Type I nor Type II			
Total number of schools	82	120	202

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003, and data on AYP status were taken from state report cards.

Note. $N = 202$. $\chi^2(2) = 15.47$. $p < .001$.

Table 3
School Characteristics by the Largest Racial/Ethnic Group in the School

School Characteristic	White ^a	African American ^b	Hispanic or Latino ^c	Significant Difference
% Title I eligible (%)	27	54	75	***
Average enrollment	917.42	1,102.07	2,030.73	ac***, bc***
Average student–teacher ratio	15.76	16.88	18.23	
Average percentage minority (%)	24	87	88	ab***, ac***
Average percentage free or reduced lunch eligible	43	61	67	ab***, ac***
Location of school ^d				
Urban (%)	19	62	63	
Suburban (%)	33	25	18	
Rural (%)	48	13	20	
Total number of schools	79	76	40	
% making adequate yearly progress	53.16	34.21	32.50	

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003, and data on AYP status were taken from state report cards.

Note. $N = 202$.

a. $\chi^2(4) = 44.24$, $p < .001$.

^b $p < .10$, two-tailed. ^{*} $p < .05$, two-tailed. ^{**} $p < .01$, two-tailed. ^{***} $p < .001$, two-tailed.

Table 4
**Descriptive Statistics for Number of Subgroups Required
 to Make Adequate Yearly Progress (AYP)**

	Met AYP		Did Not Meet AYP	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of subgroups to meet	3.24	1.44	4.39	1.57
Total number of schools	82		120	

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003, and data on AYP status were taken from state report cards.

Note. $N = 202$. $t(200) = 5.272$, $p < .001$.

School Characteristics

As shown in Table 1, the low-performing high schools that did and did not make AYP share some common characteristics. On average, they have similar poverty levels and qualify for Title I funds at similar rates.

There are three school-level characteristics, however, that distinguish low-performing high schools that made AYP from those that did not. First, high schools that made AYP appear to be better resourced. On average, they have one teacher for every 15.8 students, compared to one teacher for every 17.2 students in the schools that did not make AYP. This difference is considerable. In a school of 1,500 students, it would amount to eight additional teachers, or a staffing increase of 9%. It is also potentially significant. Recent research indicates that high schools with student–teacher ratios of 15 to 1 or less are much more likely to have the resources necessary to implement comprehensive, research-based high school reforms than are high schools with greater ratios (Balfanz, 2006).

The second major difference between the low-performing high schools that made AYP and those that did not is size. On average, those that made AYP are considerably smaller, with 525 fewer students. Finally, the low-performing high schools that made AYP are more likely to be rural high schools and less likely to be central city high schools. In particular, low-performing high schools in the rural South made AYP at high rates (70% of 41 schools).

If smaller, rural, and comparatively better-resourced schools did better in making AYP, the opposite appears to be true for large, majority-minority, high-poverty low-performing high schools. Table 2 shows that only 11 of the 54 large, majority-minority, high-poverty schools in the sample made AYP (20%), compared to 71 of the 148 schools that did not match these criteria (48%).

Schools that have Blacks or Hispanics as the largest racial/ethnic group make AYP about one third of the time, whereas schools with Whites as the majority group make AYP 53% of the time (Table 3). When Black or Hispanic students are the majority in a low-performing high school, the high school is

Table 5
**Adequate Yearly Progress (AYP) Status by Racial/Ethnic
 Minority Subgroup Composition of the School**

Subgroup Composition	# of Schools	% of Total Sample	% Met AYP	% Did Not Meet AYP
Had to meet AYP for at least one racial/ethnic minority	155	76.73	34.19	65.81
Black	103	50.99	36.89	63.11
Hispanic	42	20.79	33.33	66.67
American Indian or Native Alaskan	4	1.98	25.00	75.00
Asian or Pacific Islander	3	1.49	0.00	100.00
Did not have to meet AYP for any racial/ethnic minority	44	21.78	61.36	38.64
Missing data on race/ethnicity breakdown of AYP	6	2.97	66.67	33.33

Source. State report cards.

Note. $\chi^2(5) = 11.96, p < .05$.

also typically large (enrolling more than 1,100 students), more than 85% minority, serves a high-poverty population (more than 60% free lunch), is commonly in an urban area (62%), and has higher student-teacher ratios (16.9 and above). When White students are the majority in a low-performing high school, the school is smaller (enrollments less than 920), is less than 25% minority, serves a population with a lower concentration of poverty (43% free lunch), has a lower student-teacher ratio (15.8), and is most commonly in a rural (48%) or suburban (33%) area.

State Differences in Implementing NCLB

How difficult it is for a low-performing high school to make AYP is determined by several factors that are influenced by state policy differences in how NCLB is implemented.

Low-performing high schools that made AYP have less subgroup accountability. One factor is the number of subgroups that must make AYP. Here, school characteristics and state policy choices in implementing NCLB intersect. One of the central features of NCLB and AYP is that schools have to show not only continual overall progress toward proficiency for all students but also progress for ethnic/racial and economic subgroups as well (e.g., African American, Latino, special education, economically disadvantaged students, etc.). States have some latitude, however, in establishing how many students in a school need to belong to a subgroup for a school to be responsible for improving the achievement and graduation rates of the subgroup. In our sample, states required from 30 to 50 students in a subgroup for that subgroup to

be counted in a school's AYP calculation. These relatively small state differences can have large impacts on schools when they intersect with state and urban–rural differences in the size of schools and the racial/ethnic and economic composition of student populations. Large and diverse schools may have eight or more subgroups that need to make AYP, whereas smaller schools can have half as many. Thus, the apparent AYP advantage of fewer students in a high school may, in part, reflect the fact that high schools with fewer students may face less subgroup accountability. Table 4 shows that, on average, schools that made AYP had to do so for 25% fewer subgroups.

The AYP advantage of fewer subgroups can be vividly seen in Table 5. Low-performing high schools without racial/ethnic subgroups made AYP 61% of the time. Schools with at least one subgroup that had to meet AYP made it only 34% of the time. More than half of the low-performing high schools without racial/ethnic subgroup accountability are rural, and nearly all have less than 500 students. Of the 15 rural high schools in the South that did not have to meet AYP for any racial/ethnic minorities, 87% made AYP.

Low-performing high schools that did not make AYP faced higher levels of NCLB sanctions and interventions. NCLB is designed so that each year a school does not meet AYP there will be higher levels of scrutiny, interventions, and, ultimately, sanctions. This is, in part, based on the theory that the higher the level of intervention and sanction a school faces, the more it will be motivated to do whatever is necessary and possible to meet AYP. Consequently, schools facing greater consequences for not making AYP might be more likely to achieve it than schools for which the intervention and sanctions are small. An alternative view is that schools that have repeatedly failed to make AYP have done so in part because they lack the capacity to improve themselves or because the level of improvement required to make AYP is greater than even a school that puts forth effort and implements effective reforms can achieve. This second theory predicts that schools facing high levels of NCLB sanctions and intervention might be less likely to achieve AYP.

To test these competing theories, we used each low-performing school's NCLB improvement status in 2004–2005 to classify schools into a three-category scale, indicating minor, moderate, and major levels of NCLB improvement pressure. The minor NCLB improvement pressure category encompasses schools that are not in any sort of NCLB improvement status. The moderate NCLB improvement pressure category includes schools that are at risk or under watch for improvement (meaning they are on alert and that they could be placed in Improvement Year 1 in 2005–2006), schools that are making progress (schools that were in improvement in 2003–2004 but that are improving), and, finally, schools that are in Year 1 Improvement. The major NCLB improvement pressure category includes schools that are in their 2nd through 7th years of improvement status, schools that are under corrective action, and schools that are restructured, restructuring, or planning to restructure.

Table 6
**Comparison of Adequate Yearly Progress (AYP) Status
 by No Child Left Behind Act (NCLB) Improvement Sanctions**

	# of Schools	% of Total Sample	% Who Met AYP	% Who Did Not Meet AYP
Minor	100	50.00	66.00	34.00
Not in any improvement				
Moderate	36	18.00	36.11	63.89
At risk or under watch	7	3.50		
Improved or making progress	4	2.00		
In improvement Year 1	23	11.50		
In improvement, unspecified ^a	2	1.00		
Major	64	32.00	4.69	95.31
In improvement Year 2	17	8.50		
In improvement Years 3 to 7	14	7.00		
In improvement, unspecified ^a	18	9.00		
Corrective action	12	6.00		
Restructuring, restructured, or plan to restructure	3	1.50		

Source. State report cards.

Note. $N = 200$. Two schools from Iowa were missing NCLB improvement status. $\chi^2(2) = 61.079$, $p < .001$.

a. Schools in Improvement Unspecified were recoded, such that schools making AYP in 2003–2004 or 2002–2003 are in moderate and schools not making AYP in 2003–2004 and 2002–2003 are in major

The NCLB improvement status of a school in our sample in 2004–2005 was not determined only by its absolute level of performance. State policy choices in implementing NCLB also played a significant role. High schools with the same absolute level of performance found themselves in very different improvement statuses. These depended in part on if their state grandfathered in existing state accountability systems (if a school was found to have been low performing under state accountability systems for several years before NCLB, it could begin NCLB already in a needs improvement category), how low or high a state set the baseline achievement levels for AYP regardless of any improvement schools might have demonstrated (each state sets a baseline achievement level from which incremental progress toward proficiency is measured; if a school does not reach the baseline it does not make AYP, even if it has made improvement toward the baseline, unless the gains are large enough to trigger the safe harbor provision), and, finally, how much improvement the state requires in a given year to reach AYP (states set their own paths to proficiency, with some selecting even incremental gains each year and others expecting smaller gains in the initial years and larger gains in the later years).

In this sample, half of the schools are in the minor improvement category, 18% are in the moderate category, and 32% are in the major category

Table 7

Descriptive Statistics for State Standards and Performance on Exams

	Met Adequate Yearly Progress (AYP)		Did Not Meet AYP	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Required math proficiency level in 2004–2005	45.47	16.58	44.39	16.44
Required English proficiency level in 2004–2005	54.39	20.38	48.86	19.54
State-level eighth grade math achievement				
% proficient on No Child Left Behind exam	54.71	20.84	53.01	19.08
% proficient on National Assessment of Educational Progress (NAEP) exam	24.61	6.08	25.18	5.53
% basic on NAEP exam	64.34	7.56	64.08	7.34
Total number of schools	82		120	

Source. U.S. Department of Education NAEP State Profiles; State Report Cards.

Note. *N* = 202. No significant differences between groups.

(see Table 6). Current NCLB improvement status is inversely related to high schools making AYP. Of the low-performing high schools in the minor improvement pressure category, 66% made AYP, compared to only 5% of the high schools facing the highest levels of intervention or sanction. Equally telling is the fact that 80% of the schools in the sample that made AYP in 2004–2005 are in the minor improvement pressure category, whereas slightly more than half of the schools that did make AYP are in the major improvement category. In sum, low-performing high schools facing the lowest levels of NCLB improvement pressure more often than not made AYP, whereas only 3 of the 64 schools facing the strongest interventions or sanctions were able to make AYP in 2004–2005.

Low-performing high schools that made AYP are concentrated in states in which it appears easier to reach the required proficiency levels. Because states establish their own performance standards, design their own assessments, and establish the pace at which students must improve to reach 100% proficiency, the difficulty of reaching NCLB proficiency goals in a given year varies considerably from state to state. The ability of a school to reach NCLB proficiency goals can be influenced by at least four factors.

The first is the percentage of students required to demonstrate proficiency in a given year. In 2004–2005, the math proficiency levels required by the 34 states in our sample greatly varied, from a low of 17.5% in Missouri to a high of 75% in Tennessee. However, where states set their proficiency goal in 2004–2005 does not appear to be a determining factor in whether a low-performing high school made AYP (Table 7). In fact, low-performing high schools that made AYP faced slightly higher proficiency standards than did schools that did not make AYP.

A second factor is the academic skill of its entering students. It is not possible, however, to measure the entering academic skills of the high school students attending the low-performing high schools in our sample. What are available are data on the overall level of eighth grade achievement in the state. The percentage of students in a state scoring at proficient or basic on the most recent eighth grade NAEP mathematics exam did not affect the likelihood that a school would make AYP, nor did the percentage proficient on the 2003–2004 eighth grade NCLB mathematics exam. As seen in Table 7, the statewide eighth grade performance on NCLB and NAEP mathematics exams is about equal, on average, for low-performing high schools that did and did not make AYP.

Although neither the percentage of students needing to demonstrate proficiency in mathematics in 2004–2005 nor the overall level of mathematics achievement among a state's eighth graders appear to significantly influence whether a low-performing high school makes AYP, it is still possible that these two factors considered together would be significant. To examine this third possibility, we compared the percentage of eighth graders who scored proficient and above on a state's NCLB mathematics exam in 2003–2004, with the percentage of high school students required to obtain proficient levels in mathematics in 2004–2005 for a high school to make AYP. In other words, what may be predictive is where the proficiency level needed for high schools to make AYP is set relative to the proficiency of students in eighth grade. We found that 60% of the Southern schools that made AYP were in states in which the 2004–2005 high school mathematics proficiency goal was at least 10 percentage points lower than the percentage of eighth graders scoring proficient on the NCLB exam in 2003–2004. By comparison, only 33% of the Northern and 8% of the Western low-performing high schools that made AYP were in a state with similar conditions. This indicates that the high rate at which low-performing high schools in the South made AYP may, in part, be related to the fact that proficiency goals were more often below existing proficiency levels.

A fourth factor that can influence a high school's ability to reach NCLB proficiency levels is the difficulty of the test used to establish proficiency in mathematics and English. We examine this by comparing the percentage of eighth grade students scoring at proficient levels on their state NCLB mathematics exam to the percentage of students scoring at proficient and at basic on the most recent eighth grade NAEP exam. When comparing the percentage proficient on the NCLB exam to the percentage proficient on the NAEP exam, only 5 of the low-performing high schools in our sample are located in states (Missouri) where the NCLB tests appear to be more difficult than the NAEP exam. Of the schools, 42 come from states in which the NCLB and NAEP exams appear to be of similar rigor. However, 149 of the schools are in states in which the NCLB exam appears to be easier than the NAEP exam. In these states, the percentage of students who are proficient on the eighth grade NCLB math exam is more than 10 percentage points greater than the percentage of students who are proficient on the NAEP eighth grade math exam.

Table 8
**Odds Ratios From Logistic Regression of Selected Variables
 on Adequate Yearly Progress (AYP) Status in 2004–2005;
 Explaining Rural Advantage**

	Model I	Model II
Urban (vs. rural)	0.29**	0.37
Suburban (vs. rural)	0.48 [†]	0.44
How challenging is it to make AYP?		
# subgroups to meet	—	0.62*
Math standard	—	0.98
English language arts (ELA) standard	—	1.03
Math: National Assessment of Educational Progress basic vs. No Child Left Behind proficient	—	1.77
What are the student body characteristics?		
Total enrollment	—	1.02
% free or reduced lunch	—	0.94
% minority	—	0.99
What is the resource level?		
Student–teacher ratio	—	0.97
Promoting power	—	1.04
Historical performance or incentive to improve?		
Moderate improve	—	0.44
Major improve	—	0.04**
<i>N</i>	202	168
Pseudo <i>R</i> ²	.047	.362
χ^2	12.95**	82.29**

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003; data on ELA standards, math standards, and AYP status were taken from state report cards.

[†]*p* < .10, two-tailed. **p* < .05, two-tailed. ***p* < .01, two-tailed.

There also is a clear association between NCLB test difficulty in the eighth grade and the likelihood that low-performing high schools will make AYP. The percentage of low-performing high schools making AYP increases as the NCLB eighth grade mathematics exam gets easier. Missouri, which appears to have the most rigorous exam, seems to be an anomaly in this trend, but it also has the lowest proficiency target in 2004–2005 (17.5%).

Overall, there are clear indications that low-performing high schools that made AYP tended to be in states where high school proficiency goals were substantially below existing middle grade proficiency levels and/or the NCLB proficiency exams themselves were less rigorous.

Which School Characteristics and State NCLB Implementation Choices Affect the Chances That a Low-Performing High School Will Make AYP?

Low-performing high schools that made AYP are distinctive. As seen in the differences elucidated above, they are smaller, are better resourced, have

smaller concentrations of minority students, are predominately located in Southern states and/or rural regions of the nation, face less subgroup accountability and NCLB improvement pressures, and tend to be in states in which it appeared easier to reach required proficiency levels in 2004–2005. What we do not know is which, if any, of these factors are causative. In other words, just because these factors are associated with low-performing high schools that make AYP does not mean that they are the reasons the schools are making AYP. They may simply co-occur with other identified or unidentified factors that are significant.

Rural high schools, for example, tend to be smaller, and smaller high schools can have lower student–teacher ratios. Hence, it may be any one of the factors that is important, or all three of them may be linked to another factor that is actually promoting the success of some high schools over others. These factors could include less subgroup accountability, easier state tests, and differences in the educational challenges that the low-performing high schools in different states face. As a result, it may be that state policy differences in NCLB implementation, and not differences in school characteristics, ultimately determine why some low-performing high schools are making AYP and others are not. Ideally, neither school characteristics nor differences in how states are implementing NCLB would be important factors in making AYP. The intent of the law is to measure and promote genuine school improvement, through increased academic achievement and graduation rates for all the students in a high school.

To gain more insight into if, and which of, the distinctive characteristics of low-performing high schools that made AYP were significant factors, we examined a series of logistic regression models. Table 8 shows logistic regression results from two of these models. Model I estimates the effect of the location of the school on AYP. Rural schools seem to have an advantage when it comes to making AYP. Suburban schools are 52% less likely to make AYP than are rural schools, whereas urban schools are 71% less likely to make AYP than are rural schools.

Model II shows what happens when other variables that distinguish low-performing high schools that make AYP from those that do not are entered into the model. The rural high schools' advantage in making AYP disappears when differences in subgroup accountability, the difficulty of state exams, the state proficiency levels for meeting AYP, the student body characteristics and student–teacher ratio, and the high school's NCLB improvement status are added to the model. The results show that if urban and suburban schools faced the same situation as rural schools, there would be no significant difference in the chances of making AYP by the location of the school.

Instead of rural location, the important predictors of AYP status become the number of subgroups required to meet AYP and NCLB improvement status. Each additional subgroup lowers the odds of making AYP by 38%. Therefore, schools that face less diversity (fewer subgroups) have an easier

time making AYP. Being in major versus minor NCLB improvement reduces the chances of making AYP by 96%.

Additional models (available on request) systematically examined the impact of the locale, student body characteristics, school resources, and AYP challenge and pressure variables in various combinations and formulations. These models confirmed the central finding: Subgroup accountability and NCLB improvement status were such strong predictors that once they were added to the model none of the other examined variables (all of which have a significant association by themselves) remained significant.

NCLB improvement status, in turn, is affected by several state policy decisions regarding the implementation of NCLB, including the extent to which state accountability systems were “grandfathered in,” systematically placing more high schools under major pressure and scrutiny at the start of NCLB. Most significant, however, is the combination of state decisions on the difficulty of the high school accountability tests, where initial high school baselines were set relative to existing eighth grade achievement levels and the pace of improvement expected in the initial years of NCLB. Together, these influence how far existing levels of performance are from the levels needed to achieve AYP in low-performing high schools. This can be seen in the fact that the distance a high school’s 2003–2004 mathematics and English achievement levels, and their graduation rates, are from the levels needed to achieve AYP in 2004–2005 is a strong predictor of NCLB improvement status.

In short, low-performing high schools with the lowest degrees of AYP difficulty—those that face less subgroup accountability and have to make smaller gains or no gains at all to achieve AYP—tend to make AYP, whereas high schools with higher degrees of difficulty—those that have more subgroups and need to make substantial achievement and graduation gains—seldom make AYP. Which circumstance a low-performing high school finds itself in is not a function simply of its initial performance levels but also of state policy choices in implementing NCLB at the high school level.

The fact that subgroup accountability and NCLB improvement status appear to be the significant predictors of meeting AYP among low-performing high schools undermines the reliability and validity of AYP as a measure of school improvement among these schools. States that set high subgroup minimum counts will likely see more of their low-performing high schools make AYP, as will states with smaller high schools. But when these states report greater numbers of low-performing high schools making AYP, it does not mean that they are succeeding at meeting the intent of NCLB. Rather, they may be more able to dodge it. Their subgroups not included in the NCLB calculations could have declining outcomes, and as long as these were offset by improvements among other students, they would still make AYP.

In addition, given the variation in how states are implementing NCLB, the fact that low-performing high schools facing the highest level of NCLB sanctions and interventions almost never make AYP does not necessarily mean that these high schools are improving less rapidly or at lower rates than

the schools that make AYP. The low-performing high schools facing the highest levels of NCLB sanctions tend to be larger and more diverse (hence more subgroup accountability) and located in states in which it appeared to be tougher to meet proficiency standards in 2004–2005. Thus, two low-performing high schools in different states could have had the same actual level of improvement or nonimprovement during the first 3 years of NCLB and be facing very different interventions and sanctions in 2004–2005. One school could have made AYP each year and in effect been told it was doing fine. The other school could have been facing the highest level of NCLB sanctions because it had never achieved AYP and was in a state that used prior performance on state accountability measures to place schools in NCLB improvement status at the start.

The fact that high schools facing the least NCLB improvement pressure made AYP at relatively high rates is harder to interpret. It could indicate that these schools have been making steady gains throughout NCLB and, as a result, have been making AYP each year. Alternatively, as these schools are clustered in the rural South where they face less subgroup accountability, easier tests, and lower proficiency targets relative to existing proficiency levels, this could indicate that these schools need less improvement to make AYP. Thus, again, because of differences in how states are implementing NCLB, we cannot assume that low-performing schools that consistently make AYP are improving more rapidly or at greater rates than are schools that do not.

In School-Level Case Studies, Does AYP Emerge as a Valid Indicator or Effective Stimulus of School Improvement in Low-Performing High Schools?

Beyond the variables that could be statistically tested, several other state differences in NCLB implementation exist that case studies of the schools in our sample show make it impossible to answer the most basic question: Are low-performing high schools getting better under NCLB? State-to-state differences in the grade in which students are tested, how graduation rates are measured, and where the initial baselines are measured all work to undermine the ability to conclude that a low-performing high school that made AYP is actually performing better or improving faster than a school that did not make AYP.

Two practices stand out.

First, a number of states test in the 11th and 12th grades and have minimal graduation rate levels or gain goals. In these states, high schools with low graduation rates and minimal or no improvement can make AYP by improving the achievement levels of only the students who make it to the 11th or 12th grade.

For example, a New York City high school made AYP in 2004–2005 with seemingly impressive proficiency levels of 72% in math and 80% in English on the state's 12th grade test. This, however, is paired with a 58% cohort

graduation rate and an 81% attendance rate. This indicates that only 58% of the entering freshmen graduated and only 41% graduated proficient in mathematics and 46% in English. Yet for all practical purposes, because this school made AYP, it is being sent the signal that it is doing fine and that it should keep focusing on making students who survive to the 12th grade proficient rather than on the nearly 50% of students who are dropping out with weak academic skills.

A school in Missouri is an even stronger example of how current implementations of NCLB at the high school level can obscure more than they reveal. This school made AYP with proficiency levels of 21% in mathematics and 25% in English, based on modest gains of 8 percentage points in mathematics and 4 percentage points in English. Its graduation rate, however, *declined* 12 percentage points to 77%. The school made AYP, however, because 77% was above the minimum required to meet AYP in Missouri in 2004–2005. The signal being sent to this school is that fewer graduates is okay as long as proficiency levels keep rising on the state tests given in the 10th and 11th grades. This in turn provides a strong incentive to retain students in ninth grade or push them out.

A second practice that plays a large role in making it nearly impossible to use AYP results to determine if the nation's lowest-performing high schools are improving is that each state sets its own baseline from which high schools are supposed to progress until 100% of students demonstrate proficiency on state tests. Some states have set initial baseline pass rates on state tests at 20% or even lower. In California high schools, for example, in 2002–2003 and 2003–2004, only 11% of students had to score proficient in English and 10% in math for a school to make AYP. In 2004–2005, the bar was raised to 22% in English and 21% in math. From 2006–2007 on, the bar will steadily progress toward 100% in 2013–2014. Other states, however, set their initial or early baselines at 40% or higher. In Pennsylvania, for example, proficiency rates of 45% in mathematics and 54% in reading were required to make AYP in 2004–2005. In short, more than twice as many students needed to be proficient in Pennsylvania than in California for a high school to make AYP.

One result of this wide divergence in initial and early baselines is that a low-performing high school in California with stagnant proficiency rates in the low 20s could make AYP yearly from 2002–2003 to 2006–2007. By contrast, there is a high school in Pennsylvania that has seen a 15 percentage point increase in its math proficiency, a 20 percentage point increase in its English proficiency, and a 30 percentage point increase in its graduation rate over 4 years. Yet it is in Corrective Action 2, 1 year away from possibly being turned into a charter or seeing its faculty replaced because, despite these gains, it has not reached the baseline set by the state (and has not consistently met safe harbor provisions each year in each AYP student demographic category).

The same is true for graduation rates. In Georgia, for example, a high school could make AYP in 2004–2005 with a 60% graduation rate. In our sample, eight Georgia high schools made AYP in 2004–2005 and seven did not.

Four of the high schools that made AYP, however, had graduation rates of around 60%; one actually had a rate in the 50s but made AYP through a confidence interval, and another saw its graduation rate decline from 67% to 60%. Georgia is far from alone. In a recent publication, the Education Trust reports that no fewer than 34 states had AYP goals for high schools that were lower than the states' reported graduation rates (Hall, 2005). Moreover, in 37 states, any microscopic (0.1%) gain, or even no progress, enables a high school to make AYP for graduation rates.

There are, however, at least a dozen states with more ambitious graduation rate baselines and growth targets. California, which has a low test score baseline, has a much higher graduation rate minimum of 83%. Thus, a high school with a 61% graduation rate could make AYP in Georgia but fall far short in California. In addition, a high school with a very low initial graduation rate of say 40% in a state with a high baseline graduation rate could improve its graduation rate by 30 percentage points over 4 years—more than high schools making AYP in many states—and still find itself facing corrective action.

In sum, the case studies indicate that there is an Alice in Wonderland character to current implementations of the NCLB accountability framework: for high schools, up is down, and down is up. Some high schools that are making AYP and by implication are being told they are doing fine have extremely low, or even declining, graduation rates. Other high schools are making significant improvements in both achievement proficiency levels and graduation rates and are facing the most extreme NCLB sanctions because their starting points were so far below the baselines established by their state. As a result, it is not possible to use the AYP indicator to determine how many or to what extent the nation's lowest-performing high schools are improving. This is deeply problematic because it means NCLB is not achieving one of its core missions.

Problems with the implementation of the NCLB accountability framework for high schools are not limited to making it impossible for the public at large to know if low-performing high schools are improving. Even more problematically, these problems encourage teachers and administrators in low-performing schools to act in ways that are counterproductive to the intent of the law. This is most clearly seen in the undermining of the purpose of the safe harbor provisions. These provisions are intended to provide a means to acknowledge substantial improvement that falls short of yearly achievement goals. Reducing the percentage of students who are not proficient by 10 percentage points can free a school from the sanctions associated with not making AYP, the rationale being that if a school makes substantial improvement but falls a little short of ambitious improvement goals, it should not be penalized.

In low-performing high schools, with existing proficiency levels a great distance from their states' AYP levels, such as the 25% of high schools in our sample with mathematics proficiency levels of 20% or less, reaching safe harbor becomes the only feasible yearly achievement goal. When this is combined

with the fact that high school students are typically tested only in one grade for AYP, a perverse situation occurs. The most logical course for the low-performing high school is to focus all its available resources and reform efforts on a very small number of students—those students who are close to proficient in the tested grade.

Consider the following example. In one high school in our sample from Pennsylvania, only 5% of the students are proficient in mathematics. The current state achievement target is 45%. To reach this target, the school would need to make nearly an order of magnitude improvement in 1 year (the equivalent of learning how to run 60 miles per hour rather than 6). But to make safe harbor it needs to see only a 10 percentage point reduction in the number of 11th graders who are not proficient. This school has a nearly 50% dropout rate and, as a result, many fewer 11th and 12th graders than 9th and 10th graders. Thus, although there are 1,500 students in the school, only 250 are 11th graders, 10% of which is 25 students. NCLB in this particular application is not prodding the school to improve the education of its 1,500 students but rather to focus all its efforts on 25 students in 11th grade.

Discussion and Policy Implications

Our analyses uncover major shortcomings in AYP as an indicator of improvement, or persistent failure, in our nation's low-performing high schools. We found that 40% of the nation's low-performing high schools made AYP and that these schools tended to be better resourced, smaller, Southern, and less urban than those that did not make AYP. More fine-grained analyses, however, reveal that whether a particular school makes AYP depends on how much subgroup accountability it faces and its NCLB improvement status. Thus, state differences in how NCLB is being implemented at the high school level make it impossible to determine if the nation's low-performing high schools are getting better or which schools are making the greatest strides toward fulfilling the intent of the law. In addition, examples show how some features of current implementations of NCLB actually provide incentives or means to evade or even operate counter to the law's intent. Examples illustrate how low baseline requirements or minimal improvement targets for graduation rates may offer incentives for schools to improve achievement scores and reach AYP by pushing students out in the 9th or 10th grades. They also show how low-performing schools that are making meaningful improvements in both achievement and graduation measures can fail to make AYP (and invite sanctions), whereas similar schools that demonstrate far less improvement make AYP in other states. Equally troubling is the incentive produced through the safe harbor provision to target only a small number of students for instructional improvement, enabling schools with high concentrations of needy students to avoid more comprehensive reforms that could reach all students.

How Can NCLB Realize Its Potential in Low-Performing High Schools?

NCLB's desire to bring focused and sustained attention to low-performing high schools and provide them with incentives to improve could be an effective tool for school improvement. Transforming the nation's dropout factories into powerful engines of human capital would have a tremendous impact on the nation's economic and social well-being. As this discussion has illuminated, however, it is unlikely that the NCLB accountability framework, as currently implemented, will have this impact on the nation's lowest-performing high schools.

To better understand how to shape a more effective NCLB approach to low-performing high schools, it is first necessary to briefly examine what we know about why some high schools have very high dropout rates and low achievement levels, where they are located, and what it will take to dramatically improve them. Here, there are four key points.

There are three tiers of high schools in the United States. Broadly speaking, as indicated by the national distribution of promoting power, there are three tiers of high schools in the United States. Somewhere between 10% and 20% seem to function quite well. In these schools, nearly every student graduates, and many take and succeed in Advanced Placement courses. They are not limited to affluent suburbs. They include schools such as Baltimore Polytechnic, Central High School in Philadelphia, and the Bronx High School of Science, each of which is primarily attended by minority students. In the middle, the majority of high schools function well to average for some students but not for others. Then, there is the 12% to 15% that are the focus of this article. They do not work well for anyone and produce about half of the nation's dropouts. Every state has high schools in all three tiers, but the percentage in each tier considerably varies across states. At the low end of the spectrum, there are a few states with only a handful of low-performing high schools; at the other end, 30% to 40% of high schools in some states are low-performing (see the state summary table in the appendix for details).

Many low-performing high schools face an incredibly high degree of educational difficulty. Close to half the nation's low-performing high schools are concentrated in about 50 of its central cities. This is not an accident. Central cities often combine neighborhoods with concentrated and intergenerational poverty with a two-tier system of high schools. In these cities, students at or above grade level obtain access to the city's selective high schools or high school programs. Almost by definition this means that students who attend a city's neighborhood high schools, predominately located in its poorest neighborhoods, have skills that are below grade level. Many also have loosening attachments to schooling and worsening attendance problems. Thus, it is common for neighborhood high schools in central cities to be attended almost exclusively by students who have multiple risk factors for low achievement

and failure to graduate. For example, in one of the schools in our sample, only 15% of the ninth graders are first-time ninth graders, on age (i.e., have not been held back), not in special education, and no more than two grade levels behind in mathematics or reading skills. This means that 85% of the school's 500 ninth graders need intensive academic and social support to succeed. Similar profiles are found in the 14 other high-poverty neighborhood high schools in the school district (Neild & Balfanz, 2001).

In low-performing high schools, many students fall off the path to graduation in an entirely predictable manner. In high schools where the majority or near majority of students do not graduate, there is a clear sequence that most students who do not graduate follow (Allensworth & Eaton, 2005; Neild & Farley, 2004). They enter ninth grade with academic skills typically 2 to 4 years below grade level. A considerable number are already overage for the grade, which means they are legally only 1 or 2 years away from being able to drop out in many states. They also have weakening attendance habits and will miss a month or more of the ninth grade, and 20% to 30% might miss 10 of the first 30 days of school. Sporadic attendance, combined with poor preparation, leads to first semester course failure. The typical grade in the school may well be a *D*. Discouraged by their first-semester grades, some students conclude they will not pass ninth grade this year and are absent more frequently. Others will try harder but not raise their grades enough to earn promotion to 10th grade. Perhaps 30% to 40% of the ninth grade will be retained.

Most will attempt to repeat ninth grade the following year, but, absent major interventions to improve their attendance and academic skills, they will do no better than the first time and soon enter a rapid path toward dropping out. They may try transferring to another school or even attending an alternative school before they do drop out, but by this point their ultimate fate is fairly set. In short, although poverty and personal issues requiring social service attention are powerful contributors and are often the underlying reason for some of the students' behaviors, the direct reason most students fail to graduate is that they do not acquire enough credits to be promoted to the 10th or 11th grade or, ultimately, to graduate. And the main reason they do not earn enough credits is that they do not attend school often enough and do not have the reading and mathematics skills to pass their courses.

Improving a low-performing high school requires comprehensive reforms that are not fast, easy, or cheap. Enough is known about transforming low-performing, high-poverty high schools to effect substantial improvements in many of them. Working models, success stories, and independent rigorous evaluations exist (e.g., Kemple, Herlihy, & Smith, 2005; Legters, Balfanz, Jordan, & McPartland, 2002; Quint, Bloom, Black, & Stephens, 2005). The challenge is to develop the capacity, know-how, and will to implement what is known to work in all the high schools in need. First and foremost, it needs to be

recognized that truly comprehensive reform is required. A dominant focus on one or even several levers of improvement is not enough to address the degree of educational challenge that exists in low-performing high schools. Increased personalization and student outreach, high standards, intensive instructional programs to close achievement gaps, improved teacher quality, professional development and teacher supports, engaging school programs, and strengthened connections between high schools and colleges and employers are all needed in large, sustained, coordinated measures.

Patience, commitment, and resources are also required. Some critical factors can be quickly improved within the 1 or 2 years allowed by the current NCLB accountability framework. In low-performing high schools, it is possible to achieve significant 1-year improvements in student attendance, reductions in suspensions, improvements in course passing rates, and improvements in promotion between grade levels (Kemple et al., 2005). Comprehensive reform efforts at the district and school levels require time to plan and several years of implementation before major achievement gains can be realized.

Significantly raising student achievement typically takes more time for two reasons. First, it typically requires coordinated improvements in at least five areas: student attendance, engagement and effort, the instructional program (often both the course sequence students take and the instructional materials used), the extra help opportunities available to students with below-grade-level skills, and, finally, teacher and administrator effectiveness and support. In addition, based on our experience working with more than 50 high schools engaged in comprehensive reforms, 10% to 25% of the students may need improved social services as well. If effort, focus, or skill falter in any one of these areas or factors outside of the school's control have an effect or if a high school lacks sufficient resources to mount simultaneous reforms in all these areas, then progress toward significant achievement gains can be stalled or muted. Second, these coordinated efforts need to occur in every grade. If a student enters high school with reading and mathematics skills at the fifth grade level, strong and coordinated reform efforts may be able to bring up this student's skills to a seventh or even eighth grade level by the end of ninth grade. Despite making significant achievement gains, this student still may not have the prerequisite skills needed to pass an end-of-course algebra test in ninth grade or succeed in geometry in 10th grade. Sustained efforts will be needed throughout high school to bring this student up to grade level by 11th or 12th grade. Thus, even state-of-the-art reforms could take 4 years to enable the students who typically enroll in low-performing high schools to reach NCLB proficiency levels.

Three Proposals to Help NCLB Realize Its Intent With Low-Performing High Schools

Proposal 1: Reconceptualize safe harbor so it focuses low-performing high schools on the key points where students fall off the graduation path and encourages them to implement strategic schoolwide reforms.

At its core, the NCLB accountability framework is an ambitious attempt to use incentives and sanctions to change behavior. But as any economist or parent will tell you, getting the signals right is very tricky business. Carrots and sticks can work, but they need to be the right ones for the situation and involve a shared understanding of the desired outcome between the two parties involved.

Safe harbor as it is conceptualized is sending the wrong signals. It is encouraging low-performing high schools to focus their reform efforts on a very few students rather than on improving the whole school. Moreover, improvements in academic achievement and graduation rates in practice do not tend to happen in steady yearly increments. Nor should we really want them to in low-performing high schools. In a high school where less than 10% of students are proficient in mathematics and reading and less than 50% are graduating, do we really want to spend 5 years establishing that it needs a major transformation or longer if it manages to make incremental improvements in a few of those years?

An alternative might be to base safe harbor around significant yearly improvements in the percentage of students earning promotion from one grade to the next and taking a rigorous sequence of high school courses. The high school course sequence promoted by the U.S. Department of Education State Scholars program might be a good starting point, along with giving students the option to include a coherent sequence of high-quality career and technical education courses. Chris Swanson's (2004) Cumulative Promotion Index might provide one model of how progress from grade to grade could be measured using existing enrollment data already collected by the education department. Basing safe harbor on significant increases in the percentage of students earning on-time promotion from grade to grade and taking a rigorous sequence of high school courses would focus low-performing schools on improving the education of every student in every grade. It would also direct their reform efforts toward two of the major school-level variables that affect both graduation rates and achievement levels (Allensworth & Eaton, 2005).

Proposal 2: Use NCLB reform as an opportunity to solve the conundrum of Title I funding for high schools and acknowledge that different high schools face greatly different degrees of educational difficulty.

NCLB aims to provide both sanctions and supports. Yet, to date, the supports have been underdeveloped and largely focused on governance issues as opposed to the comprehensive organizational, instructional, and professional development and teacher support reforms that many low-performing high schools need. Equally significant is the fact that NCLB's sanctions and supports are supposed to be directed at schools that receive Title I funding. Yet most of the nation's lowest-performing high schools do

not receive Title I funds, even though they educate primarily high-poverty students. In our random sample of low-performing high schools, for example, only 47% were receiving Title I funds, even though 73% had 40% or more of the student body receiving free or reduced-price lunch. This means that the federal program for providing supplemental support to schools that face the challenge of educating students affected by poverty is not reaching many of the nation's high schools with the greatest need for additional support.

One way to resolve both problems would be to establish a separate stream of Title I funds for high-poverty high schools. Funds would be distributed by a formula that factors in both the poverty rate and the degree of educational difficulty faced by the high school. Educational difficulty would be defined in part by the number of entering students who are overage for grade, have failed courses in the middle grades, and have below-grade-level skills and/or have weak attendance habits. Finally, continuation of funding should be contingent on high schools implementing comprehensive, evidence-based reforms that address student attendance, behavior, and engagement, provide intensive and sustained extra help to those with below-grade-level skills, increase available social supports, and enable teachers and students to develop and use the skills needed to teach and learn rigorous academic material.

Proposal 3: Act now to transform or replace the lowest-performing high schools.

Our data indicate that a significant number of low-performing schools will not be improved through accountability systems and the standards movement alone. Our experience indicates that they lack the sufficient human, organizational, instructional, and financial resources to reform themselves, regardless of the amount of reform pressure put on them. Nor do we need an improved accountability system to identify these high schools. We know today who they are and where they are located (Balfanz & Legters, 2004). What is needed is a coordinated federal, state, and local effort to provide the vision, resources, tools, training, and technical assistance required to transform or replace the approximately 15% of high schools that produce most of the nation's dropouts. Thus, part of revising NCLB should involve providing the means and methods to do so. This could involve state and districts working together to provide struggling high schools with technical assistance teams either directly or by contracting with established third-party school-reform organizations with proven track records. It will involve making sure that the technical assistance teams and high schools have the necessary human and financial resources to implement, institutionalize, and sustain comprehensive reforms. It will also need to ensure that states, districts, and technical assistance teams are backed with necessary statutory power to enact needed reforms that challenge existing staffing, scheduling, and other regulatory structures.

Conclusion

Given the current interest in high school reform and with NCLB's approaching reauthorization, now is an opportune time to examine the reliability and validity of AYP as a measure of school performance. This is especially the case at the high school level and in particular for low-performing high schools that drive the low national graduation rate for minority students and the growing number of dispossessed young adults who are neither employed nor in school. These high schools must be specifically targeted for reform if the American high school is to fulfill its pivotal role as the means by which children who grow up in poverty can become adults who lead the nation.

The costs of inaction are high. High school dropouts in today's economy are far more likely to be unemployed, suffer from health problems, be dependent on social services, or be in prison than are their peers who graduate. The benefits of improving low-performing high schools are enormous. A recent study finds that our nation can recoup \$45 billion in lost tax revenues, health care expenditures, and social service outlays if we cut the number of high school dropouts in half (Levin, Belfield, Muennig, & Rouse, 2007).

Providing all students with access to a high-quality, standards-based education is the primary intent of NCLB. We fully embrace the spirit of NCLB, yet our research shows that this landmark legislation is falling short of its intentions at the high school level. Rather than effectively and consistently identifying and stimulating improvement in low-performing high schools, AYP has created a confusing landscape where improvement in some low-performing high schools is deemed inadequate, whereas even less improvement in other schools is considered adequate. If this continues, AYP will work against the spirit of NCLB by creating incentives for low-performing high schools to push out students and forgo costly, but ultimately more effective, comprehensive reforms in favor of test preparation for a targeted few.

We offer several proposals to address these shortcomings—reconceptualizing safe harbor for high schools so that it focuses low-performing high schools on the strategic schoolwide reforms, a hard look at how Title I funding is distributed to low-performing high schools, and focused efforts to provide comprehensive reforms to the most challenged high schools. These proposals attempt to strike a balance between the need for uniform standards designed to ensure that all students graduate from high school prepared for success in college, career, and civic life and the high degree of educational challenge faced by low-performing high schools. Such changes would increase the effectiveness of NCLB and more closely align implementation of the law with its stated purpose of ensuring equal access to a high standard of education for all.

Appendix

Table A1
**Descriptive Statistics for School-Level Variables, Analysis
of Selected U.S. High Schools' Adequate Year
Progress (AYP) Status (2004–2005)**

Variable	Description	% Distribution ^a		
		or <i>M</i>	<i>SD</i>	<i>n</i>
School type				
Vocational or technical school		0.99		
Magnet school		10.23		176
Charter school		0.99		
School location				
Urbanicity	Large city (250,000+)	26.24		
	Midsize city (< 250,000)	16.83		
	Urban fringe, large city	15.35		
	Urban fringe, midsize city	4.95		
	Small town (2,500 to 24,999)	6.93		
	Rural, outside CBSA	15.84		
	Rural, inside CBSA	13.86		
	Urban (large or midsize city)	43.07		
	Suburban (fringe or town)	27.23		
	Rural	29.70		
School enrollment				
Total minority students	Range 0 to 4,813	865.70	880.90	
Total enrollment	Range 56 to 4,855	1,212.80	938.80	
Student–teacher ratio				
Ratio	Range 5.1 to 56	16.65	5.16	192
Categorical	Less than 15	35.42		192
	15 to 19.99	47.92		
	20 or more	16.67		
School resources: Title I				
Title I eligible		47.24		199
Eligible for schoolwide Title I		75.25		101
School composition				
% free or reduced lunch	Range 10.2 to 99.6	55.61	21.48	200
% total minority	Range 0 to 100	62.96	34.98	
Largest group in school	White	39.11		
	African American	37.62		
	Hispanic	19.80		
	American Indian or Alaskan Native	2.48		
	Asian or Pacific Islander	0.99		

Source. U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002-2003

Note. *N* = 202, unless otherwise noted. CBSA = core-based statistical area.

a. Percentage distribution is for valid *n* for each variable.

Table A2
**Descriptive Statistics for School Performance and
 Adequate Yearly Progress (AYP) Status, Analysis of Selected
 U.S. High Schools AYP Status (2004–2005)**

Variable	Description	% Distribution		
		or <i>M</i>	<i>SD</i>	<i>n</i>
Promoting power				
Class of 2003	Range 0.16 to 0.60	0.50	0.10	
AYP testing				
% proficient math 2004–2005	Range 6.1 to 96.0	50.34	22.87	195
% proficient math 2003–2004	Range 2.9 to 96.0	45.78	21.60	201
% proficient English				
language arts (ELA) 2004–2005	Range 6.0 to 97.0	56.94	25.09	195
% proficient ELA 2003–2004	Range 5.8 to 97.0	53.84	24.64	201
AYP other indicator				
% graduated 2004–2005	Range 41.6 to 100.0	76.81	14.40	189
% graduated 2003–2004	Range 30.0 to 100.0	73.37	15.10	188
AYP status				
Made AYP 2004–2005		40.59		
Made AYP 2003–2004		39.00		200
Safe harbor (SH)				
Tried to make SH 2004–2005		22.92		192
Met SH 2004–2005		13.02		192
Subgroup breakdowns for AYP				
Student body as whole		53.96		
met AYP				
Had to meet AYP for				
minority group		77.89		199
Largest racial minority met AYP		34.69		196
Total goals required to	Range 2 to 34	14.97	6.83	197
meet AYP				
# of goals met	Range 0 to 33	11.44	6.13	197
# of subgroups reported	Range 1 to 8	3.93	1.62	
for AYP				
No Child Left Behind				
improvement status				
Status 2004–2005	Not in improvement	50.00		200
	At risk or under watch	3.50		
	Improved or making progress	2.00		
	In improvement, Year 1	11.50		
	In improvement, Year 2	8.50		
	In improvement, Years 3 to 7	7.00		
	Improvement, unspecified	10.00		
	Corrective action	6.00		
	Restructured or restructuring	1.50		
Status 2003–2004	Not in improvement	66.67		156
	At risk or under watch	3.21		
	Improved or making progress	0.64		
	In improvement, Year 1	14.10		
	In improvement, Year 2	5.77		

Table A2 (continued)

Variable	Description	% Distribution		
		or <i>M</i>	<i>SD</i>	<i>n</i>
	Improvement, unspecified	5.13		
	Corrective action	1.28		
	Restructured or restructuring	1.28		
In any improvement 2004–2005		50.00		200
Incentive to improve 2004–2005 ^a	Minor	50.00		200
	Moderate	18.00		
	Major	32.00		

Source. State report cards.

a. For improvement status, motivation to improve (categorical): 1 = minor (includes schools not in improvement), 2 = moderate (includes schools at risk or targeted for improvement, those schools making progress, or schools in improvement Year 1), 3 = major (includes schools in improvement Years 2 to 7, schools under corrective action, or schools that are restructuring or restructured or planning to restructure).

Table A3
**Descriptive Statistics for State-Level Variables, Analysis
of Selected U.S. High Schools Adequate Yearly Progress
(AYP) Status (2004–2005)**

Variable	Description	% Distribution		
		or <i>M</i>	<i>SD</i>	<i>n</i>
Grade of test administration				
# grades tests are given in	One grade	66.83		
	Two grades	21.78		
	Three grades	5.94		
	All grades	5.45		
Which grade are tests given	All: 9th to 12th	5.45		
	9th and 10th only	9.90		
	10th to 12th	5.94		
	10th and 11th only	8.91		
	11th and 12th only	2.97		
	10th only	33.66		
	11th only	25.74		
	12th only	7.43		
State standards for AYP				
Graduation standard is absolute	No, improvement is accepted	82.98		188
	Yes, must make threshold	17.02		
Graduation threshold 2004–2005	State has no threshold	3.30		182
	Range 50 to 90	96.70		
Math standard 2004–2005	Range 17.5 to 75	44.84	16.46	182
English language arts (ELA) standard 04-2005	Range 13.7 to 90	51.17	20.02	182
Rigor of state AYP exams ^a				

(continued)

Table A3 (continued)

Variable	Description	% Distribution or <i>M</i>	<i>SD</i>	<i>n</i>
% proficient or better on National Assessment of Educational Progress (NAEP): Elementary reading	Range 10 to 43	28.43	5.64	
% basic or better on NAEP: Elementary reading	Range 31 to 74	60.10	6.94	
% proficient on AYP: Elementary reading	Range 29 to 91	65.50	17.20	
% proficient or better on NAEP: Grade 8 math	Range 6 to 38	24.95	5.75	
% basic or better on NAEP: Grade 8 math	Range 29 to 76	64.18	7.42	
% proficient on AYP: Grade 8 math	Range 12 to 88	53.72	19.80	
Comparison NAEP versus AYP ^a				
ELA: Proficient versus proficient	AYP equal to NAEP	7.65		196
	AYP easier	92.35		
ELA: Basic versus proficient	AYP more difficult	16.33		196
	AYP equal to NAEP	45.92		
	AYP easier	37.76		
Math: Proficient versus proficient	AYP more difficult	2.55		196
	AYP equal to NAEP	21.43		
	AYP easier	76.02		
Math: Basic versus proficient	AYP more difficult	43.37		196
	AYP equal to NAEP	51.53		
	AYP easier	5.10		

Source. U.S. Department of Education National Assessment of Educational Progress (NAEP) and State Report Cards.

Note. *N* = 202, unless otherwise noted.

a. Comparison of NAEP versus AYP is the difference between the percentage proficient or basic on the NAEP and the percentage proficient on the AYP exam in each state.

Table A4
Sample Comparison by Selected School Characteristics

School Characteristics	Random Sample of		
	Total High Schools ^a	Low-Performing High Schools ^b	Low-Performing High Schools ^c
School type ^d			
% regular	99.10	98.18	99.00
% vocational or technical school	0.88	1.72	1.00
School location ^{e*}			
% urban	16.89	45.47	43.07

Table A4 (continued)

School Characteristics	Total High Schools ^a	Low-Performing High Schools ^b	Random Sample of Low-Performing High Schools ^c
% suburban	41.07	32.81	27.23
% rural	42.04	21.72	29.70
School enrollment			
Average enrollment ^f	919.77	1,245.82	1,212.79
Student demographics			
Average percentage minority ^g	27.07	63.76	62.96
Average percentage of students eligible for free or reduced-price lunch ^{h*}	30.96	50.84	55.61
Average student-teacher ratio ⁱ	16.25	17.26	16.65
School resources			
Percentage of schools eligible for Title I ^j	28.20	44.30	46.53

Source. Data on school characteristics were taken from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data 2002–2003.

a. $n = 14,292$.

b. $n = 2,030$.

c. $n = 202$.

d. $\chi^2(1) = 0.616, p > .05$.

e. $\chi^2(2) = 8.101, p > .05$.

f. $t(201) = -0.500, p > .05$.

g. $t(201) = -0.326, p > .05$.

h. $t(191) = -1.640, p > .05$.

i. $\chi^2(1) = 0.885, p > .05$.

j. $t(199) = 3.140, p < .05$.

* $p < .05$, two-tailed (significant difference between the low-performing high schools and the random sample of low-performing high schools).

Notes

This article was originally presented in a forum sponsored by the Civil Rights Project at Harvard University and the Chief Justice Earl Warren Institute for Race, Ethnicity and Diversity, titled *The No Child Left Behind Act: How Does it Effect High School Reform?*, at the Hart Senate Office Building in Washington, D.C., October 14, 2005. The authors wish to thank these organizations for their leadership and initial dissemination of this research. We also thank Center for Social Organization of Schools Communications Director Mary Maushard for her editorial assistance in preparing the final article.

The authors' research background and theoretical orientation involves using the tools and methods of the social sciences to translate research findings into effective intervention for low-performing schools. This research was motivated by the desire to understand how No Child Left Behind affects the nation's lowest-performing high schools and to use quantitative and qualitative research findings to identify means by which federal accountability and intervention systems can be used to improve them.

¹Adequate yearly progress (AYP) formulas widely vary by state but typically require schools to demonstrate that a certain proportion of their students (overall and in every eligible subgroup) are performing at proficient levels on state assessments and are graduating in the standard number of years (or are showing expected gains in graduation rates).

Schools may achieve AYP through a safe harbor provision by reducing the percentage of students who are not academically proficient by 10 percentage points.

²The sample was drawn using the random seed function of the SPSSx statistical software. The sample provides a confidence interval of ± 6.54 percentage points, meaning that results may be generalized to the larger national population of 2,030 low-performing high schools with 95% confidence within this range. Given the labor-intensive nature of culling data from individual district and state report cards, we were unable to collect data from the 700-school sample that would be required to achieve the desired CI standard of ± 3 percentage points.

³Nearly 94% of the 2,030 low-performing high schools across the nation are located in the 34 states included in the sample. Virtually all of the states that were not included in the sample had very few low-performing high schools. Three states that were not included in the sample, however, had more. These are Indiana (20), Oklahoma (16), and Washington (25). Given the heavy concentration of low-performing high schools in the states that are represented in the sample and the national focus of our analyses, the sample offers a reasonable degree of generalizability to the national population of low-performing high schools.

References

- Allensworth, E., & Eaton, J. (2005). *The on-track indicator as predictor of high school graduation*. Chicago: Consortium of Chicago School Research.
- Balfanz, R. (2006). What does it cost to operate a high school organized into small learning communities? When are additional resources needed? How can efficiencies be achieved? In U.S. Department of Education (Ed.), *Critical issues in development and implementation of high school small learning communities* (pp. 151–186). Washington, DC: U.S. Department of Education.
- Balfanz, R., & Legters, N. (2004). Locating the dropout crisis: Which high schools produce the nation's dropouts? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 57–84). Cambridge, MA: Harvard Education Press.
- Balfanz, R., & Legters, N. (2005). *The graduation gap: Using promoting power to examine the number and characteristics of high schools with high and low graduation rates in the nation and each state*. Baltimore: Johns Hopkins University, Center for Social Organization of Schools.
- Center on Education Policy. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington, DC: Author.
- Commission on No Child Left Behind. (2007). *Beyond NCLB: Fulfilling the promise to our nation's children*. Washington, DC: Aspen Institute.
- Darling-Hammond, Linda. (2006). No Child Left Behind and high school reform. *Harvard Educational Review*, 76(4), 642.
- Forum on Educational Accountability. (2007). *Redefining accountability: Improving student learning by building capacity*. Cambridge, MA: Author.
- Hall, D. (2005). *Getting honest about grad rates*. Washington, DC: Education Trust.
- Kemple, J., Herlihy, C., & Smith, T. (2005). *Making progress towards graduation: Evidence from the talent development high school model*. New York: MDRC.
- Legters, N., Balfanz, R., Jordan, W., & McPartland, J. (2002). *Comprehensive reform for urban high schools: A talent development approach*. New York: Teachers College Press.
- Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2007). *The costs and benefits of an excellent education for all of America's children*. New York: Columbia University, Teachers College.
- Losen, D. (2005). Graduation rate accountability under the No Child Left Behind Act. In G. Sunderman, J. S. Kim, & G. Orfield (Eds.), *NCLB meets school realities: Lessons from the field* (pp. 105–119). Thousand Oaks, CA: Corwin Press.

- Neild, R., & Balfanz, R. (2001, August). *An extreme degree of difficulty: The demographics of the ninth grade in non-selective high schools in Philadelphia*. Paper presented at the annual meeting of the American Sociological Association, Chicago.
- Neild, R., & Farley, E. (2004). Whatever happened to the class of 2000? The timing of dropout in Philadelphia's schools. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 207–220). Cambridge, MA: Harvard Education Press.
- Orfield, G., Losen, D., Wald, J., & Swanson, C. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis*. Cambridge, MA: Harvard University, Civil Rights Project.
- Quint, J., Bloom, H., Black, R., & Stephens, L. (2005). *The challenge of scaling up educational reform: Findings and lessons from First Things First*. New York: MDRC.
- Sunderman, G., Kim, J. S., & Orfield, G. (Eds.). (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.
- Swanson, C. (2004). *Who graduates? Who doesn't? A statistical portrait of public high schools graduation, class of 2001*. Washington, DC: Urban Institute, Education Policy Center.

Manuscript received January 4, 2007

Revision received April 29, 2007

Accepted June 28, 2007