

No Child Left Behind and High School Reform:

Linda Darling-Hammond

Civil Rights advocates hailed the 2001 reauthorization of the ESEA, optimistically entitled “No Child Left Behind” (NCLB), as a step forward in the long battle to improve education for those children traditionally left behind in American schools – in particular, students of color and those living in poverty, new English learners, and students with disabilities. The broad goal of NCLB is to raise the achievement levels of all students, especially underperforming groups, and to close the achievement gap that parallels race and class. The Act intends to do this by focusing schools’ attention on improving test scores for all groups of students, providing parents’ more educational choices, and ensuring better qualified teachers. This paper looks at how various aspects of NCLB support or undermine the current national movement to reform high schools – a movement that advances similar goals but approaches them from a different perspective. It proposes specific amendments to the Act that could help to achieve the goals of high-quality, equitable education for all youth.

The Need for High School Reform

Of all the ways in which urban school children are being left behind, their experiences in large, factory-model high schools are, arguably, the most egregious. In fact, in many such schools, young people are not only left behind but actively thrown overboard. In urban areas, dropout rates from large comprehensive high schools are typically 50% or more. These schools are structured as huge warehouses, often housing 3000 or more students in an organization focused more on the control of behavior than the development of community. With a locker as their only stable point of contact, a schedule that cycles them through a series of 7 or more

overloaded teachers, and a counselor struggling to serve the "personal" needs of several hundred students, teenagers struggling to find connections have little to connect to. Heavily stratified within, and substantially dehumanized throughout, most students experience such high schools as non-caring or even adversarial environments where "getting over" becomes important when "getting known" is impossible. For adults, the capacity to be accountable for the learning of 150 to 200 students daily – students whom they do not share with other teachers – is substantially constrained by the factory model structure that gives them little control over or connection to most of what happens to the students they see only briefly.

For more than 40 years, large urban high schools have been critiqued for their impersonal structures; their fragmented curricula; their segregated and unequal program options; and their inability to respond effectively to different student needs (Barker and Gump, 1964; Goodlad, 1984; Lee, Bryk, and Smith, 1993; Powell, Farrar, and Cohen, 1985; Sizer, 1984). A number of recent studies have found that, other things equal, smaller schools appear to produce higher achievement (Haller, 1993; Howley, 1989; Howley & Huang, 1991); lower dropout rates (Pittman & Haughwout, 1987); lower rates of violence and vandalism (Garbarino, 1978; Haller, 1992); more positive feelings about self and school, and more participation in school activities (Fowler, 1992; Green & Stevens, 1988; Howley & Huang, 1991; Lindsay, 1982, 1984). These outcomes generally appear more pronounced for students who are traditionally lower-achieving (Lee & Smith, 1993; 1995).

A number of small schools serving high-need students have experienced striking success, graduating more than 90% of their students and sending equal shares to college (see, for example, Bensman, 1987; Darling-Hammond, 1997; Darling-Hammond, Aness, & Ort, 2002). School size is not the only thing that may make a difference in student achievement, however, A

number of studies have found that, all else equal, schools have higher levels of graduation and academic achievement – especially for traditionally low-performing students – when they create smaller, more personalized units in which students see a smaller number of teachers over a longer period of time and teachers work together with a smaller cohort of students (Braddock & McPartland, 1993; Gottfredson & Daiger, 1979; Lee, Bryk, and Smith, 1993; Wehlage, et al., 1989). Researchers suggest that in such “communitarian” schools, students are better known, and faculty develop a more collective perspective about their work (for reviews see Lee, Bryk, and Smith, 1993; Newmann & Wehlage, 1995).

In response to these kinds of findings, major reforms of comprehensive high schools are underway in many cities, in part with the assistance of the federal government through the Small Schools Act and charter school development funds and with the help of philanthropists like the Gates Foundation, the Carnegie Corporation of New York, and others. NCLB has a complex relationship with these reforms, both helping and hindering the efforts that are underway to remake the most dysfunctional element of the American schooling system.

Elements of High-Performing Urban High Schools

Not all small schools are equally successful. Those that have implemented fewer personalizing features and less intensive changes in instruction have not produced significant benefits (McMillan, Sipe, & Wolf, 1994; Raywid, 1990, 1995; Wehlage, Smith, & Lipman, 1992). Schools that are dramatically more effective have instituted important changes in organization and instruction. In addition to smaller learning communities where students are well-known, research suggests that at least the following four elements are critically important:

1. Personalization achieved through teams of teachers working with shared groups of students – usually numbering no more than 80 and sometimes over multiple years – and

through “advisories” in which each teacher takes responsibility for about 15 students for whom he or she serves as advocate, counselor, and primary family contact. These smaller pupil loads are achieved, in part, through offering longer block classes, some of them interdisciplinary (e.g. Humanities courses that combine English and social studies) which allow students to teach fewer students for longer blocks of time.

2. Well-qualified teachers supported by ongoing peer collaboration and professional development. Successful schools have teachers who have both solid academic backgrounds and who are fully prepared and certified for teaching, including skills for teaching students with special needs, English language learners, and the wide range of other students they encounter (for a review, see Darling-Hammond, 2000). Furthermore, these teachers create strong, coherent curriculum by having time in their schedules to plan together around both subject matter and students and to pursue ongoing professional development that hones their abilities to reach all students.
3. A common core curriculum organized around performance-based assessment, which engages students with work that resembles what they will do outside of school and challenges them intellectually. Most highly effective urban schools require students to complete portfolios and performance tasks in which they must conduct significant research, design and carry out science experiments, amass and analyze evidence, apply math skills to real-world problems in engineering, physics, topography and other fields, and demonstrate competence in writing and the arts. Students revise their work in response to feedback guided by performance standards, and they defend their work before committees of teachers and external judges, much like a dissertation.

4. Supports for struggling students – Students are not tracked in these schools, but are supported in a variety of ways to meet the demands of an intellectually engaging and challenging curriculum. Teachers use of a wide repertoire of instructional strategies to adapt to students’ needs. Most schools also routinely offer a wide range of supplementary services, like after school and Saturday homework support and tutoring for all students who need additional help. They also use special education models that push-in expert teachers to the core classroom and provide resource room supports for students to be assisted in completing the same challenging work that other students are assigned, rather than pull-out classrooms where students are kept busy with workbooks and lower level tasks. These strategies differ from those in traditional high schools by avoiding unnecessary tracking and by integrating supplementary supports into the core curriculum, rather than offering disconnected services that fragment, rather than concentrate, effort.

These practices are not only found in case studies tracking small groups of schools (e.g. Darling-Hammond, Aness, & Ort, 2002; Newmann & Wehlage, 1995; Wehlage et al., 1989), but also in large-scale studies. For example, a recent study of 820 high schools in the National Education Longitudinal Study (NELS) found that schools that had restructured to personalize education and develop collaborative learning structures for adults and students produced significantly higher achievement gains that were also more equitably distributed (Lee & Smith, 1995). The schools’ practices included keeping students together in the same advisory group over multiple years, establishing smaller functional units within schools, forming teaching teams, assuring common planning time for teachers, involving staff in schoolwide problem solving, involving parents, and fostering cooperative learning.

Furthermore, in a study of more than 2,000 students within 23 restructured schools, Newmann, Marks, and Gamoran (1995) found higher levels of achievement on complex performance tasks for students who experienced what these researchers termed “authentic pedagogy” – instruction focused on active learning in real-world contexts calling for higher-order thinking, consideration of alternatives, extended writing, and an audience for student work. A recent analysis of NELS data found that students in restructured schools where “authentic instruction” was widespread experienced greater achievement test gains (Lee, Smith, & Croninger, 1995).

High School Reform and No Child Left Behind

The efforts of No Child Left Behind to support more equitable education have leveraged important attention to school reform and to the relative success of students of color and low-income students who have traditionally been poorly served by comprehensive high schools. This spotlight on the achievement of different groups of students within large schools has been useful to focus attention on inequalities not only in the cities, but also in suburban and rural communities where deeply entrenched tracking systems relegate most of those who attend “integrated” schools to the lower tracks where they often receive less challenging and lower quality instruction from less well-qualified teachers.

At the same time, the complicated rules that have accompanied No Child Left Behind have unintentionally made it more difficult for many heroic schools in low-income neighborhoods to do their work well and to keep the neediest students in school and moving toward productive futures. Two areas of the law require particular attention:

- The definition and development of “highly qualified teachers,” and
- The design of testing and accountability regulations.

Each of these areas is described below, followed by recommendations for amendments that will enable more high schools to leave no child behind.

Ensuring Highly Qualified Teachers

One of the most important aspects of No Child Left Behind is that it requires all schools to provide “highly-qualified teachers” to all students by 2006. This requirement – that all teachers be fully certified and show competence in the subject areas they teach as well as demonstrated teaching skills – is intended to correct the longstanding problem that schools serving our neediest students typically have the least experienced and well-qualified teachers, even though such students need our most skilled teachers if they are to learn what they need to know. And it is a problem that can be solved. What often looks like a teacher shortage is actually mostly a problem of getting teachers from where they are trained to where they are needed and keeping teachers in the profession, especially in central cities and poor rural areas. More than 30% of beginners leaving teaching within five years, and those with the least preparation and the least access to mentoring leave soonest (Darling-Hammond & Sykes, 2003).

In low-income schools suffering from even higher turnover rates, producing more teachers – especially through fast-track routes that tend to have high attrition – is like spending all our energy filling a leaky bucket rather than fixing it. Yet, NCLB also encourages the creation of alternative certification pathways – some of which skirt key elements of teacher learning – and regulations allow teachers to be deemed “highly qualified” as soon as they enter such programs, rather than when they complete them and demonstrate they are competent. Perversely, on the other hand, many well-prepared and highly successful teachers are deemed *not* highly qualified because they teach in multiple subject areas and do not have majors or have passed tests in all of them. The interdisciplinary teaching that is so central to many successfully

redesigned high schools – as well as to small schools in many rural and urban areas – is jeopardized by the current administration of the law.

NCLB’s emphasis on highly qualified teachers has been generally helpful and productive – and there is evidence that many states are investing in the preparation and recruitment incentives needed to produce and distribute well-prepared teachers where they are needed. However there are three issues that need attention if the implementation of the Act and schools’ efforts to improve are to be mutually reinforcing.

1. The law and regulations need to create appropriate allowances for multi-disciplinary and interdisciplinary teaching, which are a necessity in many small schools and districts and an advantage in reforming secondary schools. The Act’s current regulatory strategy is to require a major or passage of a subject matter test in each subject taught, in addition to full certification and demonstration of teaching skills. While this seems straightforward, it turns out to be highly problematic in small schools in remote rural areas – where a single teacher may teach every grade level and subject area – and in small high schools – where interdisciplinary teaching allows both the curriculum integration that makes learning more authentic and allows teachers to have a smaller pupil load, so that they can individualize instruction. Small schools that use interdisciplinary configurations and block schedules can cut pupil load from 150 or more students per teacher to as few as 40 or 50 students per teacher, without additional funding, thus allowing much greater attention to pupil learning (Darling-Hammond, 1997).

Most states have developed sensible certification laws for handling these kinds of real-world situations in ways that evaluate subject matter knowledge and teaching skills appropriately, and some have developed certification rules that specifically take into account interdisciplinary or multi-disciplinary teaching. The federal government should delegate to the

states the ascertainment of subject matter competence for teachers in these categories, through their certification systems, rather than trying to specify from Washington the only means for teachers to meet the intent of the law.

2. The administration of the Act needs to be strengthened to ensure that teachers are not labeled as “highly qualified” until they have completed preparation and demonstrated that they have mastered essential teaching skills. While the NCLB is generally viewed as overly prescriptive regarding how subject matter knowledge is to be demonstrated, it is far too lax regarding the demonstration of teaching ability, allowing individuals to be deemed “highly qualified” the moment they *enter* an alternative certification program rather than when they have finished a program and demonstrated their competence. While some excellent alternative certification programs have been constructed, the range of program quality is extremely wide. Unfortunately, many programs do not provide student teaching or coursework in essential areas like student learning and development, specific subject matter teaching methods, including the teaching of reading, or the teaching of special needs students. All of these are areas of learning that are associated with greater teacher effectiveness (Darling-Hammond, 2000). Furthermore, many place candidates in classrooms as teacher of record after a few weeks of summer training and long before they have encountered much of the coursework the program offers. Promised mentoring does not always appear. This happens almost exclusively in high-need schools serving the most disadvantaged students.

For these reasons, many alternative routes have very high attrition rates (Darling-Hammond & Sykes, 2003). Candidates who do not receive student teaching and critical areas of coursework are twice as likely to leave teaching after the first year, adding to the instability that plagues so many high-need schools. To call teachers who have not completed their training

“highly qualified” is not only dishonest, it is a disservice both to teachers recruited into lower-quality programs and to their students. Prospective teachers rely on those who regulate preparation to ensure that the programs they are recruited into are adequate to prepare them to succeed. When that does not happen, they flounder and feel inadequate, often leaving in despair rather than continuing in the profession. Although the regulations call for alternative routes to meet specific standards, many programs are not accredited and few states enforce the federal standards. This disadvantages all schools that hire such teachers and the students they serve.

NCLB should recognize all teachers as “highly qualified” only when they have *completed* a solid program of preparation, and should require that alternative certification programs, like other programs, meet accreditation standards that incorporate the Act’s regulatory standards regarding acceptable alternatives (regulations which are currently unenforced). In addition, as described below, greater incentives are needed to develop high-quality preparation programs that prepare teachers in high-need communities to work in redesigned high schools and to subsidize candidates so that they can afford to take the time they need to be well-prepared.

3. Greater federal supports and incentives are needed to recruit and prepare highly qualified teachers and to distribute them to the schools where they are most needed. There is no need for large numbers of schools in the United States to experience teacher shortages or to have to hire individuals who lack preparation to teach. There are actually at least 3 or 4 times as many credentialed teachers in the U.S. as there are jobs, and many states and districts have surpluses. Not surprisingly, however, teachers are less likely to enter and stay in teaching where salaries are lower and working conditions are poorer. They are also more than twice as likely to leave if they have not had preparation for teaching and if they do not receive mentoring in their early years on the job. These are problems that can be solved. States and districts that have

increased and equalized salaries to attract qualified teachers, created strong preparation programs so that teachers are effective with the students they will teach, and provided mentors have shown how we can fill classrooms with well-prepared teachers (Darling-Hammond & Sykes, 2003).

But solving this problem everywhere requires a national agenda. The distributional inequities that lead to the hiring of unqualified teachers are caused not only by disparities in pay and working conditions, but also by interstate barriers to teacher mobility, inadequate recruitment incentives to distribute teachers appropriately, and fiscal conditions that often produce incentives for hiring the least expensive rather than the most qualified teachers. And while the nation actually produces far more new teachers than it needs, some specific teaching fields experience real shortages. These include teachers for children with disabilities and those with limited English proficiency as well as teachers of science and mathematics. Boosting supply in the fields where there are real shortfalls requires targeted recruitment and investment in the capacity of preparation institutions to expand their programs to meet national needs in key areas.

While No Child Left Behind sets an expectation for hiring qualified teachers, it does not yet include the policy support to make this possible. In other high-achieving nations we consider peers or competitors, teachers receive a much more extensive preparation (usually 2 to 3 years of graduate level teacher education) entirely at government expense, including tuition and living expenses. Schools that support teachers' clinical training are overstaffed to ensure that they have adequate supports and governments support high-quality intensive mentoring in the first years of teaching (Darling-Hammond, in press). The U.S. has adopted none of these policies on a wide scale, and schools – especially those serving high-need students -- suffer from the inconsistent training and lack of support their teachers receive.

The federal government should play a leadership role in providing an adequate supply of well-qualified teachers just as it has in providing an adequate supply of well-qualified physicians for the nation through the Medical Manpower Act and the Health Professions Education Assistance Act. These have supported medical training, created and strengthened teaching hospitals, provided scholarships and loans to medical students, and created incentives for physicians to train in shortage specialties and to locate in underserved areas. Similar federal initiatives in education were effective during the 1960s and '70s but were eliminated in the 1980s. We need a federal teacher policy that will:

- recruit new teachers who prepare to teach in high-need fields and locations, through service scholarships and forgivable loans that allow them to receive high quality teacher education and pay it back through service in teaching;
- reduce barriers to interstate mobility by supporting the development of a national license based on a performance assessment of teaching skills, so that teachers can more easily get from places with surpluses to areas with shortages (National Academy of Education, 2005);
- strengthen teachers' preparation through incentive grants to schools of education to create professional development schools, like teaching hospitals, to train prospective teachers in urban areas and to expand and improve programs to prepare special education teachers, teachers of English language learners, and other areas where our needs exceed our current capacity; and
- improve teacher retention and effectiveness by ensuring they have mentoring support during the beginning teaching stage when 30 percent of them drop out (for a discussion, see Darling-Hammond & Sykes, 2003; NAE, 2005).

Some universities have designed and launched new and redesigned high schools to serve as professional development schools (PDS) for the training of their teachers. These PDS sites allow new teachers to learn to practice effectively in these kinds of schools specifically – and provides a critical pipeline for staffing and expanding high-performance high schools that are quite different from the organizations they are replacing. The recently drafted TEACH Act includes provisions to support such professional development schools and has introduced other elements of such a Marshall plan for teaching. However, a sustained effort to build a national infrastructure has not been mounted. For the equivalent of one week’s combat costs in Iraq, the nation could completely eliminate teacher shortages and produce more competent teaching force, providing top quality preparation for enough new teachers annually to fill all of the vacancies currently filled by underprepared teachers -- and mentor all of the new teachers who are hired over the next five years. (For details, see Darling-Hammond & Sykes, 2003). To ensure that schools can “highly qualified teachers,” this kind of focused initiative is needed.

Fixing Testing and Accountability Provisions

The goals of No Child Left Behind are to improve achievement for all students, to enhance equity, and to ensure more qualified teachers. However, its complex regulations for showing “Adequate Yearly Progress” toward test score targets aimed at “100% proficiency” have created a bizarre situation in which most of the nation’s public schools will be deemed failing within the next few years – even many that already score high and those that are steadily improving from year to year. Ironically, states that use more ambitious tests and have set higher standards for themselves will experience greater failures than those with low standards, and many have abandoned assessments that measure critical thinking and performance, just as the labor market increasingly demands these kinds of skills.

The accountability provisions of NCLB have been the subject of much analysis and considerable protest. In particular, the requirement that schools make “adequate yearly progress” toward annual test score goals for every school – until 100% of students score at the “proficient” level on state tests by the year 2014 – was set without an understanding of what this goal would really mean. Recent studies have suggested that, even under the most optimistic scenarios, at least 80% of schools in most states will have failed to achieve “AYP” by 2014 (see, for example, Wiley, Mathis, & Garcia, 2005) and in diverse states like California, 99% of schools are expected to “fail” by this date (Packer, 2004).

One fundamental problem is that the Act’s goals are unrealistic. Using a definition of proficiency benchmarked to the National Assessment of Educational Progress (NAEP), one leading measurement expert has calculated that it would take schools more than 160 years to reach such a target in high school mathematics if they continued the fairly brisk rate of progress they were making during the 1990s (Linn, 2003). In addition to these unrealistic goals there are several especially problematic aspects of NCLB for diverse urban schools, especially high schools, which create strong disincentives for high quality instruction and for efforts to keep the most vulnerable students in school. These include the following:

- Disincentives for using intellectually ambitious performance assessments;
- A “diversity penalty” experienced by schools serving many groups of high-need students, especially for assessing the progress of English language learners; and
- Incentives for pushing students out of school in order to boost test scores.

The necessary changes to address these problems are described below.

4. The law and regulations should encourage states and schools to use performance assessments that motivate ambitious intellectual work and are better measures of serious

high school learning. As noted earlier, performance assessments that require students to evaluate and solve complex problems, conduct research, write extensively, and demonstrate their learning in projects, papers, and exhibitions have proven key to motivating students and attaining high levels of learning in redesigned high schools. These kinds of assessments are the norm in European and Asian high schools, whose examination systems rely on essays and oral examinations, as well as student work products, rather than multiple choice tests. Although No Child Left Behind explicitly calls for the use of multiple measures to evaluate student and school progress, it has been administered in ways that reduce the options for states to maintain robust assessment systems that go beyond multiple choice tests, and does not encourage states to evaluate schools based on multiple measures of learning that include performance tasks. Thus, one of the first perverse consequences of NCLB was that many states that created forward-looking assessment systems during the 1990s have begun to abandon them, since they are more costly than machine scored multiple choice tests, and do not fit the federal mandate for annual testing that allows students and schools to be ranked and compared.

For example, NCLB's test requirements and costs caused Maryland to drop its sophisticated performance assessment system and Maine to eliminate performance assessments in some fields, as well as its teacher scoring process which provided strong professional development. Oregon fought to get the Department of Education to allow it to use its sophisticated computer-based adaptive testing system for the purposes of both diagnosis for instruction and standards-based assessment it was designed to serve. States like Nebraska that previously used only performance assessments to evaluate student learning have been forced to adopt norm-referenced standardized tests to meet the law's requirements (Erpenbach et al., 2003). States that have built systems relying on multiple measures including performance

assessments at the school level have received strong discouragement from the Department of Education. Relatively few states now encourage high schools to engage in performance assessments or acknowledge such assessments in state accountability systems, thus reducing the incentives for schools to focus on higher order thinking and performance skills. The administration of NCLB has pushed states back to the lowest common denominator in testing, undoing progress that had been made to improve the quality of assessments and delaying the move from antiquated norm-referenced, multiple-choice tests to more thoughtful systems that measure and develop important kinds of performance and learning.

Analysts have raised many concerns about how the law's requirements are leading to a narrower curriculum; to test-based instruction that ignores critical real world skills, especially for lower-income and lower-performing students; and to less useful and engaging education. These influences on assessment not only reduce the chances that schools will be able to focus on helping students acquire critical thinking, research, writing, and production abilities; they also reduce the opportunities that students who learn in different ways and have different talents will have to show what they have learned.

Amendments to the law and regulations should encourage rather than discouraging the use of diagnostic assessments and high-quality state or local performance assessments as a key part of state accountability systems aimed at improving curriculum and teaching. Unless more federal funds are available to support high-quality annual testing that includes performance components, the requirement for annual testing should be relaxed so that states can afford to maintain such high-quality assessments. And evidence about school progress should be expanded to include information from multiple measures, including performance assessments and information about student learning opportunities and progression through school.

5. The law needs to be amended so that AYP calculations accurately assess student progress – including the progress of English language learners and students with disabilities -- and do not penalize small schools serving the most diverse student bodies.

NCLB requires that schools be declared “failing” if they fail to meet test score targets for each subgroup of designated students annually. It requires the largest gains from lower performing schools, although these schools serve needier students and generally have fewer resources than those serving wealthier and higher scoring students.

Two separate teams of researchers have found that schools serving poor, minority, and LEP students and those with a greater number of subgroups for which they are held accountable experience what researchers have called a “diversity penalty” (Novak & Fuller, 2003; see also, Sunderman and Kim, 2004), even when they show large test score gains for low-income and minority students. This occurs because schools must meet test participation rates and test score gains for each subgroup on each test to “make AYP”, with each racial/ethnic and income group, plus English language learners and students with disabilities, counted separately. Thus, a diverse school responsible for several tests each year might need to meet each of more than 30 separate targets, while a homogenous school serving few low-income students or English language learners might need to show progress in only 5 or 6 categories.

Of two schools with identical overall gains, one may easily make adequate yearly progress while the other, more diverse school, does not. Even if students in every single subgroup make gains, the gains made by one group may be smaller than required, or a single subgroup may have 94% of students take one test in one grade rather than the required 95%. In some small schools, the absences of just a couple of students – or requested waivers from tests by a couple of parents -- can cause participation rates to dip below 95%. In one small school

with a large immigrant population in California, for example, the long month-long absence of two migrant students who returned to Mexico with their parents was enough to cause the school, which experienced large test score gains for all subgroups, to fail to make AYP. Small schools with many subgroups serving transient populations are especially vulnerable to this problem.

Perverse outcomes are frequently reported. For example, in schools closing the achievement gap, the gains made by African American and Latino students or low-income students may be quite steep while the gains made by high-scoring white students or higher income students who are already near the test ceiling are not as large, so the school is penalized for beginning to close the achievement gap (Darling-Hammond, 2004). In schools that serve special education students, allowing more than 1% of them to be tested in accordance with their individualized education plans will cause the school to fail to make AYP, even if all groups have improved, including special needs learners.

Schools that serve large numbers of new English language learners (what the law calls “Limited English Proficient” (LEP) students) and students with disabilities are subject to the most nonsensical rules, which guarantee that they cannot ultimately meet the law’s standards. Since students are assigned to these subgroups *because* cannot meet the proficiency standard, and they are typically removed from the subgroup when they do meet the standard, these schools will not ever be able to meet the 100% proficiency benchmark the law has set.

For example, section 9101(25)(D) of NCLB defines a LEP student as one “whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual — (i) *the ability to meet the State's proficient level of achievement* on State assessments described in section 1111(b)(3)....” It seems not to have occurred to policymakers that ordering schools to show 100% proficiency for students in a

subgroup that by definition scores below that level on state tests creates an impossible goal. Furthermore, as students gain proficiency in English, they are transferred out of this subgroup; thus, it is impossible for 100% of this subgroup ever to reach proficiency. For schools and districts that serve substantial numbers of LEP students, this imposes a ceiling on their overall performance as well as the performance of the subgroup. At some point it will be impossible to make the required gains because of how this subgroup is defined under law. Some advocates have suggested that States be able to count scores of students who are classified as LEP in the AYP calculations for this subgroup as long as they stay in a school (even after they become proficient in English – and presumably come within reach of achieving proficiency on the state content tests). However, the US Department of Education has not approved this definition (Erpenpach, Forte-Fast, & Potts, 2003).

The same issues pertain to the testing of students with disabilities and to the schools that serve them. Many such students who cannot demonstrate their learning on grade-level tests have individualized education plans that prescribe different assessments for charting their progress, including “instructional level” tests. The Department of Education has ruled that using such tests is permissible only if the results are counted as “non-proficient” or, for one year only, if they apply to fewer than 1% of all test-takers. In addition to the fact that this appears to violate special education laws, schools that serve large numbers of special education students will always be penalized in their AYP rankings. Furthermore, because disabilities are correlated with poverty (which is linked to poor prenatal and childhood health care, low birth weight, poor nutrition, lead poisoning, maternal substance abuse, and many other conditions that predict learning problems), this rule punishes schools that serve large numbers of low-income students.

Amendments to the law and regulations need to relax subgroup participation rates to a more reasonable level (e.g. 90%), credit gains in each subgroup without applying mechanistic rules for targets, allow the appropriate IEP testing of special needs learners in all cases, require appropriate modes of testing for all English language learners – in line with published professional testing standards -- and either abandon the impossible 100% proficient rule for this subgroup or modify it to pertain only to ELL learners who have had a reasonable amount of time to learn English (e.g. at least 3 years), keeping all of these ELL learners in the subgroup throughout their school careers for the purpose of calculating progress.

6. The flawed system of measurement – which tracks group averages rather than value-added learning gains – need be changed so as to eliminate current incentives to push out students in order to raise scores. Perhaps the most adverse, unintended consequence of NCLB’s accountability strategy is that it undermines safety nets for struggling students rather than expanding them, and it creates incentives for such students to be kept out or pushed out of school. As low-scoring students disappear, test scores go up. Because the law does not require tracking student progression and graduation rates, and because it requires inappropriate testing of ELL and special needs students, the most expedient thing for schools to do to get their scores up is to allow or even encourage such students to leave.

Table 1 shows how this operates. At “King High School,” average scores increased from the 70th to the 72nd percentile between the 2002 and 2003 school year, and the proportion of students in attendance who met the standard (a score of 65) increased from 66% to 80% -- the kind of performance that test-based accountability systems, including NCLB, celebrate and reward. Looking at subgroup performance, the proportion of Latino students meeting the standard increased from 33% to 50%, a steep increase. However, *not a single student* at King

improved his or her score between 2002 and 2003. In fact, the scores of every single student in the school went *down* over the course of the year. How could these steep improvements in the school's average scores and proficiency rates have occurred? A close look at Table 1 shows that the major change between the two years was that the lowest-scoring student, Raul, disappeared. As has occurred in many states with high stakes-testing programs, students who do poorly on the tests – special needs students, new English language learners, those with poor attendance, health, or family problems – are increasingly likely to be excluded by being counseled out, transferred, expelled, or by dropping out.

Table 1
King Middle School: Rewards or Sanctions?
The Relationship between Test Score Trends and Student Populations

	2003-04	2004-05
Laura	100	90
James	90	80
Felipe	80	70
Kisha	70	65
Jose	60	55
Raul	20	
	Ave. Score = 70 % meeting standard = 66%	Ave. Score = 72 % meeting standard = 80%

If this school had been judged using a “value-added” index that looked at the changes in individual students’ scores from one year to the next, it would have been clear that the students’ scores decreased by 8 percentile points on average rather than registering an apparent, but illusory, gain caused by changes in the student population. Recent studies have found that systems that reward or sanction schools based on average student scores (rather than looking at the growth of individual students) create incentives for pushing low-scorers into special education so that their scores won't count in school reports (Allington and McGill-Franzen, 1992; Figlio & Getzler, 2002), retaining students in grade so that their grade-level scores will look better (Jacob, 2002; Haney, 2000), excluding low-scoring students from admissions

(Darling-Hammond, 1991; Smith et al., 1986), and encouraging such students to leave schools or drop out (Haney, 2000; Orfield & Ashkinaze, 1991; Smith et al., 1986). Studies have linked dropout rates in many states and cities to the effects of grade retention, student discouragement, and school exclusion policies stimulated by high stakes tests (Haney, 2002; Advocates for Children, 2002; Jacob, 2001; Lilliard & DeCicca, 2001; Heubert & Hauser, 1999; Orfield & Ashkinaze, 1991; Roderick et al., 1999; Wheelock, 2003).

Meanwhile, steep increases in test scores have often occurred in schools with the highest retention and dropout rates. For example, Wheelock (2003) found that, in addition to increasing dropout rates, Massachusetts high schools receiving state awards for gains in 10th grade pass rates on the state tests showed substantial increases in prior year 9th grade retention rates and in the percentage of “missing” tenth graders. A similar relationship between 9th grade retention and dropouts and school rankings was found in a Texas study (Heilig, 2005). Schools that work hard to keep struggling students in school are disadvantaged by accountability systems like those encouraged by NCLB. Paradoxically, NCLB’s requirement for disaggregating data and tracking progress for each subgroup of students increases the incentives for eliminating those at the bottom of each subgroup, especially where schools have little capacity to improve the quality of services such students receive.

The consequences for individual students who are caught in this no-win situation can be tragic, as most cannot go on to further education or even military service if they fail these tests, drop out, or are pushed out to help their schools’ scores look better. The consequences for society are also tragic, as such policies lead to more students leaving school earlier – some with only a 7th or 8th grade education -- without the skills to be able to join the economy. These students join what is increasingly known as a “school-to-prison pipeline” carrying an increasing

number of undereducated youth almost directly into the criminal justice system. Indeed, prison enrollments have tripled since the 1980s and the costs of the criminal justice system have increased by more than 600% (while public education spending grew by only 25% in real dollars). More than half of inmates are functionally illiterate and 40% of adjudicated juveniles have learning disabilities that were not diagnosed or treated in school (Darling-Hammond, 2004). The annual costs of incarceration are 3 to 5 times the cost of educating the same individuals in schools years earlier, and the annual costs of dropouts are extremely high. For example, Rumberger estimates the partial costs of one year of dropouts in California at \$91 billion – counting only lost governmental income and incarceration costs. Increasingly, this growing strain on the economy deflects resources away from the services that could make people productive. If test scores are increased by pushing students out of school, the end result is not higher levels of education in the society.

Addressing this problem will require using value-added measures of individual student progress rather than cross-sectional averages that compare one year's average scores to the next. It will also require valuing keeping students in school as part of the accountability system, greater investments in improving the capacity of schools to teach – not just test -- struggling students, and appropriate means for assessing students with special needs.

While these are troubling aspects of the law's implementation, one could also argue, quite legitimately, that at least some of the schools identified as "needing improvement" (a designation that changes to "failing" if targets are not met after 3 years) indeed are dismal places where little learning occurs, or are complacent schools that have not attended to the needs of all of their students – schools that need to be jolted into action to change. And, it is fair to suggest that underserved students in such schools deserve other choices if they cannot change.

These important arguments are part of the law’s theory of action: that low quality schools will be motivated to change if they are identified and shamed, and that their students will be better served if given other educational options. These outcomes may in fact occur in some cases. The problem is that the law actually works in many other cases to label schools as failing even when they are succeeding with the very students the law wants to help, and it creates incentives that can reduce the quality of education such schools can provide.

How might the goal of improving schools actually, paradoxically, undermine them? Evidence suggests that applying “failing schools” labels to low-scoring schools that serve low-income students actually reduces the schools’ ability to attract and keep qualified teachers. For example, in North Carolina, analysts found that the state labeling system made it more difficult for the neediest schools to attract or retain the high-quality teachers (Clotfelter et al., 2003). Florida’s use of aggregate test scores, unadjusted for student characteristics, in allocating school rewards and sanctions led to reports that qualified teachers were leaving the schools rated “D” or “F” in droves, to be replaced by teachers without experience or training (DeVise, 1999). As one principal queried, “Is anybody going to want to dedicate their lives to a school that has already been labeled a failure?” NCLB’s approach of labeling schools and threatening staff dismissals has been reported as a disincentive for qualified staff to stay in high-need schools when they have options to teach in better resourced and better regarded schools with more affluent students.

Fixing NCLB

If we are to achieve the noble goals of NCLB, the law must be amended so that states have flexibility and encouragement to use thoughtful performance assessments and so that tests are used diagnostically for informing curriculum improvements rather than for punishing students or schools. Progress should be evaluated on multiple measures – including such factors

as attendance, school progress and continuation, course passage, and classroom performance on tasks beyond multiple choice tests – and gains should be evaluated with “value-added” measures showing how individual students improve over time, rather than school averages that are influenced by changes in who is assessed.

Targets should be based on sensible goals for student learning that also ensure appropriate assessment for special education students and English language learners and credit for the gains these students make over time. While progress for subgroups of students should be reported, these reports should include evidence about continuation and success in school as well as academic achievement for members of each group. Determinations of school progress should be constructed to reflect a better grounded analysis of schools’ actual performance and progress rather than a statistical gauntlet that penalizes schools serving the most diverse populations. These reporting changes should be designed to ensure that schools that are identified as failing are indeed those that are offering poor education, not those merely caught in a mathematical mousetrap. And progress should be gauged against sensible benchmarks for success. As policy analyst Bruce Fuller (2004) notes of the law’s current 100% proficiency standard:

Would government ever require automakers to produce emissions-free cars in the space of a decade, then shut down companies that failed to meet a pie-in-the-sky goal? Of course not! Better to set demanding yet pragmatic standards and require clear signs of progress. Schools should be rewarded for elevating achievement levels by some degree, rather than penalized for not meeting an absolute, unrealistic standard. The ideal level of proficiency for all -- just like emissions-free cars -- could then be approached gradually, over time.

Most important, schools that are struggling should receive intensive help to strengthen their staffs and adopt successful programs. Definitions of “highly qualified teachers” should ensure that candidates are fully prepared and have demonstrated competence before they receive such recognition. At the same time, such definitions should not create straitjackets that prevent

the inter-disciplinary teaching that can create more engaging and coherent curriculum and that enables small schools to personalize instruction. Full funding of NCLB should include supports and incentives for preparing well-qualified teachers and getting them to the schools where they are needed, including a major federal initiative to underwrite strong preparation and recruitment incentives for well-qualified teachers who will teach in high-need schools.

In addition to incentives for recruiting and retaining high quality teachers in the places where they are most needed, fixing No Child Left Behind will require a new approach to measuring and supporting school success. This approach should:

- Replace the counterproductive federally-mandated “AYP” formula with less rigid and more instructionally useful state accountability systems designed to support and assess student progress through multiple measures, including performance assessments, attendance, and student continuation in and progress through school.
- Encourage rather than discouraging the use of diagnostic assessments and high-quality state or local performance assessments as a key part of state accountability systems aimed at improving curriculum and teaching rather than punishing schools.
- Evaluate gains using "value-added" approaches that assess the progress of individual students, not changes in average student scores that penalize schools that serve the neediest students or encourage schools to push out low-scoring students.
- Appropriately assess the progress of English language learners and students with disabilities based on professional testing standards and ‘count’ the gains of these students throughout their entire school careers, rather than only for the time they are classified in these categories.

These changes will reward the efforts of those high schools that have redesigned themselves to better serve the students who are routinely left by the way side in their adolescent years, rather than penalizing or obstructing their efforts to keep students in school and enable them to learn in rigorous and relevant ways. If we really want to Leave No Child Behind, the law should create “two-way accountability” – accountability to parents and children for the quality of education they receive as a means for greater learning for all.

References

- Advocates for Children (2002). Pushing out at-risk students: An analysis of high school discharge figures -- a joint report by AFC and the Public Advocate. Retrieved [October 10, 2005] from: <http://www.advocatesforchildren.org/pubs/pushout-11-20-02.html>
- Allington, Richard L. and Anne McGill-Franzen (1992). Unintended Effects of Educational Reform in New York, Educational Policy, 6 (4): 397-414.
- Bensman, D. (1987). Quality education in the inner city: The story of the Central Park East Schools. NY: Report to the New York Community Trust.
- Braddock, J.H. & McPartland, J.M. (1993). The education of early adolescents. In L. Darling-Hammond (ed.), Review of research in education, Vol. 19, pp. 135-170. Washington, D.C.: American Educational Research Association.
- Clotfelter, C., Ladd, H., Vigdor, J., & Diaz, R. (2003, February) (revised). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? Paper prepared for the annual meeting of the American Economic Association, Washington, DC.
- Darling-Hammond, L. (1997). The right to learn: A blueprint for creating schools that work. San Francisco: Jossey-Bass.
- Darling-Hammond, L. (1991). The Implications of Testing Policy for Quality and Equality, Phi Delta Kappan, November 1991: 220-225.
- Darling-Hammond, L. (2000). Teacher Quality and Student Achievement, Educational Policy Analysis Archives, 8, (1). <http://epaa.asu.edu/epaa/v8n1>.
- Darling-Hammond, L. (2004). What happens to a dream deferred? The continuing quest for equal educational opportunity. In James A. Banks (ed.), Handbook of Research on Multicultural Education, 2nd Edition, pp. 607-630. San Francisco: Jossey-Bass, 2004.
- Darling-Hammond, L. (2004). From "Separate but Equal" to "No Child Left Behind": The Collision of New Standards and Old Inequalities. In Deborah Meier and George Wood (eds.), Many Children Left Behind. NY: Beacon Press, 2004
- Darling-Hammond, L. (in press). Teaching as a Profession: International lessons in teacher preparation and professional development. Phi Delta Kappan.
- Darling-Hammond, L., Aness, J., & Ort, S. (2002). Reinventing High School: Outcomes of the Coalition Campus School Project. American Educational Research Journal, 39, (3): 639-673.

Darling-Hammond, L. & Sykes, G. (2003). Wanted: A National Teacher Supply Policy for Education: The right way to meet the 'highly qualified teacher' challenge. Educational Policy Analysis Archives, Vol. 11, No. 33 (September 2003). <http://epaa.asu.edu/epaa/v11n33/>.

DeVise, D. (1999). A+ Plan Prompts Teacher Exodus in Broward County. *Miami Herald*, Nov. 5, 1999.

Erpenpach, W.J., Forte-Fast, E., & Potts, A. (2003, July). Statewide Educational Accountability under NCLB. Washington DC: Council for Chief State School Officers.

Figlio, D.N. & Getzler, L.S. (2002, April). Accountability, ability, and disability: Gaming the system? National Bureau of Economic Research.

Fowler, W.J. (1992, April). What do we know about School Size? What should we know? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Fuller, B. (2004). Only the Politicking Gets an 'A', The Washington Post. Washington, D.C.: Feb 1, 2004. pg. B.04.

Garbarino, J. "The Human Ecology of School Crime: A Case for Small Schools." In E. Wenk (ed.), School Crime, pp. 122-133. Davis, CA: National Council on Crime and Delinquency.

Goodlad, John I. (1984). A place called school. New York: McGraw-Hill.

Gottfredson, G.D. and Daiger, D.C. (1979). Disruption in 600 schools. Baltimore, MD: The Johns Hopkins University, Center for Social Organization of Schools.

Green, G. & Stevens, W. (1988). What research says about small schools, Rural Educators, 10 (1): 9-14.

Haller, E. J. (1992). High School Size and Student Indiscipline: Another Aspect of the School Consolidation Issue? Educational Evaluation and Policy Analysis, 14 (2), 145-156.

Haller, E.J. (1993). Small Schools and Higher-Order Thinking Skills. Journal of Research in Rural Education, 9 (2), 66-73.

Haney, W. (2000). The myth of the Texas miracle in education. Education Policy Analysis Archives, 8 (41): <http://epaa.asu.edu/epaa/v8n41/>

Haney, W. (2002). Lake Woebeguaranteed: Misuse of test scores in Massachusetts, Part I. Educational Policy Analysis Archives, <http://epaa.asu.edu/epaa/v10n24/>

Heilig, J. V. (2005). An analysis of accountability system outcomes. Stanford University.

Heubert, J. and Hauser, R. (eds.) (1999). High stakes: Testing for tracking, promotion, and graduation. A report of the National Research Council. Washington, D.C.: National Academy Press.

- Howley, C.B. (1989). Synthesis of the effects of school and district size: What research says about achievement in small schools and school districts. Journal of Rural and Small Schools, 4 (1): 2-12.
- Howley, C.B. & Huang, G. (1991). Extracurricular participation and achievement: School size as possible mediator of SES influence among individual students. Resources in Education, January 1992.
- Jacob, B.A. (2001). Getting tough? The impact of high school graduation exams. Education and Evaluation and Policy Analysis 23 (2): 99-122.
- Jacob, B. A. (2002). The impact of high-stakes testing on student achievement: Evidence from Chicago. Working Paper. Harvard University.
- Lee, V.E., Bryk, A.S., and Smith, J.B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (ed.), Review of research in education, Vol. 19, pp. 171-268.
- Lee, V.E. & Smith, J.B. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. Sociology of Education, 66 (3): 164-187.
- Lee, V.E. & Smith, J.B. (1995). Effects of high school restructuring and size on gains in achievement and engagement for early secondary school students. Madison: Wisconsin Center for Education Research, University of Wisconsin.
- Lee, V.E., Smith, J.B., & Croninger, R.G. (1995). Another look at high school restructuring: More evidence that it improves student achievement and more insight into why. Issues in restructuring schools. Madison: Center on Organization and Restructuring of Schools, University of Wisconsin.
- Lilliard, D. & DeCicca, P. (2001). Higher standards, more dropouts? Evidence within and across time. Economics of Education Review, 20(5): 459-73.
- Lindsay, P. (1982). The Effect of High School Size on Student Participation, Satisfaction, and Attendance." Educational Evaluation and Policy Analysis, 4, 57-65.
- Lindsay, P. (1984). High School Size, Participation in Activities, and Young Adult Social Participation: Some Enduring Effects of Schooling. Educational Evaluation and Policy Analysis, 6 (1), 73-83.
- Linn, Robert L. (2003). "Accountability: Responsibility and Reasonable Expectations." Educational Researcher, 32 (7): 3-13.
- McMullan, B.J., Sipe, C.L., & Wolf, W.C. (1994). Charter and student achievement: Early evidence from school restructuring in Philadelphia. Philadelphia: Center for Assessment and Policy Development.
- National Academy of Education (2005). A good teacher in every classroom: Preparing the highly qualified teachers our children deserve. San Francisco: Jossey-Bass.
- Newmann, F. M., Marks, H.M. & Gamoran, A. (1995, April). Authentic pedagogy and student performance. Paper presented at the meeting of the American Education Research Association, San Francisco.

- Newmann, F.M. & Wehlage, G.G. (1995). Successful school restructuring. Madison, WI: University of Wisconsin.
- Novak, J. & Fuller, B. (2003, December). Penalizing Diverse Schools? Similar test scores but different students bring federal sanctions. Berkeley, CA: Policy Analysis for California Education.
- Orfield, G., and Ashkinaze C. (1991). The Closing Door: Conservative Policy and Black Opportunity. Chicago: University of Chicago Press, p. 139.
- Packer, J. (2004, July 28). No Child Left Behind and adequate yearly progress fundamental flaws: A forecast for failure. Paper presented at the Center for Education Policy Forum on Ideas to Improve the Accountability Provisions. Washington, DC.
- Pittman, R., & Haughwout, P. (1987). Influence of High School Size on Dropout Rate. Educational Evaluation and Policy Analysis, 9: 337-343.
- Powell, F.C., Farrar, E., and Cohen, D.K. (1985). The shopping mall high school: Winners and losers in the educational marketplace. Boston: Houghton Mifflin.
- Raywid, M.A. (1990). Successful schools of choice: Cottage industry benefits in large systems. Educational Policy, 4 (2): 93-108.
- Raywid, M.A. (1995). The Wadleigh complex: A dream that soured. In W. Boyd, R. Crowson, & H. Mawhinney (eds.), The politics of education and the new institutionalism: Reinventing the American school. Philadelphia: Falmer Press.
- Roderick, M., Bryk, A.S., Jacob, B.A., Easton, J.Q. & Allensworth, E. (1999). Ending Social Promotion: Results from the first two years. Chicago: Consortium on Chicago School Research.
- Sizer, T. (1984). Horace's compromise: The dilemma of the American high school. Boston: Houghton Mifflin.
- Smith, F., et al. (1986). High School Admission and the Improvement of Schooling. NY: New York City Board of Education.
- Sunderman, G., & Kim, J. (2004). Inspiring Vision, Disappointing Results: Four Studies on Implementing the No Child Left Behind Act. Cambridge, MA: Harvard Civil Rights Project.
- Wehlage, G.G., Rutter, R.A., Smith, G.A., Lesko, N., and Fernandez, R.R. (1989). Reducing the risk: Schools as communities of support. Bristol, PA: Falmer Press.
- Wehlage, G.G., Smith, G. & Lipman, P. (1992). Restructuring urban schools: The new futures experience. American Educational Research Journal, 29 (1): 51-93.
- Wheelock, A. (2003). School Awards Programs and Accountability in Massachusetts: Misusing MCAS Scores to Assess School Quality.
<http://www.fairtest.org/arn/Alert%20June02/Alert%20Full%20Report.html>.
- Wiley, E.W., Mathis, W. J., Garcia, D. R. (2005). The impact of the adequate yearly progress requirement of the federal "No Child Left Behind" Act on schools in the Great Lakes region. Tempe, AZ: Educational Policy Studies Laboratory, Arizona State University.