

Gamoran, Adam

## **School accountability, american style: dilemmas of high-stakes testing**

*Schweizerische Zeitschrift für Bildungswissenschaften 29 (2007) 1, S. 79-94*



Quellenangabe/ Reference:

Gamoran, Adam: School accountability, american style: dilemmas of high-stakes testing - In: Schweizerische Zeitschrift für Bildungswissenschaften 29 (2007) 1, S. 79-94 - URN: urn:nbn:de:0111-opus-41820 - DOI: 10.25656/01:4182

<https://nbn-resolving.org/urn:nbn:de:0111-opus-41820>

<https://doi.org/10.25656/01:4182>

in Kooperation mit / in cooperation with:

**Schweizerische  
Zeitschrift  
für Bildungswissenschaften**  
**Rivista svizzera  
di scienze dell'educazione**  
**Revue suisse  
des sciences de l'éducation**

<http://www.rsse.ch/index.html>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)

Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

# School accountability, american style: dilemmas of high-stakes testing

**Adam Gamoran**

Support for this paper was provided by the School of Education at the University of Wisconsin-Madison. Findings and conclusions are those of the author.

*No Child Left Behind (NCLB) is a new federal law in the U.S. that requires states (if they wish to receive federal education funds) to set standards for student learning, design assessments to measure progress towards standards, and to hold districts and schools accountable for results. Accountability through high-stakes standardized testing is the latest development in a long move towards standards-based reform in U.S. education, a reform process that has enjoyed bipartisan support. It occurs in the context of what has been a «loosely coupled» system with little formal supervision and no national testing. NCLB confronts four dilemmas that any system of testing and accountability must address: It sets high standards; uses absolute targets to assess progress; accepts the narrowness of standardized tests in only two subjects as the price to be paid for clear and specific standards; and attempts (perhaps with limited success) to identify mechanisms to ensure equal opportunity to meet the standards. As a result of this approach, NCLB has galvanized the attention of educators and motivated teachers and students to improve; but it also leads to disproportionate sanctions for schools serving disadvantaged populations, mixes up effective and ineffective schools, and poses an unrealistic time frame for success.*

A new federal law has raised the stakes for high standards in U.S. education: The «No Child Left Behind» Act of 2001 (signed into law in January, 2002) calls on each of the 50 states to set absolute targets for student performance, and to hold schools accountable for results. What is No Child Left Behind (NCLB)? Will it achieve its goals? Do higher standards invariably come at a cost of greater inequality? The purpose of this essay is neither to promote nor to denigrate NCLB. Instead, it aims to identify the dilemmas that NCLB confronts, explain how those challenges are addressed, and discuss the likely consequences of NCLB's particular responses to the difficulties of raising standards.

To fully understand the aims and strategies of NCLB, it is essential to consider two aspects of the context in which the law emerged. First, the organization of schooling in the U.S., which differs in important respects from that of most advanced nations, constitutes a unique situation for standards-based reform. Second, the particular historical development of standards-based reform in U.S. education has set the stage for NCLB. With these contextual conditions in mind, we will examine the dilemmas that NCLB addresses, and the consequences that may lie ahead from this effort to improve U.S. education through testing and accountability.

## The Context of U.S. Education

Formal education in the U.S. has long operated as a «loosely coupled system» (Weick, 1976). With the term «loosely coupled,», analysts have intended to portray a system in which administrative linkages are weak, both between political jurisdictions (federal, state, local) and within them. U.S. school systems are decentralized, with little federal control over or funding of education. Most of the funding for schools comes from a combination of state support and local property taxes, with federal funds covering only about 10% of the cost of public schools (Odden and Picus, 2003). Privately-run schools receive virtually no public funding from federal, state, or local sources. Education is constitutionally a state responsibility, and in practice most of the authority for establishing and operating school systems has been left to local education authorities, or «school districts,» which are governmental units generally aligned geographically with municipalities or groups of municipalities or rural areas, though (until recently, at least), largely independent of municipal governments. In such a context, enacting meaningful federal reform that makes a difference at the local level obviously entails challenges that do not arise in more tightly linked systems.

### Schooling in the Loosely Coupled System

Within the school district, loose coupling means that activities and events taking place in one part of the system do not reverberate in clearly patterned ways elsewhere (Gamoran and Dreeben, 1986). Formal commands are relatively few, and formal inspection and supervision are relatively rare. Beginning teachers are usually observed and often carefully evaluated, but once they pass their initial probationary period, supervision has tended to be perfunctory rather than substantive. There is no formal system of inspection in U.S. schools as, for example, is commonplace in British schools (Wilson, 1996). Moreover, unlike most economically advanced nations, the U.S. has had no formal leaving examinations.<sup>1</sup> College-bound students do take examinations as part of the university application process, but these examinations are delivered by private firms and they are *deliberately disconnected from the secondary curriculum*. The rationale behind this

approach is that due to local control over school curricula, no single examination would be appropriate across the nation or even the state. According to this view, an examination that does not reflect any particular curriculum is insensitive to differences in learning opportunities across schools, giving students from different schools equal chances to succeed. In practice, of course, student performance on these examinations (the SAT, which stands for Scholastic Achievement Test, and the ACT, which signifies American College Testing) does reflect students' learning opportunities in high school. Thus, inequalities may be present, but they are obscured by the lack of attention to possible links between learning opportunities and outcomes. Inequalities in the U.S. have also been hidden by the lack of attention to test scores more generally.

Although the system has been loosely coupled, some management tasks are nonetheless fulfilled. A system of «ritual classification» - the sorting of students and teachers into schools, grade levels, and classrooms - has operated as an organizing principle, and a «logic of confidence» has governed school administration - that is, as long as visible signs of disarray are absent, classes are assumed to be operating properly (Meyer and Rowan, 1977, 1978). In the U.S. system, close attention has been paid to *who gets taught by whom*, but until recently, little attention has been given to *what gets taught* once the classroom door is closed. Coordination occurs through the common socialization of teachers, who may hold shared understandings of what is to take place within classrooms, despite the absence of inspection (Weick, 1982). The allocation of resources from the district to the school and classroom also supports coordination of efforts within schools without much use of formal commands or supervision (Gamoran & Dreeben, 1986).

The result of the U.S. approach to organizing schooling has been a system of high attainment but low performance: levels of high school completion and college enrolment have traditionally been among the highest in the world (although many other nations now eclipse U.S. rates of high school completion), whereas performance on international assessments is average at best, particularly in the secondary years. There is substantial variability in student performance, which tends to lie within schools more than between schools, and is tied to social and economic background. Secondary students in the U.S. have easy access to post-secondary schooling (many institutions are non-selective, requiring only high school completion; see Grodsky 2002), but articulation with the labor market is poor (Rosenbaum, 2001).

### **The Move to Standards-Based Reform**

Conditions of low average performance on international tests, high rates of inequality, and poor articulation with the labor market were important stimuli to standards-based education reform in the U.S. Standards-based reform did not begin with NCLB, but has a history that goes back over 25 years. A distinctive feature of the development of standards-based reform in the U.S. is its bipartisan

history. During a time of polarized federal government, education reform is almost unique in that it has had strong backers from both the Democratic and Republican parties (Loveless, in press, provides a fuller historical background).

A landmark in the move to standards-based reform was the publication of *A Nation at Risk*, from a bi-partisan commission appointed by President Ronald Reagan, a Republican (National Commission on Excellence in Education, 1983). *A Nation at Risk* decried the low educational standards in U.S. education, which the authors famously characterized as a «rising tide of mediocrity,» and declaring that if such a system had been foisted on the nation by a foreign power, it would be considered an act of war. As a result, many states increased their course work requirement for high school graduation (Wilson & Rossman, 1993). This effort was followed in the late 1980s under the leadership of the National Governors Association, which was at that time led by a relatively little known Democratic governor from the state of Arkansas named Bill Clinton. Building on the efforts of the governors, President George H.W. Bush (i.e. the elder President Bush), a Republican, created the National Education Goals Panel in 1990, whose purpose was to develop national goals for U.S. education. The culmination of the goals reform occurred with the return of the Democrats to the executive branch under President Bill Clinton in 1992. In 1994, Clinton signed the «Goals 2000» law, which set forth eight goals for U.S. education, including such empty platitudes as «All children will begin school ready to learn,» and impossible aims such as «The U.S. will be first in the world in mathematics and science» by the year 2000.

Not only did Goals 2000 espouse unrealistic aims, but there were no sanctions attached to any state, district, or school that failed to meet the high standards. It is this context into which NCLB has emerged. With bipartisan support, NCLB is an effort to *raise standards* and *hold schools accountable* for results, with specific requirements and sanctions that far exceed anything that has been previously attempted in U.S. education.<sup>2</sup>

### The Requirements of NCLB

No Child Left Behind is the 2002 reauthorization, signed by the current president, George W. Bush (a Republican) of the Elementary and Secondary Education Act (ESEA), a federal law that first came into existence in 1965 under President Lyndon B. Johnson (a Democrat). (Goals 2000 was the 1994 reauthorization of the same law.) Since education is a matter of state rather than federal authority, the U.S. government cannot compel states to follow NCLB. However, since the 1960s the federal government has tied federal funding to states' conformity to ESEA, and so far no state has refused the funding. Although federal funds account for less than 10 percent of education spending in the U.S., the federal contribution is nonetheless essential. Thus, states are effectively bound by the demands of NCLB. In this sense, NCLB represents a strong challenge to a system that has traditionally been loosely coupled.

Under NCLB, each state is required to set standards for student learning, and to develop assessments to measure student performance with respect to the standards.<sup>3</sup> Students are tested every year from grades 3-8, and once in high school, in reading and mathematics.<sup>4</sup> Student performance is reported at three levels at least, «basic,» «proficient,» and «advanced,» and each school and district is held accountable for the proportion of students scoring «proficient» and above. Results are reported not only for district and school averages, but for the average within each subgroup within districts and schools. Subgroups for which percent proficient must be reported include:

- Whites
- Blacks
- Latinos
- Asian/Pacific Islanders
- American Indian/Alaskan Native
- Students on free/reduced-price lunch (an indicator of poverty)
- English language learners
- Students with disabilities

The requirement to break out scores by demographic subgroups means that schools cannot hide a low-achieving subgroup behind a high average score. Subgroups must reach a minimum size to be counted (the size varies by state but is generally around 40 students), but given the minimum number, schools or districts will fall short of standards if *any one* of the subgroups fails to meet the required level of proficiency. Schools cannot escape the accountability measures by increasing the portion of students in special education, since students with disabilities are held to the same standards (up to 3 percent of special education students can be given an alternative assessment). Nor can schools avoid accountability by failing to administer the assessments to low-achieving students, because they also fall short of standards if fewer than 95 percent of their students are assessed.

No Child Left Behind contains additional requirements intended to reduce problems of inequality. As of 2006-07, states are required to place «highly qualified teachers» in every classroom; that is, teachers who hold college degrees, are certified by states, and exhibit subject-matter competence. In the past, schools with more low-income students have tended to have lower proportions of teachers who met these requirements. In addition, special funds are available to enhance teaching, but only if teachers engaged practices supported by «scientifically-based research,» that is, research supported with rigorous methodology, generally involving experimental designs.

Sanctions under NCLB are primarily aimed at schools (though funded primarily by districts). After two years of failing to meet standards, schools are required to submit a plan for improvement, offer professional development to teachers, and allow students to transfer to any public school that has not failed to meet standards. After three years, schools must provide free tutoring to low-

achieving, low-income students. After four years, «corrective action» may be taken, ultimately resulting in the closing of schools, the dismissal or transferring of staff, the reconstitution of schools as public charter schools free of district governance, and so on.

It should be clear that NCLB differs from its standards-based reform predecessors in important ways. It sets clear standards, requires performance assessments, and holds schools and districts accountable for progress towards meeting standards. Even as NCLB is being implemented across the nation, it has provoked widespread debate, with supporters and detractors from a wide range of perspectives (Loveless, in press). What are the prospects of this innovative law to raise performance levels and reduce inequality in U.S. education? One way to address this question is to identify the dilemmas confronted by NCLB, pinpoint the law's responses to the dilemmas, and consider alternatives.

### **Dilemmas of High-Stakes Testing under No Child Left Behind**

In my view, any system of test-based accountability must confront four basic dilemmas. NCLB has adopted distinctive responses to these dilemmas, which are reflected in the likely consequences of NCLB. The dilemmas are:

1. How high should the standards bar be set?
2. What counts as progress towards standards?
3. What do standardized tests measure?
4. Is it fair to set high standards for all when opportunities for learning are unequal?

By understanding how NCLB responds to these dilemmas, we can identify NCLB's likely consequences, and consider how revisions to the law, which may occur when it undergoes reauthorization in 2007 or soon thereafter, might change the outcomes.

#### **How High Should the Standards Bar Be Set?**

If the purpose of standards-based reform is to create incentives that will lead to better performance, then clearly the standards must be set high enough to create such incentives. If standards are low, as has been the case with «minimum competency tests» in the U.S., then they carry little relevance for most students and teachers. Thus, the drive to use standards to motivate improved performance should result in setting standards that are very high.

Yet if standards are set high, many will fail. Introducing a new set of standards that causes many schools to be labeled as failing, or many students to be prevented from advancing or graduating, introduces significant political difficulties. This is particularly the case in a system like that of the U.S., in which students from disadvantaged minority groups are overrepresented among those who score

low on standardized tests. Even if high standards motivate schools and students to improve over time, many will fail when the system is first implemented. Thus, the framers of accountability systems face a dilemma: set the bar high, and risk creating a large, new class of failures, or set the bar low, and give up on the chance to create incentives for improvement among most of the targets of the accountability system.

NCLB's response to this dilemma is to set the bar very high: by the year 2013-2014, each school in the nation must have 100 percent of its students scoring above the «proficient» threshold on state accountability tests. Districts and schools are held accountable for these results, and many districts, in turn, are holding students accountable by basing decisions about promotion and/or graduation, in part or in whole, on student performance on standardized assessments.

What are the consequences of NCLB's approach to setting the standards bar at a high level? While it is still early to judge the outcomes of NCLB, initial indications both suggest changes that are both consistent with NCLB's aims, and point to other directions that might be characterized as unintended consequences. On the one hand, NCLB has clearly gotten the attention of educators across the nation. It seems that hardly a discussion takes place without mention of NCLB. Moreover, a study of NCLB's implementation has found that among teachers' responses to NCLB is a desire to seek new approaches to instruction and new avenues for professional development (Hamilton, Stecher & Berends, 2005). In principle, this may lead to improved classroom instruction. Another intended consequence of NCLB that is clearly occurring is that inequalities that were once obscured have now been revealed. In particular, schools with high achievement overall but large gaps between subgroups can no longer hide their low-achieving disadvantaged group behind their high schoolwide averages. Thus, NCLB has focused attention on inequality within as well as between schools, just as its framers intended.

At the same time, the imposition of high standards has created tension and conflict in many schools and districts. Anecdotal accounts indicate that being listed as a «school in need of improvement» is demoralizing to staff, leading to transfers. Finding leaders for struggling schools is apparently becoming more difficult - while many leaders may relish the challenge of improving a low-performing school, when standards are set so high that the likelihood of success seems small, good leaders may be in short supply. Moreover, if the most successful students in a struggling school take advantage of NCLB's year-2 sanctions to transfer out, that will make the challenge of meeting standards even greater. In addition, as Kane and Staiger (2002) have shown, more diverse schools have greater chances of failing to meet the standards simply by virtue of having to meet more targets, that is, a target for each of its demographic subgroups. Thus, by taking a firm stand on the importance of high standards, NCLB has produced both increased motivation for improvement and disparate sanctions on schools populated by low-income, low-achieving students.



### What Counts as Progress towards Standards?

Under NCLB, the measure of success is called «Adequate Yearly Progress» (AYP), an indicator of whether a school is keeping pace with a schedule under which 100 percent of students will meet the proficiency threshold by 2013-14. Under NCLB's rules, each state set a baseline level of performance in 2001-02, based on the percent proficient in that year in the school in which the student at the 20th percentile in the state's achievement distribution was located. From that baseline, each school must bring all students to proficiency in the course of 12 years. Baseline levels vary widely across states, from as low as 20 percent proficient to as high as over 60 percent, but all states must reach the same end point. (Generally, variation in baseline proficiency levels do not reflect differences in students' cognitive skills across states, but rather indicate differences in the difficulty levels of state assessments; see Porter, 2006).

One approach to fixing the amount of required yearly progress is to divide the gap between the baseline and 100 percent by 12 years, and then to take the result as the amount of progress that must be achieved in each year. Many states, however, have selected a «stair-step» approach, in which no progress is expected for the first two years, then percent proficient is set at a new plateau, which remains for two more years before moving to a higher threshold, and so on. The «stair-step» approach can only be maintained for a few years, however, until the end point is approached and a large increase in proficiency levels must occur every year. In my state of Wisconsin, for example, the baseline proficiency level in mathematics was 37 percent proficient, and that level remained for two years, so that a school was considered making adequate yearly progress in mathematics if 37 percent or more of its students (in all subgroups over the minimum size) scored at or above the proficiency threshold. The target was raised to 47.5 percent in 2004-05, and will rise to 58 percent in 2007-08 and 68.5 percent in 2010-11. Thus, for the first 10 years of NCLB, the required mathematics proficiency levels go up by 3.5 percentage points per year. After that, however, the percentages must rise by 10.5 points *every year* for three years to reach 100 percent by 2013-14.

This approach to setting AYP levels, which has been adopted by many states, pushes forward the day of reckoning, in which large numbers of schools will fail to make AYP. Even so, the number of such schools in Wisconsin, which was small in the early years of NCLB, is already starting to rise, as it is in many other states.

Among the problems with this approach to counting progress towards standards, three stand out. First, as Linn (2003) and Metz (in press) have shown, expecting all schools to bring 100 percent of students across the proficiency threshold is unrealistic. To the best of my knowledge, *no* school in the U.S. has *ever* made the degree of progress that the *typical* school will have to make to meet the demands of NCLB.5 Bringing a school from, say, 60 percent to 100 percent of students proficient on a test with high standards is an immense challenge whose feasibility has yet to be demonstrated.

Second, the use of absolute targets to identify «progress» recognizes schools with high and low achievement levels, but fails to distinguish between effective and ineffective schools. Clearly, a school with high test scores that reflect an advantaged student population may have contributed little to their students' abilities to score proficient. Yet a school with low test scores due to, for example, a disadvantaged population or an ineffective feeder school, may be contributing greatly to student performance, yet fail to make «adequate yearly progress» due to the increasingly high standards. Linn (2003) demonstrated that this situation is entirely plausible. Thus, an ineffective school with high achievers may be designated as succeeding and an effective school with low achievers may be designated as failing, under a system that relies on attaining absolute targets to indicate progress.

A third problem with the AYP system is that it recognizes improvements only at the proficiency threshold, rather than throughout the achievement distribution. A rationale for focusing on bringing students to proficiency is that it will lead schools to direct resources to its low achievers. In fact, however, schools may have incentives to emphasize improvement among students who are just below the cutoff, rather than among all low achievers. Bringing struggling students from very low scores to moderately better scores may be a major accomplishment, but it receives no credit towards AYP if students do not reach the proficiency target. For schools with low-achieving populations, several years of improvement below the threshold may be necessary, but under NCLB's time frame, a school at the bottom of the distribution that is improving might be disbanded and reconstituted before enough students reach the ever-increasing proficiency threshold.<sup>6</sup>

### **What Do Standardized Tests Measure?**

NCLB requires testing in two subjects: reading and mathematics. (As of 2006-07, science is also tested in selected grades, but no sanctions are attached to science performance.) Many writers have raised concerns about the exclusive reliance on multiple-choice tests in just two subjects to hold schools accountable (e.g., McNeil, 2000). The emphasis on reading and mathematics, for example, is likely to drive away time for subjects such as art, music, physical education, and social studies. While test items are challenging, they are not «authentic,» in the sense of reflecting real-world problems or asking students for original ideas. Generally, state tests have few multi-step problems and offer little opportunity to assess higher-order thinking. This approach to testing, which is *not* mandated by NCLB but which has been adopted by the states, reflect both cost considerations, and a skepticism about the reliability of more open-ended assessments.

Any high-stakes assessment creates incentives for «teaching to the test.» The question, therefore, is whether the test measures something valuable, and whether «teaching to the test» means teaching a curriculum that matters, rather than simply instructing students how to respond to a particular assessment in-

strument. In these regards, testing under NCLB faces a dilemma. On the one hand, the tests are narrow and targeted towards the lower end of cognitive skill demands. On the other hand, one may argue that improving student performance in these areas - in particular, raising the performance of low-achieving students in basic skills - is an important step towards reducing inequality in U.S. education.

### High Standards without Equal Opportunity?

Inequality of opportunity is pervasive in U.S. education, where schools have different funding levels, are staffed by teachers with varying levels of preparation, and have students who come from unequal home environments and neighborhoods (see, e.g., Odden and Picus, 2003). Against this background, is it reasonable to hold all schools to the same performance standards? Once again, standards-based reform efforts face a dilemma. If all schools are held to the same standards, that may be unrealistic. Yet if different standards are set for schools, for example, with different student populations, that may be viewed as a capitulation to lower expectations for disadvantaged students, which President Bush (2004) and others in the administration refer to as the «soft bigotry of low expectations.» By setting the bar lower for students from disadvantaged background, we may doom them to never catching up to their more advantaged peers.

NCLB has a response to the argument that opportunities are too unequal to expect low-achieving students to catch up and reach proficiency by 2014. Provisions that mandate qualified teachers in every classroom, instruction based on «scientific research,» tutoring, and the option to transfer are all expected to narrow the opportunity gaps. Yet the amount of progress that is required under NCLB far outstrips the impact of any known intervention, in these and other areas. After all, most of the variation in student achievement is tied to conditions outside schools; while schooling matters for student learning, its power to equalize is modest next to the disqualifying forces in the wider society (for recent arguments along these lines see Rothstein, 2004).

Value-added achievement analysis is an alternative approach to examining school performance (e.g., Meyer, 1996, 2003). Rather than examining absolute targets, value-added analysis examines the contributions of schools (or teachers) to student performance, taking into account their prior performance levels and other relevant characteristics, e.g. race/ethnicity and economic status. A value-added approach to accountability focuses on the *effectiveness* of schools rather than the schools' achievement status at a given point in time. While gain scores are less reliable than raw scores, they are more valid as indicators of what the school has contributed to student learning, and thus arguably represent more appropriate criteria for accountability.

Despite these apparent advantages of a value-added approach, NCLB has rejected it in favor of reliance on progress towards absolute targets. One reason for

this rejection may be the greater simplicity of absolute targets. Whereas absolute targets require a simple calculation of means, value-added analyses are complex, particularly if attempts are made to use many years of test score data, to take account of the nested structure of schools, to adjust for test unreliability, and so on. NCLB's stated reason for rejecting value-added approaches, however, is to avoid the «soft bigotry of low expectations,» that is, a refusal to recognize anything other than common standards for all students, fearing that acceptance of divergent standards would mean that lower achieving students will remain permanently behind.<sup>7</sup>

Another alternative, as yet untried on a national level, would be to examine *both* progress towards an absolute target *and* value-added performance. If each school were evaluated on both criteria, then schools that are effective might be rewarded rather than sanctioned, even if (or perhaps especially if) their students are low-performing on an absolute scale. Under this calculation, only schools that are both low-performing and ineffective would face severe sanctions such as reconstitution.

## Conclusion

No Child Left Behind is a dramatic attempt to introduce rigorous accountability into U.S. education. Given the long tradition of local control and loose coupling, this effort faces strong challenges from many constituencies. Nonetheless, educators and administrators are clearly responding. NCLB faces the inevitable dilemmas boldly, but it is not clear if its goal of bringing all students to a common standard is feasible.

With respect to the dilemma of where to set the bar, NCLB aims high. As a result, schools with disadvantaged student populations are increasingly being identified as failing to meet standards, at higher rates than occurs for schools with more advantaged populations. Unfortunately, at this point there is no evidence the purported remedies in NCLB (improved teacher qualifications, scientifically-based teaching, tutoring, and choice) are likely to reverse this trend. Confronting the dilemma of what is tested under NCLB, the decision to focus exclusively on reading and mathematics means that other subjects may be ignored, yet it is hard to argue against knowing whether students can perform basic math operations and read with comprehension at least as a starting point for accountability.

As to the dilemma of how to measure progress, NCLB focuses exclusively on absolute targets, rather than on growth or on school effectiveness. This creates the potentially perverse situation in which a school that is effective with a low-achieving population may be judged as failing to make «adequate yearly progress,» while a school that is actually ineffective with a more advantaged population is nonetheless judged successful. A value-added approach would coun-

teract this confusion, but it would also risk low-achieving students never catching up to their higher-achieving peers. NCLB is aware that educational opportunities are unequal, and offers some strategies for overcoming these inequalities, but so far there is no evidence of interventions powerful enough to eliminate the gaps associated with social and economic disadvantage.

Where does this leave standards-based accountability in the U.S.? Due to bipartisan support, it is very likely that NCLB will be reauthorized when it comes up again (perhaps more likely after the next presidential election in 2008, rather than in 2007). However, it may be modified. Two revisions seem to me especially likely. First, if the law is not reauthorized until late in 2008 or even in 2009, it will be obvious by then that no state is on track to have all its schools with 100 percent of students proficient by 2014. In response, the time frame may be adjusted to set a more realistic timetable. Second, the focus on progress towards absolute standards may be tempered with recognition of the benefits of a value-added approach to estimative school effectiveness. A compromise, in which schools are rated on effectiveness as well as progress towards an absolute target, would be a sensible new direction - though good sense may be too much to hope for when policy-making is determined on the basis of political expediency rather than practical exigencies.

### Notes

- 1 An exception is the state of New York, which has long operated a system of Regents examinations for high school graduates. Bishop (1998) has argued that the Regents examinations have provided enhanced incentives for student learning.
- 2 Although NCLB has enjoyed bipartisan support, it has also been the target of bipartisan criticism. For example, two states are currently suing the federal government over NCLB: Connecticut, one of the nation's most liberal states, and Utah, one of the most conservative. See Loveless (in press) for details on resistance to NCLB.
- 3 For more details on NCLB, see the U.S. Department of Education web site, [www.ed.gov](http://www.ed.gov), and for state implementation, see the web sites of the various state departments of education.
- 4 Beginning in 2006-07, students are also tested in science once in grades 3-5, once in grades 6-8, and once in high school, but the science assessments are not part of the accountability system, i.e. there are no sanctions based on science performance.
- 5 I have made this assertion in talks all across the U.S. and indeed around the world, and have yet to be contradicted. I would welcome the chance to hear of such a school, so that I could change my assertion to «Only one school has ever...» or even «Only 10 schools have ever...» The point about the unrealistic expectations of NCLB would still hold.
- 6 NCLB has a «safe harbor» provision by which a school that closes the gap by 10 percent between its percent proficient in one year and the target it must reach in the next is counted as having made AYP even if it falls below the target, but it is clear from Linn's (2003) analysis that even this «safe harbor» will be out of reach for many schools.
- 7 Recently, the U.S. Department of Education granted a «waiver» from NCLB's AYP provision to five states (Tennessee and North Carolina in 2006, and Arkansas, Delaware, and, with conditions, Florida in 2007) to allow them to calculate AYP on the basis of «growth models.» Whereas value-added can be calculated from statistical models that are commonly referred to as «growth models,» the growth models allowed under this waiver are

not value-added models at all. A more appropriate term (as proposed by the Tennessee department of education) is «projection model,» in that the alternate models project forward in time, calculate whether a student's achievement trajectory would be sufficient to bring him or her to proficiency by 2014 even if s/he is not proficient at present; if so, the student may be counted towards making AYP in the current year. In practice, this approach varies little from the standard AYP calculation.

## References

- Bishop, J. (1998). *Do curriculum-based external exam systems enhance student achievement?* CPRE Research Report Series RR-40. Philadelphia, PA: Consortium for Policy Research in Education.
- Bush, G. W. (2004). *President's remarks in Minneapolis, Minnesota*. Retrieved December 1, 2006, from <http://www.whitehouse.gov/news/releases/2004/10/20041030-8.html>
- Gamoran, A., & Dreeben, R. (1986). Coupling and control in educational organizations. *Administrative Science Quarterly*, 31, 612-632.
- Goals 2000 (1994): Educate America Act, Pub. L. No. 103-227.
- Grodsky, E. (2002). Constrained opportunity and student choice in American higher education. Madison, WI: Unpublished doctoral dissertation, Department of Sociology, University of Wisconsin-Madison.
- Hamilton, L., Stecher, B., & Berends, M. (2005). *Teachers' responses to standards-based accountability*. RAND Working Paper WR-259-EDU. Santa Monica, CA: RAND.
- Kane, T., & Staiger, D. (2002). *Racial subgroup rules in school accountability systems*. Paper presented at the conference on Taking Account of Accountability, Harvard University, Cambridge, MA.
- Linn, R. L. (2003). Accountability: Responsibility and Reasonable Expectations. *Educational Researcher*, 32, 7, 3-13.
- Loveless, T. (In press). The peculiar politics of No Child Left Behind. In A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for «No Child Left Behind.»* Washington, DC: Brookings Institution Press.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge/Falmer.
- Metz, M. H. (2005). Symbolic uses of the «No Child Left Behind Act»: Reaffirmation of equality of educational opportunity or de-legitimization of public schools? In A. Sadovnik, J. O'Day, G. Bohrnstedt & K. M. Borman (Eds.), *No Child Left Behind and the reduction of the achievement gap: Sociological perspectives*. New York: Routledge.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83, 340-363.
- Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer (Ed.), *Environments and organizations* (Pp. 78-109). San Francisco: Jossey Bass.
- Meyer, R. H. (1996). Value-added indicators of school performance. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (Pp. 197-224). Washington, DC: National Academies Press.
- Meyer, R. H. (2003). *Value-added indicators: Do they make an important difference?* Evidence from Milwaukee. WCER Working Paper 2003-5. Madison, WI: Wisconsin Center for Education Research.
- National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Odden, A., & Picus, L. (2003). *School finance: A policy perspective*. Third edition. New York: McGraw-Hill.

- Porter, A. C. (2006). *NCLB lessons learned: Implications for reauthorization*. Paper presented at the conference on Standards-Based Education Reform and Children in Poverty, University of Wisconsin-Madison, Madison, WI.
- Rosenbaum, J. E. (2001). *Beyond college for all: Career paths for the forgotten half*. New York: Russell Sage.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. Washington, DC: Economic Policy Institute.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1-19.
- Weick, K. E. (1982). Administering education in loosely coupled systems. *Phi Delta Kappan*, 63, 673-676.
- Wilson, B., and Rossman, G. (1993). *Mandating academic excellence: High school responses to state curriculum reform*. New York: Teachers College Press.
- Wilson, T. J. (1996). *Reaching for a better standard: English school inspection and the dilemma of accountability in American schools*. New York: Teachers College Press.

**Key words:** No Child Left Behind (NCLB), equal opportunity law, institutional responsibility, high-stakes standardized testing, effective schools, educational efforts

## Die Verantwortung der Schule nach US-amerikanischem Vorbild: Dilemmata durch Formen des 'high stakes'-Testing

### Zusammenfassung

*No Child Left Behind* (NCLB) ist ein neues Bundesgesetz in den U.S.A., das von den Staaten, die Bundesmittel im Bildungsbereich beantragen wollen, verlangt, die Standards auf dem Gebiet der Schulkompetenzen festzulegen; die Bewertungen zu definieren, um die in Bezug auf die Standards gemachten Fortschritte zu messen, sowie Bezirke und Schulen in Bezug auf die Ergebnisse zur Verantwortung zu ziehen. Das Wecken des Verantwortungsbewusstseins (*Accountability*) durch strenge standardisierte Tests stellt die letzte Phase der Entwicklung eines langen Weges in Richtung einer auf Standards basierenden Reform der amerikanischen Bildungspolitik dar. Ein Reformprozess, der sowohl auf die Unterstützung der Mehrheits- als auch der Oppositionsparteien rechnen konnte. Er nimmt feste Form in einem Kontext an, wo früher das System nur sehr lose verknüpft war, wo es nur lückenhafte Kontrollen durch die Behörden gab und keine landeseinheitliche Bewertung vorgesehen war. NCLB muß sich mit vier verschiedenen Problemen auseinandersetzen, die jedes Bewertungs- und Prüfungssystem anpacken muß. Das heißt, hohe Standards festlegen; genaue Ziele für die Bewertung des Fortschritts vorgeben; die Beschränkung der Standardtests auf nur zwei Fächer akzeptieren als Preis für die Sicherung von spezifischen Standards; ferner versucht es (wenn auch mit wenig Erfolg) die Mechanismen zu identifizieren, welche Chancengleichheit gewährleisten sollen, die geforderten Standards zu erreichen. Dank diesem Vorgehen war es dem NCLB Gesetz mög-

lich, die Aufmerksamkeit der Erziehungsverantwortlichen zu gewinnen, aber auch die Lehrer und die Studenten anzuspornen, sich zu verbessern. Jedoch führt dieses Gesetz auch zu unverhältnismäßigen Sanktionen für diejenigen Schulen, die bereits eine benachteiligte Bevölkerung haben; es trifft keine Differenzierung zwischen effizienten und ineffizienten Schulen und legt einen unrealistischen Zeitrahmen fest, um zum Erfolg zu gelangen.

**Schlagworte:** No Child Left Behind (NCLB), Standards, Evaluation, institutionelle Strukturen und Effekte, bildungspolitische Rahmenbedingungen

## **Responsabilité de l'école selon le modèle américain : Dilemmes d'une évaluation aux enjeux essentiels**

### **Résumé**

*No Child Left Behind* (NCLB) est une nouvelle loi fédérale américaine qui demande aux états désireux d'obtenir des allocations fédérales dans le domaine de l'éducation d'élaborer des standards relatifs aux compétences scolaires, de définir des évaluations pour mesurer les progrès effectués par rapport aux standards mêmes et de responsabiliser les districts et les écoles en matière de résultats. La prise en charge de la responsabilité (*accountability*) par de sévères tests standardisés rentre dans le cadre des plus récents développements d'un long parcours de réforme de l'instruction américaine basée sur des standards; ce processus de réforme 'amélioration a pu compter tant sur le support de la majorité que de celui de l'opposition. Cette responsabilisation se dessine dans un contexte avec un système peu formel de supervision officielle et l'absence d'évaluations nationales. NCLB se confronte à quatre dilemmes que chaque système d'évaluation devant attester de sa responsabilité (*accountability*) doit affronter: définir des standards élevés ; adopter des objectifs absolus pour évaluer les progrès ; accepter l'étroitesse des tests standards dans deux matières seulement, le prix à payer pour assurer des standards clairs et spécifiques ; et, même si le succès en est limité, chercher à identifier les mécanismes qui peuvent garantir les mêmes opportunités pour atteindre les standards. Grâce à ce type d'approche, la loi NCLB a galvanisé l'attention des éducateurs, mais a aussi stimulé les enseignants et les étudiants à s'améliorer. Cependant, elle entraîne également de sanctions disproportionnées pour les écoles au service des populations défavorisées car aucune distinction n'est faite entre les écoles efficaces et inefficaces. De ce fait, cette loi propose un cadre temporel irréaliste pour en assurer le succès.

**Mots clés:** No Child Left Behind (NCLB), loi sur l'instruction, responsabilité institutionnelle, standards, évaluation, enjeux éducatifs



## Responsabilità della scuola secondo il modello americano: Dilemma dei test di alto livello

### Riassunto

*No Child Left Behind* (NCLB) è una nuova legge federale americana che chiede agli stati che desiderano ottenere degli stanziamenti federali nell'ambito dell'istruzione, di dettare gli standard relativi alle competenze scolastiche, di definire le valutazioni per misurare i progressi fatti rispetto agli standard stessi e di responsabilizzare distretti e scuole in materia di risultati. L'assunzione di responsabilità (*accountability*) mediante severi test standardizzati rientra nell'ambito dei più recenti sviluppi di un lungo cammino verso una riforma dell'istruzione americana basata sugli standard; un processo di miglioramento che ha potuto contare sia sul supporto della maggioranza che dell'opposizione. Essa prende forma nel contesto di quello che era un sistema piuttosto approssimativo, in cui esisteva una scarsa supervisione ufficiale e non si prevedevano delle indagini nazionali. NCLB si confronta con quattro dilemmi che ogni sistema di valutazione e *accountability* deve affrontare: definire degli standard elevati, adottare degli obiettivi assoluti per valutare i progressi, accettare la limitatezza dei test standard solo in due materie quale prezzo da pagare per assicurare degli standard chiari e specifici, inoltre cercare (magari con un successo limitato) di identificare i meccanismi che possono assicurare pari opportunità nel rispondere ai requisiti voluti. Grazie a questo tipo di approccio, la legge NCLB è riuscita a galvanizzare l'attenzione degli educatori, ma anche a stimolare insegnanti e studenti a migliorarsi; tuttavia, essa è anche all'origine di sanzioni sproporzionate per le scuole al servizio delle popolazioni svantaggiate, non fa distinzioni tra scuole efficaci ed inefficaci e propone una cornice temporale non realistica per il successo.

**Parole chiave:** No Child Left Behind (NCLB), legge sull'istruzione, responsabilità istituzionale, standards, valutazione, sforzi educativi.