

Examining the Assumptions Underlying the NCLB Federal Accountability Policy on School Improvement

Ellen Forte
edCount, LLC
Washington, DC

Many wonder whether the No Child Left Behind Act of 2001 (NCLB) will realize its goal of improving achievement among low-performing students in high-poverty schools. An examination of assumptions that underlie the accountability and school improvement aspects of this federal policy suggests that it will not. In spite of the coherence of the systemic reform framework in which NCLB is based, flawed rules for placing schools in improvement status, mismatches between actual needs and pre-prescribed service types, and major gaps in state and local capacity for designing and providing services mean that the implementation of school improvement may be insurmountably limited by the NCLB policy basis. Following this analysis, five recommendations for the next reauthorization of our nation's federal education policy are offered.

At the end of the day, the No Child Left Behind Act of 2001 (NCLB) is supposed to be about improving achievement among low-achieving students in high-poverty schools. School improvement, as a status and as a process, is understood as a critical mechanism for reaching that goal. Many wonder whether this policy has been effective. A definitive answer may never be forthcoming, but a set of lessons could emerge if we understand the rationale behind the policy and how the policy manifests in practice.

As the latest reauthorization of the Elementary and Secondary Education Act of 1965 (ESEA), NCLB represents the current U.S. federal education policy. In some ways, the era of NCLB has offered the best of times for school improvement efforts—for example, more attention and common expectations across schools that serve low-income students and those that serve more advantaged communities. But, in other ways, NCLB has proven the worst of times for the improvement of schools: flawed rules for placing schools in improvement status, mismatches between actual needs and pre-prescribed service types, and capacity for designing and providing services that lags well behind the numbers and needs of schools in improvement. Thus, the process of school improvement may be as limited as it is strengthened by the NCLB policy basis.

Clearly, NCLB offers remarkable points for comparison and contrast in a discussion of how school improvement currently fits within the contexts of federal education policy and approaches to school improvement in general. Next, the NCLB policy logic is illustrated to provide a framework for an ensuing exploration of the school improvement aspects of this policy as they manifest in practice.

A PICTURE OF THE NCLB POLICY

In 1965, then-President Johnson signed into law the landmark ESEA legislation that established, for the first time, federal funding for public education combined with a federal policy specifically to support educational opportunities for students from high poverty communities (Cross, 2004; Jennings, Stark Rentner, & Kober, 2002). Over the years, this policy was modified and reauthorized a number of times. With regard to the present discussion, the most notable of these modifications involved the requirements for states to monitor and evaluate the academic achievement of students in the programs receiving this federal funding. From the beginning through the 1980s, these requirements were limited and states were allowed to use one set of achievement tests in schools serving high poverty communities and another set (or nothing) for other schools.

The 1994 reauthorization of the ESEA, known as the Improving America's Schools Act (IASA), represented a

Correspondence should be addressed to Ellen Forte, edCount, LLC, 5335 Wisconsin Avenue NW, Ste. 440, Washington, DC 20015. E-mail: eforte@edCount.com

paradigm shift in federal education policy. Drawing upon the elements of the systemic reform model (Smith & O'Day, 1991), IASA required states to (a) establish common, statewide standards for all students in reading and mathematics in the 3 to 5, 6 to 8, and high school grade ranges; (b) implement statewide assessments aligned to these standards in at least three grades each for reading and mathematics; and (c) implement a statewide accountability system for evaluating school-level performance. NCLB, the next subsequent reauthorization of ESEA, extended these same concepts to include Grades 3 through 8 for standards and assessments in reading and mathematics, added the requirements for standards and assessments in at least three grades for science, and established a specific set of rules for states' accountability systems.

The logic behind this approach is as follows: If we (a) define clearly what students should know and be able to do as well as the level to which students should be able to demonstrate what they know and can do in each content area and grade level (Performance Standards), and (b) use assessments aligned to these expectations, we can (c) use these assessment scores to inform accountability decisions meant to improve school functioning and, as a result, improve student achievement.

At a very gross level, the logic underlying the NCLB policy could be illustrated as indicated in Figure 1. The idea is that student achievement increases in challenged schools when students are served better by schools that have been made better through a school improvement process. Although it is understood that the conditions underlying school performance and change mechanisms are far more complex than this simple illustration and the NCLB policy logic suggest, one can use this illustration to deduce some basic assumptions underlying the NCLB policy that, once identified, can form the basis for evaluating the effectiveness of the policy in achieving its student achievement goals.

First among these assumptions is that the mechanism by which schools are identified for improvement yields accurate results. Second, the NCLB policy model assumes that the consequences, sanctions, resources, and supports (e.g., school choice, provision of supplemental education services, reconstitution) can be characterized as improvement efforts and are appropriately assigned to and effectively implemented within identified schools. Third, these improvement efforts lead to more effective service of students and measurable, meaningful increases in student achievement. Each of these assumptions is considered in turn next.

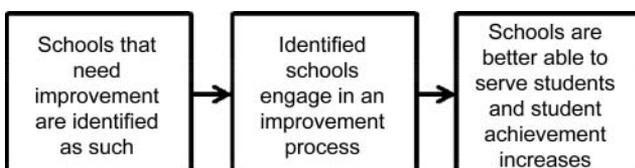


FIGURE 1 An illustration of the No Child Left Behind policy logic.

ASSUMPTION 1: SCHOOLS ARE APPROPRIATELY IDENTIFIED FOR IMPROVEMENT STATUS

Under NCLB, schools are identified for improvement via application of an adequate yearly progress (AYP) algorithm that is tightly prescribed in the NCLB legislation and ensuing regulations. AYP is not known for its elegance; President Bush's own education advisor, Sandy Kress, once described it as "Rube Goldbergesque" (Toch, 2001). In-depth analyses of this algorithm are available elsewhere, but two questions can serve to frame an understanding of how specific nuances of this algorithm determine which, how many, and whether the right schools are identified for improvement via AYP.

First, which schools is AYP designed to identify? Although AYP is often misunderstood as a means for revealing whether a school needs improvement as a function its effectiveness, NCLB implicitly defines the goal of school improvement around the concept of achievement status rather than effectiveness (Carlson, 2006; Gong, 2002; Linn, 2008a, 2008b; Raudenbush, 2004). That is, the question meant to be addressed by the AYP algorithm that is used to identify schools for improvement under NCLB most closely aligns with that posed in Model B in Figure 2. AYP answers the question, "Does the percentage of students at this school who scored in or above the proficient level reach the target for this year?" and does not address whether a school is "effective in supporting student learning and progress at an appropriate rate in this school" (Model C) or "becoming more effective in supporting student learning and progress over time" (Model D).

This achievement status model is manifested in the AYP algorithm through the criteria for student performance on standardized assessments in reading/language arts and mathematics and the criterion of 95% for participation in each of the assessments, which is meant to ensure that the percent proficient indicators reflects all students' performance. NCLB imposed the method by which states established a pattern of performance targets that must increase every 1 to 3 years until they reach 100% proficiency in the 2013–2014 school year. One might think that these increasing performance targets mean that the question AYP addresses has something to do with growth. To the contrary: AYP neither imposes expectations nor gives credit for increases in school-level scores across time or for individual student growth. Change scores, which compare school-level scores across two points, and growth models, which consider changes in the scores for individual students over time, are accompanied by their own technical strengths and limitations (Choi, Goldschmidt, & Yamashiro, 2006), so no claim is made here that either is clearly preferable to the AYP model per se. Further, the U.S. Department of Education's own evaluation (U.S. Department of Education, 2009a) of the Growth Model Pilot Program it established in the 2005–2006 school year revealed that in the two states allowed to implement growth

Criterion	Question	
	How good is this school?	Is this school getting better?
Achievement (in relation to standards)	Model A What percentage of students meets the state standards?	Model B Is the achievement level in this school improving in relation to the standards over time?
Effectiveness (in relation to past performance of students)	Model C Is the school effective in improving students' achievement from one year to the next? Are individual students making expected progress from grade to grade?	Model D Is the school becoming more effective? Are individual students learning at faster rates as they progress through the school?

FIGURE 2 Criteria for defining school quality in an accountability model. Figure derived from Gong (2002) and Carlson (2006).

models that year, only seven schools in one state made AYP solely as a result of the growth model. The growth model made no difference for any school in the other state. Now that the U.S. Department of Education has allowed 13 more states to implement growth models as part of their AYP analyses, resulting in far more variation in the types and nuances of these models, future evaluations of the impact may be different. The important point here is that one must distinguish among school identification models and the perspectives that underlie them because of their implications for which schools are in improvement status.

Given that AYP is about achievement status rather than effectiveness, one might further assume that its application results appropriately in the identification of only the lowest scoring schools. This is neither true nor desirable and leads to the second framing question: Does AYP identify the schools that actually need to improve and would benefit from state intervention to do so? There are at least three basic reasons why the answer is likely closer to no than yes. AYP probably over- and misidentifies schools (Kane & Staiger, 2002), resulting in a pool of identified schools that is larger than, and does not entirely overlap with, the schools on which limited state resources would most effectively be focused because its performance targets are unrealistic, the AYP algorithm involves an untenable number of conjunctive decisions, and the percentage proficient statistic is a poor indicator of school quality.

As previously stated above, the AYP algorithm addresses achievement status rather than effectiveness and compares school-level results to prespecified annual targets that increase regularly to 100% in the 2013–14 school year. Starting points vary across states, but if the initial percent proficient in the trajectory were 45% in 2002–2003, the minimum annual increase necessary to reach the end point would be 5 percentage points. Linn and others (Linn, 2005b, 2008a, 2008b; Linn & Haug, 2002) have pointed out that the increases required of schools under AYP, especially low-performing schools, can be well beyond those that could reasonably be expected

of schools even if they actually may be improving: There is no research evidence to support the degree or rapidity of the score increases required of schools under NCLB (Kim & Sunderman, 2004; Linn, 2005b, 2008a, 2008b; Mintrop & Sunderman, 2009), and these unrealistic targets result in an increasing number of false positives over time (R. Hill & DePascale, 2003; Linn & Haug, 2002). Thus, schools that are relatively low-performing at a given time but where test scores, and perhaps also student learning, are increasing at a greater-than-expected rate would be identified because of their low achievement status even though the school may be legitimately improving without the benefit of state interference or sanction. The consequences of this approach is the identification of a relatively large and growing number of schools and an attendant dilution of resources available to support improvement interventions in schools that really need them.

A second key aspect of the AYP algorithm with implications for the identification of schools for improvement is the requirement for a positive outcome in every one of between five and 37 separate comparisons. Contrary to what the NCLB authors may have thought, more calculations using the same data does not result in greater reliability of the conjunctive outcome in the way that triangulation methods might; analyzing outcomes for each of several student groups does not necessarily enhance transparency or decision precision.

Further, a school is identified for improvement when it misses one or more of its AYP targets in each of 2 consecutive years. This 2-year rule is not quite as supportive of decision accuracy as it might first appear. For example, if Elm Elementary had met all AYP targets in 2006–07, missed only the reading performance target for the English language learner student group in 2007–08, and missed only the reading participation target for African American students in 2008–09, Elm Elementary would be identified for improvement. It would seem important to consider other information about this school before declaring that it needs

improvement. States are allowed to use additional information, but only if that results in the identification of a greater number of schools, not to remove schools from this sanction list (section 200.19(d)(1)).

This prescriptiveness may have been introduced as a reaction to states' apparent lack of attentiveness to accountability requirements under the previous version of ESEA. When AYP was introduced in the Clinton administration's version of ESEA, the IASA, states had flexibility around the definition and combination of AYP indicators as long as the definition of progress mostly reflect performance on the a state's tests. Unfortunately, few states developed or employed rigorous AYP models under IASA, perhaps because many were slow to warm to the notion of Title I rules applying to all schools and perhaps because the development of a large-scale, test-based model that can quickly and accurately classify schools into improvement categories poses significant technical and practical challenges.

NCLB predefined its AYP indicators and model in the legislation text. The indicators must reflect assessment performance (typically the percentage of students who score in or above the proficient achievement level) and assessment participation rates (95% or higher), calculated separately for reading and mathematics, for each of up to nine student groups (nine groups times two content areas times two indicators equals 36 indicators) plus include graduation rates for high schools and another academic indicator for elementary and middle schools (a 37th indicator). Thus, AYP could potentially encompass up to 37 indicators for a school or a district if the student population were very large and very diverse. States are allowed to set a minimum group size (minimum n) for the calculation of AYP,¹ and few schools have student populations that meet these minimums in more than a few groups. However, AYP involves five indicators at a minimum (when only calculating AYP for the "all students" group) and 17 indicators with a reasonable set of four student groups (e.g., all students, African American students, White students, students with disabilities).

The number of indicators is a matter of great concern to those who understand the implications of multiple conjunctive decisions² for the reliability of an overall judgment (Gong, 2002; Kane & Staiger, 2002; Linn, 2003, 2005b, 2006, 2008a; Linn, Baker, & Betebenner, 2002; Marion et al., 2002; Ryan & Shepard, 2008). It also concerns those who recognize the punitive implications for large and highly diverse schools. Prior to IASA, U.S. federal education policy allowed for different academic expectations for different stu-

dents and schools. Although IASA introduced the concept of "all" with regard to students and school, states could condition "all" such that many English language learners and students with disabilities could be excluded from testing and, therefore, could practically be excluded from the content and performance expectations and from AYP. Thus, school-level outcomes did not necessarily reflect the performance of all students and, because schools were not required to report participation rates, it was impossible to know the degree to which these outcomes were affected by the exclusion of students, especially low-performing students like the ones ESEA was meant to support. The implication of including all students in assessments and AYP could be the identification of more schools for improvement, particularly schools that serve students with disabilities, students from low-income backgrounds, and English language learners (Gong, 2006; Kim & Sunderman, 2004).

Although this would seem consistent with the original intent of ESEA and the presumed NCLB intent not to leave any student behind, the practical effects are likely much different. Schools that serve large and diverse student populations are more likely to be identified for improvement even if their most-challenged students are doing better than similar students who attend smaller, less diverse schools (Kim & Sunderman, 2005; Mintrop & Sunderman, 2009; Simpson, Gong, & Marion, 2005). As notable as the impact of a large number of student groups on the likelihood a large, diverse school will be identified for improvement is the degree to which the performance of some student groups is not revealed in AYP. An analysis of AYP outcomes for schools in five states revealed that with a minimum n of 30, the national average for this parameter (Forte & Erpenbach, 2006), scores for between 89.2% and 96.6% of special education students statewide were not reflected in a special education subgroup result for AYP (Simpson et al., 2005), rather defeating the purpose of the multiple subgroup approach to AYP and underscoring the extra "burden" of diversity in large schools.

It is very likely that scores for some students do not reflect their knowledge and skills as is assumed under the NCLB AYP model. For example, English language learners who have not yet achieved a level of English language proficiency that would allow them to demonstrate in English what they know and can do academically may be at a significant disadvantage on the tests that contribute to AYP (Abedi et al., 2005; Abedi & Dietel, 2004; Kopriva, 2008). Regardless of whether schools serve enough ELLs to meet the minimum n requirements, inclusion of scores for which there could be little evidence to support the assumed meaning can lead to the overidentification of schools that serve ELLs.

Further exacerbating this situation is the use of the percent proficient statistic itself and the error inherently associated with variations in year-to-year student samples (R. Hill & DePascale, 2003; Kane & Staiger, 2002; Marion et al., 2002). The authors of the NCLB legislation may have avoided use of a mean score out of fear that its inherent

¹These minimum n s range from one to 100 with a mode of 30 (Forte & Erpenbach, 2006).

²Multiple conjunctive decision models are those in which an outcome is based on two or more criteria that must each be met. Conjunctive models can be contrasted with compensatory models, in which an outcome is based on multiple criteria, but meeting some of these criteria can compensate for not meeting others.

balance of low and high performance would allow schools to offset the impact scores of lower performing students by those of higher performing students. Unfortunately, their solution ignores most of the information in student score distributions and relies on a judgment about the operational meaning of proficiency that varies widely across states and tests. The many problems with the percent proficient statistic are well documented, and it is widely understood among evaluators and researchers that this statistic provides, at best, a questionable statement about school quality (Ho, 2007; Koretz, 2008b; Linn, 2005b, 2008a, 2008b; Linn & Haug, 2002; Martineau, 2006; Popham, 2004).

The preceding discussion was designed to point out that the set of schools identified for improvement under the NCLB rules is probably (a) larger in number than and (b) does not entirely overlap with the set of schools that actually need or could at least benefit from state intervention (see Figure 3). As a result of the use of the AYP model to identify schools, too many schools are identified for improvement, some of the schools identified do not need state intervention, and it is likely that some schools that are not identified actually do need intervention but escape identification due to deficiencies of the identification process. It cannot be known how many and which of the nearly 15,000 schools in improvement status as of the 2006–07 school year, a number that is nearly 21/2 times the number identified in the 2002–03 school year, actually need improvement and could benefit from the types of supports available under NCLB or state-specific systems.

Thus, the rationale for the assumption that the identification mechanism yields appropriate results is critically weak. Empirical evidence that the mechanism works as intended is missing altogether.

ASSUMPTION 2: CONSEQUENCES ASSOCIATED WITH SCHOOL IMPROVEMENT STATUS

A second set of assumptions underlying the NCLB school improvement policy logic is that the consequences, sanctions, resources, and supports can be characterized as improvement efforts and are appropriately assigned to and effectively implemented within identified schools. Analysis of this set of assumptions starts here with a description of the NCLB requirements for identified schools and a brief summary of the evidence regarding the implementation and effectiveness of these requirements.

An annual AYP determination must be made for every school, regardless of whether a school receives Title I funds, and will result in identification if targets are missed in each of two consecutive years. As previously noted, the missed targets do not have to be the same ones across these years. Many states have been allowed to limit identification to cases where targets were missed twice consecutively in the same content area, but states are not allowed to further restrict

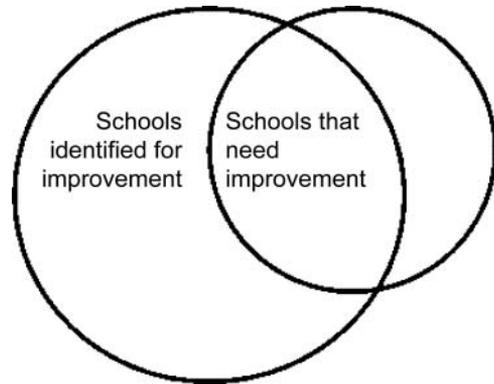


FIGURE 3 Possible relationship between schools identified for improvement and schools that need improvement.

missed targets to within the same student group (Forte Fast & Erpenbach, 2004).

Schools that do not receive Title I funds may or may not be subject to consequences associated with their identification; the decision about whether and what consequences to apply is left to states. Once a Title I–funded school is identified for improvement, it must embark on a predefined path of sanctions (see Figure 4) that begins with the development and implementation of its school improvement plan. This plan must address “the academic issues that caused it to be identified for school improvement” (U.S. Department of Education, 2006, p. 8) and describe specific strategies to improve academic achievement for all students that are “based on scientifically based research” (U.S. Department of Education, 2006, p. 9). As specified in the NCLB legislation and regulations, this plan must also identify additional indicators the school will use to monitor progress, how the school will improve parental involvement and professional development quality, and its program for teacher mentoring.

In addition to implementing the school improvement plan, a school in its 1st year of improvement status must offer its students the option of choosing a different public school and must set aside 20% of its Title I dollars to support the costs associated with choice, such as transportation of students to their chosen other school.

School choice is the first of three controversial sanctions applied to school in improvement status that will be considered here in the order in which they are imposed on schools in improvement status, the others being provision of supplemental educational services and reconstitution. The effectiveness of school choice as required under NCLB in improving school quality or student achievement is impossible to determine because its effective implementation has been, and will likely continue to be, negligible. Out of the nearly 7 million students who were eligible for school choice in the almost 11,000 schools in improvement status in the 2006–07 school year (U.S. Department of Education, 2009b), only 1%

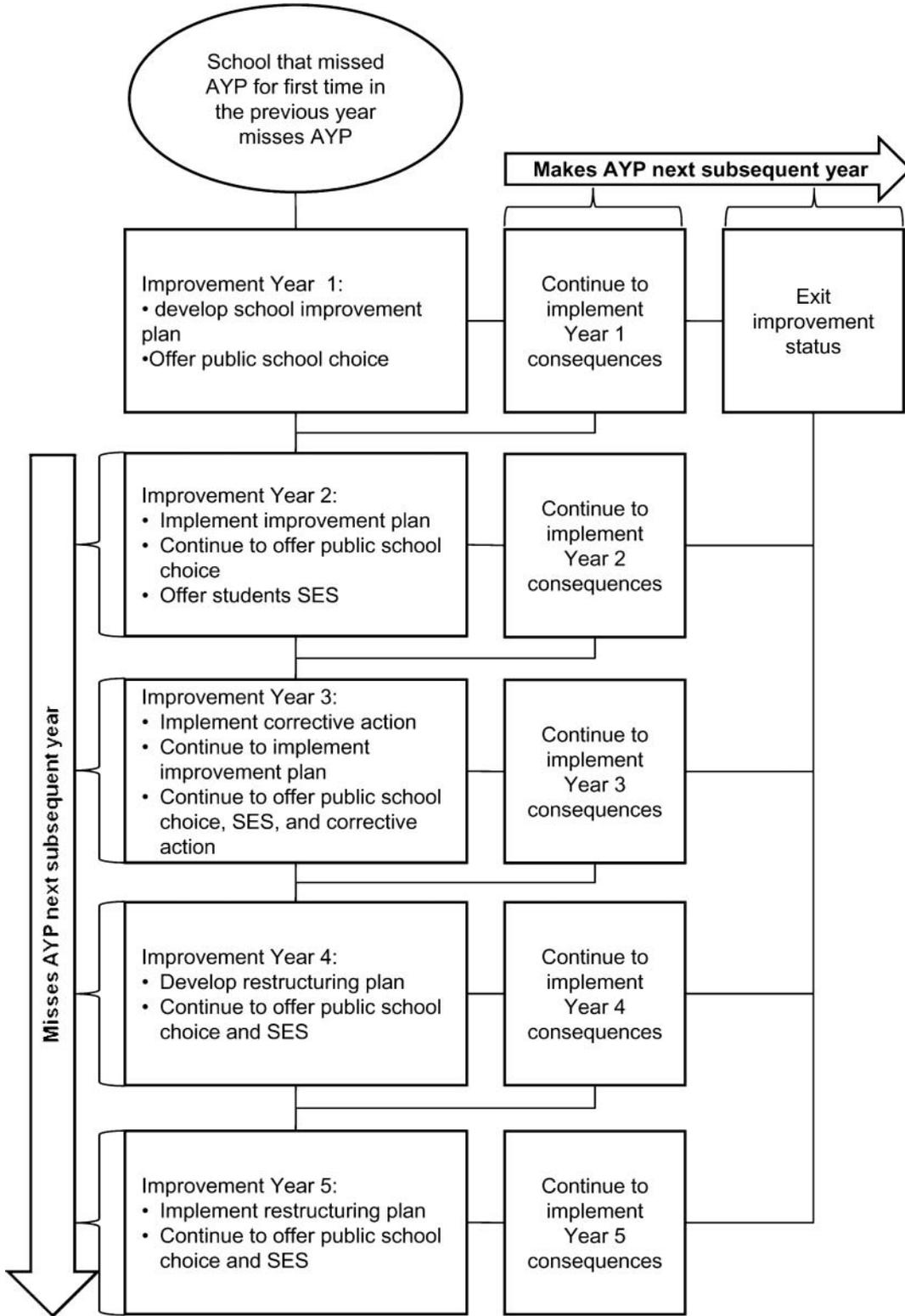


FIGURE 4 No Child Left Behind school improvement sanctions and timeline. *Note.* AYP = adequate yearly progress; SES = supplemental educational services.

actually chose and attended a different school (Stark Rentner et al., 2006; Vernez et al., 2009).

Poor communication of the choice option to parents of eligible students and unavailability of other school options are only slightly to blame for the extremely low rate at which this option is exercised (Kim & Sunderman, 2004). The primary reasons that parents and students choose to stay put involve their personal experiences in and connections to their existing schools and teachers (Vernez et al., 2009), suggesting a school's improvement status is not of major concern to parents. In fact, it is possible that a negative culture within a school could result from, rather than lead to, identification for improvement (Mathis, 2009; Mintrop, 2004; Mintrop & Trujillo, 2005). Parents consistently report greater satisfaction with their own children's schools than with public schools in general and favor public school choice more as a concept than an action one would actually take (Elam, 1995; Rose & Gallup, 2006, 2007). Thus, it is highly unlikely that public school choice, which does not directly effect change in school functioning and found its way into NCLB as one of several political compromises that ensured passage of this legislation (Cross, 2004), rather than on the basis of any evidence that it improves the quality of schools or students' achievement, will ever be viable as a school improvement strategy.

If a school makes AYP while in its 1st year of improvement status, it must continue to implement its improvement plan and to offer public school choice. If that school misses AYP that year, it must offer supplemental educational services (SES) to its students. As defined by the U.S. Department of Education (2009b), SES encompass "additional academic instruction [services that] may include . . . tutoring, remediation and other supplemental academic enrichment services" (p. 1) that are offered in outside of the regular instructional program and by other individuals or organizations that students encounter at their schools. That is, SES can mean a wide range of service types, intensities, and delivery mechanisms (e.g., in-person, online, after school, on weekends) that are meant to improve academic achievement.

In terms of participation, SES has proven more successful than public school choice: 17% of students eligible for SES services in the 2005–06 school year participated in these services (Vernez et al., 2009). Parents may feel more comfortable taking advantage of SES than public school choice because it can be seen as "extra" support that is free and directly addresses their child's academic needs (Vernez et al., 2009). A dramatic increase in the number of SES providers between 2003 and 2007 (Vernez et al., 2009) and stronger frameworks for approving, monitoring, and evaluating SES providers (Ross, Harmon, & Wong, 2009) may improve students' access to higher quality SES. Therefore, although recent independent evaluations reveal only small, if any, statistically significant effects of SES on student achievement (Chicago Public Schools, 2007; Zimmer, Gill, Razquin, Booker, & Lockwood, 2007), SES could ultimately yield some success in improving student achievement. However, it must be noted

that, as is the case with public school choice, it may not be appropriate to call SES a school improvement strategy because it effects no direct change on school functioning. It treats the students rather than the school.

In Year 3 of improvement status, a school continues to offer public school choice, SES, and also faces corrective action by its district, such as changes in school leadership or curriculum (Mintrop & Trujillo, 2005). These sanctions continue into Years 4 and 5, when restructuring requirements take effect. Restructuring can mean that the school is converted to a charter school or has its operations outsourced to a private educational management organization, that some or all of the staff are replaced, or that responsibility for school management is turned over to the state education agency (SEA). As of the 2007–08 school year, more than 3,500 schools were in restructuring, an increase of more than 50% over the 2006–07 school year (Center on Education Policy, 2008). Restructuring may improve some aspects of school functioning in some cases (Garland, 2003; Malen, Croninger, Muncey, & Redmond-Jones, 2002), but general statements about the effectiveness of restructuring in terms of improving overall school functioning or student achievement are severely limited due to great variations in intervention type, quality of implementation, and fit with school context and needs (Mathis, 2009).

Thus, there is little to no evidence that the three most visible sanctions NCLB imposes on schools in improvement status are effective and some possibility that they may be conceptually unrelated to the notion of enhanced school functioning. The NCLB school improvement consequence that gets the least press, but which may most legitimately be characterized as an improvement effort and have the most promise for directly addressing specific problems in school performance, is the school improvement plan. However, limitations in fulfilling the role of the SEA in support of these plans and an imbalance in the components necessary for meaningful change are hindering this promise.

Within 3 months of being identified, a school must develop and submit to its district a comprehensive plan (U.S. Department of Education, 2002: Federal Register, 34 CFR Part 200, section 200.41(a) through section 200.41(c)) that identifies and spells out policies and practices to address the "specific academic issues that caused the school to be identified for school improvement," including measureable academic objectives for "all groups of students"; specifies the roles and responsibilities of school, district, and state in implementing the plan; incorporates strategies for enhanced parent involvement; and includes an assurance that at least 10% of the school's Title I, Part A, funds will be used to support "high quality professional development to the school's teachers, principal, and, as appropriate, other instructional staff" that focuses on "improving academic achievement and encompasses a teacher mentoring program." Districts are required to support schools in developing their school improvement plans and can leverage assistance from the SEA, institutions

of higher education, and private entities to do so (Federal Register, 34 CFR Part 200, section 200.40(b)); within 45 days of receipt of the plan from a school, a district must implement a peer review process to review the plan, support any necessary revisions to it, and approve it (Federal Register, 34 CFR Part 200, section 200.41(d)). This is a tall order for districts, especially those with many identified schools. SEAs are meant to provide support, but this support is not necessarily forthcoming because of capacity limitations at the SEA level.

The mandated role of the SEAs in supporting specific school improvement efforts was significantly strengthened with NCLB (Rhim et al., 2007); SEAs are required to establish “statewide systems of support” (NCLB section 1117(a)(5)(a)) that may include school support teams, deployment of teachers and administrators with a track record of success in Title I schools, or other approaches. It is this expanded SEA role that both intrigues and worries education policy analysts. Whereas the authority and responsibility of districts with regard to school support is traditional and relatively unquestioned, direct involvement of SEAs in school-level actions represents a significant shift from customary policy and practice customs (Hannaway & Woodroffe, 2003; Lusi, 1997; Sunderman & Orfield, 2006). Given the rapidly increasing numbers of schools in improvement across an increasing number of districts (U.S. Department of Education, 2009c) and the need for coherent, statewide, service-oriented assistance for school improvement efforts, as opposed to a focus on compliance with bureaucratic rules, this new role of the state seems to be warranted (Council of Chief State School Officers, 2006; Dwyer et al., 2005; Redding, 2007; Reville et al., 2005).

Manifestation of this new role ranges across states from cursory observance to funded commitment. A study in which 13 states were surveyed about their processes for reviewing school improvement plans, a basic and relatively unobtrusive level of state involvement in schools, only seven had such processes in place and three reported that they did not even collect these plans (DiBiase, 2005). On the other hand, some states, such as Alabama, Kentucky, Ohio, Tennessee, and Washington, have created sophisticated, coordinated systems of support that involve large-scale, intensive school interventions that support implementation of school improvement plans that are tailored to the specific needs of the schools (Redding, 2007; Stecher et al., 2008). These systems expand the team of school improvement professionals both within and outside of the SEAs. Within these SEAs, school improvement teams have reached beyond federal programs units to include educators with expertise in areas such as curriculum and instruction, special education, English as a second language and bilingual education programs, and early reading programs. To leverage expertise outside the SEAs, school support teams in these states have established partnerships with state professional associations and networks of distinguished educators and consultants who have the experi-

ence and credibility to serve on the front line of improvement efforts within needy schools (Kerins, Perlman & Langdon, 2009). In general, however, SEAs have extremely limited capacity to operate as required under NCLB (Archer, 2006; Dwyer et al., 2005; Education Commission of the States, 2002; Carlson Le Floch, Boyle, & Bowles Therriault, 2008b; Rhim, Hassel, & Redding, 2007; Reville, 2004).

In addition to philosophical and sometimes legislative obstacles to expansions in the centrality, breadth, and sophistication of the SEA (Sunderman & Orfield, 2006), the NCLB vision of the SEA coincided with increasingly poor economic circumstances for many state governments that has led to staff downsizing and program reduction (Carlson Le Floch, Boyle, & Bowles Therriault, 2008a, 2008b; Minnici & Hill, 2007). Increases in accountability, assessment and data reporting requirements, as well as more direct involvement in school support, have required both expansion and transformation of the scope and type of work required of state-level educators. When asked to counterweigh their departmental strengths and weaknesses, staff in most SEAs have reported overall capacity limitations due to financial, temporal, technological, organizational, personnel-related, or the lack of relevant expertise (Carlson Le Floch et al., 2008a). Similarly, SEA staff reported inadequacies in organizational capacity (44 states) and personnel numbers and expertise (43 states) due to limited funding (Center on Education Policy, 2007). Increased collaboration among SEAs and partnerships among SEAs and other entities may help (Unger et al., 2008), but it remains to be seen whether some SEAs have enough internal capacity to build upon and whether other entities that the U.S. Department of Education has identified as resources, such as universities and federally funded comprehensive and regional centers charged with supporting states, themselves have the capacity to address the depth and breadth of SEA needs. This lack of SEA capacity in the NCLB era is particularly troubling given that the architects of the original ESEA legislation included funding under Title VI to support capacity building within SEAs; this early understanding of the need for centralized strength to enable local policy implementation was counteracted in subsequent reauthorizations that removed this federal support for SEA capacity building (Cross, 2004).

These severe capacity limitations at the SEA level, in combination with the sanction-orientation and prescriptiveness of the NCLB policy, make it difficult for states to establish promising comprehensive support systems that draw upon all three important components of effective school reform: incentives, capacity, and opportunities to effect change (P. T. Hill & Celio, 1998). Obvious incentives within any NCLB-related school improvement system are the avoidance of restructuring and the embarrassment associated with the public reporting of poor performance scores (Mintrop, 2004). Although other, more positive types of external incentives are possible, such as the contingent availability of grants and positive public recognition, organized statewide examples of

these incentives are rare and are likely to remain so given current limitations on SEA human and fiscal capacity (Rhim et al., 2007).

Whatever the incentives, their connection to meaningful, sustained changes in school functioning is hindered in the absence of improved school capacity (Malen & Rice, 2004; Massell, 1998; Mintrop & Trujillo, 2005). For example, publicly reporting a school's poor academic performance may garner staff attention but may lead to incoherent stabs at change and to demoralization rather than to actual improvements in leadership, management, programming, or instruction in schools with limited capacity and no coherent support for improving that capacity (Finnigan & Gross, 2007; Fullan, 2003; Mintrop, 2004). Likewise, neither incentives nor improved capacity is sufficient to effect reform in the absence of legitimate opportunities to change (Mintrop & Trujillo, 2005; Rhim et al., 2007). In the tightly prescribed system of sanctions and compliance under NCLB, opportunities and adequate timelines for change may be particularly difficult to find (Mintrop & Sunderman, 2009; Sunderman, Tracey, Kim, & Orfield, 2004).

ASSUMPTION 3: SCHOOL IMPROVEMENT EFFORTS LEAD TO INCREASES IN STUDENT LEARNING

As we have seen, the methods for placing schools in improvement status and both the model for and the capacity to implement school improvement requirements are significantly flawed. Thus, the first two of the three assumptions identified in the proposed illustration of the NCLB policy logic just offered appear to be faulty. However, in spite of these failed assumptions about how NCLB is meant to work, is student learning improving in the context of the NCLB school improvement policy?

There is some evidence that statewide assessment scores have increased during the period in which NCLB has been in effect. A 50-states analysis of the percentages of students scoring in or above the proficient level of performance on statewide assessments in reading and mathematics between 2002 and 2008 revealed that these percentages increased over this period in 83% of the cases and trends favored increases over decreases (Center on Education Policy, 2009). Other studies using statewide assessment scores suggest similar outcomes (Center on Education Policy, 2008; The Education Trust, 2004, 2005). However, definitions of proficiency vary so greatly across states that apparently similar changes in the percent proficient across states may actually reflect very different degrees of mean changes in score distributions (Ho, 2007). Further, when other indicators are considered, such as scores on the National Assessment of Educational Progress (NAEP), trends are less positive than those demonstrated on state assessments during the NCLB period (Fuller, Wright, Gesicki, & Kang, 2007; Lee, 2006, 2007) and do

not increase in slope over the years prior to NCLB or the years since NCLB was enacted (Lee, 2007; Mintrop & Sunderman, 2009). This suggests that the increases in state assessment scores may reflect changes that do not reflect or support student learning, such as inappropriate test preparation practices and the narrowing of curricula (Koretz, 2008a, 2008b; Koretz & Barron, 1998). On the other hand, the rates at which traditionally lower achieving student groups, such as students with disabilities and English language learners, have been included in the samples on which NAEP scores are based have increased over time (National Center for Education Statistics, 2009) and remain lower than the inclusion rates for state assessments (Board on Testing and Assessment, 2004). These differences across time and assessment complicate the comparisons between performance and trends on NAEP and on state assessments (Haertel, 2003); this is particularly unfortunate given the original focus of Title I improving achievement among low-achieving students.

Even if one concludes that student learning has improved under NCLB, one cannot yet claim that this results from the school improvement requirements of the NCLB policy. Public school choice, which is so rarely exercised that its impact on student learning may not be discernable, and SES, which may be associated with improved student achievement (Zimmer, Gill, Razquin, Booker, Lockwood, Vernez, et al., 2007), are strategies that separate students from schools rather than promote school improvement as a means of improving student achievement. Restructuring and the implementation of school improvement plans are meant to address school improvement directly, and there is some evidence that the direction of some additional resources to schools and districts may be associated with improved student outcomes (Mintrop & Trujillo, 2005; Sunderman & Orfield, 2006). On the whole, however, the evidence that improved student achievement are associated with these efforts is inconclusive. Thus, it appears that the NCLB school improvement policies as they are implemented may not represent mechanisms for improving student achievement.

CONCLUSIONS

The assumptions underlying the NCLB policy logic hold that schools in need of improvement can be identified via a large-scale algorithm, that pre-established sanctions applied to these schools will lead to their improvement and that these improvements in identified schools will yield increases in student achievement. This argument is compelling for its simplicity and apparent rationality, but its assumptions seem to lack merit. Inadequacies in the algorithms used to identify schools for improvement status, the disconnect between the consequences associated with school improvement status and the strategies that could address the specific needs of an individual school, and the chasm between the capacity to support identified schools and the capacity within states

education agencies to provide this support all contribute to the gap between the goal and the reality. In other words, the NCLB approach to school improvement does not seem tenable. So what is a country to do?

Certainly, the next reauthorization of the ESEA will provide a great opportunity to address the shortcomings of the current version of the law and could do so without losing sight of the fundamental commitment to supporting high-quality educational opportunities for all students. All those who remain in the reauthorization conversation understand that an approach that does not include large-scale systems of assessments and accountability will not be politically viable. However, simple “tweaks” to the current system, such as allowing greater flexibility in minor aspects of AYP or switching the order of public school choice and SES sanctions will have no impact on school functioning or student achievement. Thus, the question is not how to smooth NCLB’s rough edges but how to create a fundamentally different model for improving schools that encompasses large-scale assessments and accountability systems without imposing unworkable limits restrictions on state and local autonomy? Addressing this big question in full is beyond the scope of this article and the capacity of its author. But five recommendations are offered based on the school improvement argument explored here.

First, the systemic reform model (Smith & O’Day, 1991) on which NCLB and its predecessor, the Improving America’s Schools Act of 1994, were based provides a strong foundation for reform and should be revisited as part of the preparations for reauthorization. NCLB undeniably retained the first of three critical components of this model, a common, comprehensive set of expectations and goals, as manifested in requirements for statewide systems of standards and assessments. NCLB partially retained the second component, which involves the provision of clear guidance and support for administrators and other educators in achieving the common expectations and goals, because it did not pair the concept of guidance with a means for building necessary capacities for implementation at the state and local levels. Perhaps this failure to recognize the importance of capacity to the implementation of the policy led also to the replacement of the third component of systemic reform, flexibility in governance and decision making, with a “state-proofed” prescription for evaluating schools and imposing standardized sanctions. The next four recommendations flow from this first.

Second, after recognizing the necessity of adequate capacity at the state level to the implementation of the policy, the next version of ESEA should return to its roots and commit to strengthening and restructuring SEAs as service providers. This commitment should include funding for individual states, prioritize collaboration among states, and provide for evaluations of the quality of services offered by comprehensive and regional centers.

Third, the next version of ESEA should encourage states, or consortia of states, to develop models for evaluating

schools that address school effectiveness rather than achievement status. That is, models for identifying schools that need intervention and for monitoring school progress should address questions such as: Is the school effective in improving students’ achievement from one year to the next? Are individual students making expected progress from grade to grade? Is the school becoming more effective? Are individual students learning at faster rates as they progress through the school? The answers to these questions should include, but not be restricted to, consideration of large-scale assessment data.

Fourth, the next version of ESEA should require states, or consortia of states, to develop school improvement systems that encompass incentives, mechanisms for state and local capacity-development, and opportunities to innovate. Allowing for concurrent strategies to ensure that students are being served, perhaps through well-monitored SES, could assuage anxieties that have pre-empted the patience necessary to implement reforms adequately and observe measureable manifestations of school improvement. To ensure that these state- or consortium-specific approaches do not fall short of the equity and excellence goals that underlie our federal education policy (Forte, 2007), states/consortia should be required to engage in formative and summative evaluations of implementation and effectiveness of these systems rather than simply base their design on research. States should be allowed to use school evaluation models that draw upon a range of data and be required to evaluate the degree to which these models effectively identify and address schools’ needs.

It is important to note that although the previous recommendations highlight the notion of collaboration among states, they are not directly addressed by the current Common Core State Standards Initiative being implemented by the National Governors’ Association, the Council of Chief State School Officers, Achieve, ACT, and the College Board (<http://www.corestandards.org/>). This initiative may lead to opportunities for states to collaborate on assessments and even some components of accountability systems but at present does not include plans for the collaborative development of student or school evaluation models or the pooling of data for such purposes. Individual states may resist this level of coinvolvement because it would greatly facilitate interstate comparisons; on the other hand, collaboration may free up resources that could then be made available for school improvement efforts.

Finally, innovation in school evaluation and improvement has been stifled in recent years by the narrow and inflexible interpretation of general principles for school reform. If the next reauthorization of ESEA is to improve prospects for schools that serve our nation’s most challenged students, its implementation should prioritize plausibility and promise that is supported by research and evaluation over legislated prescriptions based primarily on rhetoric.

REFERENCES

- Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., & Mirocha, J. (2005). *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives*. Los Angeles, CA: University of California at Los Angeles, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., & Dietel, R. (2004). Challenges in the No Child Left Behind Act for English language learners. *Phi Delta Kappan*, *85*, 782–785.
- Archer, J. (2006). Proposed grants for improvement highlight key new role for states. *Education Week*, *25*(24), 1.
- Board on Testing and Assessment. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments* (J. A. Koenig, & L. F. Bachman, Eds.). Washington, DC: The National Academies Press.
- Carlson Le Floch, K., Boyle, A., & Bowles Therriault, S. (2008a). *Help wanted: State capacity for school improvement*. Washington, DC: American Institutes for Research.
- Carlson Le Floch, K., Boyle, A., & Bowles Therriault, S. (2008b). *State systems of support under NCLB: Design components and quality considerations*. Washington, DC: American Institutes for Research.
- Carlson, D. (2006). *Focusing state educational accountability systems: Four methods of judging school quality and progress*. Dover, NH: The Center for Assessment. Retrieved from <http://www.nciea.org/publications/Dale020402.pdf>
- Center on Education Policy. (2007). *Has student achievement increased since No Child Left Behind?* Washington, DC: Author.
- Center on Education Policy. (2008). *A call to restructure restructuring: Lessons from the No Child Left Behind Act in five states*. Washington, DC: Author.
- Center on Education Policy. (2009). *Is the emphasis on "proficiency" short-changing higher- and lower-achieving students?* Washington, DC: Author.
- Chicago Public Schools. (2007). *SES tutoring programs: An evaluation of year 3 in the Chicago Public Schools*. Chicago, IL: Author.
- Choi, K., Goldschmidt, P., & Yamashiro, K. (2006). *Exploring models of school performance: From theory to practice*. Los Angeles, CA: University of California at Los Angeles, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Council of Chief State School Officers. (2006). *Consortium for School Improvement policy brief: State support to schools in need of improvement*. Washington, DC: Author. Retrieved from http://www.ccsso.org/content/pdfs/SSSNI_FINAL.pdf
- Council of Chief State School Officers. (2009). *Schools identified for improvement, or not making AYP, by state: 2006-7*. Retrieved from <http://www.ccsso.org/content/pdfs/Schools%20identified%20for%20improvement%202006-07%20AYP.pdf>
- Cross, C. (2004). *Political education: National policy comes of age*. New York, NY: Teachers College Press.
- DiBiase, R. W. (2005). *ECS policy brief: State involvement in school restructuring under No Child Left Behind in the 2004-05 school year*. Denver, CO: Education Commission of the States. Retrieved from <http://www.ecs.org/clearinghouse/64/28/6428.pdf>
- Dwyer, C., with Bormand, J., Cevoen, L., Fafard, M. B., Frankel, S., Harvell, C., & Zarlengo, P. (2005). *Leadership capacities for a changing environment: state and district responses to the No Child Left Behind Act of 2001*. Providence, RI: Education Alliance at Brown University.
- Education Commission of the States. (2002). *State intervention in low-performing schools and school districts*. Denver, CO: Author.
- The Education Trust. (2004). *Measured progress: Achievement rises and gaps narrow, but slowly*. Washington, DC: Author.
- The Education Trust. (2005). *Stalled in secondary: a look at student achievement since the No Child Left Behind Act*. Washington, DC: Author.
- Elam, S. (1995). *How America views its schools: The PDK/Gallup polls, 1969-1994*. Bloomington, IN: Phi Delta Kappa.
- Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, *44*, 594–629.
- Forte, E. (2007, November). *Assessment and accountability post-NCLB: The next generation*. Invited keynote address for the Annual Conference of the California Educational Research Association, Dana Point, CA.
- Forte, E., & Erpenbach, W. J. (2006). *Statewide educational accountability under the NCLB Act—a report on 2006 amendments to state plans*. Washington, DC: Council of Chief State School Officers.
- Forte Fast, E., & Erpenbach, W. J. (2004). *Revisiting statewide educational accountability under NCLB: An examination of states' initial requests for amendment under the No Child Left Behind Act of 2001*. Washington, DC: Council of Chief State School Officers.
- Fullan, M. (2003). *Change forces with a vengeance*. London, UK: Routledge Farmer.
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher*, *36*, 268–278.
- Garland, L. (2003). *Navigating treacherous waters: A state takeover handbook*. Lanham, MD: Scarecrow.
- Gong, B. (2002). *Designing school accountability systems: Toward a framework and process*. Washington, DC: CCSSO.
- Gong, B. (2006, April). *Valid OR reliable?: Subgroups in NCLB's AYP*. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.
- Haertel, E. H. (2003). Including students with disabilities and English language learners in NAEP: Effects of differential inclusion rates on accuracy and interpretability of findings. [Paper commissioned and prepared by the National Assessment Governing Board]. Retrieved from ERIC database. (ED500430)
- Hannaway, J., & Woodroffe, N. (2003) Policy instruments in education. *Review of Research in Education*, *27*, 1–24.
- Hill, P. T., & Celio, M. B. (1998). *Fixing urban schools*. Washington, DC: Brookings Institution Press.
- Hill, R., & DePascale, C. A. (2003, June). *Adequate yearly progress under NCLB: Reliability considerations*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL. Retrieved from http://www.nciea.org/publications/NCMEconf_RHCD03.pdf
- Ho, A. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice*, *26*(4), 11–20.
- Jennings, J., Stark Rentner, D., & Kober, N. (2002). *A new federal role in education*. Washington, DC: Center on Education Policy.
- Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*, 91–114.
- Kerins, T., Perlman, C., & Langdon, J. (2009). Results from the statewide systems of support survey: Five case study states. In T. Kerins, C. Perlman, & S. Redding (Eds.), *Coherence in statewide systems of support* (pp. 47–92). Lincoln, IL: Academic Development Institute.
- Kim, J. S., & Sunderman, G. L. (2004). *Does NCLB provide good choices for students in low performing schools?* Cambridge, MA: The Civil Rights Project at Harvard University.
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, *34*(8), 3–13.
- Kopriva, R. (2008). *Improving testing for English language learners*. New York City, NY: Routledge.
- Koretz, D. (2008a). A measured approach. *American Educator*, Fall, 8–39.
- Koretz, D. (2008b). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

- Koretz, D., & Barron, S. (1998). *The validity of gains in scores on the Kentucky instructional results information system (KIRIS)*. Washington, DC: RAND.
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Lee, J. (2007). *The testing gap: Scientific trials of test-driven school accountability systems for excellence and equity*. Charlotte, NC: Information Age.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3–13.
- Linn, R. L. (2005a). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives*, 13(33). Retrieved from <http://epaa.asu.edu/epaa/v13n33/>
- Linn, R. L. (2005b). Fixing the NCLB accountability system (CRESST Policy Brief 8). Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation, Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Linn, R. L. (2006). Validity of inferences from test-based educational accountability systems. *Journal of Personnel Evaluation in Education*, 19(1–2), 5–15.
- Linn, R. L. (2008a). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40, 699–711.
- Linn, R. L. (2008b). Toward a more effective definition of Adequate Yearly Progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 27–42). Thousand Oaks, CA: Corwin.
- Linn, R. L., Baker, E., & Betebenner, D. (2002). *Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001* (CSE Report No. 567). Los Angeles, CA: University of California at Los Angeles, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from ERIC database. (ED467440)
- Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 24(1), 29–36.
- Lusi, S. F. (1997). *The role of state departments of education in complex school reform*. New York, NY: Teachers College Press.
- Malen, B., Croninger, R., Muncey, D. & Redmond-Jones, D. (2002). Reconstituting schools: “Testing” the “theory of action.” *Educational Evaluation and Policy Analysis*, 24(2), 113–132.
- Malen, B., & Rice, J. K. (2004). A framework for assessing the impact of education reforms on school capacity: Insights from studies of high-stakes accountability initiatives. *Educational Policy*, 18 631–660.
- Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers. Retrieved from <http://www.ccsso.org/content/pdfs/AYPpaper.pdf>
- Martineau, J. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31(1), 35–62.
- Massel, D. (1998). *State strategies for building capacity for education: progress and continuing challenges* (Research Report Series RR-41). Philadelphia, PA: Consortium for Policy Research in Education.
- Mathis, W. (2009). *NCLB’s ultimate restructuring alternatives: Do they improve the quality of education?* Boulder, CO, and Tempe, AZ: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from <http://epicpolicy.org/publication/nclb-ultimate-restructuring>
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- Minnici, A., & Hill, D. (2007). *Educational architects: Do state education agencies have the tools necessary to implement NCLB?* Washington, DC: Center on Education Policy.
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn’t work)*. New York, NY: Teachers College Press.
- Mintrop, H., & Sunderman, G.L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement – and why we may retain it anyway. *Educational Researcher*, 38(5), 353–364.
- Mintrop, H., & Trujillo, T. (2005). Corrective action in low-performing schools: Lessons for NCLB implementation from state and district strategies in first generation accountability systems. *Education Policy Analysis Archives*, 13(48), 1–30.
- National Center for Education Statistics. (2009). *NAEP inclusion policy*. Retrieved from <http://nces.ed.gov/nationsreportcard/about/inclusion.asp>
- Popham, W. J. (2004, July). *Ruminations regarding NCLB’s most malign provision: Adequate Yearly Progress*. [Paper prepared for the Center on Education Policy Forum] Retrieved from <http://www.cep-dc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documentid=216&documentFormatId=3509>
- Public Law 107-110-Jan. 8, 2002 115 Stat. 1499.
- Raudenbush, S. (2004). *Schooling, statistics, and poverty: Can we measure school improvement? The Ninth Annual William H. Angoff Memorial Lecture*. Princeton, NJ: Education Testing Service. Retrieved from <http://www.ets.org/Media/Research/pdf/PICANG9.pdf>
- Redding, S. (2007). An evolution in American education. In S. Redding & H. Wahlberg (Eds.), *Handbook on statewide systems of support* (pp. 57–74). Lincoln, IL: Academic Development Institute.
- Reville, S. P. (2004). High standards + high stakes = high achievement in Massachusetts. *Phi Delta Kappan*, 85(8), 591.
- Reville, S.P., Coggins, C., & Candon, J. (2005). *Reaching capacity: A blueprint for the state role in improving low performing schools and districts*. Boston, MA: Rennie Center for Education Research & Policy.
- Rhim, L., Hassel, B., & Redding, S. (2007). State role in supporting school improvement. In S. Redding & H. Walberg (Eds.), *Handbook on statewide systems of support* (pp. 21–56). Lincoln, IL: Academic Development Institute.
- Rose, L. C., & Gallup, A. M. (2006). The 39th annual Phi Delta Kappa/Gallup poll of the public’s attitudes toward the public schools. *Phi Delta Kappan*, 88(1), 41–56.
- Rose, L. C., & Gallup, A. M. (2007). The 39th annual Phi Delta Kappa/Gallup poll of the public’s attitudes toward the public schools. *Phi Delta Kappan*, 89(1), 33–48.
- Ross, S., Harmon, J., & Wong, K. (2009). *Improving SES quality state approval, monitoring, and evaluation of SES providers*. Retrieved from <http://www.centerii.org/survey/downloads/Improving%20SES%20Quality.pdf>
- Ryan, K., & Shepard, L. (2008). *The future of test-based educational accountability*. New York, NY: Routledge.
- Simpson, M. A., Gong, B., & Marion, S. (2005). *Effects of minimum cell sizes and confidence interval sizes for special education subgroups on school-level AYP determinations*. Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved from <http://www.nciea.org/publications/MinNCLBGSMMMS05.pdf>
- Smith, M., & O’Day, J. (1991). *Putting the pieces together: Systemic school reform* (CPRE Policy Brief). New Brunswick, NJ: Eagleton Institute of Politics.
- Stark Retner, D., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S. & Zabala, D. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.
- Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A. E., McCombs, J.S., & Naftel, S. (2008). *Pain and gain: Implementing*

- No Child Left Behind in three states, 2004-2006*. Washington, DC: RAND.
- Sunderman, G. L., & Orfield, G. (2006). Domesticating a revolution: No Child Left Behind reforms and state administrative response. *Harvard Educational Review*, 76(4), 526–556.
- Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Toch, T. (2001, November). Bush's big test. *Washington Monthly*. Available from <http://www.washingtonmonthly.com/>
- Unger, C., Lane, B., Cutler, E., Lee, S., Whitney, J., Arruda, E., & Silva, M. (2008). *How can state education agencies support district improvement? A conversation amongst educational leaders, researchers, and policy actors*. Providence, RI: The Education Alliance at Brown University.
- U.S. Department of Education. (2002). *Title I—Improving the academic achievement of the disadvantaged, Final regulations, 34 CFR Part 200*. Retrieved from <http://www.ed.gov/legislation/FedRegister/finrule/2002-4/120202a.html>
- U.S. Department of Education. (2006). *LEA and school improvement: Non-regulatory guidance*. Retrieved from <http://www.ed.gov/policy/elsec/guid/schoolimprovementguid.pdf>
- U.S. Department of Education. (2009a). *Evaluation of the 2005–06 growth model pilot program*. Retrieved from <http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc>
- U.S. Department of Education. (2009b). *State and local implementation of the No Child Left Behind Act: Volume VII – Title I school choice and supplemental education services: final report*. Retrieved from <http://www.ed.gov/rschstat/eval/choice/nclb-choice-ses-final/choice-ses-final.pdf>
- U.S. Department of Education. (2009c). *Title I implementation: Update on recent evaluation findings*. Retrieved from <http://www.ed.gov/rschstat/eval/disadv/titlei-implementation/titlei-implementation.doc>
- Vernez, G., Naftel, S., Ross, K.E., Carlson Le Floch, K., Beighley, C., Gill, B. . . . O'Day, J. (2009). *State and local implementation of the No Child Left Behind Act: Volume VII — Title I school choice and supplemental educational services: Final report*. Santa Monica, CA: RAND.
- Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J. R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I – Title I schools choice, supplemental educational services, and student achievement*. Washington, DC: RAND.
- Zimmer, R. W., Gill, B. P., Razquin, P., Booker, K., Lockwood, J. R., Vernez, G., & O'Day, J. (2007). *Do Title I school choice and supplemental educational services affect student achievement?* Washington, DC: RAND.

Copyright of Educational Psychologist is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.