

CRESST Policy Brief 8

Fixing the NCLB Accountability System



Robert L. Linn

Robert L. Linn is co-director of the National Center for Research on Evaluation, Standards, and Student Testing and former president of the American Educational Research Association and the National Council on Measurement in Education.

The No Child Left Behind (NCLB) Act is praiseworthy for the special attention it gives to improved learning for children who have been ignored or left behind in the past. The emphasis on closing the achievement gap is certainly commendable, as is the encouragement given to states to adopt ambitious subject matter standards and enhance teacher quality. NCLB's focus on students with low achievement seems to have had some short-term positive effects. The percentage of schools meeting Adequate Yearly Progress (AYP) targets increased in 2003-04 from the year before in most states, and the recently released National Assessment of Educational Progress (NAEP) long-term trend scores have shown some narrowing of achievement gaps.

Given the positives, we might conclude that NCLB is working, and hence no changes are needed at this point. Unfortunately, the accountability system has some fundamental problems that threaten to undermine its central goals over the next few years. Dissatisfaction with some of the accountability provisions led the U.S. Department of Education to make some changes in NCLB accountability requirements last year, with more on the way this year.¹ The changes, however, are what Jim Popham calls "edge-softening" and do not deal with NCLB's fundamental problems,² which include expectations, targets, state proficiency levels, reporting, and the safe harbor provision. The remainder of this policy brief describes each problem and offers proposals for improvement.



FUNDAMENTAL PROBLEMS AND PROPOSALS FOR IMPROVEMENT

Expectations

The most serious problem is that the NCLB expectations for student achievement have been set unrealistically high, requiring that by the year 2014, 100% of students must reach the proficient level or above in math and reading. Based on current improvement levels and without major changes in the definition of adequate yearly progress (AYP), almost all schools will fail to meet NCLB requirements within the next few years.

Using test scores from the National Assessment of Educational Progress (NAEP) as an example, we note that in 2003, no state or large district had anything close to 100% of their students performing at the basic NAEP achievement level, much less the NAEP proficient level, in either Grade 4 or Grade 8 in either reading or mathematics. In 2003, New Hampshire, traditionally one of the highest performing states, still had 25% of its students performing at the below basic achievement level in fourth-grade reading (see Figure 1). On the same assessment, the District of Columbia reported 69% of its students at the below basic level. California, with the largest student population of any state, had just 21% of its students at or above the proficient level in 2003.

2003 New Hampshire 4th-Grade NAEP Reading Achievement Levels

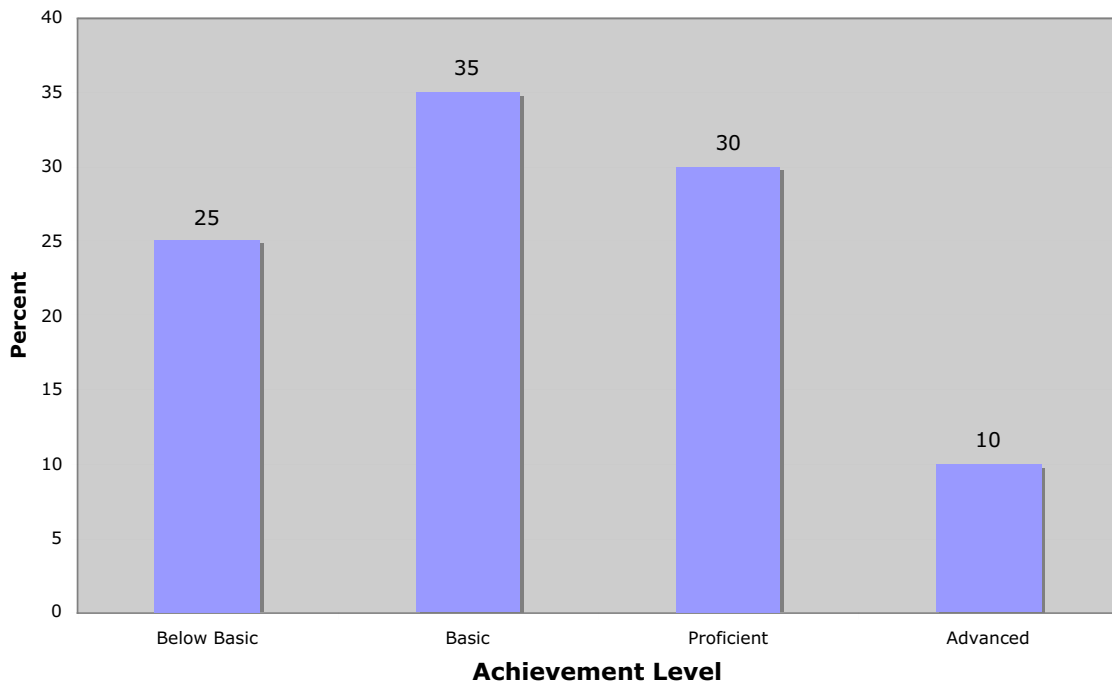


Figure 1

Furthermore, the law requires unrealistically rapid rates of improvement when compared to incremental improvements on earlier NAEP assessments. In fourth-grade mathematics, for example, the percentage of students at the proficient level or above on NAEP would have to have an annual improvement rate 3.9 times faster than the rate of increase between 1996 and 2003 (see Figure 2). For eighth-grade mathematics, the rate of improvement would need to be 7.5 times faster. Such rapid acceleration of math achievement is unrealistic.

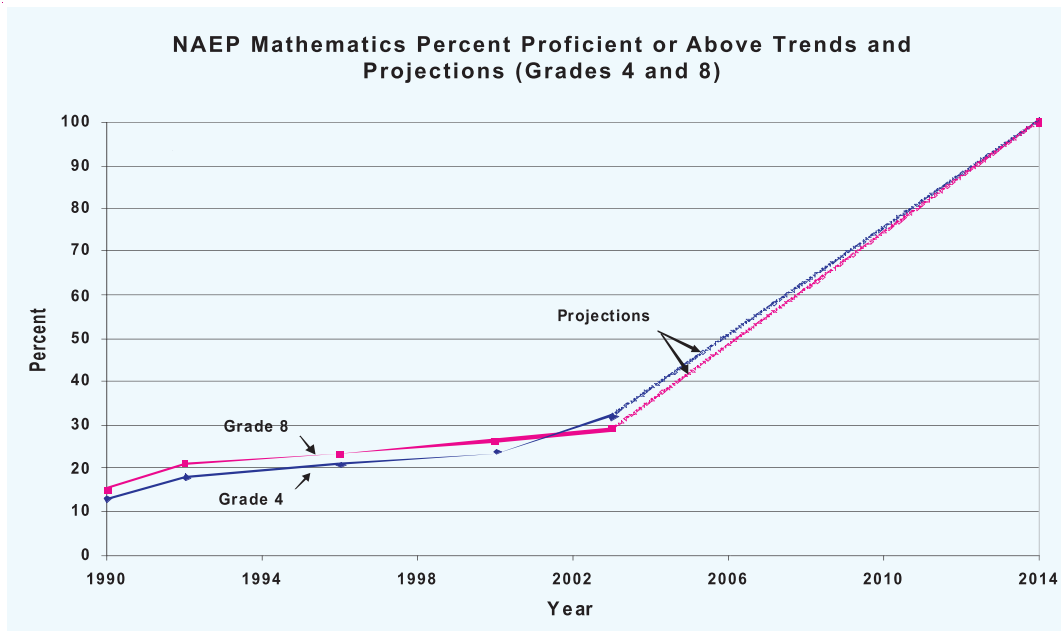


Figure 2

In reading, the rate of increase would require an even more unlikely jump. In fourth grade, 31% of the students scored at the proficient level or above on the 2003 reading NAEP assessment compared to 29% in 1998. At Grade 8, 32% were proficient or above in both 1998 and 2003, representing no increase in 5 years. To reach NCLB goals by 2014, student performance would need to increase from essentially a flat line to at least 7 percentage points every year for 9 years.

Further complicating the high expectations problem is that proficiency levels on state assessments vary substantially. Some state proficiency levels are similar to NAEP, others are more lenient, and a few states are more stringent. Missouri, for example, is more stringent. In 2003, 21.3% of Missouri’s eighth-grade students scored at the proficient level or above on the Missouri mathematics assessment. In the same year and grade level, 28% of Missouri students scored at the proficient level or above on NAEP mathematics. The Missouri 3-year trend of percentage proficient or above on its own Grade 8 mathematics test is virtually flat, 21.1% in 2002, 21.3% in 2003, and 22.9% in 2004. Reaching 100% proficient or above by 2014 is not realistic.

California achievement levels are somewhat more lenient than NAEP. In 2003, 31% of California’s eighth-grade students scored at the proficient level or above on the California English-language arts test, whereas only 22% scored proficient or above on the 2003 NAEP reading assessment. As with Missouri, eighth-grade California students’ language arts performance was virtually flat. Scores for 2001, 2002, 2003, and 2004 were 32, 32, 31, and 33 percent proficient or above respectively. Flat-line performance does not augur well for reaching 100% by 2014.

I’ve previously argued that performance goals “mandated by the accountability system should be ambitious, but also should be realistically obtainable with sufficient effort.”³ At the very least, there needs to be an existence proof. That is, there should be evidence that the goal does not exceed one that has previously been achieved by the highest performing schools. For example, if the top 10% of schools in a state improved an annual average of 3% proficient or above each year in the past 5 years, then 3% might be the annual state goal. That would be a major challenge to the vast majority of schools, but might be a target that is within reach with sufficient effort.



Fixed Targets

Although the “P” in AYP stands for progress, it is important to recognize that, with the exception of the “safe harbor” provision discussed later, the comparison of performance in a given year to a fixed target—known as the annual measurable objective (AMO)—is the sole determinant of whether a school or district makes AYP. For example, Nevada’s elementary school annual measurable objective in 2004-05 is 45.4% proficient or above in mathematics. A low-performing school that made a dramatic improvement in the prior year, let’s say increasing from 30% proficient or above to 40%, would still fall below the state objective and be placed on a watch list. Meanwhile, a higher performing school whose scores had actually dropped, say from 65% proficient or above to 55%, would still meet its annual measurable objective because it remained above the 45.4% proficient level or above (see Figure 3). While setting fixed targets makes goals easy to understand, it may produce distorted results and does not necessarily treat schools fairly, especially those that are making substantial progress. Indeed, schools that make substantial progress yet fail to reach ever increasing fixed targets may lose incentive to make further improvements.

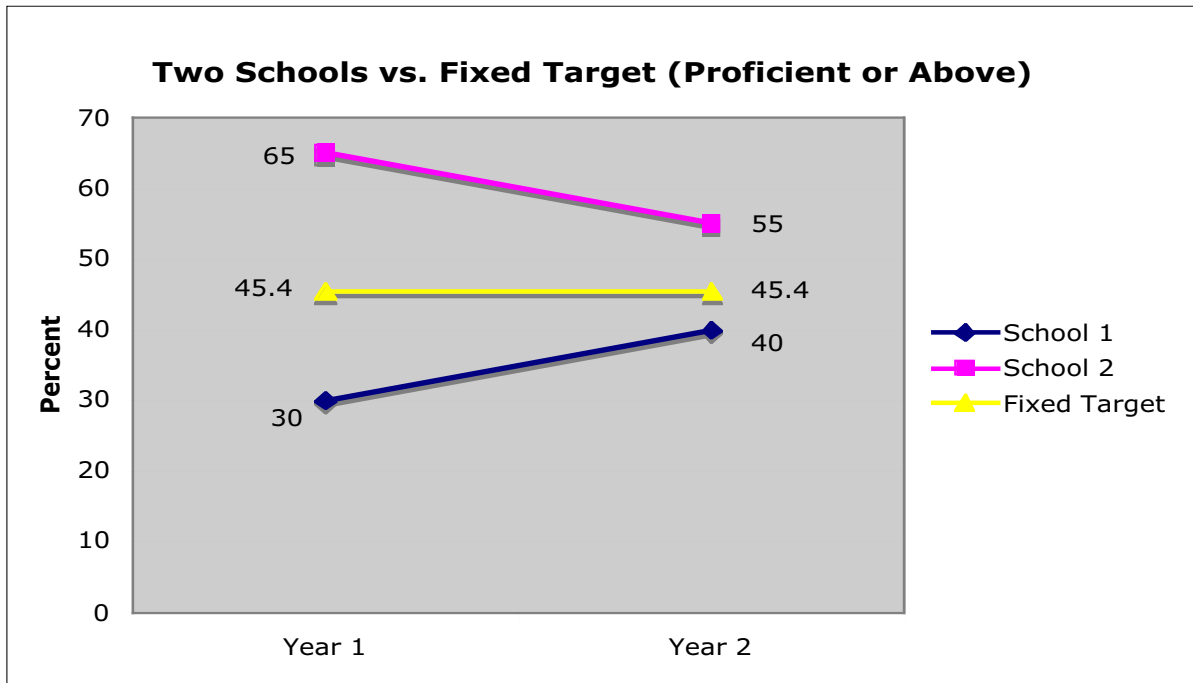


Figure 3

Solutions to this problem are quite simple. Indeed, many state accountability systems already use both status and improvement to measure progress. A typical method is to compare successive groups of students, for example, fourth-grade students in 2005 compared to fourth-grade students in 2004. With the NCLB requirement to test students in Grades 3 through 8 every year, some states are measuring changes longitudinally, tracking improvement from one grade to the next grade, for example, from third grade to fourth to fifth grade and beyond. Tennessee has used the longitudinal method for a number of years, and their “value-added” approach to evaluating schools has attracted a great deal of attention.



Either successive or longitudinal approaches, or both, should be included in determination of AYP.⁴ U.S. Secretary of Education Margaret Spellings has indicated that she is open to the possibility of using growth in the determination of AYP. This is a step in the right direction; however, allowing growth to determine AYP will be of limited value if the expectation of 100% proficient or above by 2014 is maintained.

Definition of Proficient

NCLB requires states to set “challenging student academic achievement standards.” The problem is that states have set the achievement standards in ways that vary greatly in stringency, especially compared to NAEP. Figure 4 shows this variation, arbitrarily using reading results from the first four states in the alphabet, Alabama, Alaska, Arizona, and Arkansas. Note the very large differences in achievement level scores between NAEP tests and state tests in 2003. The smallest difference is in Arkansas, 34 percentage points, and the largest difference is in Alabama, 55 percentage points. Gaps are similar for most states, in both mathematics and reading, and in eighth grade as well as fourth.

Compared to NAEP performance standards, most states have set very lenient standards, such that a large majority of their students perform at the proficient level or above on their assessments. Figure 4 also suggests that most states will be far below the 100% proficient goal in 2014 when measured by NAEP.

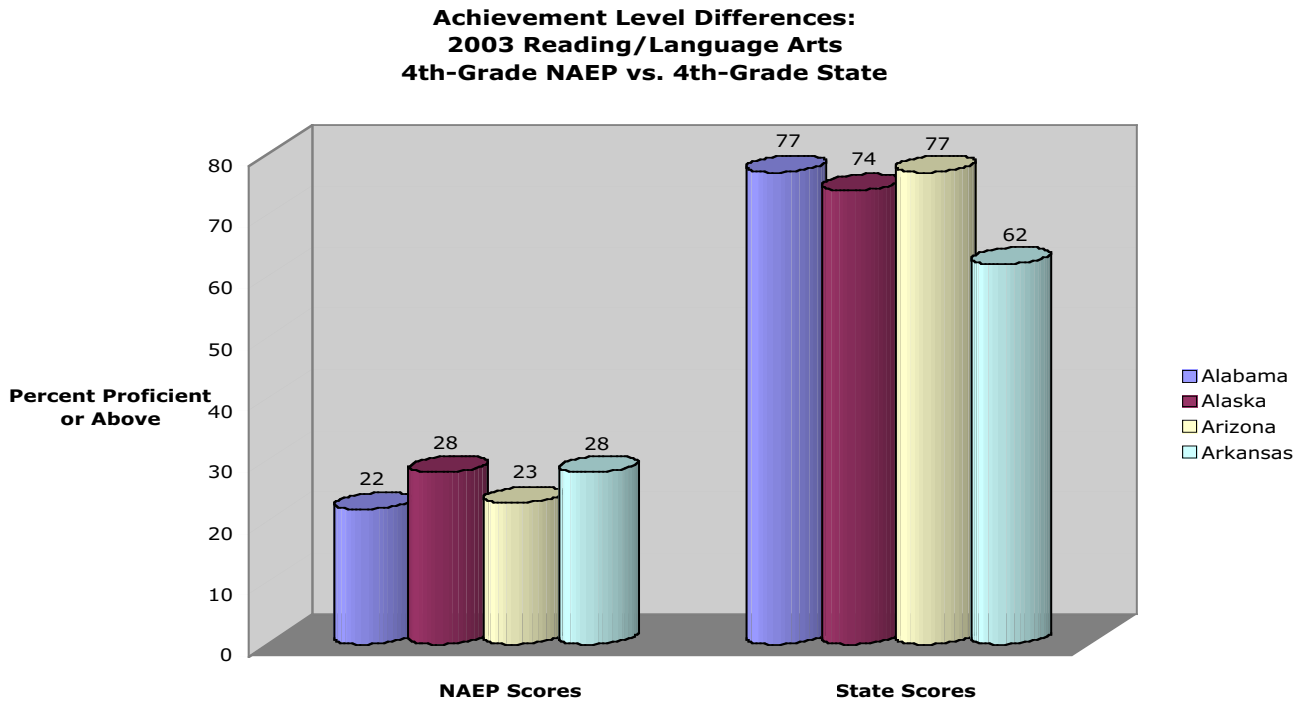


Figure 4



Performance standards differ substantially between states as well. For example, Figure 5 shows only modest differences in eighth-grade NAEP mathematics scores between Colorado and Missouri students. In 2003, 34% of Colorado students scored proficient or above on NAEP compared to 28% for Missouri.⁵ But on their own state tests in 2003, 67% of Colorado students scored proficient or above compared to just 21.3% in Missouri. The difference between the 67% and the 21.3% clearly has more to do with how rigorously states set their performance standards than with real differences in achievement between students. With each state setting its own performance standards and without an external, independent test, the word “proficient” loses meaning.

Different State Performance Standards

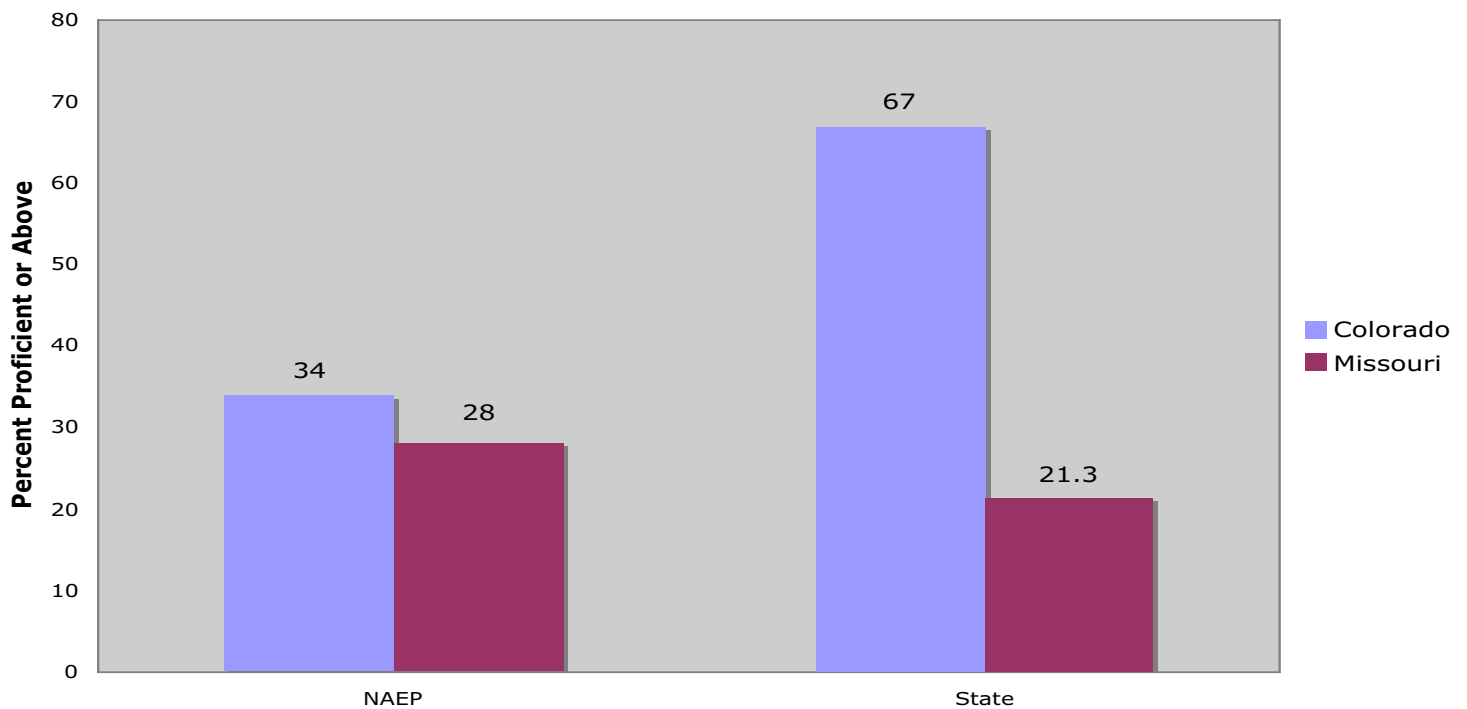


Figure 5

Fortunately, there are several preferable approaches to reporting results in terms of percent proficient or above. One simple approach is to define the standard or cut score on a state assessment to be equal to the median score in a base year, presumably 2002. The percentage of students scoring above that constant cut score would then be used to monitor improvement in achievement with target increases set at reasonable levels, for example, 3% per year. Let’s say the median score in fifth-grade mathematics was 50% for a state in 2002. In 2003, the target would have been 53%; in 2006, the target would be 62%; and in 2014, the target would be 86%. This would represent a gigantic improvement in the achievement of the state’s students, but might not be totally unrealistic, and surely is not as implausible as 100% proficient or above.



Another alternative would be to use what Jim Popham has called grade-level descriptions.⁶ “At-grade-level” might correspond more closely to the “basic” than the “proficient” level in most states. Using past experience, targets could be set that would bring the achievement of an ever increasing percentage of students up to the “at-grade-level” standard.

Disaggregation

As stated earlier, a positive feature of NCLB is its emphasis on groups of low-achieving students who have too often been ignored in the past. The accountability system attempts to assure adequate attention to these groups of students by requiring the separate reporting of results for economically disadvantaged students, students with disabilities, limited English proficient students, and by race/ethnicity. Such disaggregated reporting of results provides a mechanism for monitoring the achievement of lower performing groups and narrowing achievement gaps.

However, reporting more groups increases the number of ways that schools can fail to meet AYP. If just one group fails, the entire school misses its goal. This places schools with many student subgroups at a distinct disadvantage, usually urban schools with large numbers of Latino, African American, ELL, or special needs students. For example, in 2003, more than 75% of the 937 California schools reporting results for two subgroups made their AYP target. In the same year, only 39% of the 291 schools reporting results for six subgroups made their AYP target.⁷

A straightforward fix to the over-identification of schools as not meeting AYP would be to change the safe harbor provision as discussed next.

Safe Harbor

NCLB includes a safe harbor provision. If a subgroup of students in a school falls short of the AYP target, the school can still meet AYP if (a) the percentage of students who score below the proficient level is decreased by 10% from the year before, and (b) there is improvement for that subgroup on other indicators. In application, the safe harbor provision helps very few schools because the bar is set so high. The 10% decrease in students scoring below proficient far exceeds progress even in schools with exceptional achievement improvements. Not surprisingly, only a tiny fraction of schools have met AYP through the safe harbor provision since NCLB’s inception. A more realistic criterion is a 3% reduction in the below proficient category. Resetting the safe harbor provision to a more realistic rate would solve many of the problems caused by expanded subgroup reporting, yet still promote significant achievement increases.



CONCLUSION

NCLB has the potential to make substantial contributions to the achievement of students who have lagged behind and been ignored in the past. Some features of the NCLB accountability system, however, need to be modified if the praiseworthy goals of NCLB are going to be achieved. I offer three suggestions.

- 1. The most important modification is to set realistic performance targets for adequate yearly progress, rewarding effort with success. The need for more realistic goals applies to both the safe harbor provision of the law and to the annual performance targets.**
- 2. AYP should be determined by a consideration of growth in achievement and not just status in comparison to a fixed target.**
- 3. The current definitions of proficient achievement established by states lack any semblance of a common meaning. Alternatives to defining proficiency should be considered that would provide more meaningful and comparable achievement targets.**

Notes

- ¹See Olson, L. (2005). Requests win more leeway under NCLB. *Education Week*, 24(42), p. 1.
- ²Popham, W. J. (2004). Shaping up the 'No Child' Act: Is edge-softening enough? *Education Week*, 23(38), p. 40.
- ³Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.
- ⁴See Hoff, D. J. (2005). States to get new options on NCLB Law. *Education Week*, 24(31), pp. 1, 38. Olson, L. (2005). States hoping to grow into AYP success. *Education Week*, 24(37), pp. 15, 20.
- ⁵Colorado uses the partially proficient level for state reporting as the proficient level for purposes of NCLB.
- ⁶Popham, W. J. (2004). *Ruminations regarding NCLB's most malignant provisions: Adequate yearly progress*. Retrieved August 8, 2005, from <http://www.ctredpol.org/pubs/Forum28July2004>
- ⁷Based on Novak, J. R., & Fuller, B. (2003). *Penalizing diverse schools?* (PACE Policy Brief 03-4). Berkeley: University of California, Policy Analysis for California Education.



Center for Research on Evaluation,
Standards, and Student Testing
Eva L. Baker, Co-Director
Robert L. Linn, Co-Director
Joan L. Herman, Co-Director
Daniel Koretz, Associate Director
Ronald Dietel, *CRESST* Line Editor
Katharine Fry, Editorial Assistant

UCLA Center for the Study of Evaluation
CSE / CRESST
GSE&IS BLDG MAILBOX 951522
LOS ANGELES CA 90095-1522
ADDRESS SERVICE REQUESTED

Presorted Standard
U.S. Postage
PAID
Permit No. 400
Laguna Beach, CA

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this publication do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.