

Rethinking the No Child Left Behind Accountability System

Robert L. Linn

National Center for Research on Evaluation, Standards, and Student Testing

University of Colorado at Boulder

Paper prepared for a forum on No Child Left Behind sponsored by the Center on Education Policy, Washington, DC, July 28, 2004. Work on this paper was partially supported under the Educational Research and Development Centers Program, PR/Award Number R305B96002, as administered by the Institute for Education Sciences, U.S. Department of Education. The findings and opinions expressed 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.

Rethinking the No Child Left Behind Accountability System

The No Child Left Behind (NCLB) Act is the centerpiece of President Bush's educational agenda. NCLB was passed by an overwhelming bipartisan majority in both houses of Congress. It provides over 12 billion dollars in financial assistance to states, districts, and schools with high percentages of poor children to help improve student achievement. NCLB has much that is worthy of praise. It is particularly praiseworthy for its emphasis on all children and special attention it gives to improving learning for children that have been too often ignored or left behind in the past. The emphasis on closing the achievement gap is certainly praiseworthy. The encouragement that NCLB gives to states to adopt ambitious subject matter standards is also noteworthy.

Accountability is a prominent feature of NCLB. The NCLB accountability system is intended to contribute to improving education for all children by identifying schools that are in need of improvement so that corrective action can be taken. The system requires that all students achieve at the proficient level or higher (as defined by states) in mathematics and reading/English language arts by 2014.

Test-based educational accountability systems are often the subject of controversy. Hence, it is not surprising that the NCLB accountability system has been the focus of a great deal of controversy during the past two years. Dissatisfaction with some of the accountability provisions has led the U.S. Department of Education to make some changes in NCLB accountability requirements in recent months. The changes, however, are what Jim Popham (2004) recently referred to as "edge softening" and do not deal with the fundamental problems of the accountability system. This leads to two questions. What are the fundamental problems? How can the accountability system be improved so that the worthwhile goals of NCLB can be realized?

Fundamental Problems and Proposals for Improvement

The NCLB accountability system has several fundamental problems that need to be addressed. Five of the most serious problems are identified below and proposals for improvement are provided.

Unreasonable Expectations

The most serious problem is that the expectations for student achievement have been set unrealistically high and, as a consequence, almost all schools will fall short of the adequate yearly progress (AYP) targets within the next few years, unless major changes are made in the definition of AYP.

NCLB requires states to participate every other year in the National Assessment of Educational Progress (NAEP) reading and mathematics assessments at grades 4 and 8 starting in 2003. Although the use of state-level NAEP results are not specified in the law, it is reasonable to think of those results as providing some kind of benchmark for state assessments. In 2003, no state or large district had anything close to 100% of their students performing at the **basic** level, much less the **proficient** level at either grade 4 or grade 8 in either reading or mathematics.

Furthermore, the law requires an unrealistically rapid rate of improvement. This becomes apparent when the required changes are compared to a variety of different types of evidence, including historical data on change in student performance on state testing programs or gains that have been made on NAEP. In mathematics, for example, the percentage of students at the proficient level or above on NAEP would have to have an annual rate of improvement between 2003 and 2014 that is 2.3 times as fast at grade 4 as the rate actually realized between 2000 and 2003. At grade 8, the rate of improvement in the percentage of students at the proficient level or above in mathematics would need to be 6.5 times as rapid between 2003 and 2014 as it was between 2000 and 2003. Such rapid acceleration of achievement trends is unrealistic. In reading, the rate of increase in percentage proficient or above is an even more unrealistic jump. The essentially flat trend lines in the percentage of

students who are proficient or above in reading at both grades 4 and 8 would have to suddenly accelerate and maintain that previously unseen rate of improvement for the next decade.

I've previously argued that performance goals "mandated by the accountability system should be ambitious, but also should be realistically obtainable with sufficient effort" (Linn, 2003, p.4). At the very least, there needs to be an existence proof. That is, there should be evidence that the goal does not exceed that that has previously been achieved by the highest performing schools. For example, if the best performing 10% of schools in a state had rates of improvement in the percentage of students achieving at the proficient or above during the past 5 years that averaged 3% per year, then adequate yearly progress might be defined as a 3% increase in the percentage of students achieving at the proficient or above each year. That would be a great challenge to the vast majority of schools, but might be a target that is within reach with sufficient effort. As is discussed below, a target based upon an effect size statistic would be even better.

Definition of Proficient

NCLB requires states to set "challenging student academic achievement standards" that "describe two levels of high achievement (proficient and advanced ... [and] a third level of achievement (basic)" (NCLB, 2001). The problem is that states set the student academic achievement standards and they have done so in ways that vary greatly in stringency. Some states have set very lenient standards such that 80% or more of their students perform at the proficient level or above on their assessments while other states have set such stringent standards that 20% or less of their students perform at that level. Moreover, the differences in stringency have essentially no relationship to differences in the performance of students in the various states on NAEP. A consequence of this huge variability in stringency of state standards is that, in the context of NCLB, the word "proficient" has become meaningless. Certainly, reporting results in terms of percent proficient on state assessments lacks comparability from state to state.

As is discussed below, tracking the percentage of students above a cut score whether it is called proficient or something else is not necessarily the best way to monitor progress. However, if the percentage of students who are above a cut score on a state assessment is to be used, the cut score should be more meaningful than the state established proficient levels which lack any semblance of a common meaning across states. There are several approaches that would be preferable to reporting results in terms of percent proficient or above. One simple approach would be 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, e.g., 3% per year. With a target increase of 3% a year, the proportion of students scoring above the 2002 median would need to increase from 50% in 2002 to 86% in 2014. That would represent a gigantic improvement in the achievement of the nation's students, but might not be totally unrealistic, and surely is not as poorly defined as 100% proficient or above given the huge state-to-state variability in the meaning of proficient.

Another alternative would be to compare change over time to the current variability of student performance. This is usually done by dividing the change in mean scores by the standard deviation of scores, a measure often called an "effect size." The use of effect size does not suggest any acceptance of the current distribution of performance as adequate. Rather, effect size statistics are used routinely because they make available a variety of evidence that can be used to gauge the rate of change—e.g., historical data on trends, international comparisons, and evaluations of the effects of past educational interventions.

An effect size for 2003 would be equal to the difference in the mean achievement score in 2003 and the mean in 2002 divided by the standard deviation in 2002. Using effect size statistics for the top performing 10% of schools, an annual target increase in effect size could be set. This might result in an annual target increase in effect size of, say, .05. Thus, a school would need to have an effect size of .6 when 2014 was compared to the 2002 base year. An effect size of .6 would mean that average student in 2014 would be performing at a level

equivalent to roughly the 73rd percentile in the 2002 distribution. That would represent a larger increase in student achievement of the nation's students than has been achieved in any similar period of time during the last 50 years. Although effect size statistics are seemingly less transparent to the public than the percentage of students scoring above a cut score, effect size statistics are widely used and are certainly less complicated to explain than the confidence intervals now used by half the states, with another 10% of the states proposing to do so, to determine AYP.

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 degree to which the goal of improving the achievement of these underserved groups of students is being achieved. It also is necessary for determining whether the achievement gap is being closed.

It is highly desirable that disaggregated reporting of student achievement results be continued. There is a problem caused by this requirement that needs to be addressed, however. Since schools can fail to meet AYP in many different ways but can meet it in only one way, it turns out that schools with a sufficient number of students in each of several targeted groups to be reported are less likely to meet AYP targets than schools of the same size and similar performance but with a homogeneous student body (e.g., nearly all students who belong to one racial/ethnic group). A large part of this problem is caused by the year-to-year instability of results with barely enough students in particular subgroups for the results to be reported. The straight-forward solution to this problem is to allow schools to aggregate their subgroup results across two or three years – a solution that has already been allowed for the calculation of participation rates.

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 (1) the percentage of students who score below the proficient level is decreased by 10% from the year before, and (2) there is improvement for that subgroup on other indicators. In application, very few schools that would not otherwise make AYP do so because of the safe harbor provision. The very small percentage of schools that are saved by the safe harbor provision is due to the fact that the 10% decrease in students scoring below proficient sets a very high bar in comparison to what is achieved by even high performing schools. If a provision is desired to allow schools to meet AYP by showing decreases in the percentage of students scoring below the proficient level, then consideration should be given to alternative criteria such as an above average decrease in the percentage of students scoring below the proficient level from one year to the next. This would likely lead to a criterion closer to a 3 or 4% reduction in the below proficient category from one year to the next rather than the current 10% criterion.

Change as Well as Status

Although the NCLB accountability system might appear to focus on change, in many ways, it actually focuses on status. Schools where students are achieving at relatively high levels in the baseline year (2002), for example, might actually show a decline in achievement in 2003 and 2004 but still make AYP. Schools with very low achievement in the baseline year, on the other hand, will routinely fail to meet AYP even if they show rather sizeable year-to-year gains in student achievement. This is so because, with the exception of the safe harbor provision, NCLB is concerned with status in a given year rather than change from one year to the next. Consequently, schools that have a high achieving level to begin with have a relatively easy time meeting AYP without any gains in achievement, at least in the first few years. On the other hand, schools with initially low achieving students would have to have extraordinary improvement in achievement to meet AYP. Consequently, many schools that are actually

showing considerable progress, and deserve recognition for the gains they are making, fail to meet AYP because of their initial low performance.

A change in the NCLB accountability system that would allow schools to meet AYP either because their current achievement met a target or because the improvement in achievement met an improvement target seems desirable. This might be accomplished with a less stringent safe harbor criterion as was discussed above. Consistent with proposals above, both the current year achievement target and the improvement target should be set in light of what has been shown to be possible by high performing schools.

An alternative way of evaluating change in achievement that is attractive to several states is the use of longitudinal student records to track the growth in achievement for individual students. Analytical procedures, commonly referred to as value-added models, are used to estimate the school effects on student growth. Consideration should be given to the possibility of allowing states to use results of value-added analyses to provide evidence of improved achievement. The value-added results could be used, possibly in combination with status measures, to satisfy AYP requirements.

Conclusion

NCLB has the potential to make substantial contributions to the achievement of students who have lagged behind and too often 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. The most important modification is to set performance targets for judging adequate yearly progress that are more reasonable and for which there is a realistic hope that they might be achieved given sufficient effort. The need for more realistic goals applies to both the safe harbor provision of the law and to the annual performance targets. 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. Refinements are needed in the reporting of results for subgroups of students that would allow for year-to-year fluctuations in results.

Finally, changes to AYP requirements should be made that would allow schools to get credit for gains in achievement as well as absolute performance in a given year.

References

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.

No Child Left Behind Act of 2001m Pub. Law No. 107.110.

Popham, W. J. (2004). Shaping up the 'No Child' Act: Is edge-softening enough? *Education Week*, 23(38), May 26, p. 40.