

Toward a More Effective Definition of Adequate Yearly Progress

Robert L. Linn

National Center for Evaluation, Standards, and Student Testing

University of Colorado at Boulder

Toward a More Effective Definition of Adequate Yearly Progress

The Elementary and Secondary Education Act (ESEA) of 1965 was the main educational component of President Johnson's "Great Society" program. The central aim of ESEA is to provide aid to schools for the education of economically disadvantaged children. It is the principal federal law affecting elementary and secondary education throughout the country (Hess & Petrilli, 2006, p. 9). During the first three decades the focus of ESEA was on the distribution of funds and on educational inputs. That began to change with the 1994 reauthorization of ESEA by the *Improving America's Schools Act (IASA) of 1994* when a shift started to take place to give greater attention to student achievement outcomes. Under IASA, however, the shift in focus from inputs to outcomes did not have real teeth. IASA held schools accountable for student achievement, but the penalties for schools that fell short of expectations were less severe than they are under the most recent reauthorization of ESEA, the No Child Left Behind (NCLB) Act of 2001. NCLB not only strengthened the focus on student achievement, but added the teeth by strengthening the consequences for school failure and by holding schools and school districts accountable for the specific achievement of poor, minority, limited English proficient students, and students with disabilities.

Accountability is a central feature of NCLB. Adequate yearly progress (AYP) is the measure that is used to hold schools and school districts accountable under NCLB. "The law's fundamental dictate is that all schools and districts 'make AYP'" Hess & Petrilli, 2006, p.33). Schools that meet AYP requirements are assumed to be functioning well and enhancing student academic achievement. Schools that fail to make AYP are presumed to be

falling short of expectations. Schools that do not make AYP for 2 years in a row are identified as “needs improvement” and are subject to sanctions. The schools must develop a school improvement plan, offer supplemental services such as tutoring, and must allow students the option of transferring to another public school within the district that is not in the needs improvement category. The sanctions become increasingly severe for schools that continue to fall short of AYP targets for a 3rd, 4th, or 5th consecutive year.

To make AYP schools and districts must meet or exceed AYP targets that are set in terms of the percentage of students who are proficient or above in mathematics and reading or English language arts. Those targets are set each year at levels that increase over the years at rates that lead to the 100% proficiency goal by 2013-2014. States must set the proficient academic achievement level as well as at least two other levels (basic and advanced) on their reading or English language arts and mathematics assessments. NCLB provides a general description of proficient academic achievement and encourages the setting of the standard at a high level. In practice, however, states have set the proficient achievement bar at levels that vary widely in stringency (Linn, 2003b, in press). For example, in 2005 only 35% of the grade 4 students in Missouri were at the proficient level or above according to the Missouri definition of proficient achievement in reading whereas 89% of the grade 4 students in Mississippi were found to be proficient or above according to the Mississippi definition of proficient achievement in reading (Olson, 2005). Such a huge discrepancy can only be explained by the difference in the stringency of the definitions of the proficient achievement level in Missouri and Mississippi.

Limitations of AYP

The definition of AYP has several serious limitations. First, the expectation of universal proficiency by 2013-2014 is not a goal that can be achieved unless proficiency is watered down to correspond to a low level of achievement (Linn, 2003a). Second, as was briefly mentioned above, the definition of proficient achievement varies wildly from state to state. Indeed, the variation is so large that “proficient” achievement lacks any semblance of a common meaning across states. Consequently it is not meaningful to talk about universal proficiency as though it implied a common high level of student achievement. Third, AYP is limited by the almost exclusive focus on current achievement in a given year in comparison to a fixed target rather than attending to gains in achievement. Fourth, because there are many hurdles to clear to make AYP states have introduced a number of different conditions for subgroup reporting that make it easier for schools to make AYP and those conditions have undermined the fundamental concept. Finally, the narrow focus on state assessments of achievement in mathematics and reading or English language arts has potentially negative consequences. Each of these five limitations will be considered in greater detail and suggestions will be offered for changing the determination of AYP in ways that will address each of the limitations.

The Unrealistic Expectation of Universal Proficiency

Twenty-nine percent of the nation’s 8th grade students performed at the proficient level or above in mathematics on the National Assessment of

Educational Progress (NAEP) in 2003 (Braswell, Dion, Daane, & Jin, 2005).¹ That was up from 26% in 2000 and from 23% in 1996 (Braswell, et al., 2005). The lack of progress in the percentage of 8th grade students who were proficient in mathematics between 2003 and 2005 does not bode well for reaching 100% proficiency by 2013-2014. Nor is the 6 percentage point increase between 1996 and 2003 encouraging. Even if the rate of increase were to double, from an average the just under 1% per year between 1996 and 2003 to an average of 2% per year, less than half (47%) of the 8th grade students would be proficient in mathematics in 2013-2014 according to NAEP.

The picture is somewhat brighter for 4th grade mathematics because 4th grade students have made greater gains in percent proficient or above according to the NAEP definition of proficient achievement than 8th grade students. The percentage of 4th graders who were proficient or above was 21% in 1996, 24% in 2000, 32% in 2003 (Braswell, et al., 2005) and 35% of 4th graders in public schools were proficient or above in 2005 (Pirie, Grigg, & Dion, 2005). Although the improvement in the mathematics achievement of 4th grade students has been substantial and fairly steady with increases in the percentage of students who were proficient or above averaging approximately 1.6% per year, a continuation at that rate would still leave half (51%) the 4th graders performing below the proficient level in mathematics in 2014.

The NAEP mathematics assessment results make the 100% proficiency goal appear to be out of reach despite the fact that the gains in mathematics achievement have been substantially greater than the gains in reading. The prospects for reading are even more discouraging than they are for

¹ In other words, the grade 8 NAEP mathematics proficient cutscore was at the 71st percentile of US 8th graders in 2003. In 2005 proficient achievement on the grade 8 NAEP mathematics assessment was also at the 71st percentile (29% proficient or above) (Pirie, Grigg, & Dion, 2005)

mathematics. The trend lines for NAEP reading assessments are best described as essentially flat. Thirty percent of the 4th grade public school students and 29% of the 8th grade public school students were proficient or above on the 2005 NAEP reading assessments (Perie, Grigg, & Donahue, 2005). The corresponding 1998 results for all students were 29% at grade 4 and 32% at grade 8 (Donahue, Daane, & Jin, 2005). That is, there was no increase in the percentage of students who were proficient or above in reading at grade 4 from 1998 to 2005 and the corresponding percentage for 8th grade students was slightly lower in 2005 than it was in 1998.

Trends on NAEP over the past several years provide ample reasons to doubt that the 100% proficiency goal is obtainable even with the best of efforts or the belief that the rate of improvement would be twice as great in the future as it has been in recent years (see also, Lee, 2006;Linn, 2003a). NAEP's definition of proficient achievement is admittedly ambitious, but ambitious academic achievement standards are exactly what are called for by NCLB. Even if the NAEP basic achievement were used, the goal of 100% by 2014 does not seem realistic. The percentage of public school students who performed at the basic level or above in reading increased by only 4 points (from 58% to 62%) from 1998 to 2005 at grade 4 and was unchanged at grade 8 (Pirie, Grigg, & Donahue, 2005). Gains in the percentage of students performing at the basic level or above were greater in mathematics, but not large enough to make 100% a reasonable goal by 2014. At grade 8 the percentage of public school students at the basic level or above went from 62% in 2000 to 67% in 2003 and to 68% in 2005. The corresponding percentages at

grade 4 were 64% in 2000, 76% in 2003, and 79% in 2005 (Pirie, Grigg, & Dion, 2005).

For another perspective on the idea of universal proficiency, it is useful to consider international assessments such as the international assessment of mathematics achievement conducted by the International Association for evaluation of Educational Achievement (IEA). IEA's 2003 Third International Mathematics and Science Study (TMSS) assessment of mathematics included 46 countries at grade 8 and 25 countries at grade 4 (Mullis, Martin, & Foy, 2005). Although U.S. students performed above the international average at both grade levels, they did not perform nearly as well as students from some of the other participating countries. At grade 8, students from the Republic of Korea had an average level of mathematics achievement in the "knowing cognitive domain" that was higher than any other country. Singapore had the highest average at grade 4 (Korea did not participate at that grade level). In every country, however, there was considerable variability in student achievement.

Percentiles of achievement in mathematics at grade 8 are shown in Table 1 for 4 selected countries, Korea, Japan, the Netherlands, and the United States which had average achievement scores that ranked 1st, 5th, 9th, and 14th, respectively among the 46 participating countries. The mathematics achievement of students from Korea, Japan, and the Netherlands is clearly better than that of U.S. students. A score of 557 is at the 75th percentile in the U.S. and might be used as a rough proxy to define proficient achievement. (This means that 75% of American students scored below proficient). Note that a score of 557 is above the 50th percentile in the Netherlands and substantially

above the 25th percentile for students in Japan and Korea. In other words, no country had even three-fourths, much less all, of their students at the proficient level or above when that level is defined by the 75th percentile in the U.S.

< Insert Table 1 about here >

Comparative international results for grade 4 are shown in Table 2 in a manner that parallels the grade 8 results in Table 1. The medians for the Netherlands and the United States are quite similar as would be expected given the proximity of their international ranks. The distribution of performance is more spread out in the United States than it is in the Netherlands. Thus the 95th percentile is lower and the 5th percentile is higher in the Netherlands than in the U.S. The averages for Japan and Singapore are considerably higher than the U.S. average. Nonetheless, as can be seen, more than a quarter of the students in Singapore score below U.S. 75th percentile and more half the students in Japan score below that level. As was true at the 8th grade, no country is even close to having all of its students at the proficient level or above when proficient performance is defined to be equal to the 75th percentile in the U.S. If a small country such as Singapore with a relatively homogeneous population and a strong emphasis on education still has more than a quarter of its students below a reasonable proxy for proficient performance corresponding to the U.S. 75th percentile, then it is hard to imagine that the goal of 100% proficiency is at all realistic. Tying AYP targets to this unrealistic goal for 2013-2014 will make the annual targets less and less obtainable as we approach that date.

< Insert Table 2 about here >

Definitions of Proficient Achievement

As has already been noted, NCLB encourages states to set the proficient academic achievement standard at a high level. This is consistent with the standards-based reforms of the last decade that have consistently encouraged standards to be set at ambitious levels. That is the context in which the basic, proficient, and advanced academic achievement standards (called achievement levels) were set on NAEP in the early 1990s and for many states that set their achievement standards before the enactment of NCLB. Given the context in which academic achievement standards were being set prior to NCLB it is not surprising that, as was done for NAEP, a number of states also set their standards at ambitious levels. Of course, there are no consequences for students or schools for performance that is below the proficient level on NAEP, and prior to NCLB there were few, if any, consequences for students or schools that performed below the proficient level on most state assessments. The context suddenly changed when NCLB was signed into law by President Bush in January 2002.

As has already been discussed, NCLB introduced clear consequences for schools where the percentage of students who scored at or above the proficient level on a state assessment was less than the target percentage required to make AYP. The consequences for failing to make AYP, together with the knowledge that the annual targets for the percentage of students who are proficient or above has to increase from year to year on a trajectory that would reach 100% by 2-13-2014, led a few states to reconsider their academic achievement standards. In addition, states that had to introduce new assessments and set new academic achievement standards post NCLB were

operating in a radically different context than existed before law was enacted. Not surprisingly, states that set academic achievement standards since 2002 have generally set them at more lenient levels than states that set their standards before 2002.

The effect of the change in context is clearly illustrated by the actions in Colorado following the enactment of NCLB. Colorado had established academic achievement standards before NCLB became law. Four levels of achievement called advanced, proficient, partially proficient, and unsatisfactory were set on the Colorado Student Assessment Program (CSAP). The four performance levels are still used to report on student achievement to schools, districts, and parents. For purposes of NCLB, however, Colorado uses only three levels of achievement. The four CSAP levels used for state purposes are collapsed into three levels for determining AYP. The state's unsatisfactory level is relabeled basic, the partially proficient and proficient levels are collapsed into a single level called proficient while the highest level of achievement retains its label of advanced.

The lower level of achievement required for a student to be called proficient for purposes of NCLB than for state purposes makes a substantial difference in the percentage of students who are identified as proficient in Colorado. In reading in 2006, for example, 90% of 4th grade students reached the proficient level or above for purposes of NCLB, but 22% had CSAP scores in the partially proficient category. Thus, 68% rather than 90% of the grade 4 students were reported to be proficient or above in reading on state reports to schools, districts, and parents. Similar differences were obtained for other grade levels and for mathematics. For example, 75% of the 8th grade students

were reported to be proficient or above in mathematics in 2006 using the NCLB performance levels but only 45% reached the proficient level or above when the partially proficient level was reported separately to schools. Districts, and parents.

States that introduced new assessments after 2002 on which academic performance standards had to be set did not have to collapse levels of achievement to have standards that were more lenient post NCLB than they were before that time. Recognizing that the definition of proficient achievement has real consequences for schools and is not merely an aspiration, states established academic achievement standards that were less stringent than the ones that states had established before NCLB was enacted.

The stringency of academic achievement standards depends on a number of factors, including the context in which the standards are set, the uses that are to be made of the standards, the method that is used to set standards, and the judges that participate in the standard setting process (Glass, 1978; Jaeger, 1989; Linn, 2003b; in press). Consistent with the uncertainties surrounding standard setting, there is a broad professional consensus that "...there is NO true standard that the application of the right method, in the right way, with enough people will find" (Zieky, 1995, p. 29).

Because standard setting is subject to the many sources of variability, such as the influence of context and the fact that states set their standards at different times, there is tremendous variability in the stringency of the standards from state to state. Olson (2005) reported the percentage of students who scored at the proficient level on the individual state assessments and on NAEP

at grades 4 and 8 in 2005 for 47 states.² The average percent proficient or above on the state assessments in reading and mathematics at grades 4 and 8 were more than twice as large as the average percent proficient or above on NAEP (Linn, in press).

Not only were the state proficient standards less stringent on average than the NAEP standards, but the percentages of students reported to be proficient or above on the state assessments also were considerably more variable from state to state than the corresponding percentages on state NAEP. The ratios of the variances for the percentage of students who were proficient or above according to the 47 state assessments to the variances of the corresponding percentages for state-by-state NAEP results in 2005 were 6.61, 6.67, 5.36, and 6.02 for grade 4 reading, grade 8 reading, grade 4 mathematics, and grade 8 mathematics, respectively. The differences in variances between NAEP and state assessments are not due to actual differences in achievement between states, but are due instead to differences in the stringency of proficiency standards between states. Furthermore, there is little relationship between the percentages proficient or above on the state assessments and the corresponding percentages proficient or above on NAEP (Linn, in press).

The percentages proficient or above for a number of individual states on their own assessments make no sense when compared to other things that are known about education and student achievement in those states. On the grade 8 NAEP 2005 mathematics assessment, for example, the percentage of public school students performing at the proficient level or above was somewhat higher in Missouri (26%) than in Tennessee (21%) (Perie, Grigg, & Dion, 2005,

² The closest grade was used for the few states that did not have assessments in place at either grade 4 or 8 in 2005.

p. 16). On their own grade 8 state mathematics assessments, however, a whopping 87% of Tennessee students were reported to be proficient or above whereas only 16% of Missouri's students were reported to have performed at the proficient level or above (Olson, 2005, p. S2). The large discrepancy in percentages on the state assessments for Missouri and Tennessee cannot reasonably be explained by differences in student achievement in mathematics at grade 8. The most obvious explanation of the discrepancy is that the proficient academic achievement standard is much more stringent for the Missouri grade 8 mathematics assessment than it is in Tennessee. Other examples could be presented to reinforce the conclusion that the stringency of state academic achievement standards is varies greatly from state to state, A consequence of the variability in stringency is that "proficient" achievement has no common meaning across states. Thus, even if the unrealistic goal of 100% proficient or above could be achieved, it would mean radically different things in different states.

Current Status vs. Improvement

The "adequate yearly progress" label seems to imply that student achievement has improved during the course of a year. However, with the exception of the safe harbor provision,³ that allows schools that would not have done so otherwise to make AYP if the percentage of students scoring below the proficient level is reduced by 10% or more from the previous year, the determination of AYP does not depend on improvement from one year to the

³ In practice, the safe harbor provision saves few schools because a reduction of the percentage of students who perform below the proficient level by 10% is much larger than gains that are normally realized from one year to the next. In the last couple of years, however, states have begun using confidence intervals around the reductions in the percentage of students below the proficient level, thereby effectively lowering the 10% figure and increasing the number of schools that are saved by the safe harbor provision.

next. Rather, it depends on a comparison of the percentage of students who are at the proficient level or above in a given year to a target percentage known as the annual measurable objective (AMO) for that year.

The comparison of current student performance to the AMO makes it relatively easy for schools where students who are already achieving at high levels to exceed the AMO and make AYP. Indeed, a school where students have been achieving at high levels can have a decline in achievement from one year to the next, and still make AYP. Schools serving students who start the school year with achievement that is far below the AMO for a given year, on the other hand, must have dramatic gains in achievement in a given year to make AYP. Most of the latter schools will fail to meet AYP even if they show rather sizeable year-to-year gains in student achievement because they start the year so far below the AMO. Thus, the current AYP system provides an advantage to schools serving students who are already achieving at high levels and puts schools serving initially low achieving students at a substantial disadvantage.

In contrast to the NCLB accountability system that, with rare exceptions, relies on a comparison of current status to a fixed annual target, most state accountability systems give substantially more weight to year-to-year improvement in student achievement. The measurement of improvement is done in two ways. Some states (e.g., North Carolina and Tennessee) track individual student achievement longitudinally and use gains in achievement to hold schools accountable. Other states (e.g., California and Kentucky) compare the performance of successive cohorts of students (e.g., 4th graders in 2005 and 4th graders in 2006) to measure improvement and some states (e.g., Colorado and Florida) use a combination of current performance and

improvement in performance to hold schools accountable. Giving schools credit for year-to-year improvement in student achievement puts schools that start the year with quite different levels of student achievement on a more equal footing than a system that relies almost exclusively on current achievement.

In response to concerns about the current approach to AYP, Secretary of Education Spellings (2005) announced a pilot program that let states propose ways of using a growth model to make AYP determinations. Several “core principles” that must be met for a proposal to be approved were specified by Secretary Spellings in a letter to the Chief State School Officers. The first principle stated that the growth model “must ensure that all students are proficient by 2013-14 and set annual goals to ensure that the achievement gap is closing for all groups” (Spellings, 2005). Because, as was discussed above, the 100% proficient goal is unrealistic, this principle severely limits the utility of growth models for determining AYP.

The pilot program proposals submitted by North Carolina and Tennessee were approved for implementation of growth model pilots in 2005-2006 (Spellings, 2006). Six other states (Alaska, Arkansas, Arizona, Delaware, Florida, and Oregon) that submitted proposals were told that they would get early consideration for possible implementation in 2006-2007 if they submitted revised proposals.

Although the pilot program is a step toward the inclusion of improvement in student achievement in the determination of AYP, it is currently limited to only a few states. It is also severely limited by the maintenance of 100% proficient goal in 2013-2014 that is both unrealistic and unequal from state to state due to the lack of a common definition proficient academic achievement. The

constraints imposed by the 100% proficiency requirement may be the reason that the implementation of the growth models to determine AYP in North Carolina and Tennessee did not result in a major changes in the AYP status of schools in 2006 (Olson, 2006).

Multiple Hurdles

Unlike accountability systems in most states that use a compensatory approach that allows superior achievement in one content area to make up for sub-par performance in another content area, NCLB uses a conjunctive approach whereby schools must have achievement that meets or exceeds percentage proficient or above targets in both reading or English language arts and in mathematics. Actually there are more than just the two achievement hurdles that must be cleared in order to make AYP. At a minimum, a school must clear 5 hurdles. In addition to the two percentage proficient or above targets, a school must assess at least 95% of their eligible students in each subject area and exceed the performance target for the other academic indicator selected by the state (typically attendance for elementary and middle schools and graduation rate for high schools).

For large schools with diverse student bodies the number of hurdles that must be cleared to make AYP can be substantially greater than the minimum of 5. Marion, White, Carlson, Erpenbach, Rabinowitz, and Sheinker (2002) have shown that the number of hurdles to be cleared could be as large as 37. The larger number of hurdles is due to the requirements for disaggregated reporting of subgroup performance. Four hurdles (2 for subgroup participation rates and 2 for subgroup achievement in reading/English language arts and mathematics) are added to the 5 for the school as a whole for each subgroup that is large

enough to require disaggregated reporting. Thus, if there are 8 subgroups of sufficient size, the school would have to clear a total of 37 hurdles. Although few schools are large and diverse enough to reach the maximum, many large schools have to clear 21 or 25 hurdles because they have 4 or 5 subgroups that are large enough to require disaggregated reporting.

Making schools accountable for the achievement of subgroups identified by NCLB is clearly consistent with the NCLB goal of closing gaps in achievement for the identified subgroups. It is also clear, however, that NCLB's multiple hurdle approach makes it considerably more difficult for large schools with diverse student bodies to meet AYP requirements than it is for small schools or schools with homogenous student bodies (Kim & Sunderman, 2005; Linn, 2005; Novak & Fuller, 2003).

States have responded to the challenges schools face in making AYP in a number of ways. They have increased the minimum number of students in a subgroup that is required for disaggregated reporting. They have introduced the use of confidence intervals for the percentage of students who are proficient or above and for determining the year-to-year change in the percentage of students who score below the proficient level for purposes of safe harbor calculations (Center on Education Policy (CEP), 2005; Sunderman, 2006). These changes make it easier for a school to make AYP, but they also make the definition of AYP more complicated and less transparent.

Narrow Focus on State Reading and Mathematics Assessments

There is no question that reading and mathematics are critically important. The narrow focus on these two subjects as measured by state assessments, however, can lead to distortions in the curriculum and instruction

that students receive. As was evident from responses at recent public hearings on NCLB (Public Education Network, 2006) there is widespread concern that the focus of the NCLB accountability system is too narrow. A substantial proportion of the public believes that there is too much emphasis on a single assessment as the determining factor for AYP. Opinions expressed at the hearing favored a reduction in the emphasis on state reading and mathematics assessments coupled with an increased reliance on information from formative assessments and evaluations.

The focus on reading and mathematics together with the high stakes attached to making AYP has led schools to increase the time spent on these subjects at the expense of other subjects such as science and social studies that are also important parts of education. Although states will be required to have science assessments in place starting in 2007-08, no clear use of the science assessment has been specified. Moreover, they are only required at one grade level in each grade level span (elementary, middle, and high school). Thus, it is not clear that the addition of science assessments will lead to any real changes in the NCLB accountability system.

Most of the districts (71%) that participated in a survey conducted by the Center on Education Policy (CEP) (2006) reported a reduction in the time devoted to at least one other subject to allow more time to be devoted to reading and mathematics (p.89). Although the additional time spent on reading and mathematics instruction may enhance achievement in those subjects, it comes at the expense of other important subjects. The report of reduced time spent on non-tested subjects is consistent with results reported in other studies. Sunderman, Tracey, Kim, & Orfield (2004), for example, found that a substantial

majority of teachers in the two districts that they surveyed reported that AYP requirements caused some teachers to de-emphasize or neglect content in untested topics and to increase the amount of time spent on classroom activities specifically designed to prepare student for state mandated assessments.

The great emphasis on performance on just two assessments, one in reading or English language arts and one in mathematics, coupled with the sanctions for schools that fail to make AYP, can not only narrow instruction to those subjects but distort the teaching of reading and mathematics. Drill and practice on topics covered on the tests in a predictable manner and frequent practice on benchmark tests consisting of items with the formats that mirror the items on the state assessment can lead to score inflation, i.e., “a gain in scores that substantially overstates the improvement in learning it implies” (Koretz, 2005, p. 99).

Suggestions for Improving the Determination of AYP

The determination of AYP could be improved by addressing the five limitations of the current approach that were discussed above. First, the unrealistic expectation of 100% proficiency should be replaced by a goal that is still ambitious, but realistically obtainable with sufficient effort on the part of educators and students (Linn, 2003a). One way to select a goal that is both ambitious and realistically obtainable is to look at accomplishments of schools that have shown substantial gains in student achievement in the past. For example, schools that rank among the top, say 20%, of all schools in terms of the gains their students have made over a period of 4 or 5 years could establish the goal for all schools. The goal would be more realistic than the 100%

proficiency goal since 20% of the schools have managed to make those gains already. The goal also would be ambitious for the majority of schools that had not realized such large gains in student achievement in the past.

Second, the notion of proficient academic achievement should either be modified so that it is defined to have a common meaning from state to state, or it should be replaced by another marker of achievement. A uniform definition a target achievement level could be realized by defining a cutscore on each state assessment that was equal to the median achievement of students in a base year (e.g., 2002). Although there would be a small variation in the stringency of the median due to differences in the achievement of students in different states and due to the differences in the state assessments, the variation would be tiny in comparison to the state-to-state variation in the definitions of proficient achievement. Using the average annual gains made by the top 20% of the schools, the annual target could then be established for the percentage of students scoring above the median performance in the 2002 base year. This might lead to an annual target increase of, say, 3%. With a 3% gain per year, the proportion of students scoring above the 2002 base year median would need to increase from 50% in 2002 to 86% in 2014. Such a goal is clearly ambitious but it is also much more realistic than the current 100% proficient goal. Moreover, it would provide a reasonably uniform definition of target achievement across states.

Of course, not all schools would start with half their students scoring above the state median. Schools where only a quarter or less of their students score above the state median would have to have extraordinary improvement to meet the targets set for all schools. Hence, there would also need to be a safe

harbor type of provision. Instead of meeting the absolute target of the percentage of students above the base year median, schools could qualify as making AYP if they showed substantial improvement each year (e.g., an increase of 4 or 5% of their students scoring above the base year median for the state).

Third, the way in which AYP is determined should be expanded to allow schools that show substantial increases in student achievement to meet the requirements rather than relying almost entirely on a comparison of current achievement to an annual target. Improvement could be evaluated either by computing growth for individual students with a longitudinal data system such as allowed for states approved for the NCLB pilot program or by comparing achievement of successive cohorts of students.

Allowing schools to make AYP either by meeting an absolute target or by making substantial gains in student achievement from the previous year does not give up on the same goals for students. Such a system would not dilute the goals for some groups of students, rather, by allowing improvement as well as current status as ways of making AYP, it would make it more realistic for schools with a large percentage of students with low achievement initially to get credit for demonstrating substantial gains in student achievement that would eventually lead to high levels of performance.

Fourth, the conjunctive, multiple hurdle approach to determining AYP should be replaced by a compensatory system that would allow superior achievement in one subject to make up for achievement that is somewhat below a target level in another subject. States generally use some form of a compensatory system in their own accountability systems and a move in that

direction by NCLB would make the federal and state accountability systems more compatible. Monitoring the achievement of subgroups should continue and superior achievement for one subgroup should not make up for below par achievement for another subgroup. However, the volatility due to small numbers of students in particular subgroups should be addressed by allowing schools to aggregate subgroup results over two or three years rather than requiring results to meet targets every year.

Finally, the measures used for accountability should be broadened to include more subjects and assessment information obtained from sources other than state assessments. Formative assessments and professional judgments of student achievement by educators could be used to supplement the information that is provided by state assessments. The additional measures could be easily accommodated in a compensatory system. The additional measures in a composite index would be likely to reduce the use of practices that result in inflated test scores such as narrow teaching to the specific content and formats used on state assessments.

Combining teacher produced ratings of student achievement with state assessments and other assessment results would require that the teacher scores be reported in a common metric such as a 1 to 5 scale. Concerns that teachers might report inflated ratings would need to be addressed, but the potential gain in information would be worth the added effort needed to obtain and use teacher produced scores. A common set of district-selected benchmark assessments, formative classroom assessments selected by teachers, and systematic teacher ratings of student accomplishments would supplement the information about student achievement that is provided by state

assessments. Students that may not do well on a standardized state assessment would have other ways of demonstrating what they know and are able to do. With a broader array of measures teachers would not have the same pressure to devote so much time to narrow test preparation for poor and minority students, but could instead spend the time on broader instructional goals.

Table 1				
Percentiles of Achievement in Knowing Cognitive Domain For the 2003 TIMSS Mathematics Assessment at Grade 8 (Mullis, Martin, & Foy, 2005)				
Country				
Percentile	Korea Rep of	Japan	Netherlands	United States
95	717	677	611	623
75	650	611	563	557
50	599	567	522	509
25	539	519	480	462
5	444	443	422	397
International Rank Among 46 Countries				
	1	5	9	14

Table 2

Percentiles of Achievement in Knowing Cognitive Domain
 For the 2003 TIMSS Mathematics Assessment at Grade 4
 (Mullis, Martin, & Foy, 2005)

Country				
Percentile	Singapore	Japan	Netherlands	United States
95	784	696	617	657
75	697	620	566	584
50	633	566	531	529
25	563	512	494	474
5	442	424	440	396
International Rank Among 25 Countries				
	1	4	7	8

References

- Braswell, J. S., Dion, G. S., Daane, M. C. and Jin, Y. (2005). *The nation's report card: Mathematics 2003*. (NCES 2005-451). U.S. Department of Education, Institute for Education Sciences, National Center for Statistics. Washington, DC: U.S. Government Printing Office.
- Center on Education Policy. (2006). *From the Capitol to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author, March. Available at: <http://www.ctredpol.org/>.
- Donahue, P. L., Daane, M. C., and Jin, Y. (2005). *The nation's report card: Reading 2003* (NCES 2005-453). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Elementary and Secondary Act of 1965, Public Law No. 89-10.
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, 15, 237-261.
- Hess, F. M. & Petrilli, M. J. (2006). *No Child Left Behind Primer*. New York: Peter Lang Publishing, Inc.
- Improving America's Schools Act of 1994, Public Law No. 103-382.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 485-514). New York: Macmillan.
- 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.
- Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in*

- accountability testing. Yearbook of the National Society for the Study of Education* (pp. 99-118), Vol. 104, Part I.
- Lee, J. (2006). *Tracking achievement gaps and assessing 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.
- Linn, R. L. (2003a). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32, No. 7, 3-13.
- Linn, R. L. (2003b, September 1). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives*, 11 (31).
- Linn, R. L. (2005, June 28)). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives*, 13(33).
- Linn, R. L. (in press). Performance standards: What is proficient performance? In C. E. Sleeter (Ed.). *Educating for democracy and equity in an era of accountability*. New York: Teachers College Press.
- Marion, S. T., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, A. & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. A paper series: Implementing the state accountability requirements under the No Child Left Behind Act of 2001. Washington, DC: Council of Chief State School Officers.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2005). *IEA's TIMSS 2003 international report on achievement in the mathematics cognitive domains. Findings from a developmental project*. International Association for evaluation of

Educational Achievement (IEA). Newton, MA: TIMSS & PIRLS

International Study Center, Lynch School of Education, Boston College.

No Child Left Behind Act of 2001, Public Law No. 107.110.

Novak, J. R. & Fuller, B. (2003). Penalizing diverse schools? Similar test scores, but different students, bring federal sanctions. *Policy Brief, December(3-4)*, Berkeley, CA, University of California Berkeley: Policy Analysis for California Education, available online at <http://pace.berkeley.edu>.

Olson, L. (2005). Room to Maneuver. *Education Week*, Special pull out section: A progress report on the No Child Left Behind Act, pp. S1-S6, December 14.

Olson, L. (2006). As AYP bar rises, more schools fail: percent missing NCLB climbs amid greater testing. *Education Week*, 2(4), 1, 20, September 20.

Perie, M., Grigg, W., & Dion, G. (2005). *The nation's report card: Mathematics 2005* (NCES 2006-453). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Perie, M., Grigg, W., & Donahue, P. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Spellings, M. (November 21, 2005). Letter to Chief State School Officers, announcing growth model pilot program, with enclosures. Available at: <http://www.ed.gov/nclb/landing.jhtml>.

Spellings, M. (May 17, 2006). Press release. Secretary Spellings approves Tennessee and North Carolina growth model pilots for 2005-2006.

Available at

<http://www.ed.gov/news/pressreleases/2006/05/05172006a.html>.

Sunderman, G. L. (2006). *The unraveling of No Child Left Behind: How negotiated changes transform the law*. Cambridge, MA: The Civil Rights Project at Harvard University.

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.

Zieky, M. J. (1995). A historical perspective on setting standards. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments n Statistics*. (pp. 1-38). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.