

6-28-2005

Education Policy Analysis Archives 13/33

Arizona State University

University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/coedu_pub

 Part of the [Education Commons](#)

Scholar Commons Citation

Arizona State University and University of South Florida, "Education Policy Analysis Archives 13/33 " (2005). *College of Education Publications*. Paper 568.

http://scholarcommons.usf.edu/coedu_pub/568

This Article is brought to you for free and open access by the College of Education at Scholar Commons. It has been accepted for inclusion in College of Education Publications by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Copyright is retained by the first or sole author, who grants right of first publication to the **Education Policy Analysis Archives**. **EPAA** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by H.W. Wilson & Co.

Volume 13 Number 33

June 28, 2005

ISSN 1068-2341

Conflicting Demands of No Child Left Behind and State Systems: Mixed Messages about School Performance¹

Robert L. Linn

University of Colorado at Boulder

Citation: Linn, R. L. (2005, June 28). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives*, 13(33). Retrieved [date] from <http://epaa.asu.edu/epaa/v13n33/>.

Abstract

An ever-increasing reliance on student performance on tests holds schools and educators accountable both to state accountability systems and also to the accountability requirements of the No Child Left Behind (NCLB) Act of 2001. While each state has constructed its own definition of Adequate Yearly Progress (AYP) requirements within the confines of NCLB, substantial differences between the accountability requirements of many state systems and NCLB still have resulted in mixed messages regarding the performance of schools. Several features of NCLB accountability and state accountability systems contribute to the identification of a school as meeting goals according to NCLB but failing to do so according to the state accountability system, or vice versa. These include the multiple hurdles of NCLB, the comparison of performance against a fixed target rather than changes in

¹ I thank the editor, Sherman Dorn and three unknown reviewers for helpful comments on an earlier draft of this paper. The work reported herein 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. Accepted under the editorship of Sherman Dorn. Send commentary for *EPAA* about this article to Casey Cobb (casey.cobb@uconn.edu).

achievement, and the definition of performance goals. The result of these features is a set of AYP measures that is inconsistent both with existing state accountability systems and also with state NAEP performance. Using existing achievement to set the cut-score measured by AYP and using the highest-performing schools to set the year-to-year improvement standards would improve the NCLB accountability system.

Keywords: accountability; Colorado; Florida; Kentucky; No Child Left Behind.

Test-based educational accountability has expanded greatly during the past decade. Most states were already using tests results to hold schools accountable prior to the time that President Bush signed the No Child Left Behind (NCLB) Act into law in January 2002. NCLB not only further increased the already relatively strong emphasis in a number of states on the use of student test results as a means of holding schools accountable, but it also superimposed a new set of accountability rules that often give signals that conflict with those provided by the state accountability systems.

Mixed Messages

NCLB requires states to test students in grades 3 through 8 in mathematics and English/language arts starting no later than the 2005–2006 school year. NCLB requires each state to have adopted “challenging academic content standards and challenging student academic achievement standards” (P. L. 107–110, Section 1111(b)(1)(A)). States must also establish adequate yearly progress (AYP) goals for each year from 2002 to 2014 that culminate in the 2014 goal where all students are at or above the proficient student academic achievement standard. As discussed below, however, states still control many important system characteristics in complying with NCLB, such as the specification of content standards, the choice of assessments, and the setting of academic achievement standards.

For states with functioning assessment and accountability systems of their own, NCLB accountability has frequently been layered on as a separate system. Kentucky, for example, had a comprehensive school accountability system in place before the enactment of NCLB. The Kentucky accountability system uses tests in seven content areas (reading, writing, mathematics, science, social studies, arts and humanities, and practical living/vocational studies). The tests are administered at selected grades so that the overall testing burden is distributed across grade levels. Composite index scores are used for school accountability. The index scores are derived across content areas and include some non-test measures (e.g., attendance or graduation rate). Biennial accountability targets for the composite index scores are set for schools relative to the school’s starting position defined by the school’s accountability index score in the 1999–2000 biennium. Schools that started low have to gain more in their index score than schools that started out relatively high but all are supposed to reach an index value of 100 by the 2013–2014 biennium (Kentucky Department of Education, 2004). In the computation of the index value, students who score at the highest level (called distinguished) on a test contribute 140 points, students at the proficient level contribute 100 points, and students in various categories below proficient contribute an amount less than 100—how much less depends on how far below the proficient level the student’s score is.

NCLB imposes a quite different set of accountability requirements for Kentucky schools. Mathematics and reading must be reported separately and schools must make annual, rather than biennial, measurable objectives in each subject (not just on a composite score). No extra credit is

allowed for students scoring at the distinguished level. They are simply lumped with proficient students in the proficient or above category. School performance is compared to an absolute target, which is the same regardless of where the school started. Schools must meet AYP requirements in both reading and mathematics, not only for the student body as a whole, but for each of several subgroups (assuming there are enough students in a subgroup to be counted for NCLB accountability purposes): major racial and ethnic groups, English language proficiency status, student disability status, and economic status.

Given the differences between Kentucky’s own school accountability system and the NCLB system, it is hardly surprising that the two systems are giving mixed messages. Except for schools with special designations, seven hundred thirty of 986 (74.0%) of Kentucky schools made AYP in 2004 (Ford & Thacker, 2005). According to Kentucky’s own accountability, however, 943 of the 986 (95.6%) schools met their Kentucky biennium goals in 2003–2004. Thus, the best possible agreement between the two systems would be if all of the 730 schools that made AYP also met the Kentucky biennium goals and all of the 43 schools that fell short of the Kentucky biennium goals also failed to make AYP for a combined total agreement of the two systems of 78.3% (773 out of the 986 schools). Even in this best-case scenario, just over 20% of the schools would receive mixed messages by meeting the goal according to one accountability system, but failing to do so according to the other system.

Table 1
State and NCLB Classifications of Kentucky Schools, 2004

Met the State Goal	Met AYP Targets ^a		Total
	Yes	No	
Yes	713 (72.3)	230 (23.3)	943 (95.6)
No	17 (2.6)	26 (12.6)	43 (4.4)
Total	730 (74.0)	256 (26.0)	986 (100)

^a Cumulative percentages in parentheses.
Based on Ford and Thacker (2005).

Table 1 displays the cross-classification of the 986 schools in terms of meeting or not meeting AYP in 2004 and meeting or not meeting Kentucky’s accountability goals. As can be seen, not all the schools that made AYP also met the Kentucky goals. Hence, a quarter of the schools (247 of 986) received mixed messages that they met expectations according to one accountability system but failed to meet them according to the other system.

The mixed messages of the NCLB and the individual state’s own accountability system are not unique to Kentucky. Florida’s state accountability system has assigned letter grades of A, B, C, D, or F to schools based on the performance of students on their state assessment. The distributions of Florida school grades over the seven years between 1999 and 2005 are displayed in Figure 1. These distributions have painted quite a favorable picture of school performance. The percentage of schools receiving grades of A increased from 8.3% in 1999 to 47.6% in 2004. The percentage of schools receiving either an A or a B has also increased sharply (from 21.3% in 1999 to 68.0% in 2004), while the percentage of schools receiving D’s or F’s declined from over a quarter of the schools in 1999 (27.9%) to less than a tenth of the schools in 2004 (8.8%).

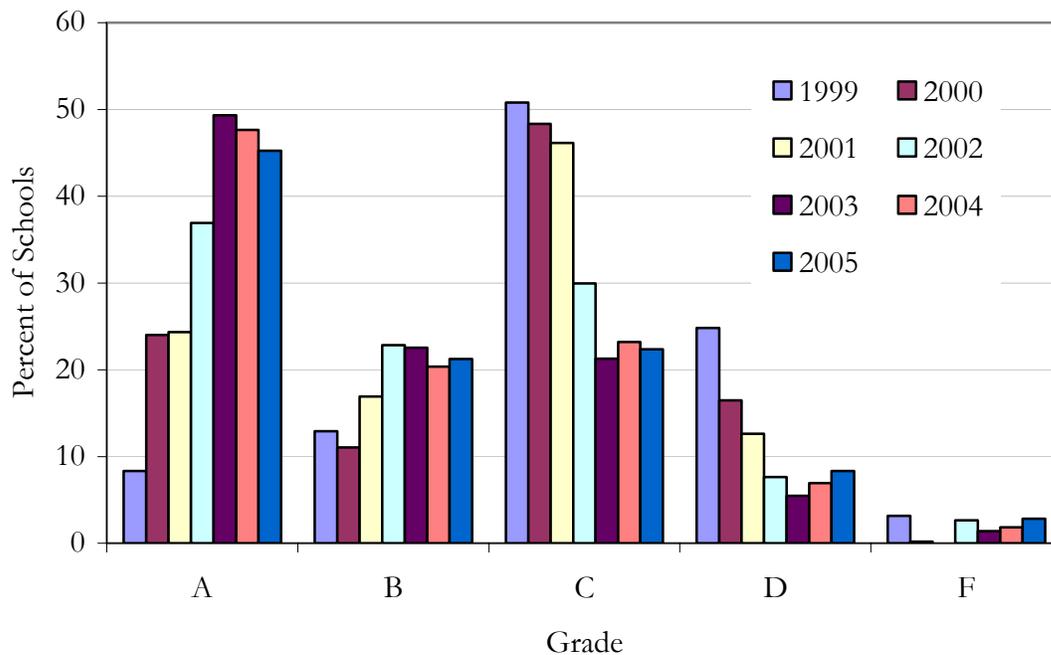


Figure 1 . Distribution of Florida School Grades by Year
 (Source: http://schoolgrades.fl DOE.org/0405/pdf/04_05page1_2.pdf)

The NCLB accountability results in Florida in two recent years have provided a sharp contrast to the positive results from Florida's own accountability system. Florida had the dubious distinction in 2003 of leading the nation as the state with the largest percentage of schools (82%) that failed to make AYP. Although there was some decline in the percentage of Florida schools that did not make AYP in 2004 (from 82 to 77%), only Alabama had an equally high percentage of schools failing to make the AYP target in 2004 (Olson, 2004, p. S6). Other southern states had more modest percentages of schools that failed to make AYP in 2004: Georgia, 20%; Louisiana, 8%; Mississippi, 24%; North Carolina, 29%; South Carolina, 44%, Tennessee, 14%; and Virginia, 25% (Olson, 2004, p. S6). As will be discussed in greater detail below, the variation by state makes little sense in comparison to other information about student performance by state such as that provided by the National Assessment of Educational Progress (NAEP), but it is nonetheless clear that the state's own accountability system and NCLB are giving quite a mixed picture in Florida. Fifty six percent of the 1262 schools in Florida that received an A in 2004 failed to make AYP.

The mixed messages provided in Kentucky and Florida are repeated in varying degrees in a number of other states. Colorado, for example, has an academic performance rating system that assigns schools to one of five graded performance categories called Unsatisfactory, Low, Average, High, and Excellent. Figure 2 displays the percentage of schools that made AYP in 2003 by school type and academic performance rating. As can be seen in Figure 2, there is a clear relationship between the Colorado academic performance rating of a school and the likelihood that the school will meet AYP. The relationship is far from perfect, however. Consequently, a substantial number of schools are receiving mixed messages. For example, 21.9% of the schools rated "Unsatisfactory" and 47.5% of the schools rated "Low" made AYP, while 13.7% of the schools rated "high" failed to make AYP.

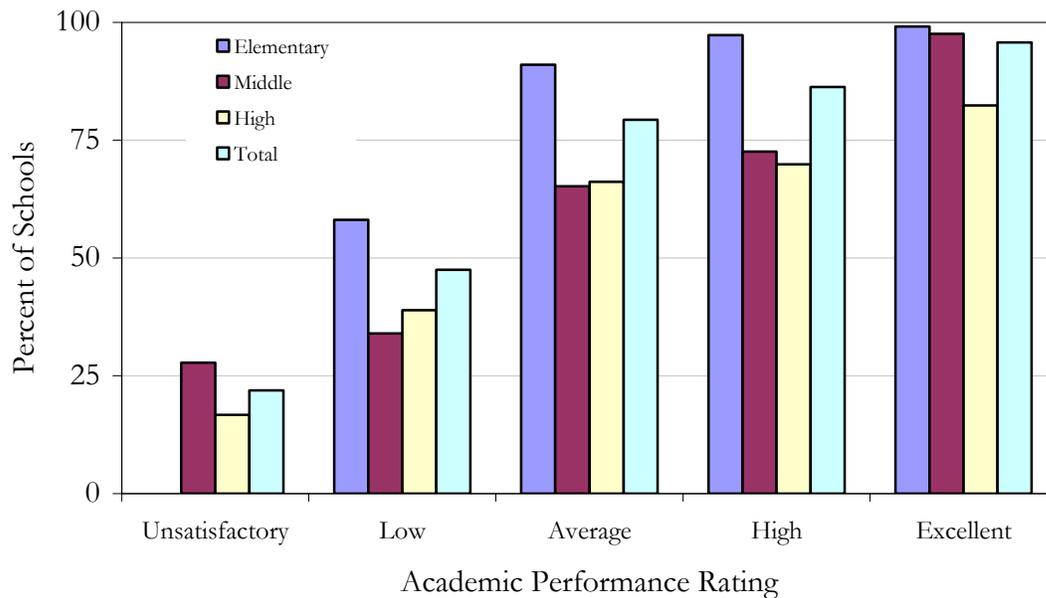


Figure 2
 Percentage of Colorado Schools Making AYP in 2003 by School Type and Academic Performance Rating (Source: <http://www.ced.state.co.us/>)

Although the summary statistics provide clear evidence that mixed messages are being given by NCLB and state accountability systems, it is, as was clearly illustrated by Dillon (2004), the failure of prestigious suburban high schools to meet AYP requirements that seems to have caused the most consternation. Dillon quotes people such as Representative Judy Baggert of Illinois, who “helped write the law” and former North Carolina Governor James Hunt, who has praised the law, among others who were dismayed when they learned that particular high schools that they knew to be excellent were identified for failing to make AYP. Although, as is explained below, there are a variety of aspects of the NCLB identification system that makes it likely that excellent schools will be found wanting by failing to make AYP it is nonetheless confusing to parents and the general public. The confusion is summed up well in the title of Dillon’s article “Good schools or bad? Conflicting ratings leave parents baffled” (2004).

Why Mixed Messages?

There are several features of the NCLB accountability requirements that make it likely that the results will conflict with the accountability of individual states. Some of these features also contribute to the wide state-to-state disparities in the proportion of schools that meet AYP. Three of these features, the use of absolute targets rather than improvement targets relative to a school’s starting level, the need to meet targets in both reading and mathematics rather than a composite, and the requirement of meeting targets for separate subgroups within a school, were mentioned in passing in the discussion of the Kentucky results. These and other features are elaborated in this section.

Multiple Hurdles

Unlike many state accountability systems, NCLB requires schools to clear several hurdles. The most obvious of these is that student achievement must exceed the annual measurable objective (AMO) in both mathematics and reading/English language arts. Performance on, say, reading achievement that far exceeds the AMO cannot compensate for mathematics achievement that just misses the AMO. In addition to meeting the separate AMOs for mathematics and reading/language arts, schools must have at least 95% of their eligible students participate in the assessments in each subject. The school must also meet the goal established for the “other academic indicator,” usually attendance rate for elementary and middle schools and graduation rate for high schools, required by NCLB. Thus, there are a minimum of 5 hurdles for a school with a homogeneous student body and insufficient numbers in any subgroup to be held accountable for disaggregated results.

The number of hurdles for meeting AYP expands rapidly for large schools with diverse student bodies due to the disaggregation requirements of NCLB. A school with more than the minimum number of students, designated by the state and approved by the U.S. Department of Education, for purposes of AYP in each of, say, 4 racial ethnic groups, students with limited English proficiency, economically disadvantaged students, and students with disabilities would have not 5, but 33, hurdles to clear (the 5 when all students in the school are considered as a whole, plus 16 for the 4 hurdles for each of 4 racial/ethnic groups, plus 4 for students with limited English proficiency, plus 4 for the economically disadvantaged students, plus 4 for the students with disabilities (see Table 2).

Thus, schools can meet AYP requirements in only one way, by clearing multiple hurdles, but can fall short in many different ways. Given the larger number of hurdles to be cleared by more diverse schools it is not surprising that Novak and Fuller (2003) found that schools serving more diverse student bodies were less likely to meet AYP requirements than schools serving less diverse student bodies. “[E]ven when students display almost identical average test scores schools with more subgroups are more likely to miss their growth targets under federal rules set by the No Child Left Behind Act” (Novak & Fuller, 2003, p. 1). These results are not unique to California. In Massachusetts, for example, about 48% of the schools failed to make AYP in 2002–2003, but only 9% of the 373 schools with only one subgroup of sufficient size to be used in determining AYP failed to meet the requirement, whereas 87% of the 106 schools with 6 or 7 subgroups included in the determination of AYP failed to meet the requirement (Nelson & Rosenberg, 2004a). These results are to be expected because even a school with high average achievement may be tripped up on one of the multiple hurdles, such as missing the participation rate for a particular subgroup or because students with disabilities perform below the proficient cutoff in either reading or mathematics. “In Westport, Conn., [for example] the Bedford Middle School, where test scores are often among Connecticut’s highest, was called low performing because the school failed to meet the 95 percent standard for testing for the disabled by one student” (Dillon, 2004).

Table 2
33 AYP Hurdles for a Large School with a Diverse Student Population

Demographic Group	Reading/English Language Arts		Mathematics		Other Academic Indicator
	Participation Rate	Percent Proficient or Above	Participation Rate	Percent Proficient or Above	
All Students	⊙	⊙	⊙	⊙	⊙
Racial/Ethnic Group 1	⊙	⊙	⊙	⊙	
Racial/Ethnic Group 2	⊙	⊙	⊙	⊙	
Racial/Ethnic Group 3	⊙	⊙	⊙	⊙	
Racial/Ethnic Group 4	⊙	⊙	⊙	⊙	
Economically Disadvantaged	⊙	⊙	⊙	⊙	
Students with Limited English Proficiency	⊙	⊙	⊙	⊙	
Students with Disabilities	⊙	⊙	⊙	⊙	

Note. Table modeled after Marion, White, Carlson, Erpenbach, Rabinowitz, & Sheinker (2002).

Status vs. Growth Targets

State accountability systems frequently establish performance targets based on growth, thereby taking into account previous performance as well as current status. NCLB requirements, on the other hand, with the exception of the safe harbor provision discussed below, focus only on current status in comparison to the performance target. California’s accountability system, like the Kentucky system described above, provides a good example of a carefully developed system that focuses on growth in achievement. Growth is calculated not by tracking individual students and computing indices based on longitudinal data, but by comparing successive cohorts of students (e.g., the achievement of fourth grade students in a school in 2004 compared to the achievement of fourth graders in that school in 2005). California’s system uses an academic performance index (API) that is a weighted combination of performance on tests of English language arts (including writing) and mathematics for grades two through eight. For grades 9 through 11, history-social science and science are included along with English language arts and mathematics for the weighted API. The API is scaled to have scores that range from 0 to 1000.

An API score of 800 has been selected by the State Board of Education “as the target toward which all schools should aspire” (California Department of Education, 2004, p. 29). Schools are not sanctioned for falling short of the absolute target of 800, however. Instead a school is held accountable for meeting their annual API growth target, which “is defined as five percent of the distance from the school’s API and the statewide performance target or a minimum of one point” (California Department of Education, 2004, p. 30). For example, a school with an API in 2003 of 700 would have an API growth target of 5 points (5% of 800–700) for 2004, while schools with APIs of 650 and 750 would have growth targets of 7.5 and 2.5, respectively. California’s focus on growth rather than status obviously stands in sharp contrast the federal AYP requirements.

In order for a school to meet all API growth requirements, students in the school who are members of a “numerically significant” subgroup defined by ethnicity or socioeconomic disadvantage “must achieve at least 80 percent of the school-wide annual growth target (California

Department of Education, 2004, p. 30). Thus, while the California accountability system includes some attention to subgroup performance within a school, it does so in a way that differs from NCLB in at least two significant ways. First, fewer subgroups are considered. Second, the subgroup target for improvement of performance is somewhat lower than the school-wide target, which unlike NCLB, implicitly makes some allowance for the less reliable gains in achievement for subgroups that obviously have fewer students than are available for calculating the school-wide API changes.

The one provision of the NCLB accountability system that considers improvement from one year to the next rather than only annual performance in comparison to an AYP is the, so called, 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.

Although the safe harbor provision is intended to allow schools that fall short of the AYP goal to still make AYP if they show substantial improvement, 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 even by schools where students show considerable improvement from one year to the next. Only a tiny fraction of schools actually meet AYP through the safe harbor provision because it is so extreme. In Pennsylvania, for example, only about 1% of about 780 schools that had one or more subgroups miss the AYP target made AYP because of the safe harbor provision in 2002–2003 (Nelson & Rosenberg, 2004b). The percentage of schools that were saved by safe harbor was somewhat greater in Massachusetts than in Pennsylvania. Forty-four (5%) of the 884 schools that made AYP in Massachusetts did so because of the safe harbor provision. The 44 schools that were saved by the safe harbor provision tested fewer students on average than the typical Massachusetts school, and smaller schools tend to have less stable results from one year to the next. Those schools also had fewer subgroups large enough to be considered in the determination of AYP status (Nelson & Rosenberg, 2004a).

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% reduction in the below proficient category from one year to the next rather than the current 10% criterion. Changing the safe-harbor provision from a 10% reduction in below proficient to a 3% reduction would go a long way toward solving the problems caused by the multiple hurdles created by subgroup reporting while still maintaining a focus on the improvement in performance of all subgroups.

In addition to considering growth as measured by comparison of successive cohorts of students, it would also be desirable to allow the use of estimates of growth that rely on longitudinal data with matched student records across years. Tennessee has received considerable attention for its “value-added” approach based on longitudinal data (See, for example, Sanders and Horn, 1998) and a number of states are interested in tracking individual students across years and using longitudinal results in their accountability systems.

Other Requirements for NCLB Achievement Goals

There are several important differences between the way in which school achievement goals are set for purposes of NCLB and the ways in which they are typically set in state accountability

systems. First, as was discussed in the previous section, there is the difference between status in comparison to a target used for NCLB and improvement targets used in the typical state accountability system. Second, there is the difference between NCLB's use of an absolute level of performance that is constant regardless of a school's initial status whereas states usually set targets at levels that depend on the school's performance in a baseline year or biennium. Third, NCLB and the typical state system differ in the establishment of the long-range goals and the timeline for reaching those goals.

Establishment of Proficiency Targets for 2014. As was previously noted, NCLB requires states to set challenging academic performance standards. There must be at least three performance standards for each assessment. NCLB provides only general guidelines to states for defining the academic achievement standards, specifying only that the state establish "(ii) challenging academic achievement standards that—(I) are aligned with the State's academic content standards; (II) describe two levels of high achievement (proficient and advanced) that determine how well children are mastering the material in the State academic content standards' and (III) describe a third level of achievement (basic) to provide complete information about the progress of the lower-achieving children toward mastering the proficient and advanced levels of achievement" (NCLB, P. L. 107–110, sec. 1111 (b)(1)(D)).

All students must be at the proficient level or above by 2013–2014 for schools and districts to avoid sanctions, regardless of how leniently or stringently the state defines the proficient standard. State starting points for percent proficient or above were supposed to have been established using assessment results from the 2001–2002 academic year. The state's starting point is equal to the higher of the following two values: (1) the percentage of students in the lowest scoring subgroup who achieve at the proficient level or above and (2) "the school at the 20th percentile in the State, based on enrollment, among all schools ranked by the percentage of students at the proficient level" (NCLB, P. L. 107–110, Sec. 1111 (b)(2)(E)(ii)).

Establishment of Intermediate (AYP) Targets. States must also establish intermediate goals (the AMOs described above), for AYP between the 2001–2002 starting point and the 100% proficient goal in 2013–2014. The first increase in the goal from the starting point must occur by 2004–2005, and subsequent increases must occur in not more than three years following the last increase.

State-to-State Variability in AYP Plans and Judgments

Variability in Trajectory of Intermediate (AYP) Targets

Although some states, e.g., Florida, have set their intermediate goals for AYP by using equal increments each year to move from the starting point in 2002 to 100% in 2014, a number of states have elected to use a stair-step approach to setting their intermediate goals for AYP with increases in 2005, 2008, 2011, and 2014 and static levels for intermediate years as illustrated in Figure 3 by Colorado and North Carolina. Alternatively states opted for stair steps in 2005, 2008 and 2011, but then had annual increments through 2014 as is illustrated in Figure 3 by Arizona and Louisiana.

Porter, Linn, and Trimble (2005) referred to the approach illustrated by Arizona and Louisiana as a "back-loaded trajectory" and called the approach illustrated by Colorado and North Carolina a "linear with plateaus trajectory". Porter, et al obtained information about state trajectories for 47 states. Nineteen state plans use a linear with plateaus trajectory, 24 state plans use a back-loaded trajectory, and only 4 states use a straight-line trajectory.

Variability in Proficiency Targets

A comparison of the graphs of the AYP targets for the four states displayed in Figure 3 shows that the four states have quite different starting points. North Carolina has a starting point of 74.6% proficient or above which is slightly more than 10 times as high as Arizona's starting point of 7% proficient. Colorado's starting point of 60.7% is twice as high as Louisiana's starting point of 30.1%. Yet the 2014 AYP target for all students in these four states, as well as in all the other states, is 100% proficient or above. That sort of improvement in student achievement is completely unrealistic (see, for example, Linn, 2003; 2004, and McCombs, Kirby, Barney, Darilek & Magee, 2004, for discussions of the unrealistic nature of the 100% proficient goal by 2014).

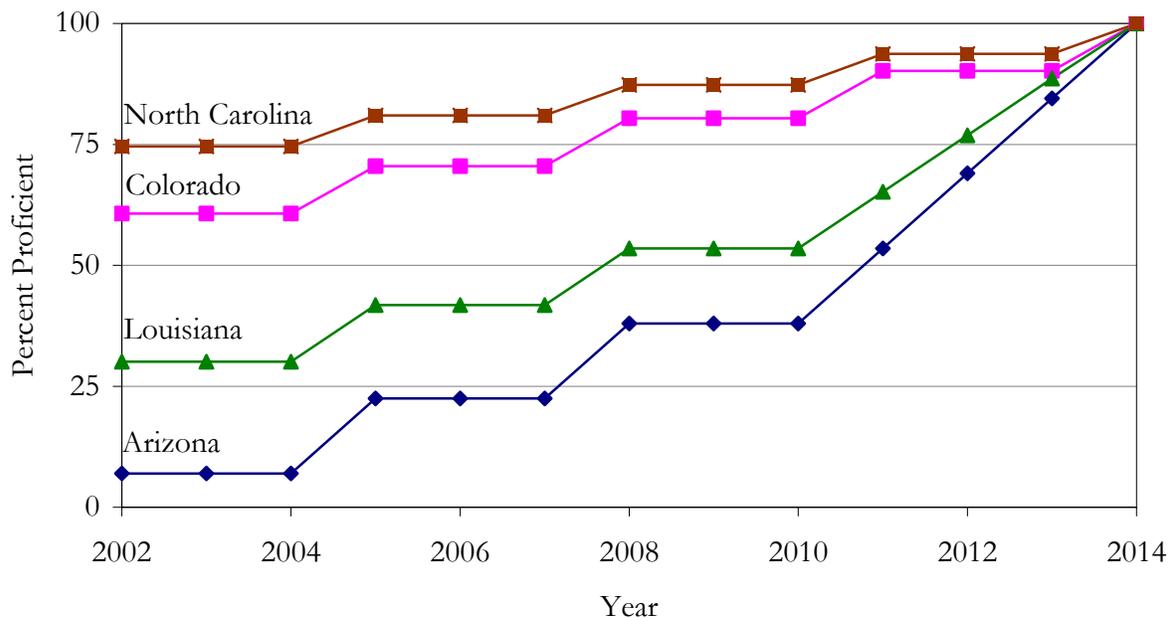


Figure 3
Intermediate (AYP) Goals for Four Illustrative States

The large variation in starting points is a result of the large between-state variability in the stringency of state performance standards. Because of concerns about state control of education and avoidance at the federal level of anything that hints of an attempt to impose a national curriculum, it is not surprising that the definition of academic content standards, the choice of assessments that are used to measure those standards, and academic achievement (performance) standards are left for the state to determine. The result is that the state performance standards and assessments are not comparable. That would not necessarily be a problem were it not for the requirement that all students reach the proficient level or above by 2014. In addition to being unrealistic, the 100% proficient goal is radically different from one state to the next.

The differences among starting levels for the four states shown in Figure 3 are much larger than the differences in actual performance of students in the eighth grade mathematics. The percentages of public school students who were at or above the proficient level on the 2003 NAEP grade 8 mathematics assessment were as follows for the four states displayed in Figure 3: Arizona, 21%; Colorado, 34%; Louisiana, 17%; and North Carolina, 32% (National Center for Education

Statistics, 2004). Although student performance on the 2003 grade 8 NAEP mathematics assessment does differ for these four states, the range from high to low in percent at the proficient level or above is 17%, which is small in comparison to the range in differences in AYP starting points of 67.6%. Moreover, the rank order of the four states in terms of AYP starting values does not match the rank order in terms of actual performance on NAEP in 2003.

There obviously are differences between the state eighth grade mathematics assessments and NAEP. States have different academic content standards that overlap imperfectly with the NAEP mathematics framework and the conditions under which the assessments are administered differ. But Congress apparently expected there to be some reasonable relationship between NAEP and state assessments since NCLB requires states to participate every other year in the state-by-state NAEP assessments in reading and mathematics at grades 4 and 8 beginning in 2003. Comparisons of state assessment results to NAEP results provide an indication of how well state results generalize. Furthermore, the work of McLaughlin and Bandeira de Mello (2002) indicates that there is a fairly substantial relationship between most state assessments and NAEP. Thus, it seems reasonable to conclude that the differences between states in terms of their percent proficient or above on NAEP and the NCLB starting values have more to do with difference in the stringency of their performance standards than to differences in the state assessments and NAEP.

It should also be noted that the steep annual increases that Arizona and Louisiana chose to use to set AYP targets for the last four years (2011 through 2014) are just the opposite of what might be expected. Past experience with test-based accountability systems has shown that larger gains are usually made in the first few years following implementation and that gains generally become smaller in later years. Moreover, common sense suggests that it likely to be much harder to realize a gain of 5% to move from 95% to 100% than from 30% to 35% proficient or above.

Variability in Percentage of Schools Meeting AYP Goals

Figure 3 illustrated the fact that states differ substantially in the starting points, as well as their intermediate AYP goals. There is also considerable state-to-state variability in the percentage of schools that met AYP in each of the first two years (2003 and 2004) of AYP reporting. Olson (2004) reported the percentage of schools making AYP for 41 states in 2003 and 44 states in 2004. The percentage of schools that met AYP goals in 2004 in a 45th state, Illinois, was obtained from the Illinois State Board of Education web site. In 2003 the percentage of schools that met AYP goals ranged from a low of 18% to a high of 95%, with an average of 65.6% for the 41 states listed by Olson (2004). For the 45 states with 2004 results, the range was from 23% to 96% with an average of 74.2%.

Olson (2004) reported the percentage of schools meeting AYP goals for **both** 2003 and 2004 for only 36 of the states. The percentage of schools meeting AYP goals was higher in 33 of the 36 states in 2004 than in 2003 and the three exceptions, Indiana, Louisiana, and Michigan, had decreases from 2003 to 2004 of only 1 or 2%. The average 2003 to 2004 increase in percentage of schools making AYP goals was 10.2% and eight states (Connecticut, Delaware, Massachusetts, Missouri, North Carolina, Pennsylvania, South Carolina, and Tennessee) had increases that ranged from 20 to 30%. These appear to be remarkable improvements in a single year. Although the increases probably reflect some real improvement in student performance, they are largely due to artifacts such as schools getting better about meeting requirements that at least 95% of the eligible students in each subgroup participate in the assessments. An even larger part of the apparent improvement can more reasonably be attributed to changes in AYP calculations that states requested and the U.S. Department approved between the 2003 and 2004 reports.

The Center for Education Policy (CEP) posted a report on their web site dated October 22, 2004 summarizing changes in state implementation of NCLB accountability rules (CEP, 2004). According to the CEP report, 47 states requested approval for change in their NCLB accountability plans. The U.S. Department of Education posted letters to 35 of those states in time to be reviewed for the CEP report “approving many, though not all of the changes” (CEP, 2004, p. 1). Many of the approved changes make it easier for schools to meet AYP goals. For example, 11 states changed the minimum group size for disaggregated reporting and 12 states introduced the use of confidence intervals, which give the benefit of the doubt to schools in cases where the percentage of students who are proficient or above is somewhat below the target value required for meeting the AYP goal.

The number of schools that meet AYP requirements increases when the minimum group size is increased and/or the use of confidence intervals is introduced (Porter, Linn & Trimble, 2005). Thus, states that increased the minimum group size in 2004 or introduced confidence intervals for the first time helped schools make AYP goals in 2004 that would not have made it without these changes in NCLB accountability plans. It is worth noting in this regard, that 4 of the states (Missouri, North Carolina, Pennsylvania, and South Carolina) that were among the 8 states showing the largest increases in percentage of schools meeting AYP goals from 2003 to 2004 were also among the 12 states that started using confidence intervals in 2004.

Meeting AYP Goals and Performance on NAEP

The substantial state-to-state variability in the percentage of schools meeting AYP makes it evident that the likelihood that a school will fail to meet AYP goals depends not only on the performance of students in the school, but also, at least in part, on the state in which the school is located. Furthermore, as can be seen in Figures 4 and 5, the percentage of schools in a state that meet AYP goals has only a weak relationship to differences among states in student performance on NAEP. Figure 4 shows the relationship of the percentage of schools within the state that met AYP goals in 2003 and the average percent proficient or above across the 2003 NAEP grade 4 reading, grade 8 reading, grade 4 mathematics, and grade 8 mathematics assessments. Figure 5 shows the relationship of the same NAEP average percent proficient or above with the percentage of schools that met AYP goals in 2004.

It is apparent from an inspection of Figures 4 and 5, that there is only a relatively weak relationship between the performance of students in a state on NAEP and the percentage of schools that meet their AYP goals. In 2003, the state with the second lowest average performance on NAEP had more than 90% of their schools make AYP, whereas only half the schools met their AYP goals in the state with the highest performance on NAEP (Figure 4).

The relationship between the average NAEP performance in 2003 and the percentage of making AYP in 2004 included a few more states and slightly higher than it was for 2003 AYP results. Nonetheless, there are some notable outliers shown in Figure 5 where a state with relatively low average performance on NAEP has a noticeably higher percentage of schools making AYP than another state with relatively high average performance on NAEP.

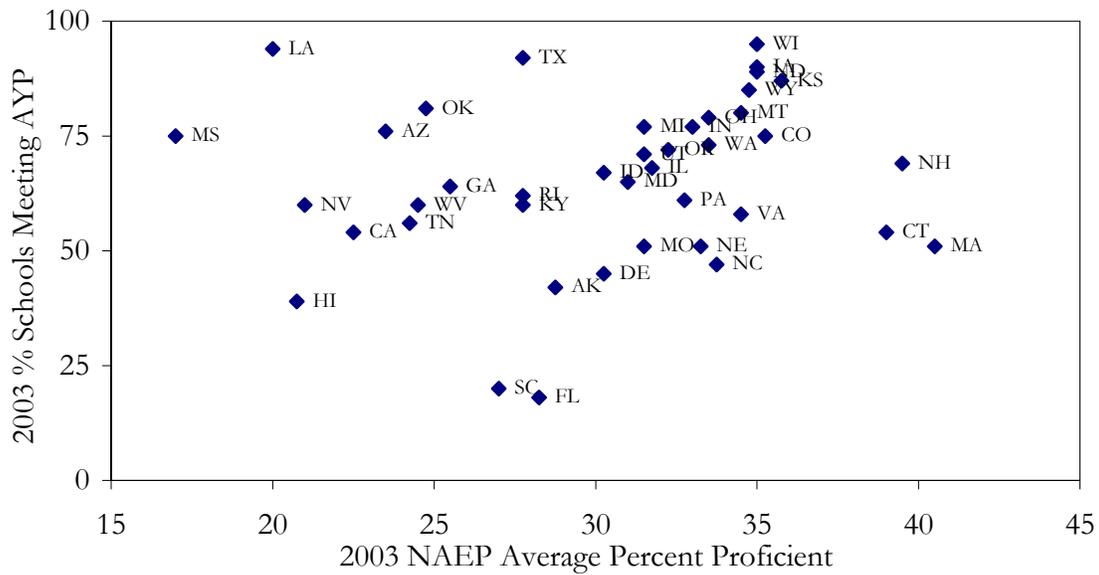


Figure 4
 Relationship of Average of State Percent Proficient on 2003 NAEP and Percentage of Schools Meeting AYP in 2003 (41 States, Correlation = .26)

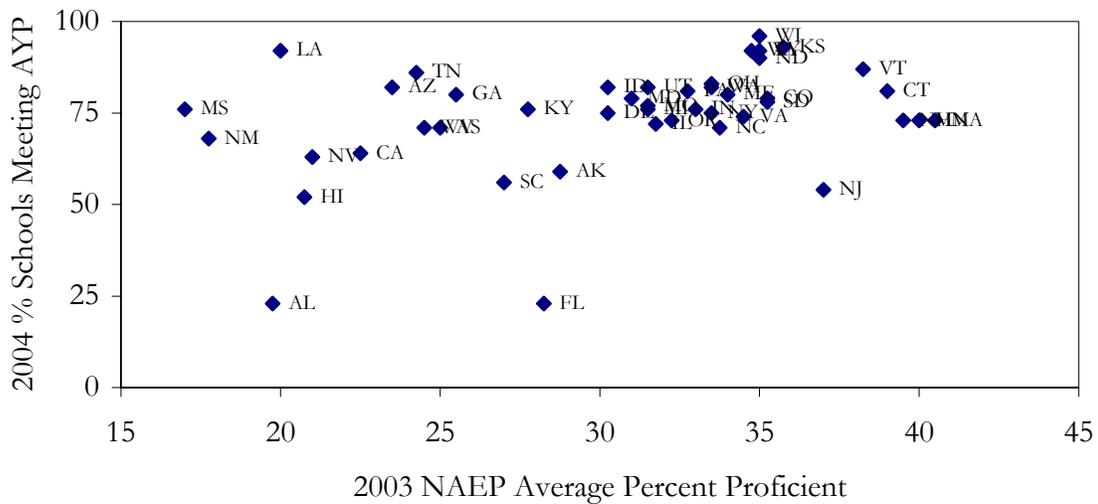


Figure 5
 Relationship of State Average Percent Proficient on 2003 NAEP and Percentage of Schools Meeting AYP in 2004 (45 States, Correlation = .35)

Results such as those shown in Figure 3, 4, and 5, clearly illustrate that it is not meaningful to compare states in terms of the performance standards or the rates at which schools in different states meet NCLB’s AYP requirements. The performance standards set by states bear little relationship to real between-state variability in student performance. Differences in performance

standards set by states as well as differences in the ways in which states, with the approval of the U.S. Department of Education, comply with NCLB accountability requirements, with regard to features such as the minimum number of students needed for determining if subgroups must meet targets, and whether or not confidence intervals are used, obscure between-state comparisons of percentages of schools meeting AYP requirements. Although comparisons among states is not one of NCLB's purposes, it is clear that the likelihood that a school will be identified as failing to meet AYP targets in a given year, or be placed in the needs improvement category, or be subject to more serious sanctions, depends to a substantial degree on the state in which the school is located and not exclusively on the effectiveness of the school.

Conclusion

Test-based accountability has become a pervasive consideration for schools and educators as a consequence of the combination of state accountability requirements and those imposed by NCLB. Because of the substantial differences in state and NCLB requirements, mixed messages that are confusing to the public are being given about school performance. The goals established under NCLB are already unrealistic for many schools that started with low performance in 2002 and will become increasing so, not only for those schools but for all schools as the increases in AYP targets occur, especially in 2005 and 2008 when many states will have big jumps in their AYP targets. If the goal for 2013–2014 remains unchanged, essentially all schools will fail to meet the unrealistic goal of 100% proficient or above, and No Child Left Behind will have turned into No School Succeeding.

Significant changes in NCLB accountability requirements are needed to avoid labeling all schools as failures. What are some of the needed changes? Possibly most important is to make the goal something that is more realistically obtainable. As noted above, NCLB requires states to participate every other year in the 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. (It should be noted that the NAEP achievement levels have been the subject of considerable criticism, in part, because they are set at levels that are higher than the performance of students in any country—see, for example, Linn, 2003.)

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 a value that has previously been achieved by the highest performing schools. For example, if the top 10% of schools in a state, in terms of sustained improvements in student achievement, had rates of improvement in the percentage of students achieving at the proficient level 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.

Saying that all students must be at the proficient level or above by 2014, but leaving the definition of proficient achievement to the states has resulted in so much state-to-state variability in the level of achievement required to meet the proficient standard that “proficient” has become a meaningless designation. Certainly, reporting results in terms of percent proficient or above on state assessments lacks comparability from state to state.

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 (Linn, 2004). 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 use what Popham (2004) has called grade-level descriptors. "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.

The NCLB insistence on a common target for all schools, regardless of where they started, is appealing in the sense that it sets the same high expectations for all, but is nonetheless counterproductive when it leaves schools with initially low performing students with no realistic hope of making the absolute target. Schools demonstrating substantial improvement should not be labeled as failing to make adequate progress, and for the reasons discussed above, NCLB's safe harbor provision turns out to be no real help to most schools in this regard due to the high hurdle that is established.

Holding schools accountable for the performance of students in subgroups that have too often been ignored in the past (e.g., racial/ethnic minorities, economically disadvantaged, limited English proficient students, and students with disabilities) is a desirable feature of NCLB. As it is implemented, however, it places large, diverse schools at a substantial disadvantage. Changing the safe-harbor provision from a 10% reduction in below proficient to, say, a 3% reduction would go a long way toward solving the problems caused by the multiple hurdles created by subgroup reporting while maintaining a focus on the improvement in performance of all subgroups.

References

- California Department of Education. (2004). *2003–2004 Academic Performance growth report: Information Guide*. Retrieved October 22, 2004, from <http://www.cde.ca.gov/>.
- Center for Education Policy. (2004). Rule changes could help more schools meet test score targets for the No Child Left Behind Act. Retrieved October 22, 2004 from <http://www.cep-dc.org/>.
- Dillon, S. (2004, September 5). Good schools or bad? Conflicting ratings leave parents baffled. *New York Times*.
- Ford, L. A. & Thacker, A. A. (2005). *Consequential impact of the No Child Left Behind Act on Kentucky's Accountability System: Phase I*. FR-05-03. Louisville, KY: HumRRO.
- Kentucky Department of Education. (2004). *2004 CATS interpretive guide detailed information on using your score reports*. Frankfort, KY: Author, available online at <http://www.education.ky.gov>.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3–13.
- Linn, R. L. (2004, July). *Rethinking the No Child Left Behind Accountability System*. Paper presented at Center for Education Policy Forum, Washington, DC. Available online at <http://www.ctredpol.org>.
- 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.
- McCombs, J. S., Kirby S. N., Barney, H., Darilek, H., & Magee, S. (2004). *Achieving state and national literacy goals, a long uphill road*. TR-180. Santa Monica, CA: Rand.
- McLaughlin, D. & Bandeira de Mello, V. (2002, April). *Comparison of state elementary school mathematics achievement standards using NAEP 2000*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- National Center for Education Statistics. (2004). *The Nation's report card: Mathematics highlights 2003*. Available online at <http://nces.ed.gov/nationsreportcard/mathematics/>.
- Nelson, H. F. & Rosenberg, B. (2004a). *AYP status and school proficiency ratings 2002–2003: Massachusetts*. Washington, DC: American Federation of Teachers.
- Nelson, H. F. & Rosenberg, B. (2004b). *AYP status and school proficiency ratings 2002–2003: Pennsylvania*. Washington, DC: American Federation of Teachers.

No Child Left Behind Act of 2001, Pub. Law No. 107-110.

Novak, J. R. & Fuller, B. (2003, December). 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. (2004). Taking Root [Special Report]. *Education Week*, 24(15), pp. S1-S10.

Popham, W. J. (2004, July). *Ruminations regarding NCLB's most malignant provision: Adequate Yearly Progress*. Paper prepared for the Center for Education Policy Forum, Washington, DC. Available online at <http://www.ctredpol.org/>.

Porter, A. C., Linn, R. L., & Trimble, S. (2005, April). *Consequences of AYP design decisions: Lessons from Kentucky*. Paper presented at the Annual meeting of the National Council on Measurement in Education. Montreal, Canada.

Sanders, W. & Horn, S. (1998). Research findings from the Tennessee value added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.

About the Author

Robert L. Linn

School of Education
University of Colorado at Boulder

Email: Robert.Lynn@colorado.edu

Robert L. Linn is Distinguished Professor of Education at the University of Colorado at Boulder and Co-Director of the National Center for Research on Evaluation, Standards, and Student Testing. He has published over 200 journal articles and chapters in books dealing with a wide range of theoretical and applied issues in educational measurement and has received several awards for his contributions to the field, including the ETS Award for Distinguished Service to Measurement, the E.L. Thorndike Award, the E.F. Lindquist Award, the National Council on Measurement in Education Career Award, and the American Educational Research Association Award for Distinguished Contributions to Educational Research. He is past president of the American Educational Research Association, past president of the National Council on Measurement in Education, past president of the Evaluation and Measurement Division of the American Psychological Association, and past vice-president for the Research and Measurement Division of the American Educational Research Association. He is a member of the National Academy of Education, a Lifetime National Associate of the National Academies, and serves on two Boards of the National Academy of Sciences.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Greg Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

David C. Berliner

Arizona State University

Casey Cobb

University of Connecticut

Mark E. Fetler

California Commission on Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Appalachia Educational Laboratory

Patricia Fey Jarvis

Seattle, Washington

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Burian-Fitzgerald
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Públicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España