

The American Review of Public Administration

<http://arp.sagepub.com/>

School Performance Measurement : Politics and Equity

William H. Miller, Brinck Kerr and Gary Ritter

The American Review of Public Administration 2008 38: 100

DOI: 10.1177/0275074007304387

The online version of this article can be found at:

<http://arp.sagepub.com/content/38/1/100>

Published by:



<http://www.sagepublications.com>

On behalf of:



[American Society for Public Administration](http://www.aspa.org)

Additional services and information for *The American Review of Public Administration* can be found at:

Email Alerts: <http://arp.sagepub.com/cgi/alerts>

Subscriptions: <http://arp.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://arp.sagepub.com/content/38/1/100.refs.html>

>> [Version of Record](#) - Jan 30, 2008

[What is This?](#)

School Performance Measurement Politics and Equity

William H. Miller

University of Illinois, Springfield

Brinck Kerr

University of Arkansas, Fayetteville

Gary Ritter

University of Arkansas, Fayetteville

Growing political pressure for increased efficiency in government has led many to support the use of performance measurement schemes. Such schemes are never neutral and frequently influence the allocation of resources and power. The increased use of performance measures in the public sector raises questions regarding the fairness of the measures and the justness of the accompanying sanctions and rewards. Focusing on public schools, this study demonstrates the need to assess governmental performance measurement schemes for their impact on equity. Three measurement schemes (No Child Left Behind, value added, and adjusted performance measures) are evaluated and compared regarding their impact on majority African American public schools in the state of Arkansas.

Keywords: *education; policy; performance; equity*

Roughly a half century ago, it again became clear that performance measurement in education was a controversial and contested process (Kelly & Miller, 1989). Measurement continues to be a political issue. In a contemporary spin of the *Brown*-related desegregation performance controversy, the Supreme Court considered whether school districts may use counts of African American students to determine desegregation performance to guide district enrollment policy (*Meredith v. Jefferson County Board of Education*, 05-915; *Parents Involved in Community Schools v. Seattle School District*, 05-908). In this article, we focus on the measurement of school-level academic performance.

The role of federal and state governments, the adequacy of measurement schemes, the tendency of measures to mandate privatization or choice alternatives, and the sanctions or rewards attached to measures all contribute to contemporary political debates about the performance and direction of U.S. public education. The push for greater accountability and increased standards, some policy makers and think tanks argue, is about the performance of *individual* children, *individual* classrooms, and *particular* school administrators and is not related to the former emphasis on *group rights*, an emphasis that was prevalent in the immediate post-*Brown* era (Gibson & Asthana, 1998; Kirsanow, 2002; Manno, 1995). The newest performance standards are often presented as positive and sincere efforts to leave

Initial Submission: January 28, 2006

Accepted: May 14, 2007

“no child behind” irrespective of the child’s race, ethnicity, or group affiliation (No Child Left Behind [NCLB], Section 6301(3); also see Hess & Petrilli, 2006; Sunderman, Kim, & Orfield, 2005). Paradoxically, Orfield (2005) notes, even though legislators were presented with a list of experts willing to testify before committees considering pending NCLB legislation in the 107th Congress (2001-2002), none of these experts on minority issues and civil rights was called or heard. Many scholars and other observers continue to be concerned with group and individual implications of academic performance measures. A good discussion of the philosophical debate over standards-based approaches may be found in a recent article concerning whether such approaches are “expanding the moral community” or “blaming the victim” (McDermott, 2007).

The purpose of this study is less philosophical. We explore the disparate impact of various education performance measurement schemes on predominantly (i.e., more than 50%) African American elementary schools in Arkansas when compared to elementary schools that are not predominantly African American. It is important to note that Arkansas schools do not represent a national sample for the purposes of statistical generalization. Although many of the conditions in Arkansas are similar to those in other states, especially smaller southern states, this study is not designed to create nationally generalizable measurement standards. Rather, the primary objective is to demonstrate how various exam-based measurement schemes might differentially affect schools serving African Americans in Arkansas. If these measurement schemes result in disparate group impacts, then we believe the resulting inequities should be evaluated as well. Relevant to the evaluation of performance measurement schemes, Levin-Waldman (2004) argues that “there is no absolute model that can drive the policy process . . . policy ultimately boils down to how data are interpreted and how those interpretations are put in the service of those interests that have a particular agenda” (p. 139). We now turn our attention to the education policy literature on test performance and public school standards and accountability.

Literature Review

One of the central debates among education policy scholars during the past several decades has concerned the character and desirability of measurement standards (Ravitch, 2000). Many scholars and practitioners favor increased standards and testing based on uniformity at either the national or the state levels (e.g., Hess & Finn, 2004; Powell, Farrar, Cohen, National Association of Secondary School Principals, & National Association of Independent Schools, Commission on Educational Issues, 1985; Ravitch, 2000; Thernstrom, 2000; Thernstrom & Thernstrom, 2003). Researchers favoring these proposals are variously characterized as concerned about efficiency, declining student performance, the relative performance of U.S. schools compared with schools in other advanced economies, a weakened curriculum, promotion of a rigorous education for all children, and the producing of students with skills required to survive in competitive market economies.

On the other hand, some observers favor standards that are flexible and sensitive to diverse student environments (e.g., Apple, 1996; Berliner & Biddle, 1995; Kozol, 2000; K. J. Meier, Polinard, & Wrinkle, 2000). Kozol (2000) is concerned that policies mandating uniform standards will place exclusive emphasis on the goals of economic productivity and efficiency.

Likewise, D. Meier, Cohen, and Rogers (2000) argue that there is too much emphasis on efficiency and economic productivity and that we need to recognize that schools do teach (and should teach) democratic values such as tolerance, compromise, inclusive decision making, and the like. Ayers (2000) observes that uniform state standards can have the effect of creating two-tiered state systems with identical standards aimed at thriving and collapsing schools. Inherent in any scheme of uniform standards are incentives for teachers to teach to the tests, a practice that many educators feel is undesirable (D. Meier et al., 2000). Apple (1996) maintains that stressing a common culture through the imposition of a uniform curriculum will have the effect of reviving traditional class, gender, and especially race hierarchies. Berliner and Biddle (1995) go as far as to question the commonly accepted notion that the performance of U.S. schools has declined during the past several decades. Their empirical analysis suggests that the public education “crisis” of the 1970s and 1980s was “manufactured” and that some individuals promoting the crisis supported the redistribution of resources from needy to privileged schools. Moreover, another key question inherent in schemes based on uniform standards is how the schemes treat economically and racially disadvantaged groups.

An important subarea of the general debate about school measurement standards focuses on the gap in test scores between Black and White students. A consistent finding in this body of literature is that Blacks in the aggregate score lower on standardized tests than do Whites and that this gap has narrowed slightly since the 1970s (Berliner & Biddle, 1995; Jencks & Phillips, 1998). Most studies in this literature focus on either the causes of or consequences of the gap in test scores.

Explanations for causes in the Black–White performance gap include racial disparities in family income and education, differences in school resources, the development of an oppositional culture, lower-quality teachers, large class sizes, and racial bias in cognitive tests (Coleman et al., 1966; Ferguson, 1998; Grissmer, Flanagan, & Williamson, 1998; Jencks & Phillips, 1998; Ogbu, 1978). Hanushek and Raymond (2005), in an empirical work, note two impacts of NCLB relevant to our study:

First, accountability as seen during the 1990s tended to help White achievement more than Black achievement. Second, the observed movement toward higher minority concentrations in schools has a detrimental effect on Black achievement, again pushing toward a wider distribution of achievement. (p. 297)

Jencks and Phillips (1998) argue that desegregation seems to have increased the test scores of Black students in the South without affecting the scores of Whites, again suggesting that the resegregation of schools will harm Blacks without harming Whites.

A very recent study by the Civil Rights Project argues that NCLB has not helped the nation and states significantly narrow the achievement gap. The project argues that “by 2014, less than 25 percent of poor and black students will achieve NAEP [National Assessment of Educational Progress] proficiency in reading, and less than 50 percent will achieve proficiency in math” (Lee, 2006, p. 11). Interestingly, Thernstrom and Thernstrom (2003, p. 273), in their argument for stronger standards, also maintain that NCLB will not sufficiently close the racial gap in test scores. Hoerandner and Lemke (2006) used estimated production functions, and they too determined that schools with large gaps could not reasonably be

expected to meet the uniform NCLB standards. They found that only 10% to 15% of the gap was “controllable” by school-level actions or policies and that 30% to 50% of the gap was caused by uncontrollable factors such as income and racial composition.

According to many scholars, the continuing Black–White performance gap has important consequences. The empirical research indicates that the Black–White performance gap is associated with racial disparities in educational attainment and income, higher dropout rates for African American children, lowered expectations for African American children, development of two-tiered school systems, the practice of teaching to the content of standardized tests, and, of course, increased likelihood that schools will be labeled as failing (Ayers, 2000; Jencks & Phillips, 1998; K. J. Meier et al., 2000; Sunderman et al., 2005). Jencks and Phillips (1998) argue that a significant reduction in the gap would do more to move the United States toward racial equality than any other politically plausible alternative.

In this article, we examine the consequences of this performance gap and its measurement from an equity perspective. Consistent with NCLB and other measurement schemes, we focus on the school as the unit of analysis. It is important to note that much of the previous empirical research on performance focuses on the students and does not use the school as the unit of analysis. Despite the importance of the impact of performance measurement on African Americans as a group, Paul (2004) argues that only a limited number of empirical and legal research projects focus specifically on the question of impact.

The Role of Equity in Performance Measurement

The literature on performance measurement in the public sector emphasizes the need to measure productivity to ensure accountability and efficiency (Courty, Heinrich, & Marschke, 2005; McEwen, 1995; Nicholson-Crotty, Theobald, & Nicholson-Crotty, 2006; Pidd, 2005). Although affirming the importance of such measures, these authors and others warn against improper, poorly developed, or misguided measures or misguided applications of measures. Undesirable outcomes include the lack of fairness in accountability regarding factors beyond agency control, distorted behavior resulting from the use of accountability measures, concentration on outcome measures rather than on outcomes themselves, and so on. There seems to be wide agreement among scholars that there is no “one best” measure of public sector performance (Courty et al., 2005; Kelly & Miller, 1989). There is also a consensus that it is desirable to use multiple measures in assessing performance, proposing reforms, providing rewards, providing sanctions, and so on.

Although the use of any particular set of indicators to influence sanctions or rewards may be controversial, there is little doubt as to the need to measure education performance. As McEwen writes (1995),

Public education is funded by taxpayers who want and have a right to know if they are getting value for their investment. Such accountability requires public information. An indicator system is a tool to focus reform and to improve accountability by providing better information about the education system’s performance. (p. 27)

As noted above, the question is not whether educational progress should be measured—most, if not all, agree that it should—but how performance should be measured, how to

interpret the results of the measurement, and how such measurements should be utilized to allocate resources and power to public schools. In the present context, NCLB uses a particular performance measurement scheme that may lead to negative sanctions including reductions in school-level resources and/or control (Farmer, 2005; Hanushek & Raymond, 2005; Harris & Herrington, 2006; Lee, 2006). In such situations, performance measures must be evaluated not only for their impact on efficiency (however it is defined) but also on their equity impact (Myers, 2002; Smith, 1995). This study, through our analysis of performance measures in public schools, demonstrates the need to assess such governmental performance measures for their impact on equity.

A short discussion of what constitutes equity in this instance might be helpful. DeClue (1988), in his discussion of desegregation in St. Louis, writes,

The *Brown* decision did two things that we must never allow ourselves to forget. Number one, it called into account the whole system of segregation. We've seen the walls tumble. And number two, *Brown*, more than anything else, said to this country that you can no longer degrade black people and use the results of that degradation to oppress them further. That to me is the essence of school desegregation cases. (p. 4)

DeClue here identifies the two aspects of equity in education that might serve as a working definition of equity for our purposes. Both aspects are examples of group, or as Fredrickson (1990) calls it, block inequality.

Following DeClue (1988) and the *Brown* understanding, and for the purposes of this article, the lack of policy equity can then be identified as having two sources. First, an inequitable policy might identify a group as a minority and then discriminate against this group simply by virtue of its minority status. Legal (de jure) segregation in education is an example of this type of inequity. The second type of inequitable policy is a policy that uses "the results of that [past] degradation to oppress them further" (DeClue, 1988, p. 4). We will focus on this latter type and so ask whether a policy may be inequitable because the policy may degrade (penalize, sanction, disadvantage) African Americans as a group in some manner by using the results of past degradations to further oppress or harm.

Context

Arkansas schools, taken on the whole, are highly segregated by ethnic group. Focusing on the schools in this study, the typical (mean) school is 19% African American. But this can be misleading. Of the schools in this study, 50% have 3% or fewer African Americans, and most of these 50% have very few African American students. Of the schools, 25% are greater than 25% African American, and 63% of the African American students go to schools that are majority African American (greater than 50% African American). It is beyond the scope of this study to determine if this continued segregation is the result of past policy and history, present policy (de jure), or some continuing nonpolicy factors (de facto). Furthermore, we are not suggesting that segregation alone causes the gap researchers find in minority educational achievement. Such questions are very complex and are well beyond the scope of this study. Rather, we would like the reader to be aware of the high level of segregation

in the state to better understand the political, socioeconomic, and civil rights issues that performance measurement in such a situation raises.

It should also be noted that the correlation between ethnicity and economic status in Arkansas schools is very high. African Americans in Arkansas have been and continue to be, on the whole, significantly poorer than Whites (U.S. Census Bureau, 2002). In our study of schools, 87% (median) of the children in majority African American schools were eligible for free and reduced lunches. This is not to understate the level of poverty among many Whites in Arkansas but rather to say that poverty is present among African Americans at a much higher level. In the majority non-African American schools in our study, 57% (median) of their children were receiving free and reduced lunches.

Method

As previously stated, we focus our research on school-level analysis. There is a good deal of research on the usefulness of various productivity measures at the classroom level. In this study, we are not exploring either student-level or classroom-level measurement or performance. We are exploring school-level models for performance evaluation. So care must be taken not to confuse individual student or classroom-level evaluation with the task of modeling school-level success or failure. For instance, the documented concern a researcher might have with reliability caused by yearly variation at the individual classroom level is less of a problem at the school level, which will most often include multiple classrooms. It is important to be cognizant of this ecological distinction.

Consistent with this school-level focus, this article does not concern itself with whether individual student achievement scores on tests are equitably measured or whether there is measurement equity when students are aggregated into ethnic groups. Though the question of the equity of individual or group-level measurement is an important subject of inquiry, our focus is on the impact of the various measurement schemes or models that utilize individual student scores to create models of school performance. There are important methodological, philosophical, and political issues in the assumptions imbedded in these various school performance modeling schemes. We argue that the way policy makers construct and utilize these models of success or failure will differentially affect minority schools.

In addition, our analysis of various performance measurement schemes will not allow us to address the very important question of whether the testing instruments are in fact truly valid measures of a student's skills or a school's performance. That is, this study will not allow us to ascertain whether schools that are highly rated by present tests are actually doing a better job at teaching kids; as a result, we cannot speak to the question of whether school-level performance measurement is truly useful or even technically possible given the present sophistication of educational measurement. However, because we know that performance measurement is an ongoing exercise having important implications for resource allocation, the critical question for this study is: What are the effects of different school performance measurement schemes with respect to minority groups and equity?

For our analysis, we selected all elementary schools in Arkansas that had both second- and fourth-grade classes in 2002 and 2004. We use second and fourth grades because many districts have middle schools that begin in the fifth grade. To be included, the school must

Table 1
Measures Utilized in Study

Proficiency measures	
Adequate yearly progress (AYP) measure of success	ACTAAP exams adjusted by Arkansas scheme for NCLB, 2002-2003 school year
Value-added measure	Change in school percentage proficient ACTAAP on exam from second (2002) to fourth grade (2004)
Adjusted performance measure	ACTAAP percentage proficient in fourth grade adjusted by regression model
Contextualizing measures for adjusted performance measures	
School poverty	Percentage free and reduced lunch 2002-2003 school year
School size	Total school enrollment in 2002-2003 school year
School rural-urban code	Eight-level classification (rural to urban), National Center for Education Statistics, 2003-2004
District per capita property tax	District property tax collected ÷ total enrollment, fiscal year 2002

Note: NCLB = No Child Left Behind; ACTAAP = Arkansas Comprehensive Testing Assessment and Accountability Program.

also have nonmissing scores for the Arkansas Comprehensive Testing Assessment and Accountability Program (ACTAAP) math and literacy exam in 2002 and 2004. In this manner, we were able to utilize measurement schemes that require multiyear testing within each school. For independent or contextualizing variables, we use 2002-2003 values as much as possible. These variables include per capita property tax, percentage of students who are eligible to receive free and reduced lunches, rural or urban status, and total school enrollment. These and other measures utilized in this study are described in Table 1.

Analytic Method

Using the Arkansas elementary schools in our sample, we generate performance ratings using each of three performance measurement schemes. We then compare outcomes for African American schools compared to all other schools. Finally, we produce a table of high-performing schools and low-performing schools for each group based on each of these measurement schemes.

Adequate Yearly Progress (AYP), NCLB Model

The first performance rating scheme we will consider is the AYP scheme mandated by the NCLB legislation. This model is based on an unadjusted single-grade-level performance measure (ACTAAP) and compares that score to a goal or standard for success. NCLB represents a “status model” for performance measurement. The stated goal or standard of NCLB is to get all schools to 100% proficiency in both mathematics and literacy by the 2013-2014 academic school year (National Conference of State Legislatures, 2005). In the meantime, schools must meet predetermined goals each year based on the percentage of the tested

population earning scores of proficient or above. The annual goal for percentage of proficient students increases each year so that it equals 100% by 2013-2014. Schools that do not reach the hoped-for improvements measured each year face increasingly strict sanctions, including mandatory use of the school's Title I funding to support student attendance at alternate schools, provision of supplemental education services, and possibly reorganization (Sunderman et al., 2005).

The AYP model includes two aspects that make it slightly different from a simple status model. The first is the safe harbor provision, which allows schools that did not meet AYP goals to earn exemptions from AYP sanctions if the schools experienced substantial "growth" toward the AYP goal. This departure from a simple status model will result in fewer schools' being identified in the years preceding 2013-2014.

However, the second unique aspect of the AYP model, often referred to as the trip wire system, will result in greater numbers of schools being identified. There are various trip wires, or performance goals, each of which must be met for a school to meet AYP goals. Failure to meet any single performance goal results in a school's failure to make AYP. The various performance goals are based on the tests taken and the subgroups attending the school. The NCLB law in Arkansas requires that the following subgroups of students each meet proficiency performance goals in both reading and math: White students, African American students, Hispanic students, limited-English-proficient students, low-income students, and special education students. A school is only held accountable for a subgroup's performance if it has more than 40 students tested in that particular subgroup. In essence, therefore, schools with more diverse populations are at greater risk of not making AYP because these schools have more opportunities to fail. For example, an Arkansas school with significant numbers of students who are White, Black, Hispanic, low income, limited English proficient, and special education have far more trip wires to avoid than does a nondiverse, advantaged school. NCLB sets up an accountability system that Losen says is "among the most race-conscious legislative remedies to racial inequity in K-12 education since Title VI of the Civil Rights Act of 1964" (Farmer, 2005, p. 443).

The end result of the AYP-NCLB process is a simple pass or fail designation provided by the state education department. That is, schools will either meet AYP or fail to meet AYP. For the AYP-NCLB status model, we will simply compare the failure rate of majority African American schools with those of the other schools around the state.

Value-Added Model

Spours and Hodgson describe value-added performance measures as measures that indicate "the difference between the state of knowledge or qualifications of a student on course entry and her/his state on exit" (Saunders, 1999, p. 236). When specifically related to testing, a more relevant definition is that value added is "a method of analyzing and reporting student test results based on improvement ('growth') in standardized test scores over two or more points in time" (Allen, Galinat, Leedham, & Leverich, 2004, p. 1). Value added has been said to be more useful for assessment than diagnostics (Balou, 2002).

To measure value-added performance, we looked at the difference between second- and fourth-grade test scores. Although it would be better to use individual student scores and only include those students who were in both second and fourth grade in a particular school,

those data are unavailable to us. However, a large proportion of students in second grade in a school will remain in the same school in fourth grade. So to test value-added change for the purposes of comparing various measures, the assumption of relative stability seems acceptable.

Consistent with what we ask of other performance schemes, the primary question for our analysis of the value-added measurement scheme is: Does this value-added measure affect majority African American schools differently than nonmajority African American schools? We answer that in two ways. We compute the change from second- to fourth-grade math scores by school and determine if there is any significant correlation between the percentage African American in the school and change in math and literacy scores. Then we compute the mean change from second to fourth grade in math and literacy proficiency for majority African American schools and nonmajority African American schools to see if there is a difference in the mean change by group.

Adjusted Performance Model

The third method we consider is an adjusted performance measure. Rubenstein, Schwartz, and Stiefel (2003) argue that “performance measures, such as test scores or success rates, are often inferior to performance measures adjusted for client and environmental characteristics, or adjusted performance measures (APMs)” (p. 607). Noting that noncontextualized performance measures often include factors over which organizations have no control, Rubenstein et al. advocate for a more sophisticated approach. Courty et al. (2005) note that in performance measurement, those being evaluated should be held accountable for their effort but not for external influences. So our value-added regression analysis includes only a school’s external or uncontrollable variables.

It is worth noting here that some advocates, especially those who worry that adjusted performance measures will result in lower standards, argue that adjusted performance measures will hurt minorities. The Murphy Commission, a conservative Arkansas advocacy group, makes this point in a report titled *Restoring Public Education’s Academic Mission: High Expectations, Academic Standards, Proven Methodologies and Curriculum* (Watson, n.d.). In one place, the commission cites an advocate for NCLB and other status models (Kati Haycock, Education Trust; <http://www2.edtrust.org/>) who argues that adjusted performance measures will result in lower standards for minority children, and so this type of measure will not serve minorities well.

We begin our analysis of adjusted performance measures by determining which uncontrollable contextualizing variables might affect test scores. These contextualizing variables (economic status, rural or urban status, financial resources available, total enrollment) are not controllable by local school policy. Consistent with our definition of equity, these variables can be used to “even the playing field” and account for the possible impact of past harm done. The use of regression analysis to determine which nondiscretionary variables affect test scores is common in research using adjusted performance measurement (Hanushek & Raymond, 2005; Rubenstein et al., 2003). Consistent with past practice, this study includes district per capita property tax, school percentage of students who are eligible to receive free and reduced lunches, school rural or urban status, and school total enrollment as contextualizing variables. Because both math and literacy exams are used in the calculation of NCLB–AYP success, both scores will be considered here also.

Table 2
Adequate Yearly Progress Status Model

School Type	<i>n</i>	% Failing
All schools	227	45
Majority African American	34	94
Non-majority African American	193	37

Finally, in the conclusion, we present the results in a manner designed to explore the percentage of schools in both groups that might be sanctioned, comparing each of the performance measurement models to the others.

The Analysis

AYP–NCLB (Status Model)

Using the AYP status model measure for the selected Arkansas schools (see Table 2), 45% (103) of all schools are not meeting the standard set by the AYP–NCLB scheme. In the subgroup of these schools that are majority African American, 94% (32) were found not to be meeting AYP standards in the 2002-2003 school year and will be or are in danger of being sanctioned. Of the nonmajority African American schools, 37% (71 schools) are failing to meet AYP standards. The group differences in the percentage of schools meeting the AYP–NCLB status model standard are both substantively and statistically significant.

A Value-Added Performance Measurement Strategy

First, the change from second- to fourth-grade math scores computed by schools was examined to determine if there was any relationship between the percentage African American in the school and change in math scores. The correlation is weak (.09) and not statistically significant. The correlation between literacy scores and percentage African American ($r = -0.16$), however, has both statistical and perhaps policy significance. Although a weak correlation, this means that as the percentage of African American students goes up, the improvement in literacy between second and fourth grade goes down.

In Table 3, we look at the mean change scores to determine if there are differences between majority African American schools and nonmajority African American schools. The mean change score in math for all schools was -6 , which represents a drop in math proficiency from second grade to fourth grade of six percentage points. The mean change for the majority African American schools was -3 and for nonmajority African American schools was -7 . The mean literacy change score for all schools was -16 . The mean for majority African American schools was -21 , and for nonmajority African American schools it was -15 .

For easy comparison with other performance measurement schemes, it is desirable at this point to have the ability to count how many African American school are meeting standards using this value-added scheme compared to other performance measurement schemes. However, this would require a normative decision process to set a standard that

Table 3
Value-Added Scheme—Mean Change Proficiency Score
From Second Grade to Fourth Grade

School Type	<i>n</i>	% Change Math (Second to Fourth)	% Change Lit (Second to Fourth)
All schools	229	−6	−16
Majority African American	34	−3	−21
Non–majority African American	195	−7	−15

Note: Cell entries are percentage point differences.

Table 4
Percentage of Schools Failing Using Value-Added Measure

School Type	<i>n</i>	Proficient in Math	Proficient in Lit	Proficient in Math and Lit
Majority African American	34	82	62	53
Non–majority African American	195	73	77	61

would indicate whether schools were acceptable after accounting for student poverty and other contextual factors. For this study, to propose such a standard would be too normative and beyond the scope of this article.

So, in Table 4, we focus on the lower quartile schools as one possible descriptive cutoff measure indicating “below acceptable.” Using this descriptive cutoff, 82% (28) of majority African American schools are above the lower quartile performance in math, and 73% (143) of the non–African American schools are above the lower quartile in math. Of the majority African American schools, 62% (21) are above the lower quartile performance in literacy and 77% (150) of the non–African American schools are above the lower quartile in literacy. Finally, when the criteria are that both math and literacy be above the lower quartile of school performance, majority African American schools succeed 53% of the time and nonmajority African American schools succeed 61% of the time. The differences are statistically significant.

A Contextualized Evaluation Strategy (Adjusted Performance Measure)

As a first step in creating the adjusted performance measure, a regression model was used to test the predictive power of external non-school-level policy variables in explaining math proficiency in the fourth grade. As anticipated, the percentage of students receiving free and reduced lunch was the most useful in predicting test scores (Gibson & Asthana, 1998; Hoerandner & Lemke, 2006; Payne, 2003). The estimates reported in Table 5 indicate that free and reduced lunch was the strongest predictor, with total enrollment explaining just a small portion of the variance in math scores. In a follow-up stepwise regression model, percentage free and reduced lunch explained 18% of the variance, whereas total enrollment explained far less (1% of the variance). An identical model was run, only this time with literacy

Table 5
Regression of Control Variables on Fourth-Grade Test Scores in 2004

Variable	Math Proficiency		Literacy Proficiency	
	Estimate	T-Score	Estimate	T-Score
Constant	78.77	8.92***	72.48	8.48***
Total enrollment (2002-2003)	-0.01	-1.84*	0.00	0.13
Property tax	0.00	0.01	-0.00	-0.25
Percentage free and reduced lunch	-0.49	-7.08***	-0.54	-8.12***
Urban-rural	-0.55	-0.84	0.10	0.15
Adjusted R^2		.18		.25
N		229		229

Note: ACTAAP = Arkansas Comprehensive Testing Assessment and Accountability Program. For the math proficiency model, the dependent variable is the ACTAAP score for mathematics. For the literacy proficiency model, the dependent variable is the ACTAAP score for mathematics.

* $p < .10$, one-tailed. *** $p < .01$, one-tailed.

scores as the dependent variable. In this case, the only statistically significant variable is the percentage of students eligible for free and reduced lunch funding. In a follow-up stepwise regression model, percentage free and reduced lunch explained 26% of the variance in literacy test scores.¹ Based on this analysis, the economic status of the students clearly is significant in explaining the variation of performance in math and literacy but does not explain the majority of the variation. Though interesting, this is but the first step in contextualizing test scores at the school level. The next step is to use the predicted scores for each school, which was derived from the regression model to adjust the test scores of each school.

Before proceeding with the construction of the adjusted performance measure, we examine unadjusted math proficiency rates. The unadjusted mean score for math proficiency in all schools in the sample reports 41% of all students proficient in math. The unadjusted mean math score for majority African American Schools shows 20% math competency, and for nonmajority African American schools the mean is 45%. Using test scores not adjusted for context, there is a very substantial difference between these two groupings of schools, with the group of majority African American schools lagging considerably behind.

The next step in the creation of the adjusted performance measure is to adjust the raw fourth-grade math scores for each school using the school's predicted or expected scores from the regression model above. In essence, the context is being taken into account when comparing schools. Returning to our primary focus, the question is whether adjusting for context, such as the extent of poverty in the school, results in more or less disparate outcomes for majority African American schools than nonmajority African American schools. Compared to each other, did these two groups of schools do better or worse than expected given the context? The percentages below are the average percentage point differences from the expected percentage of students who are proficient for both groups.

Adjusting for context, nonmajority African American schools are doing just a few percentage points worse on average than might have been expected from their context. The average for the nonmajority African American schools was two percentage points lower than their context would suggest. The majority African American schools have 11% more students

Table 6
Contextual Model (Adjusted Performance Measure)—Mean
Score Variance From Expectation

School Type	Variance From Expected Math	Variance From Expected Lit
Majority African American adjusted	11	12
Non-majority African American adjusted	-2	-2

Note: Cells are percentage point differences.

Table 7
Contextual Model (Adjusted Performance Measure):
Percentage of Schools Proficient

School Type	Schools Math Proficient	Schools Lit Proficient	Schools Proficient in Both
Majority African American adjusted	94	97	91
Non-majority African American adjusted	71	71	59

who are proficient than predicted by the model. Put simply, when adjusting for the economic circumstances of their students, majority African American schools outperform nonmajority African American schools on math test scores (see Table 6).

Repeating the same analytic process for literacy test scores, the mean unadjusted literacy proficiency score for all schools reports 40% of students as proficient. The unadjusted mean for majority African American schools reports 16% proficient, and the unadjusted mean for nonmajority African American schools shows 44% proficient. Using contextualized measures, the performance of majority African American schools compared to what their context would predict shows this group 12 percentage points higher in their performance than expected. The nonmajority African American schools report a mean two percentage points lower than expected for this group. So for both math and literacy, majority African American schools do better than the context would expect and, by this performance scheme, are “succeeding” at higher levels than are nonmajority African American schools.

Having noted the normative nature of any official standard, it is perhaps useful again to use the lower quartile schools on the contextualized measure as a descriptive cutoff for comparative purposes so that we can examine the number of majority African American schools succeeding or failing compared to other nonmajority African American schools. First, examining contextualized math scores (see Table 7), 94% (32) of the majority African American schools (34 total) are above the lowest quartile of performance when using contextualized measures. Of the nonmajority African American (195) schools, 71% (139) are above the lowest quartile of performance when using this contextualized measure. African American schools significantly outperform nonmajority African American schools in contextualized math scores using the lowest quartile as the descriptive cutoff. These differences are statistically significant.

Table 8
Percentage of Schools Not Meeting Standards by Scheme

	All Schools	Majority AA	Non-Majority AA
Adequate yearly progress (NCLB)	45	94	37
Value-added measure	41	47	39
Contextual measure	37	9	42

Note: AA = African American; NCLB = No Child Left Behind.

Turning now to literacy scores, 97% (34) of the majority African American schools (35 total) are above the lowest quartile of performance when using contextualized measures. Of the nonmajority African American (195) schools, 71% (138) are above the lowest quartile of performance for literacy scores when using this contextualized measure. African American schools significantly outperform nonmajority African American schools in literacy using the lowest quartile as the descriptive cutoff. These differences are statistically significant.

Finally, to avoid sanctions in the AYP system described above, *both math and literacy* must be above the standard, so we will now count schools that are above the lower quartile of the schools on both the contextualized math and literacy measure. Of the majority African American (34 total) schools, 91% (31) were above the 25% point. Of the non-African American majority schools (195 total), 59% (115) were above the lower 25% cutoff in math and/or literacy. When the criteria include both math and literacy scores above 25%, there is a significant drop in the percentage of nonmajority African American schools that meet the criteria. Again, differences are statistically significant.

Conclusion

Hershberg (2005) argues that the federal government should encourage states that use high-quality value-added or growth models to conduct pilot studies to determine the impact of a growth-to-standards approach. We believe this study affirms the value of such studies. Allowing that adjustments in each of the performance scheme cutoffs would create different proportions of success or failure, Table 8 summarizes the three methods and their divergent impact.

Majority African American schools, when compared to nonmajority African American schools, are most vulnerable to sanctions when using AYP criteria.

We believe that our analysis of AYP demonstrates that exclusively relying on NCLB-AYP standards to evaluate schools raises significant equity concerns. In this, we agree with many performance measurement scholars who feel that it is desirable to use multiple measures in assessing performance, proposing reforms, providing rewards, and administering sanctions. When compared to other performance measurement strategies, it is clear that the AYP model tends to identify schools with large African American enrollments at a much greater frequency than do the other two measurement strategies. When context is taken into account, majority African American schools tend to do somewhat better than non-African American schools. When the value-added measure is utilized, the ethnic composition of the school is

least important in predicting sanctions. Interestingly, the nonmajority African American schools fail at roughly the same rate in each of the three measurement schemes.

Given the way that AYP–NCLB identifies a strikingly larger number of African American schools than the other measures, a critical question is: Does the AYP–NCLB approach actually measure school performance? The literature and our analysis suggest that this status measure does not measure school performance but student performance. Because student baseline performance is so closely linked to student income status, schools will tend to be sanctioned if they have low-income African American students. This sanctioning will be the result of their student’s lower baseline status rather than the school’s ability to move students forward.

Apple (1996) believes that we are witnessing the revivification of more traditional class, gender, and especially race hierarchies. If one assumes (and this is a matter of ideological and political debate) that African Americans have a lower income because of a history of racism, then sanctioning African American schools because they are predominantly African American and low income may be using “the results of that [past] degradation to oppress them further” (DeClue, 1988, p. 4).

The AYP findings described above may not be new to decision makers. The Arkansas Supreme Court recently found the state of Arkansas in violation of its own constitution, arguing that the state did not provide “adequate” education to its children. The court raised concerns regarding equity (i.e., the distribution of resources across districts) as well. This resulted in a special legislative session charged, among other things, with increasing state funding for education. Closing some schools was suggested as a way of increasing efficiency and reducing the size of the budgetary increase required. It was suggested that status-based school performance measures be used to determine which schools to close.

The state assistant attorney general advised decision makers that a policy of closing a school and consolidating based on the school’s inability to meet certain status measurement standards would make the state vulnerable to a federal discrimination lawsuit under the Equal Protection Clause (Blomeley, 2003). Legislators interviewed postsession reported that their peers had backed away from such an approach because of fear of civil rights litigation. One legislator said,

We couldn’t take a standards-based approach because we would have shut down a bunch of small black schools and left untouched a bunch of small white schools. The state would have been in an incredibly vulnerable position if we had tried to do a full standards-based approach. (quoted in Thomas, 2006)

Consequently, school size rather than the AYP performance model became the criteria for school closings designed to cut costs (Belden, 2005; Thomas, 2006).

AYP standards

may induce states to violate the Equal Protection and Due Process rights of minority and disadvantaged students, or, at the very least, induce states to violate federal regulations prohibiting federal funding for programs with a “disparate impact” on minority students. (Nash, 2003, p. 239)

The press reports and interviews described above strongly suggest that legislative leaders understand that school performance measures are political and technical in their construction and impact. Measurement schemes determine which groups will be rewarded and which will

be sanctioned and often must include legal considerations and political ones. In one New York court, a plaintiff said,

We stand before the court . . . because the effect of the constitutional wrong visited upon the children of New York City is no less insidious than the harm that the Supreme Court condemned in *Brown* against the Board of Education. . . . In 1999 we remain a house divided. (Campaign for Fiscal Equity, 1999)

The measurement of intelligence, employment equity, poverty, representative bureaucracy, and housing desegregation are other examples of the very political nature of measurement in a society where issues of “race” are still very much present if not publicly discussed. Hoerandner and Lemke (2006), noting status measure challenges nationwide, argue that without additional state or federal funding, uniform standards will place unequal challenges and tax burdens on some localities.

Some would argue that the AYP–NCLB scheme is really best for minority students. Though the scheme labels many minority schools as failing, some argue that it will give African American students new choice options and that this result is the most beneficial outcome for individual minority students. Others argue that those minority schools labeled as failures will be disempowered by sanctions and that minority students will only be given the false hope that the market, this time, will serve them well (Marshall, 2004). We have found that although it can be argued that NCLB is designed to help individual minority students, it often tries to accomplish this by disproportionately sanctioning minority schools. Research such as ours helps make the case that as long as such dramatic race and class differences exist, performance measurement analysis will be incomplete without a thorough discussion of the equity implications of these measurement schemes.

Note

1. The R^2 level would be considered weak and subject to criticism for underspecification if the proposed model were attempting to capture all of the causal factors. It is important to remember that our purpose in this regression model, unlike most uses of regression models, is not to capture all of the variables that contribute to the prediction of test scores. We are concerned only with variables that are reasonably outside of the control of the school.

References

- Allen, R., Galinat, C., Leedham, C., & Leverich, J. (2004). *Value added assessment* (Research No. 2). Madison: Wisconsin Education Association Council Research Work Group.
- Apple, M. W. (1996). *Cultural politics and education*. New York: Teachers College Press.
- Ayers, W. (2000). The standards fraud. In J. Cohen & J. Rogers (Eds.), *Will standards save public education?* (pp. 64-69). Boston: Beacon.
- Balou, D. (2002). *Sizing up test scores*. Retrieved August 15, 2005, from <http://www.educationnext.org/20022/10.html>
- Belden, L. (2005). *Whose voices are heard at Arkansas' legislative table, and why?* Fayetteville: University of Arkansas.
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.

- Blomeley, S. (2003, November 4). School case lawyer set to sue again: Tiny Lake View district tired of waiting for state reforms. *Arkansas Democrat Gazette*, p. 1.
- Campaign for Fiscal Equity. (1999). Landmark state education financing case opens with shocking statistics of inadequacy in New York city schools. Retrieved August 12, 1999, from <http://www.cfequity.org/prday1.html>
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.
- Courty, P., Heinrich, C., & Marschke, G. (2005). Setting the standard in performance measurement systems. *International Public Management Journal*, 8(3), 321.
- DeClue, J. (1988). The St. Louis case: A personal view. *Equity and Choice*, p. 4.
- Farmer, C. J. (2005). The No Child Left Behind Act: Will it produce a new breed of school financing litigation? *Columbia Journal of Law and Social Problems*, 38(4), 443.
- Ferguson, R. F. (1998). Can schools narrow the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 318-374). Washington, DC: Brookings Institution.
- Frederickson, H. G. (1990). Public administration and social equity. *Public Administration Review*, 50(2), 228.
- Gibson, A., & Asthana, S. (1998). School performance, school effectiveness and the 1997 white paper. *Oxford Review of Education*, 24(2), 195.
- Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the Black-White score gap narrow in the 1970s and 1980s? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 182-228). Washington, DC: Brookings Institution.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297.
- Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209.
- Hershberg, T. (2005). Value-added assessment and systemic reform: A response to the challenge of human capital development. *Phi Delta Kappan*, 87(4), 276.
- Hess, F. M., & Finn, C. E. (2004). *Leaving no child behind? Options for kids in failing schools*. New York: Palgrave Macmillan.
- Hess, F. M., & Petrilli, M. J. (2006). *No Child Left Behind primer*. New York: Peter Lang.
- Hoerandner, C., & Lemke, R. (2006). Can No Child Left Behind close the gaps in pass rates on standardized tests? *Contemporary Economic Policy*, 24(1), 1-17.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution.
- Kelly, P., & Miller, W. (1989). Assessing desegregation efforts: No "best measure." *Public Administration Review*, 49(5), 431.
- Kirsanow, P. N. (2002). *Civil rights and wrongs: A new agenda for a new era* (Heritage Lecture 731). Retrieved June 2, 2006, from <http://new.heritage.org/Research/LegalIssues/HL731.cfm>
- Kozol, J. (2000). *Ordinary resurrections: Children in the years of hope*. New York: Crown.
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: Harvard University, Civil Rights Project.
- Levin-Waldman, O. M. (2004). Policy orthodoxies, the minimum wage, and the challenge of social science. *Journal of Economic Issues*, 38(1), 139.
- Manno, B. V. (1995). *In the image of the great society: Reinventing K-12 education in the Clinton administration*. Washington, DC: Hudson Institute. <http://www.edexcellence.net/foundation/publication/publication.cfm?id=214>
- Marshall, C. (2004). Social justice challenges to educational administration: Introduction to a special issue. *Educational Administration Quarterly*, 40(1), 3.
- McDermott, K. A. (2007). "Expanding the moral community" or "blaming the victim"? The politics of state education accountability policy. *American Educational Research Journal*, 44(1), 77.
- McEwen, N. (1995). Educational accountability in Alberta. *Canadian Journal of Education*, 20(1), 27.
- Meier, D., Cohen, J., & Rogers, C. (2000). *Will standards save public education?* Boston: Beacon.
- Meier, K. J., Polinard, J. L., & Wrinkle, R. (2000). Bureaucracy and organizational performance: Causality arguments about public schools. *American Journal of Political Science*, 44, 590-602.
- Myers, S. L., Jr. (2002). Presidential address—Analysis of race as policy analysis. *Journal of Policy Analysis and Management*, 21(2), 169.

- Nash, D. (2003). Improving No Child Left Behind: Achieving excellence and equity in partnership with the states. *Rutgers Law Review*, 55(1), 239.
- National Conference of State Legislatures. (2005). *Task force on No Child Left Behind: Final report*. Washington, DC: Author.
- Nicholson-Crotty, S., Theobald, N. A., & Nicholson-Crotty, J. (2006). Disparate measures: Public managers and performance-measurement strategies. *Public Administration Review*, 66(1), 101.
- Ogbu, J. U. (1978). *Minority education and caste: The American system in cross-cultural perspective*. New York: Academic Press.
- Orfield, G. (2005). *No Child Left Behind: A faculty response to President Bush's education bill*. Retrieved June 5, 2006, from <http://gseweb.harvard.edu/news/features/orfield09012002.html>
- Paul, D. G. (2004). The train has left: The No Child Left Behind Act leaves Black and Latino literacy learners waiting at the station. *Journal of Adolescent & Adult Literacy*, 47(8), 648.
- Payne, R. K. (2003). *A framework for understanding poverty* (3rd rev. ed.). Highlands, TX: aha! Process.
- Pidd, M. (2005). Perversity in public service performance measurement. *International Journal of Productivity and Performance Management*, 54(5/6), 482.
- Powell, A. G., Farrar, E., Cohen, D. K., National Association of Secondary School Principals, & National Association of Independent Schools, Commission on Educational Issues. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.
- Ravitch, D. (2000). *Left back: A century of failed school reforms*. New York: Simon & Schuster.
- Rubenstein, R., Schwartz, A. E., & Stiefel, L. (2003). Better than raw: A guide to measuring organizational performance with adjusted performance measures. *Public Administration Review*, 63(5), 607.
- Saunders, L. (1999). A brief history of educational "value added": How did we get to where we are? *School Effectiveness and School Improvement*, 10(2), 233-256.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2/3), 277.
- Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.
- Thernstrom, A. (2000). No excuses. In J. Cohen & J. Rogers (Eds.), *Will standards save public education?* (pp. 35-39). Boston: Beacon.
- Thernstrom, A. M., & Thernstrom, S. (2003). *No excuses: Closing the racial gap in learning*. New York: Simon & Schuster.
- Thomas, S. (2006). *Legislative decision making in the policy making process: Public school consolidation and the 2003 special session of the 84th Arkansas general assembly*. Unpublished doctoral dissertation, University of Arkansas, Fayetteville.
- U.S. Census Bureau. (2002). *2000 census of population and housing, summary population and housing characteristics, PHC-1-5, Arkansas*. Washington, DC: Author.
- Watson, D. (n.d.). *Restoring public education's academic mission: High expectations, academic standards, proven methodologies and curriculums*. Retrieved February 26, 2007, from <http://www.reformarkansas.org/policy/restoring.htm>

William H. Miller is an associate professor of public administration and chair of the public administration department at the University of Illinois, Springfield. His research interests include public policy analysis, education policy, minority policy, and policy research methods.

Brinck Kerr is a professor of political science and director of the Public Policy PhD Program at the University of Arkansas, Fayetteville. His research interests include ethnic and gender employment patterns at the subnational level, education policy, and inequalities in resource distribution.

Gary Ritter is an associate professor of education and public policy and holder of the Endowed Chair in Education Policy in the Department of Education Reform at the University of Arkansas. His research interests include standards-based and accountability-based school reform, racial segregation in schools, volunteer tutoring programs, program evaluation, the impact of preschool care on school readiness, and school finance.