education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal

epaa aape

Arizona State University

Volume 18 Number 21 September 10th, 2010 ISSN 1068–2341

The Black-White Achievement Gap Revisited

Henry Braun
Boston College

Lauren Chapman
Boston College

Sailesh Vezzu
Educational Testing Service

Citation: Braun, H., Chapman, L., Vezzu, S. (2010). The Black-White achievement gap revisited *Education Policy Analysis Archives,18* (21). Retrieved [date], from http://epaa.asu.edu/ojs/article/view/772

Abstract: This study examines trends in Black student achievement and in the Black-White achievement gap over the period 2000 to 2007, employing data from ten states drawn from the NAEP Grade 8 mathematics assessments. Results are obtained for three levels of aggregation: the state, school poverty stratum within the state, and schools within poverty stratum. In addition, information on the ten states' education policies for the period 1998 to 2005 was compiled. States relative ranks on the overall strength of their reform efforts were compared to their relative ranks with respect to their success in improving Black student achievement and in reducing the Black-White achievement gap. This study constitutes an extension of earlier work that considered the same issues for the period 1992 to 2000 and, thus, offers a unique comparison between the pre-NCLB era and the present one. Although the ten states certainly differed in their outcomes, the general picture at all three levels of aggregation is quite clear: The achievement gaps are substantial and the introduction of federally mandated high stakes test-based accountability through No Child Left Behind has had a very modest impact on the rates of improvement for Black students and on the pace of reductions in the achievement gaps between Black students and White students. Moreover, there was only a weak association between states' policy rankings and their rankings related to test

results. It appears there is a need for both fresh thinking on education reform and a more concerted effort to collect comprehensive longitudinal information on states' education policies.
Keywords: NAEP; NCLB; achievement gap; state education policies; hierarchical analyses.

Revisitando la brecha en el rendimiento de estudiantes negros y blancos

Resumen: Este estudio examina las tendencias en el rendimiento académico de estudiantes negros y la brecha en el rendimiento de los estudiantes negros y blancos en el período 2000 a 2007, utilizando datos de las evaluaciones NAEP de Matemáticas para el octavo grado en diez estados. Se obtuvieron resultados para tres niveles de agregación: estatal, estrato de pobreza de la escuela en el estado, y escuelas dentro del estrato de pobreza. Además, se compiló la información sobre las políticas de educación en los diez estados para el período 1998 a 2005. El ranking relativo de la fortaleza de los esfuerzos de reforma de los estados se comparó con las mejoras de rendimiento de los estudiantes negros y con la reducción de la brecha de logros académicos de los estudiantes negros-Blancos. Este trabajo es una continuación de trabajos previos que consideraron los mismos temas para el período 1992 a 2000 y, por tanto, ofrece una única comparación entre la era pre-NCLB y la actual. Aunque los diez estados difieren en sus resultados, el cuadro general en los tres niveles de agregación es bastante claro: las brechas académicas son importantes y la introducción del mandato federal "Que Ningún Niño Se Quede Atrás (por sus siglas en inglés NCLB) con sus graves consecuencias a través de pruebas estandarizadas ha tenido un impacto muy modesto en las tasas de mejora para los estudiantes negros y en el ritmo de la reducción de la brecha de rendimiento entre estudiantes negros y blancos. Por otra parte, sólo se encontró una débil asociación entre el ranking de los estados y sus clasificaciones relacionadas a los resultados de las pruebas académicas. Al parecer, hay una necesidad de que pensar ideas nuevas sobre reforma educativa y un esfuerzo más concertado para reunir datos longitudinales completos e información sobre las políticas educativas de los estados.
Palabras clave: NAEP; NCLB; brecha en el rendimiento académico; políticas estatales de educación; análisis jerárquico.

Revisitando a defasagem de rendimento academico dos estudantes negros e brancos

Resumo: Este estudo analisa as tendências no desempenho acadêmico dos alunos negros e a defasagem de rendimento academico para estudantes negros e brancos no período 2000-2007, usando dados das avaliações do NAEP de matemática para a oitava série em dez estados. Os resultados foram obtidos para três níveis de agregação: o estado, a estratificação da pobreza na escola no estado, e escolas dentro do estrato da pobreza. Além disso, nós compilamos informações sobre as políticas de educação em dez estados, para o período 1998-2005. O ranking da força dos esforços de reforma dos estados em relação às melhorias desempenho do aluno negro e reduzindo a abertura da realização para os estudantes negros e brancos. Este trabalho é uma continuação de estudos anteriores que encontraram as mesmas questões para o período 1992-2000 e, portanto, oferece um único comparação entre o pré-NCLB e atual. Enquanto os dez estados, diferem em seus resultados, o quadro geral em todos os três níveis de agregação é bastante clara: as defasagems são importantes e conquista a introdução do mandato federal No Child Left Behind (por sua sigla em Inglês NCLB), com graves conseqüências por meio de testes padronizados teve um tremendo impacto melhoria modesta das taxas para estudantes negros e do ritmo de redução da diferença de realização entre estudantes negros e brancos. Além disso, apenas uma fraca associação foi encontrada entre o ranking de estados e suas classificações relacionadas com os resultados dos ensaios acadêmicos. Aparentemente, há uma necessidade de novas ideias para pensar reforma educacional e um esforço mais concertado para recolher informações e dados longitudinais

completos sobre as políticas educacionais dos estados.
Palavras-chave: NAEP; NCLB; defasagem de rendimento academico; políticas de educação do estado; análise hierárquica.

Introduction[1]

The present focus on improving the productivity of the nation's schools has its origins in the public debate initiated by the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983). Arguably, concerns about the achievement gap between White students and Black students assumed political significance with the passage of the Elementary and Secondary Education Act in 1964. In principle, states can track their success in raising achievement overall and, specifically, their progress in reducing the achievement gap. However, the quality of the states' tests, their tendency to change tests every few years and the phenomenon of score inflation (Linn, 2000; Koretz, 2003) undermines the utility of the statistics they report. Moreover, since states are free to adopt whatever testing system they deem appropriate to their needs, comparisons among states using their own test results are essentially impossible (Feuer, et al. 1999).

Consequently, most investigators turn to the National Assessment of Educational Progress (NAEP) for credible data on student achievement that can be used both to track progress over time and to make comparisons among states. In 4th grade mathematics, for example, NAEP trends show steady improvement from 1990 to 2007: The median score increased from 214 to 242.[2] In fact, 242 is above the 80th percentile in the 1990 NAEP score distribution. Indeed, there was a substantial increase at all percentiles over this time period. A similar story holds for 8th grade mathematics: The median score increased from 264 to 283 over the period 1990 to 2007. In this grade, 283 corresponds approximately to the 70th percentile in the 1990 NAEP score distribution. Again, there was a substantial increase at all percentiles over this time period.

These aggregate score gains were mirrored by the gains made by sub-populations classified by race/ethnicity. In particular, at the 4th grade, the Black-White score gap decreased from 32 points in 1990 to 26 points in 2007, although the gap has been essentially constant since the 2003 administration.[3] By contrast, at the 8th grade, the Black-White score gap changed from 33 points to 32 points over the period.[4] To put this difference in perspective, in 2007 the average score in the highest performing state (Massachusetts) was 298 and the average score in the lowest performing state (Mississippi) was 265 – a difference of 33 points. Thus, the score gap not only appears to be large, but is rather substantial when compared to score differences among states.[5] The Hispanic-White gap in 8th grade mathematics followed a similar trend and in 2007 was slightly smaller –

[1] The research reported here was supported by a National Assessment of Educational Progress Secondary Analysis Grant (R902B070004) from the National Center for Education Statistics, Institute of Education Sciences. The authors would like to thank Bruce Kaplan and Aubrey Wang for their assistance in the early stages of this work and Liz Brophy for her help in preparing the final version of the manuscript. The authors also benefitted from the comments of the editor and the reviewers.

[2] Some care is needed in making score comparisons over such a long period. Strict maintenance of the score scale as assessments tasks are removed and replaced is difficult to accomplish. Moreover, since 1996 NAEP has permitted accommodations and this resulted in a slight scale shift from the one established in 1990 with the first "Trial State Assessment".

[3] The score gap is computed as the difference in the average scores for the two groups.

[4] Actually, the gap increased to about 40 points in 1996, remained there in 2000 and fell to 35 points in 2003.

[5] Note that 40 points is approximately the estimated difference in population means in mathematics between 4th and 8th grades.

approximately 26 points. For more statistics see National Center for Education Statistics (2007a). The Black-White gap in 8th grade mathematics is the focus of this study.

Concerns regarding the magnitude and persistence of the achievement gap have economic, moral and political dimensions. That a large and growing proportion of the population (i.e. disadvantaged minorities) appears to have, on average, substantially lower skills than the majority group raises worries both about global economic competitiveness and the equitable distribution of economic opportunity and social mobility (Kirsch, Braun, Yamamoto & Sum, 2006). The implications for the fabric of democratic society have also been addressed (Friedman, 2005). Indeed, one of the arguments for passage of the No Child Left Behind Act (NCLB) was that it would lead to a dramatic reduction in the achievement gaps.

Through the 1990s, various states adopted different education reform strategies, with the goal of improving achievement overall and reducing achievement gaps. For the most part these strategies were based on a model of standards-based reform and test-based accountability (O'Day & Smith, 1993). In many cases, accountability involved meaningful rewards and sanctions for schools. The adoption of "high-stakes testing" was controversial and numerous studies investigated the question of whether this policy was effective (Amrein & Berliner, 2002; Braun, 2004; Nichols, Glass & Berliner, 2006). The findings were not definitive, to say the least.

Amrein & Berliner analyzed state NAEP reading and mathematics data from the 1990s and concluded that once differences in exclusion rates were taken into account, there was no consistent evidence of a relationship between the imposition of high stakes testing and improvement in NAEP scores. Braun (2004) was critical of the methodology employed by Amrein & Berliner (2002) and, adopting a comparative approach, found weak evidence of a relationship for 8th grade mathematics (he did not analyze NAEP reading scores). However, in a supplementary investigation employing a form of cohort analysis, he found little evidence of a relationship. Subsequently, Nichols, et al. (2006) introduced a more sophisticated measure of test-based accountability, denoted as the Pressure Rating Index (PRI), and correlated lagged changes in the PRI with changes in NAEP scores. They found weak evidence of a positive relationship in 4th grade mathematics, but neither for 8th grade mathematics nor for reading in 4th or 8th grades.

Braun, Wang, Jenkins, and Weinbaum (2006) argued that most policy studies in this area suffered from one, if not two, serious flaws: (i) They anchored their comparisons on differences among states in testing policies, ignoring other reform initiatives undertaken by the states and (ii) They based their analyses on a one-time snapshot of states' policy profiles rather than adopting a longitudinal perspective that is more appropriate given the lags between policy adoption and implementation to scale and impact in the field. Accordingly, it was not surprising that different teams of investigators reached different conclusions. Another difficulty is that most states are quite heterogeneous with respect to student achievement, so that positive results in one sector (e.g. higher poverty schools) could be masked by negative results in another sector (e.g. lower poverty schools).

Braun et al. (2006) attempted to address these deficiencies by conducting a multi-dimensional, longitudinal education policy analysis for ten states over the period 1988 to 1998. This was coupled with an in-depth investigation of patterns in Black student and White student achievement using the 8th grade NAEP mathematics assessments administered in 1992 and 2000. Statistical analyses were conducted at three levels: the state, school poverty stratum within state, and school within stratum and state. At the last stage, the policy analysis was linked to the NAEP results in order to examine the association between differences in policy and differences in student outcomes.

Braun et al. (2006) found a modest relationship between the relative strength of a state's overall policy efforts and its success in improving Black student achievement. There was a weaker

relationship between policy efforts and success in reducing the Black-White achievement gap. On the basis of their research, Braun et al. (2006) characterized the Black-White achievement gap as pervasive, profound and persistent: Pervasive because the gap was observed in all ten states, profound because it was found at all levels of aggregation, and persistent because the median gap did not diminish over the period of the study.

With the passage of NCLB in late 2001, the education policy landscape has changed dramatically as states have struggled to interpret, and then comply with, the requirements of the Act. The question of the moment – in view of the pending reauthorization of ESEA – is whether the implementation of NCLB, along with state policies, has had a salutary effect on student achievement and, in particular, on the reduction of achievement gaps. For technical reasons, it is essentially impossible to isolate the contribution of NCLB since its impact is confounded with other initiatives undertaken by the states. One could argue, in fact, that this would be beside the point, as the implicit "theory-of-action" behind NCLB was to focus the attention of states, districts and schools on student achievement.[6]

However, one can pose a related set of questions: How does the progress (or lack of same) made by a state in the era of NCLB compare to its progress (or lack of same) in the pre-NCLB era? During the period 2000 to 2007, how similar were the states with respect to their success in increasing Black student achievement and in reducing the achievement gap?

To be truly useful to policy makers, the answers to both questions should be structured at different levels of aggregation. For example, the fact that the achievement gaps in both State A and State B have been reduced by six points from 2000 to 2007 may obscure important variations in both school demographics and student performance within each state. A more pointed and relevant comparison would be based on the observation that among high poverty schools in State A the average achievement gap between Black students and White students attending the same schools fell by eleven points, while in State B that same statistic fell by only one point.

Another, equally important, question is: For the period 2000 to 2007, what is the relationship between differences in state education policies and differences in their success in improving Black student achievement, as well as in reducing achievement gaps? Although policy makers are interested in "what works", patterns of association derived from such observational data can only suggest possible causal linkages. Nonetheless, observed patterns can be informative from both substantive and methodological perspectives.

The present study has been designed to address these questions. It comprises: (i) An extension of the longitudinal policy analysis for the ten states in the earlier study to cover the period 1998 to 2005, (ii) A comprehensive, multi-level analysis of data from the 2000 and 2007 NAEP 8th grade mathematics assessment, and (iii) The linking of the policy analysis and the quantitative analysis of student achievement on NAEP.

Thus, the study allows a comparison of trends in achievement from 1992 to 2000 and those observed from 2000 to 2007. Of course, the latter time period begins just before NCLB and ends some five years after its implementation. This should provide sufficient time for a meaningful comparison of the overall impact of education policies between the two periods.

The article proper is organized into eight sections. Following the literature review, there are sections on the data analysis strategy and the multi-level models employed in the quantitative analysis. The next section describes the state policy analysis and this is followed by two sections containing the quantitative results, descriptive and model-based. The final two sections present the results of linking the policy analysis and the quantitative analysis and a short conclusion.

---

[6] An interesting issue is the extent to which the advent of NCLB has led to greater homogenization of education policies across states.

Review of the Literature

The persistence and magnitude of the Black-White achievement gap continues to attract the attention of both researchers and policy makers. The relationship at the individual level between educational attainment and achievement on the one hand and life course (employment, wages, civic engagement, etc.) on the other, has been well documented (See Kirsch, Braun, Yamamoto, and Sum (2006) and the references therein). The societal costs associated with inadequate education have also been calculated, as have the benefit/cost ratios of various interventions (Levin, 2009). Thus, describing the achievement gaps from different perspectives and, especially, examining the relationships between policy initiatives and trends in the sizes of the gaps remains an important focus for research.

Lee (2007) conducted a wide-ranging study of achievement gaps using both NAEP and state assessment data. Among other analyses, he considers trends at the national level in the Grade 8 Black-White achievement gap in mathematics from 1992 to 2005. The gap increased from 1990 to 2000 and then decreased until 2005 (Figure 6.4). By 2005 the gap of approximately 33 points was nearly identical to what it had been in 1990. More to the point, the decrease began just as NCLB became law (and, arguably, could not have had any impact) and the rate did not accelerate with succeeding administrations. Magnuson and Waldfogel (2008) used the Long-term Trend component of NAEP to track the Black-White achievement gap in mathematics for thirteen-year-olds. The gap increased steadily from 1986 (24 points) to 1999 (32 points) and then fell to 27 points in 2004. Since most states had not fully complied with the mandates of NCLB by 2004, it would be speculative to attribute the decline to its passage.

Explanations for the existence of the gap and possible remedies abound. Data from the Early Childhood Longitudinal Study-Kindergarten Cohort (National Center for Education Statistics, n.d.) document the gap as children enter school. Hanushek and Rivkin (2006), using the same data base, attribute the growth of the gap over time to various school factors such as levels of segregation and mobility, as well as teacher quality. A resource-based argument is also propounded by Harris and Herrington (2006) who conclude that the positive impact of high-stakes accountability has been oversold. A sobering analysis is presented by Raudenbush (2009), who amasses evidence that increasing the amount and quality of schooling can be expected to reduce – but not eliminate – the achievement gap. Morris and Monroe (2009) in a study of the achievement gap in the U.S. South emphasize the importance of taking into account geographical and historic context, as well as socio-economic factors and school culture, in understanding the sources of the gap and how best to ameliorate it.

With the passage of NCLB, there have been a number of analyses to ascertain its impact. As indicated above, there is little evidence that implementation of NCLB has caused a shift in the trends that were evident prior to 2001. Lee (2007) has shown that reductions reported on state tests are typically not replicated in states' NAEP results. There is added confusion because in some of the literature, the achievement gap indicator is the difference in the percent proficient rather than the difference in means. The former is technically problematic because it is highly dependent on the location of the proficiency standard in relation to the two score distributions (Ho, 2008).

Much of the criticism of NCLB stems from the perceived unfairness of using a status measure to evaluate school effectiveness. Concerns have also been raised regarding unintended consequences such as the narrowing of the curriculum, teaching to the test and focusing attention on those students whose perceived score is near but below the proficiency standard. Although these are important and serious issues, they are not central to the thrust of this paper. The focus here is on whether there is evidence, at some level of aggregation, that there have been reductions in the

achievement gaps and that the magnitude of these reductions is positively correlated with the strength of states' reform efforts. If such is the finding, then a debate over whether the benefits of high stakes accountability outweigh the costs and unintended consequences can be initiated. In the absence of such evidence, however, the burden must fall on proponents of test-based accountability to justify the continuation of this policy.

Data and Analysis Strategy

The ten states examined here were chosen for the original study based on their participation in NAEP, the sizes of their Black student population and their varying approaches to education reform. Together they enroll nearly half of all Black students in U.S. public schools.

The data for the analyses reported here are derived from two NAEP restricted use data sets obtained from the National Center for Education Statistics. The data are derived from the NAEP administrations in the years 2000 and 2007. It should be noted, however, that the results reported for the year 2000 in Braun et al. (2006), are based on a different data set than the one employed in this study. The explanation is that two parallel assessments were conducted in 2000. In one, accommodations were not permitted, while in the other accommodations were permitted. The resulting data sets are denoted as R1 and R2, respectively. In order to maintain compatibility with the 1992 data (for which no accommodations were permitted), Braun et al. (2006) used the R1 data. Similarly, to maintain compatibility with the 2007 data (for which accommodations were permitted), this study uses the R2 data. The student samples drawn for the two assessments in 2000 are statistically equivalent and, consequently, there are minimal differences between them. Thus, comparisons of the changes between the earlier period (1992 to 2000) and the later period (2000 to 20007) are appropriate and meaningful.

The data sets used here contain individual level records with school and student information. These are necessary to compute the disaggregated descriptive statistics and to fit the multi-level models presented below. As was the case in Braun et al. (2006), schools in the NAEP sample have been categorized as lower poverty (Stratum 1 – S1) or higher poverty (Stratum 2 – S2), according to whether the percentage of enrolled students eligible for free or reduced price lunch were lower or greater than fifty percent. The categorization was based on auxiliary data collected by NAEP. In cases where the data was missing, comparable information from the 2005 Common Core of Data (CCD) was employed. In all, only five schools had to be eliminated from the analysis because they could not be categorized. For the multi-level analyses, the data set was edited to include only those schools with at least one White student and one Black student in the NAEP sample. In what follows, this data set is referred to as the "minimal reduced sample".

The quantitative analysis proceeds in two phases. The first phase employs simple weighted averages and the associated standard errors to describe and compare student achievement at three levels of aggregation: Among states, between strata within states and across states within strata, and between student groups within strata. The analysis proceeds from higher to lower levels of aggregation. The advantage is that we are then able to interpret the magnitudes of the Black-White achievement gaps in relation to gaps at other levels.

The second phase of analysis employs hierarchical linear models (HLM) to obtain alternative estimates of achievement gaps.[7] The rationale for this approach is founded on the recognition that the estimate of the Black-White achievement gap in a given year based on data from students within a stratum in a particular state confounds score differences among student groups with average score

[7] The methodology of fitting HLM is discussed in Raudenbush and Bryk (2002). The specifics of the current application are presented at greater length in Braun et al. (2006).

differences among schools. That is, Black students and White students are generally not proportionately distributed across schools in a stratum. If (say) Black students are more likely to be enrolled in schools with lower overall levels of achievement, then the usual estimate of the achievement gap will reflect, to some extent, such differences among schools.

To gauge the degree of confounding, it is possible to obtain a pooled estimate of the within school achievement gap. This involves locating the relevant data from each school and then aggregating across schools, and is accomplished most easily by fitting a suitably defined HLM. The resulting estimate can then be compared to the estimate obtained in the first phase. Further, using an expanded model, one can also estimate a pooled within school achievement gap after accounting for differences among students within each school on measured individual level characteristics associated with academic achievement. Note that only schools with at least one Black student and at least one White student in the NAEP assessed student sample from that school can contribute data to the estimation of the within school achievement gap. Depending on the stratum and state, the number of such schools may be substantially smaller than the total number of schools in the corresponding NAEP school sample.

Generating Estimates

NAEP is a complex assessment initiative. One aspect of this complexity is reflected in the design of the student sample, which is a stratified, cluster sample. Schools are stratified on a number of characteristics and probability samples of schools are drawn from each stratum. A random sample of students is then selected from each school. Because of this complexity, all reported descriptive statistics are computed using sampling weights that take account of unequal probabilities of selection, as well as school non-response. This yields approximately unbiased estimates of the corresponding population quantities.[8] In addition, standard errors are calculated using a variant of the jackknife procedure (Burke and James, 1997), to properly reflect different sources of variation.

Another aspect of complexity is the structure of the assessment itself. Because of the size of the item pool for any one assessment, each student takes only a small fraction of the pool. Accordingly, NAEP does not try to estimate a score for an individual student. Rather, its goal is to estimate how a group of students would have performed had they been exposed to the entire item pool. This is accomplished by generating a family of proficiency distributions for each student, based on the cognitive data provided by the student, in addition to a large number of characteristics drawn from the student, teacher and school questionnaires. Five members of this family are selected at random and one outcome randomly generated from each member. These five outcomes are termed "plausible values" or PV. See Allen, Johnson, Mislevy, and Thomas (1999) for more details.

Descriptive analyses (Phase 1) are conducted five times, once for each set of PV and the results averaged to obtain an estimate of the target population quantity. Differences in the results across the five analyses are used to obtain an estimate of measurement error that is combined with an estimate of sampling error to provide an estimate of the total error of the statistic. Again, appropriate sampling weights are used throughout.

Multi-level analyses (Phase 2) were carried out with a software package HLM6.06 (Raudenbush, Bryk, Chong, and Congdon, 2004).[9] As indicated above, we employed two different models. The first, simpler model can be represented as:

---

[8] The claim of unbiasedness rests on some strong assumptions, which are subject to some skepticism. See Braun and Qian (2008) for more details.

[9] HLM6.06 has an option that is adapted to the NAEP data structure; that is, it automatically carries out five analyses for the five sets of plausible values and reports the averages of the five analyses. The program also

$$y_{ij} = B_{oj} + B_1(BvsW)_{ij} + B_2(OvsW)_{ij} + e_{ij}$$
$$B_{oj} = B_{00} + r_{oj} \qquad (1)$$

where

i indexes students in school j
y = plausible value
B_{0j} = intercept for school j
B_1 = pooled within school Black-White achievement gap
B_2 = pooled within school Other-White achievement gap
BvsW = indicator for Black-White (only) contrast
OvsW = indicator for Other-White (only) contrast
e = student level residual term, assumed to be normally distributed with mean zero and common variance, independent across students
B_{00} = average intercept across schools
r = school level residual term, assumed to be normally distributed with mean zero and common variance, independent across schools.

The first equation in model (1) is the school specific regression and the second represents the school specific intercepts as drawn from a common (normal) distribution. BvsW equals 1 if the student is Black and 0 otherwise. OvsW equals 1 if the student is neither Black nor White and 0 otherwise. Interest focuses on the estimate of B_1, which is an estimate of the average difference between Black students and White students attending the same school, pooled across schools in the stratum. This is referred to below as an adjusted estimate of the achievement gap.

To obtain a fully adjusted estimate of the achievement gap, we fit model (2):

$$y_{ij} = B_{0j} + B_1^*(BvsW)_{ij} + B_2(OvsW)_{ij} + \mathbf{C}^*\mathbf{X}_{ij} + e_{ij}$$
$$B_{0j} = B_{00} + r_{0j}. \qquad (2)$$

Here, all recurring terms are defined as in model (1). Further, **C** is a vector of regression coefficients and **X** is a vector of 12 indicator variables related to four student characteristics: home environment, parental education, school lunch eligibility, and number of school days absent. For more discussion of multi-level modeling in this context, consult Braun et al. (2006). More details on the variables in **X** can be found in Appendix A.

Approach to State Policy Analysis

The policy analysis conducted for this report provides a description of the education policy efforts undertaken by each of the ten states over the period 1998 to 2005. The methodology was modeled on that of Braun, et al. (2006), which employed five policy levers (governance, education finance, curriculum and standards, teacher quality, and assessment and accountability) and constructed state profiles describing the initiatives undertaken by each state with respect to those levers. Comparisons among states were based on these profiles.

---

accommodates the use of weights. In view of the design of the NAEP sample, we employed school weights but not student weights.

For this report, state profiles were developed with the aid of archival data, obtained principally from the states' department of education websites. Additional information was culled from reports issued by the National Center of Educational Statistics (NCES), as well as from such education policy groups as the Thomas B. Fordham Institute and the Institute for a Competitive Workforce. Finally, the CCD provided certain state-level statistics. Unlike the effort documented in Braun et al. (2006), no interviews were conducted with individuals in each of the states. This constrained the amount of information available with respect to politics and governance. Furthermore, there was limited data on specific issues such as out-of-field teaching. Nevertheless, the profiles provide a concise and reasonably comprehensive picture of the education policy efforts (1998 to 2005) in the ten selected states. A sample state profile is provided in Appendix B. The other nine profiles are available upon request from the authors.

The completed state profiles were used to develop state rankings on four of the five policy levers (governance was not included in the rankings). The construction of these rankings also relied on information from the annual "Quality Counts" reports issued by Education Week, both as an additional source of information and as a check on our judgments. The logic behind our rankings is straightforward: We presume that a state's efforts in education reform are intended to improve the productivity of the education system, broadly defined. Such efforts include greater allocation of resources, strengthening regulations and the imposition of an accountability system with meaningful consequences. The greater the effort, as best we could tell from the data available, the higher the ranking we awarded. That said, we remain agnostic as to whether such efforts, singly or in combination, would result in desirable outcomes. Indeed, that is the point of the study – to determine whether states that are more highly ranked were more likely to achieve greater success on two outcomes: Raising Black student NAEP scores and/or reducing the Black-White achievement gap on NAEP. We recognize that the observed relationships are complicated by the fact that the policy intent and political calculations, as well as the quality and scope of the implementation, varied considerably across the ten states. The specifics regarding how the rankings were determined are presented in the remainder of this section.

The four policy lever ranking documents comprise tables that display information about each of the ten states, followed by a classification of the states into one of three categories: 'Leader', 'Middle Ground', and 'Laggard' (These titles were adapted from those used in the report of the Institute for a Competitive Workforce, 2007). A rationale accompanies each state's classification for each policy lever. The four ranking documents can be found in Appendices C – F. In view of the nature of the policy data we employed and the qualitative judgments we applied, we chose to group states into ordered categories rather than to assign a numerical index to each state. The latter would give the impression of greater accuracy than we could justify. At the same time, grouping states in ordered categories for both policy and outcomes still allowed us to examine patterns of association between the two.

In general, states' policies became more homogeneous over the period of interest, in part, because of the mandates of NCLB. For example, all ten states had developed standards for mathematics by the year 2005. This convergence made it more difficult to distinguish among the states. For the present study, the focus was on documenting the changes in policies and features of implementation over the period of interest. The 1998 through 2005 time period saw changes in education policies in all ten states, though the period was generally characterized more by variation in the scope and timing of specific policies, as opposed to differences in the policies themselves. Thus, while the questions posed were similar to those in Braun et al. (2006), we searched for nuances in how the states addressed education policy issues in a climate of growing accountability and increased performance expectations.

Bearing this in mind, we identified the following questions of interest:

*Education Finance***:** How has each state's proportional contribution to the cost of public elementary and secondary education changed over the selected time period, in comparison with other states in the study? How do each state's trends in total expenditures per pupil compare to the other states over the examined time period? What was the degree of financial equity (between the highest and lowest poverty districts) in the state?
*Curriculum and Standards***:** Were any new or revised curriculum/standards developed during the time period? What were the textbook adoption procedures in place? How did external policy groups rate the states' standards?
*Teacher Quality***:** Were there changes to the states' teacher licensure requirements during the examined time period? Was professional development treated as a priority in the state's teacher-related policies? What was the trajectory of each state's average teacher salary over the selected time period, in comparison with other states in the study? Were teachers encouraged to seek National Board Certification (as demonstrated by the growth in number of certified teachers during the examined period)?
*Assessment and Accountability***:** Were there changes to the state's accountability system during the examined time period? Did the state have a commitment to assessing their students in the subject of mathematics during the time of interest? Did the state make evident an effort of establishing accountability in their graduation requirements?

Analysis of Education Reform Policy Levers

Education Finance

Given the nature of the data available, states were most easily distinguished on the basis of the finance lever. Three key types of information were used to examine education finance reform: trends in the proportion of revenues for public elementary and secondary education provided by the state, trends in total expenditures per pupil, and changes in the poverty gap. The relevant data are displayed in Table 1.

Using the information from the above table, as well as the other data included in the state profiles and ranking documents (See Appendix C), New York, Maryland and Virginia were identified as the leaders in education finance reform. New York took the top spot on the finance rankings due to a combination of a dramatic increase in annual per pupil expenditures during the 1998-2005 period (NCES, n.d.; NCES, 2006; NCES, 2007b), and a reduction in the funding gap between low and high poverty districts. It is important to note, however, that NY had the largest discrepancy between high and low poverty districts throughout the entire period (Carey, 2004). On the other hand, the state's increase in per pupil expenditures between 1998 and 2005 was more than double that of either the ten state or national averages (NCES, n.d.). While the state's proportional contribution to public elementary and secondary education decreased during the 1998 through 2005 period, New York's percentage remained above that of any of the other nine states.

Maryland was also designated as a leader, mostly because of its efforts in improving school finance equity. In comparison to the other states, Maryland led in enhancing spending equity among districts. The state also was second only to New York in the increase in per-pupil expenditures during the 1998-2005 period. Virginia had the third highest increase in per pupil expenditures between 1998 and 2005 (NCES, n.d.). In addition, the gap between high poverty and low poverty districts closed by $536, the second largest reduction among the ten states in the sample (Carey, 2004; Liu, 2006).

Five states fell in the middle category. They were: South Carolina, Michigan, Tennessee, California and Kentucky. Out of these five states, three saw a decrease in the poverty gap between higher and lower poverty districts (KY, MI and SC), while the poverty gap increased in two of these middle ground states (CA and TN). All five of the states' differences in trends in total state expenditures per pupil between 1998 and 2005 were below the ten state average. In terms of the average state proportional contribution to public elementary and secondary education, three of the five states were above the ten state average (CA, KY and SC), while two were below (MI and TN).

Table 1:
Selected State Education Finance Indicators

State or Jurisdiction	Average State Proportional Contribution to Public Elementary and Secondary Education (1998-2005)[1]	Trends in Total Expenditures per Pupil: Difference in Dollars (1998 - 2005)[2,3]	Difference in Gap Between Revenues per Students in Highest and Lowest Poverty Districts (1997-2004)[4,5,6]
Nation	48.6	1688	($99)
New York	65.2	4359	$11
Maryland	38.2	2705	$529
Virginia	39.8	2536	$536
South Carolina	49.9	1893	$497
Michigan	39.0	1908	$335
Tennessee	44.4	1727	$206
California	59.2	2104	($54)
Kentucky	59.3	1495	$567
North Carolina	44.7	1248	($79)
Texas	40.7	1561	($320)
Ten State Average	48.0	2154	$116

[1] An average was calculated using information obtained from the Common Core of Data (CCD) surveys from the National Center for Education Statistics (NCES) during the 1998-1999 through 2004-2005 school years.
[2] Unadjusted U.S. dollars.
[3] Difference was calculated between the 1998-1999 and 2004-2005 school years. Taken from the NCES (n.d.b) National public education financial survey: Fiscal years 1990-2002; NCES (2007b) Expenditures for public elementary and secondary education: School year 2004-2005; NCES (2006) Current expenditures for public elementary and secondary education: School year 2003-2004.
[4] Cost-adjusted U.S. dollars, 40% adjustment for low-income students. Parentheses () indicate that there was an increase in the gap between lowest and highest poverty districts (i.e., to the detriment of the most impoverished schools).
[5] Source: Carey (2004) The funding gap 2004; Liu (2006) The funding gap 2006.
[6] Carey (2004) described highest-poverty school districts as those in the top 25% statewide determined by the % of students living below the federal poverty line, and lowest-poverty school districts as the 25% in the bottom of that distribution.

Two states were placed in the laggard category for their progress, or lack thereof, on education finance reform between 1998 and 2005. North Carolina had the smallest increase in per pupil expenditures between 1998 and 2005 of all ten states (NCES, n.d; NCES, 2006; NCES, 2007b). In addition, the state's gap between highest and lowest poverty districts increased between

1997 and 2004 (Carey, 2004; Liu, 2006). Texas also experienced an increase in the spending gap. First, the state's increase in per pupil expenditures was third from the bottom, falling below both the ten state and national averages. In addition, the state ended the period in 2005 with the lowest proportional contribution by the state (35.9%), a percentage that had steadily decreased since 1998 (NCES, n.d.).

Curriculum and Standards

During the period 1998 - 2005, with an increased emphasis on accountability, the need for high quality curriculum and standards with aligned state assessments became increasingly apparent. Thus, a number of states sought either to develop or to revise their curricula during this period. State rankings were based on data obtained from states' departments of education websites, as well as on the findings of other policy researchers (Thomas B. Fordham Institute and Institute for a Competitive Workforce).[10] Table 2 presents the relevant data.

Table 2
Selected State Curriculum and Standards Indicators

State or Jurisdiction	State has standards that are clear, specific, and grounded in content in the subject of mathematics [1,2]		State revisions in textbook policies between 2000 and 2005[3]	Fordham Institute Grades of States' Math Standards[4]		Additions/ Revisions to state curriculum or standards between 1998-2005[5]
	2000	2005		2000	2005	
California	ES MS HS	ES MS HS	Yes	A	A	Yes
Tennessee		ES MS HS	No	F	D	Yes
Virginia	ES MS HS	ES MS HS	Yes	B	C	Yes
New York	ES MS HS	ES MS HS	No	B	C	Yes
Texas	ES MS HS	ES MS HS	No	B	C	Yes
Maryland	ES MS HS	ES MS HS	No	C	C	Yes
South Carolina	ES MS HS	ES MS HS	No	B	D	Yes
Michigan	ES MS HS	ES MS HS	No	F	C	Yes
Kentucky	ES MS HS	ES MS HS	Yes	B	C	Yes
North Carolina	ES MS HS	ES MS HS	No	A	C	Yes

[1] ES=Elementary School, MS=Middle School and HS=High School
[2]Source: Quality Counts, 2001; Quality Counts, 2006
[3]Information gathered from state profiles. See example in Appendix B
[4] Source: Klein, et al., 2005
[5] Information gathered from state profiles. See Appendix B for an example.

We designated two states as leaders. California adopted new mathematics content standards in 1997, the year before the examined time period. The corresponding mathematics framework was adopted in 1998, with revisions in 2005 (California DOE, 2007). In both 2000 and 2005, the state

[10] An independent assessment of the quality of state standards for the period of interest was outside the realm of this study.

was given a grade of A from the Fordham Institute for its mathematics standards, and was the only one of the ten states to receive a grade of A from the Foundation in 2005 (Klein, et al., 2005). Tennessee, the other leader, did not fare as well in the Fordham Institute ratings, but improved from a grade of F in 2000 (Braden, et al., 2000) to a D in 2005 (Klein, et al., 2005). Tennessee was the only one of the ten states that did not have standards that were "clear, specific, and grounded" in mathematics at any level in 2000 (Quality Counts, 2001), but developed standards that fit that description in mathematics at all three levels by 2005 (Quality Counts, 2006). The Tennessee Curriculum Standards for students in grades K through 8, were developed for English/language arts, mathematics and science in 2001(Tennessee DOE, n.d. a). The high school mathematics standards were developed in 1998 and revised in 2004 (Tennessee DOE, 2006).

Five states fell in the middle classification: Virginia, New York, Texas, Maryland and South Carolina. All five of these states updated, revised or developed standards and/or curriculum during the 1998 to 2005 period. Four of the five states' Fordham Institute grade of their math standards decreased during the 2000 to 2005 period, while Maryland's stayed the same. Additionally, four of these states did not have any revisions to their textbook policies between 2000 and 2005, while Virginia did have revisions to their textbook policies during that period.

Three states (Michigan, Kentucky and North Carolina) were ranked in the laggard category for their curriculum and standards policies. A common thread is near stagnancy in their policies regarding curriculum and instruction. Michigan's original curriculum was published in 1996 (MDE, 1996), while the mathematics curriculum framework including teaching and learning activities was added in 1998 (MDE, 2007a). In Kentucky, the state's standards were implemented before 1998. However, Kentucky's Program of Studies for Grades Primary – 12, were added in 1998 to clarify the standards (KDE, 2007). In North Carolina, the state standards were also developed before the examined time period, with revisions to mathematics curriculum implemented in 1998 and again in 2003 (NC Department of Public Instruction, 2005). All three states had math standards that were considered "clear, specific and grounded" throughout the entire examination period (Quality Counts 2001; Quality Counts 2006). While the state of Michigan, according to the Fordham Institute, had standards that improved from failing with an F in 2000 to a C in 2005, the Institute for a Competitive Workforce gave the state's standards a grade of "average" for their rigor (2007, p. 33). Kentucky and North Carolina's standards each received a lower grade from the Fordham Institute in 2005 than they did in 2000 (Braden, et al., 2000; Klein, et al., 2005), and the Institute for a Competitive Workforce gave them ratings of "average" (p. 28) and "middling" (p. 44), respectively.

Teacher Quality

Since teaching excellence is considered to be one of the most important factors in improving student achievement (Darling-Hammond, 1999; Darling-Hammond & Youngs, 2002), education reforms that aim to improve the caliber of teachers are critical in addressing achievement gaps. Table 3 compares state policies which, in addition to qualitative information taken from the state profiles were viewed as most relevant in assessing the states commitment to improving teacher quality. Three states were identified as leaders: Virginia, South Carolina and California. Between the 1998-1999 and 2004-2005 school years, all three had average annual teacher salaries that increased more than both the ten-state and national averages, with California having the biggest increase among the ten states (American Federation of Teachers, n.d.). In addition, both Virginia and South Carolina placed an emphasis on professional development through their policies instituted during the period. Virginia revised its teacher licensure requirements in 1998 (Virginia DOE, 1998), and in 2004, mandated that the state's professional development be aligned with the state's standards (Virginia

DOE, 2004). In addition, the state had a commitment to financing professional development throughout the period (Quality Counts 2001; Quality Counts 2006).

Table 3
Selected State Teacher Quality Indicators

State or Jurisdiction	Changes to Teacher Certification between 1998-2005[1]	Trends in Average Teacher Salary[2] Difference Between 1998 and 2005	Number of National Board Certified Teachers[3]	
			2000	2005
Nation	**N/A**	**$7,028**	N/A	N/A
Virginia	Yes	5,248	144	905
South Carolina	Yes	6,460	371	4,445
California	Yes	11,278	785	3,377
North Carolina	Yes	6,561	2,377	9,815
New York	Yes	6,586	104	588
Tennessee	Yes	5,979	35	173
Texas	Yes	5,692	35	231
Kentucky	No	9,785	79	899
Maryland	Yes	7,683	70	660
Michigan	No	7,668	89	189
Ten State Average	**N/A**	**7,294**	409	2,128

[1]Information gathered from state profiles. See Appendix B for an example.
[2] Unadjusted U.S. dollars
[3]Source: Quality Counts, 2001; Quality Counts, 2006

South Carolina included professional development as one of the objectives in the Education Accountability Act of 1998, which specified that teachers in low performing schools who participated in professional development activities that addressed needs in their school's improvement plan could earn credits towards recertification. The act also called for the development of a professional development accountability system (SC General Assembly, 1998). California, on the other hand, emphasized teacher preparation program standards during the examined period. Legislation, passed in 1998, created multiple routes to teacher certification, all of which were to be accredited, and called for teacher preparation standards to be aligned with the state-adopted K-12 academic content standards (California Commission on Teacher Credentialing, 2001).

The four states placed in the middle classification were: North Carolina, New York, Tennessee and Texas. All four passed legislation during the period that focused on teacher certification. Additionally, in terms of trends in average teacher salary between 1998 and 2005, all four states were below the national and ten-state average differences during that time period. In terms of the number of National Board Certified teachers, North Carolina had the most teachers who met that criterion in both years, while New York, Tennessee and Texas had far fewer teachers that were Board Certified.

The remaining three states (Kentucky, Maryland and Michigan) were classified as laggards. There were few major discernable changes or additions to policies regarding teachers in any of these three states. As of 2004, Kentucky was the only state to ban out-of-field teaching (Ingersoll and

Curran, 2004). Due to the shortage of teachers in some subjects (including mathematics), however, Kentucky still continued to employ teachers with emergency certificates (Hoff, 2003). While there may have been changes in the types of certification in Maryland during the examined period, we were unable to find any evidence that indicated exactly when these changes took place. In 2004, however, the state did adopt new professional development standards (MSDE, 2004). Michigan also updated its professional development standards in 2003 (MDE, 2007b). In addition, while Maryland had the third highest increase in average teacher salary, Kentucky and Michigan had the two lowest increases of the ten states (American Federation of Teachers, n.d.). Although all three states had low numbers of National Board Certified teachers throughout the period, Kentucky and Maryland increased their numbers considerably more than did Michigan (Quality Counts 2001; Quality Counts, 2006).

Assessment and Accountability

During the examined period, "assessment" and "accountability" became two of the hottest buzzwords in education policy across the country. All ten states had accountability systems in place by the end of the examined time period, in line with the No Child Left Behind requirement that all states had to test their students in English/language arts and mathematics in grades 3-8 and one later grade, by the 2005-2006 school year. Therefore, changes in assessment and accountability were evaluated by timing of implementation (i.e., assessments implemented earlier rather than later in the period), as well as by a commitment to a more comprehensive accountability system, as opposed to simply administering the required assessments.

Three states were designated as leaders. The first, South Carolina, enacted the Education Accountability Act of 1998 (EAA), which established a performance-based accountability system within the state (SC Education Oversight Committee, 2004). In the Act, accountability was defined as "acceptance of the responsibility for improving student performance and taking actions to improve classroom practice and school performance" (SC Legislature, 1998). The EAA required the development of learning standards, to be reviewed periodically, and aligned with state assessments. The EAA also established school report cards, and required districts to create local accountability systems (SC Legislature, 1998). The Palmetto Achievement Challenge Tests (PACT), developed in accordance with the EAA, assessed students in the four core academic subjects: English/language arts, mathematics, science, and social studies. The assessments were given in grades three through eight, and were aligned with the state standards for the corresponding grades. The PACT in mathematics was first administered in 1999 (SC DOE, 2007).

Tennessee earned its classification through implementation of a variety of educational policies in the state's assessment and accountability programs during the 1998-2005 period. While the state began the period with mathematics assessments aligned with the state standards at the high school level, by the 2005-2006 school year, elementary and middle school mathematics assessments aligned with the state standards had also been developed (Quality Counts, 2001; Quality Counts 2006). According to Education Week's 2005 Quality Counts report, Tennessee is "closer than many other states to having standards-based exams in every core subject in every grade span", including standards-based tests in mathematics in elementary, middle, and high school (Skinner, 2005). The Tennessee Comprehensive Assessment Program (TCAP) Achievement Tests were first given to students in grades three through eight beginning in 1992. TCAP Secondary Assessments were developed in 1998 (Tennessee DOE, n.d. b). The High School Examinations Policy adapted from an earlier program in 2002, stated that students who were entering the 9th grade during the 2001-2002 school year had to pass the TCAP three gateway examinations, in mathematics, science and

English, in order to earn a high school diploma (Tennessee DOE, n.d. b). Between the years 2000 and 2004 the graduation rate for the state increased by 14.7% (Swanson, 2008).

The third state included in the leader category is California. California's Public Schools Accountability Act (PSAA) of 1999 emphasized three main objectives. These included (1) increased accountability at the school level, (2) imposing rewards and sanctions based on school progress, as well as (3) community involvement as a means to increase student achievement (California DOE, 2004). The Accountability Performance Index (API) is the "cornerstone" of the PSAA, and was used to measure school performance on an index that ranged from 200 to 1000 (California DOE, 2004). While the Standardized Testing and Reporting (STAR) program was authorized in 1997, the Stanford 9 was designated as the state test for administration beginning in 1998. The test was given to students in grades 2 through 8 in reading, mathematics, written expression, and spelling. Secondary students (grades 9 through 11) were tested in reading, writing, mathematics, science, and history/social science (State of California, 1998). Also during this period, the California Content Standards tests were developed. These assessments comprised California specific items only, unlike earlier assessments that consisted of a combination of Stanford 9 questions and items created specifically for the California tests (State of California, 2000; State of California, 2001).

Five states were identified as being in the middle classification: North Carolina, Kentucky, Virginia, Texas and Maryland. All five of these states made changes in their accountability system during the 1998 to 2005 time period, with the exception of North Carolina who first implemented their accountability system during the 1996-1997 school year. Both Kentucky and Virginia developed assessment systems in 1998. Texas and Maryland entered the examined time period with previously established accountability systems, however both made revisions to their respective systems during the 1998-2005 period.

New York and Michigan, identified as laggards, also revised their assessment policies in compliance with NCLB. However, neither state made any revisions prior to the mandated 2005-2006 deadline. In 2005, New York began testing all students in grades 3 through 8 in mathematics and English/language arts, in accordance with the NCLB requirement (NYSED, 2005 a). Prior to this, only students in grades 4 and 8 were assessed in those subjects. The Michigan Educational Assessment Program (MEAP) was first administered during the 1969-1970 school year, and continued to be the primary instrument of student accountability through the examined time period. The system did undergo some changes in the fall of 2005 in order to comply with NCLB requirements, including a redesign of the assessments and modification of the testing schedules with respect to content areas and grade levels (MDE, 2004).

Conclusions and Summary Rankings

As is evident from the preceding discussion, some states had a stronger commitment to education policy initiatives than did others. However, it is important to note that no state was placed in the same category across all four of the levers. Thus, according to our relative ranking on each of the four policy levers, each state placed greater emphasis on some policies than on others. Braun et al. (2006) used weights in deriving the overall ranking of states' reform policies. However, since only four levers were examined in this study rather than the original five, all four were given equal weight in the present ranking process.

The overall rankings for the ten states reflect a summary judgment of the states' efforts with respect to the four policy levers. These rankings were developed independently of the statistical analyses of the NAEP data and clearly involve a subjective element, particularly given the nature of the policy relevant information that was available. The focus of the policy component of this study was similar to that of Braun et al. (2006). That is, the rankings were determined in light of the following question:

> Given the character and quality of a state's policy efforts during the 1998-2005 period, in comparison to the other states in the study, is it reasonable to expect that over the period 2000 to 2007 it would achieve relatively greater progress in closing the achievement gap -- or increasing the scores of Black students -- in comparison to those other states? (compare Braun et al., 2006, p. 37)

To arrive at the overall rankings, we reviewed the individual questions linked to each of the policy levers in order to obtain a broader view of each state. We then revisited the states' classifications with respect to each of the four policy levers. Finally, we assigned a numerical value to each category. Those who fell in the leader category on a policy lever were assigned a score of one, middle ground a two, and laggard a three. There were also gradations within the three classifications. In order to further differentiate between the groups, especially within the middle ground where many states fell, we additionally ranked the states within that level. The state at the top of the middle level received half a point less (i.e., 1.5), whereas the state in the bottom spot of that level received an additional half a point (i.e., 2.5). States' scores on the four policy levers were summed to obtain the overall scores. On that basis the states were then classified into the three categories. The maximum score range was four to twelve. The actual range of the states in our study was five to eleven. The corresponding scores for the three classifications -- leader, middle ground and laggard – were less than seven, seven through nine and greater than nine, respectively. The results are presented in Table 4.

Table 4
Policy Rankings Overall and by Lever

Ranking	Education Finance	Curriculum and Standards	Teacher Quality	Assessment and Accountability	Overall
1	NY	CA	VA	SC	CA
	MD	TN	SC	TN	SC
	VA		CA	CA	TN
					VA
2	SC	VA	NC	NC	MD
	MI	NY	NY	KY	NY
	TN	TX	TN	VA	TX
	CA	MD	TX	TX	
	KY	SC		MD	
3	NC	MI	KY	NY	KY
	TX	KY	MD	MI	MI
		NC	MI		NC

Note: Within each of the rankings (Leader=1, Middle Ground=2 and Laggard=3) across the four policy levers, states were further ranked within each category. For example, in the Leader category of the Education Finance lever, New York was ranked above Maryland, which was ranked above Virginia. These rankings were considered to be informative, but not to be over interpreted.

As shown in the table, there were four states that were designated as leaders overall. California was considered to be a leader on three of the policy levers, showing a strong commitment towards curriculum and standards, teacher quality and assessment and accountability, while focusing less on education finance. South Carolina, Tennessee and Virginia showed a strong commitment

towards two of the policy levers, while their commitment was less evident in the other two. Three states --Maryland, New York and Texas -- had mixed records in terms of the consistency of their education polices throughout the period, showing a stronger commitment to some than others. The remaining three states --Kentucky, Michigan and North Carolina -- showed the least commitment in their efforts to improve their education policies during the examined period.

It should be borne in mind that these judgments are comparative and not absolute. It is evident from the data presented that all the states engaged in reform efforts over this period. Moreover, some states, such as North Carolina, were highly ranked for their policy efforts in the previous period and so the years 1998 to 2005 may have served as a time of consolidation. We return later to the implications of this possibility.

Quantitative Results

Preliminaries

To set the stage for the analyses to follow, we begin with Table 5 which displays the basic data for the ten states. The ten state NAEP means range from 260 (California) to 277 (Michigan), a range of 17 points. Over the period in question, all states, except Michigan, experienced an increase in NAEP scores. The largest increase was nearly 17 points (South Carolina) with the median gain being about 11 points – a little more than a point per year. For seven of the states, the changes in the exclusion rates were no more than two percent. Maryland experienced the largest increase, 4.6 percent, as well as the second largest increase in NAEP scores.

Table 5
State NAEP results

State	Achievement 2000	Change in mean achievement 2007 - 2000	Change in exclusion rate 2007 -2000
MI	277	-0.4	0.7
NC	276	7.6	-3.0
VA	275	12.9	0.3
TX	273	12.4	-2.2
MD	272	13.8	4.6
NY	271	8.7	-0.6
KY	270	8.8	2.0
SC	265	16.9	1.3
TN	262	12.4	3.8
CA	260	10.6	-2.3
Median	272	11.5	0.5

The principal aim of the study is to probe beneath the surface of these state results. Accordingly, in Table 6 we display the counts of schools and students in the NAEP samples for each state for the 2000 and 2007 administrations. In 2000, the count of schools ranges from about 70 (California) to about 110 (Virginia). In all states the number of schools in the lower poverty stratum (S1) is greater, and in some cases substantially greater, than the number in the higher poverty stratum. Turning to 2007, we note that the counts of schools in many states are considerably larger than the counts in 2000. This is mainly due to the oversampling of schools in cities participating in the Trial Urban District Assessment, which was introduced in 2002. Of course, the mandates of No Child Left Behind also played a role.

Table 6
Numbers of schools and students, by stratum and state.

		2000				2007			
		# of	# of Students			# of	# of Students		
State	Stratum	Schools	Total	White	Black	Schools	Total	White	Black
CA	S1	40	1,000	560	50	130	3400	1600	210
	S2	30	700	110	100	180	4600	500	440
KY	S1	60	1500	1330	130	40	1100	1,000	80
	S2	40	900	730	120	70	1400	1200	160
MD	S1	80	2100	1410	480	80	2100	1200	610
	S2	20	400	30	340	30	500	100	350
MI	S1	70	1600	1420	80	90	1900	1600	120
	S2	20	400	120	250	40	600	200	320
NY	S1	40	1,000	820	80	70	1600	1300	140
	S2	30	700	120	340	80	2000	300	710
NC	S1	80	1900	1410	400	70	2200	1400	510
	S2	30	600	190	320	80	1800	500	870
SC	S1	50	1200	880	320	50	1300	800	320
	S2	50	1100	450	660	60	1300	600	690
TN	S1	70	1700	1380	240	60	1400	1200	210
	S2	30	600	290	270	60	1200	600	550
TX	S1	60	1300	840	150	90	2500	1400	300
	S2	50	1000	300	150	140	4300	600	920
VA	S1	90	2200	1560	430	90	2200	1400	440
	S2	20	400	100	220	20	400	100	280

Note: Numbers have been rounded to conform with NCES publication policies.

The larger number of schools in 2007 is reflected in the larger counts of White students and Black students. For example, in 2000, there were three states (California, Michigan and New York) for which the number of Black students in S1 was less than 100. In 2007, there was only one such state (Kentucky).

Table 7 presents the stratum means for each state for 2000 and 2007, along with the change over the period. Here change is defined as the "value in 2007 minus the value in 2000". Hence, a positive value indicates a desirable outcome – namely, a gain in achievement over the period. Mirroring the results at the state level, all states, except Michigan, show an average gain for students enrolled in lower poverty schools (S1). The median gain is 11 points, with South Carolina and Tennessee having experienced the largest gains, 16 points. Similarly, all states showed an average gain for students enrolled in higher poverty schools (S2). Gains ranged from 3 points (Michigan) to 32 points (Maryland). The median gain is 15 points, corresponding to a gain of about 2 points per year. The gains in most states were close to the median. In both S1 and S2, most of the reported gains are highly statistically significant (i.e., different from zero).

Table 7
Mean achievement by stratum and state.

	Lower Poverty S1						Higher Poverty S2					
	2000		2007		2007-2000		2000		2007		2007-2000	
State	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
California	271	(3.0)	283	(1.3)	12	(3.3)	242	(2.6)	258	(1.1)	16	(2.8)
Kentucky	277	(1.5)	285	(1.7)	8	(2.2)	258	(1.8)	274	(1.3)	15	(2.3)
Maryland	280	(1.4)	291	(1.7)	11	(2.1)	233	(4.4)	265	(2.2)	32	(4.9)
Michigan	284	(1.7)	285	(1.3)	0	(2.1)	247	(3.3)	250	(3.2)	3	(4.6)
New York	285	(1.9)	290	(1.4)	5	(2.4)	253	(3.7)	267	(2.0)	15	(4.2)
North Carolina	281	(1.5)	291	(1.4)	10	(2.0)	259	(3.0)	273	(1.7)	14	(3.4)
South Carolina	273	(2.0)	290	(1.4)	16	(2.4)	255	(2.0)	273	(1.7)	18	(2.6)
Tennessee	267	(1.7)	283	(1.6)	16	(2.4)	247	(3.3)	263	(1.7)	16	(3.8)
Texas	282	(2.2)	296	(1.5)	14	(2.7)	264	(2.0)	277	(1.2)	13	(2.4)
Virginia	278	(1.4)	291	(1.2)	13	(1.9)	255	(3.2)	269	(2.8)	14	(4.3)
Median	279		290		11		254		268		15	

Analysis of Gaps

As is evident from Table 7, in both years the mean in S1 is greater than the mean in S2 for all states. Table 8 presents the stratum gaps for each state for both 2000 and 2007. The stratum gap is defined as the difference in means between S1 and S2. Here change is defined as the "value in 2000 minus the value in 2007". Hence, a positive value indicates a desirable outcome – namely, a reduction in the stratum gap. The states are listed in ascending order of the stratum gap in 2000. In that year Kentucky, Texas and South Carolina had the smallest gap (18 points) and Maryland the largest (47 points). The median gap was 23 points. Recall that the difference in means between the lowest and highest scoring states in 2000 was 17 points (see Table 5). This is the first indication that within-state differences dominate between-state differences.

Table 8
Stratum gaps in mean achievement by state (standard errors in parentheses).

State	[S1 - S2] 2000		[S1 - S2] 2007		Reduction in Gap 2000 to 2007	
KY	18	(2.4)	11	(2.1)	7	(3.2)
TX	18	(3.0)	19	(1.9)	-1	(3.6)
SC	18	(2.8)	17	(2.2)	2	(3.6)
TN	20	(3.8)	20	(2.4)	0	(4.4)
NC	22	(3.4)	19	(2.2)	4	(4.0)
VA	23	(3.5)	22	(3.1)	1	(4.6)
CA	30	(4.0)	25	(1.7)	4	(4.3)
NY	32	(4.2)	22	(2.4)	10	(4.8)
MI	38	(3.7)	34	(3.5)	3	(5.1)
MD	47	(4.6)	26	(2.8)	21	(5.3)
Median	23		21		3	

In 2007, stratum gaps range from 11 points (Kentucky) to 34 points (Michigan). The median gap was 21 points. The last column of Table 8 displays the reduction in the stratum gaps over the period. Only three states (Kentucky, New York and Maryland) experienced reductions that were statistically significantly different from 0 at the 0.05 level. Kentucky maintained its position as the state with the smallest stratum gap and Maryland relinquished its position as the state with the largest gap to Michigan.

Black-White Achievement Gaps

We begin with a return to a state-level display. Table 9 presents mean scores for White students, Black students and the differences between them for each state in both 2000 and 2007. We observe that, with the exception of White students in Michigan, in all states the average scores of both White students and Black students improved from 2000 to 2007. The achievement gap in 2000 ranged from 22 points in Kentucky to 45 points in Michigan. Again, we note that 22 points is 5 points greater than the largest between-state difference overall. The median achievement gap is 34 points, which can be compared to the median stratum gap of 23 points.

Table 9
Mean achievement for White students and Black students, and Black-White gaps, by state and year.

	2000						2007						Reduction in Gap 2000 - 2007	
	White		Black		W-B		White		Black		W-B			
State	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
CA	277	(2.2)	235	(5.1)	42	(4.7)	287	(1.3)	253	(2.1)	35	(2.2)	8	(5.2)
KY	272	(1.3)	250	(2.9)	22	(3.0)	282	(1.1)	257	(2.7)	25	(2.9)	-2	(4.2)
MD	286	(1.3)	244	(2.7)	42	(2.8)	300	(1.3)	265	(1.3)	36	(1.7)	7	(3.3)
MI	285	(1.5)	239	(3.3)	45	(3.4)	285	(1.1)	244	(2.2)	41	(2.3)	4	(4.1)
NY	284	(1.9)	251	(4.5)	33	(4.8)	290	(1.2)	258	(2.3)	32	(2.5)	1	(5.4)
NC	287	(1.4)	252	(1.4)	35	(1.9)	295	(1.2)	266	(1.6)	29	(1.9)	6	(2.7)
SC	277	(1.7)	247	(1.6)	30	(1.9)	293	(1.2)	265	(1.3)	27	(1.5)	3	(2.4)
TN	269	(1.4)	235	(3.0)	34	(3.1)	282	(1.1)	254	(1.7)	28	(2.0)	6	(3.7)
TX	287	(1.6)	253	(3.0)	34	(3.0)	300	(1.4)	271	(1.5)	29	(1.9)	5	(3.6)
VA	283	(1.3)	253	(1.8)	30	(2.1)	296	(1.5)	268	(1.4)	28	(1.8)	2	(2.8)
Median	284		249		34		291		261		29		5	

In 2007, the achievement gaps ranged from 25 points in Kentucky to 41 points in Michigan, with a median of 29 points. The last column of Table 9 contains the reductions in the achievement gap over the period.[11] Except for Kentucky, all are positive but none are statistically significant. The median reduction is five points. Thus, at best, there is weak evidence of a general reduction in the achievement gap over the period.

The examination of stratum-level patterns in the Black-White achievement gap begins with Table 10, which displays the means for White students and Black students for each combination of stratum, state and year, as well as the gains over the period. Turning to S1, we observe that in all states White students improved from 2000 to 2007. All the gains were statistically significant except in Michigan. The median gain was 13 points. Black students improved in eight states, with seven of the gains being statistically significant. The median gain was 16 points. In S2, White students improved in nine states, but the gains were only statistically significant in six states. The median gain was 9.5 points. Black students showed improvement in all ten states, with statistically significant gains in six states. The median gain was 16 points.

[11] Note that here again change is defined as the "value in 2000 minus value in 2007". Hence, a positive value indicates a desirable outcome – namely, a reduction in the achievement gap.

Comparing the strata, it appears that overall White students showed somewhat greater improvement in S1 than in S2. There was no difference overall for Black students. Interestingly, there were six states (Maryland, North Carolina, South Carolina, Tennessee, Texas and Virginia) that experienced statistically significant gains for both White students and Black students in both strata.[12]

Table 10

Mean achievement for White and Black students, by stratum and state.

	2000				2007				2007 - 2000			
	White		Black		White		Black		White		Black	
State	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
S1 (Lower Poverty)												
CA	279	(2.5)	235	(10.3)	291	(1.6)	262	(3.0)	12	(3.0)	27	(10.7)
KY	279	(1.5)	254	(2.9)	288	(1.6)	260	(4.7)	9	(2.2)	6	(5.5)
MD	287	(1.2)	256	(2.2)	302	(1.3)	267	(1.6)	15	(1.8)	12	(2.7)
MI	286	(1.6)	257	(3.8)	287	(1.2)	257	(3.9)	1	(2.0)	-1	(5.5)
NY	286	(2.0)	274	(6.9)	292	(1.3)	271	(4.2)	6	(2.4)	-3	(8.1)
NC	288	(1.6)	256	(1.7)	298	(1.5)	272	(2.1)	9	(2.2)	16	(2.7)
SC	281	(2.0)	253	(2.5)	297	(1.7)	272	(2.0)	16	(2.6)	19	(3.2)
TN	271	(1.4)	244	(4.9)	286	(1.6)	266	(2.8)	15	(2.2)	21	(5.6)
TX	289	(2.1)	259	(4.9)	304	(1.6)	276	(3.0)	15	(2.7)	17	(5.8)
VA	283	(1.3)	257	(2.3)	297	(1.5)	273	(1.8)	14	(2.0)	16	(2.9)
S2 (Higher Poverty)												
CA	269	(4.2)	235	(5.8)	274	(2.2)	246	(2.6)	5	(4.7)	12	(6.3)
KY	261	(1.9)	246	(4.6)	277	(1.3)	255	(3.3)	16	(2.3)	9	(5.6)
MD	258	(11.5)	229	(4.2)	281	(3.2)	260	(2.4)	22	(11.9)	31	(4.8)
MI	270	(3.0)	234	(3.5)	268	(4.4)	239	(2.6)	-3	(5.3)	5	(4.4)
NY	275	(5.1)	246	(4.5)	283	(2.9)	254	(2.5)	7	(5.8)	9	(5.1)
NC	277	(2.8)	246	(2.4)	287	(2.1)	261	(1.9)	10	(3.5)	15	(3.1)
SC	270	(2.7)	244	(2.2)	286	(2.0)	262	(1.5)	16	(3.3)	18	(2.7)
TN	263	(3.8)	227	(4.0)	275	(2.1)	249	(2.3)	12	(4.3)	22	(4.7)
TX	280	(1.4)	246	(3.3)	289	(2.2)	268	(1.8)	9	(2.6)	22	(3.7)
VA	276	(4.4)	243	(2.6)	285	(3.5)	260	(2.1)	9	(5.6)	17	(3.4)

[12] For Maryland and Virginia, the results for Whites in Stratum 2 are significant at the .10 level. All other results are significant at the .05 level.

We now turn our attention to Table 11, which displays the achievement gaps for each combination of stratum, state and year, along with the change in the gap over the period. In S1, the achievement gaps in 2000 ranged from 12 points in New York to 43 points in California, with a median of 28 points. In 2007, the achievement gaps ranged from 20 points in Tennessee to 35 points in Maryland, with a median of 27 points. Interestingly, the stratum-level achievement gaps are typically only slightly smaller than the state-level achievement gaps (see Table 9).[13] Finally, changes in the achievement gaps over the period varied from an increase of 9 points in New York to a decrease of 14 points in California. None of the changes achieved statistical significance. The median change was a decrease of only two points.

Turning to the results for S2, we find that in 2000 the achievement gaps ranged from 15 points in Kentucky to 36 points in Michigan and Tennessee, with a median of 32 points. In 2007, the achievement gaps ranged from 21 points (Maryland and Texas) to 28 points (California and New York). The median was 25 points. Again, we see that the stratum-level gaps are of the same magnitude as the state-level gaps. Changes in the achievement gaps over the period ranged from an increase of seven points in Kentucky to a decrease of thirteen points in Texas. Only the change in Texas reaches statistical significance. The median change is a reduction of seven points.

Table 11

Gaps in mean achievement (White mean – Black mean) by stratum and state

State Gap Reduction 2000 - 2007			S1(Lower Poverty) gap						S2(Higher Poverty) gap					
			2000		2007		2000-2007		2000		2007		2000-2007	
Mean	S.E.	State	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
8	(5.2)	CA	43	(9.3)	29	(3.1)	14	(9.8)	34	(5.6)	28	(3.1)	6	(6.4)
7	(3.3)	MD	31	(2.2)	35	(1.9)	-3	(2.9)	30	(11.3)	21	(3.7)	9	(11.9)
6	(3.7)	TN	26	(4.9)	20	(2.5)	6	(5.5)	36	(5.2)	26	(3.1)	10	(6.0)
6	(2.7)	NC	32	(2.0)	26	(2.4)	6	(3.1)	31	(3.6)	26	(2.8)	5	(4.5)
5	(3.6)	TX	30	(4.7)	28	(3.2)	3	(5.7)	34	(3.8)	21	(2.4)	13	(4.5)
4	(4.1)	MI	29	(3.8)	30	(3.8)	-1	(5.4)	36	(4.0)	29	(4.7)	7	(6.2)
3	(2.4)	SC	28	(2.3)	25	(2.3)	3	(3.3)	26	(2.9)	24	(2.1)	2	(3.6)
2	(2.8)	VA	26	(2.5)	24	(2.1)	2	(3.3)	33	(5.0)	25	(3.6)	8	(6.1)
1	(5.4)	NY	12	(7.1)	21	(3.9)	-9	(8.1)	30	(6.2)	28	(3.5)	1	(7.1)
-2	(4.2)	KY	25	(3.0)	28	(4.9)	-3	(5.7)	15	(4.6)	22	(3.4)	-7	(5.7)
5		Median	28		27		2		32		25		7	

[13] This is noteworthy inasmuch as the stratum gaps within each state are large. Consequently, one might expect the achievement gaps within strata to be considerably smaller than the state level gaps, especially since Black students are more likely to be found in S2 than in S1.

In Table 11 states are ordered by the magnitude of the reduction in the state-level achievement gap. There appears to be a weak correlation between those reductions and the reductions at the stratum level. There is more consistent evidence of a reduction in the achievement gap in S2 than in S1. However, even in S2, the median reduction corresponds to a trend of only one point per year.

Grading the States

A principal goal of this study is to examine the strength of the association between states' policy efforts and their success in improving Black student achievement and in closing the Black-White achievement gap. (Of course, there must be sufficient policy variation across states to make this exercise at all interesting.) Because it takes time for state policies to be implemented and to have impact, we evaluated the policies for the lagged period 1998 to 2005. As in Braun et al. (2006), we adopted a normative approach; that is, we ranked each state on each policy lever in relation to the other nine states.

Accomplishing the goal requires a three step process. The first step can be found in the last column of Table 4, which displays the states' classification on overall policy. The result of the second step can be found in Table 12 below, in which states are again classified into one of three categories on the basis of their outcomes on the designated indicators. The classifications are based on an evaluation of the data from Tables 10 and 11, taking into account both the relative magnitude of the change and its statistical significance. The third step involves linking states' policy rankings with their rankings on outcomes. This will be carried out in the next-to-last section.

Returning to Table 12, we focus first on the outcome "Improving Black student achievement". Four states (Michigan, New York, Tennessee, Texas) are in the same category for both strata. Only California has the maximal discrepancy: It is in the highest category for S1 but in the lowest for S2. The other five states are in adjacent categories for S1 and S2. With respect to the outcome " Closing the achievement gap". Four states (Kentucky, New York, Tennessee, Virginia) are in the same category for both strata. The other six states are in adjacent categories for S1 and S2. Thus, there is reasonable agreement in the states' relative ranks for the two strata. It is noteworthy that on all four rankings Tennessee is in the highest category and New York is in the lowest category.[14] Texas appears in the highest category three times, while Kentucky and Michigan fall in the lowest category three times.:

Table 12
Ranking states on student achievement outcomes.

	Improving Black student achievement		Closing the achievement gap	
Ranking	Lower poverty stratum (S1)	Higher poverty stratum (S2)	Lower poverty stratum (S1)	Higher poverty stratum (S2)
1	CA, NC, SC, TN, TX, VA	MD, TN, TX	CA, NC, TN	TN, TX
2	KY, MD	NC, SC, VA	SC, TX, VA	CA, MD, MI, NC, VA
3	MI, NY	CA, KY, MI, NY	KY, MD, MI, NY	KY, NY, SC

[14] This is exactly the reverse of what was found for the 1992 -2000 period, where New York was always in the highest category and Tennessee in the lowest!

Another Perspective on Achievement Gap Trends

Until this point, we have described achievement gaps solely in terms of a difference of means, which is but one way of summarizing the difference between two cumulative distribution functions (CDFs). It can be helpful to provide a more complete characterization of this difference because it illuminates how the mean difference arises. Holland (2002) provides a simple strategy that usually yields informative displays. The key idea is to compute for each of a number of selected percents (in the range of 1 to 99) the difference in the corresponding percentiles for the two distributions and then to plot that difference against the selected percent. When done for a dense set of percents, the method yields a graph that can be easily interpreted. For example, if CDF A lies entirely to the left of CDF B, then for a given percent the corresponding percentile for B will always be greater than that for A.[15] The plot will then lie entirely above the horizontal zero line. In particular, if CDF B represents a mean shift of x points relative to CDF A, the plot will be a horizontal line x points above the zero line. Deviations from a horizontal line signal differences other than a simple mean shift. This approach is now illustrated with data from three states, California, Kentucky and New York.

From Table 11, we know that in California S1 the mean achievement gap was reduced from 43 to 29 points. We observe in Figure 1, that the percentile gap in 2000 was extremely large for the lowest percents (being greater than 60 points at 10 percent) and drifted downward with higher percents (about 30 points at 90 percent).[16] By contrast, the percentile gap in 2007 was more nearly constant at around 30 points. One interpretation is that from 2000 to 2007, the lower tail of the score distribution for Black students shifted strongly rightward toward that of the White students and that this shift contributed to the mean reduction of 14 points.[17] For example, at 20 percent, the percentile difference in 2000 was about 55 points but in 2007 it was about 31 points. In California S2, the mean achievement gap was reduced from 34 to 28 points. From Figure 2, it appears that the reduction is almost entirely due to the extended lower tail of the distribution for Black students (relative to White students) in 2000 being sharply shifted rightward. For example, at 10 percent, the percentile gap was about 45 points in 2000 but only 27 points in 2007. Through most of the range of percents, the percentile gaps in both years were approximately 30 points.

By contrast, in Kentucky S1, the mean achievement gap increased from 25 to 28 points. Figure 3 indicates that the increase was mainly due to a leftward shift of the upper tail of the distribution for Black students (relative to White students) from 2000 to 2007. In Kentucky S2, the mean achievement gap increased from 15 to 22 points. From Figure 4, we see that is due to both a general mean shift of 5 to 6 points along with an extension of the lower tail of the distribution for Black students (relative to White students). In New York S1, the achievement gap increased by 9 points and this is reflected in Figure 5, which displays a general shift of that amount over most of the range of percents. New York S2 experienced a reduction of 1 point (i.e. essentially no change). Figure 6 indicates that this is the result of a tradeoff between a relative shift rightward of the lower tail of the distribution for Black students and a small shift leftward over most of the range of percents.

---

[15] The population corresponding to CDF B is said to be stochastically larger than the population corresponding to CDF A.

[16] The sampling variation of ordinate values can be quite large at extreme percents, so they should not be over interpreted.

[17] We know from Table 10 that from 2000 to 2007 mean scores for both White students and Black students increased substantially, indicating that both distributions shifted to the right.

Thus, the "Holland plots" provide the basis for a more nuanced narrative of the changes over time in the differences in the score distributions for White and Black students. Combining the Holland plots with results from the analysis of changes in the achievement gaps within schools (see next section), makes possible a more comprehensive account of the impact of state policies.

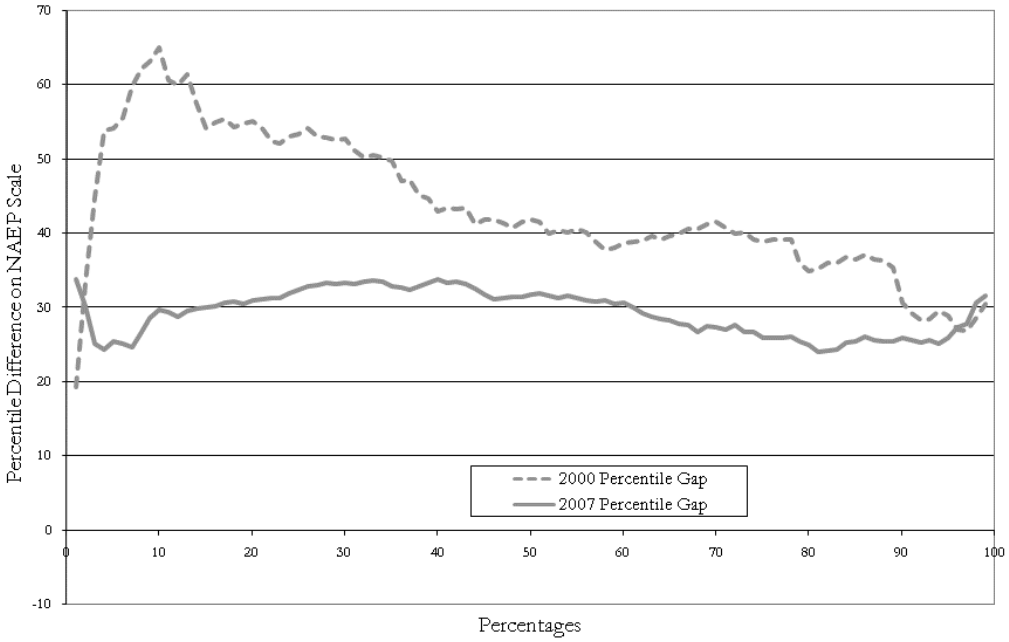

Figure 1. CA: S1: Gap in Percentiles between Whites and Blacks.

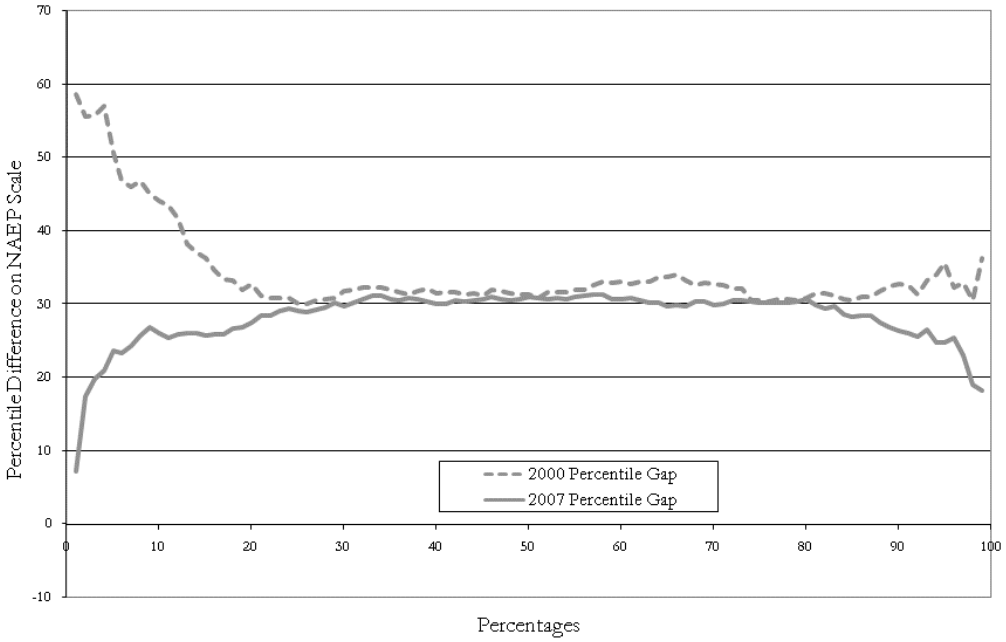

Figure 2. CA: S2: Gap in Percentiles between Whites and Blacks.

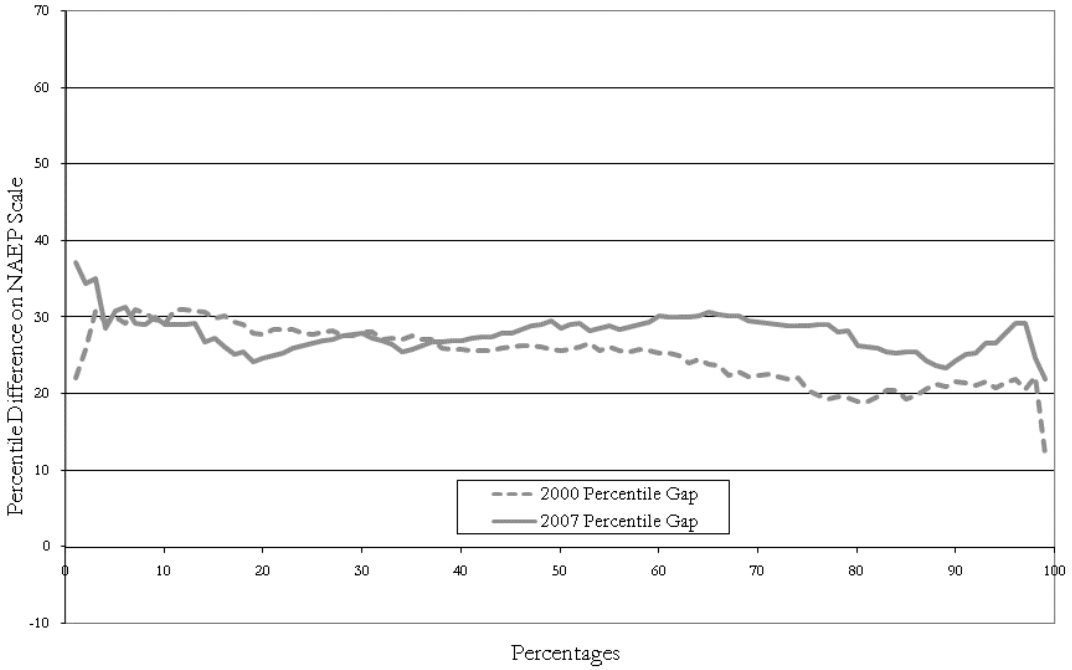

Figure 3. KY: S1: Gap in Percentiles between Whites and Blacks.

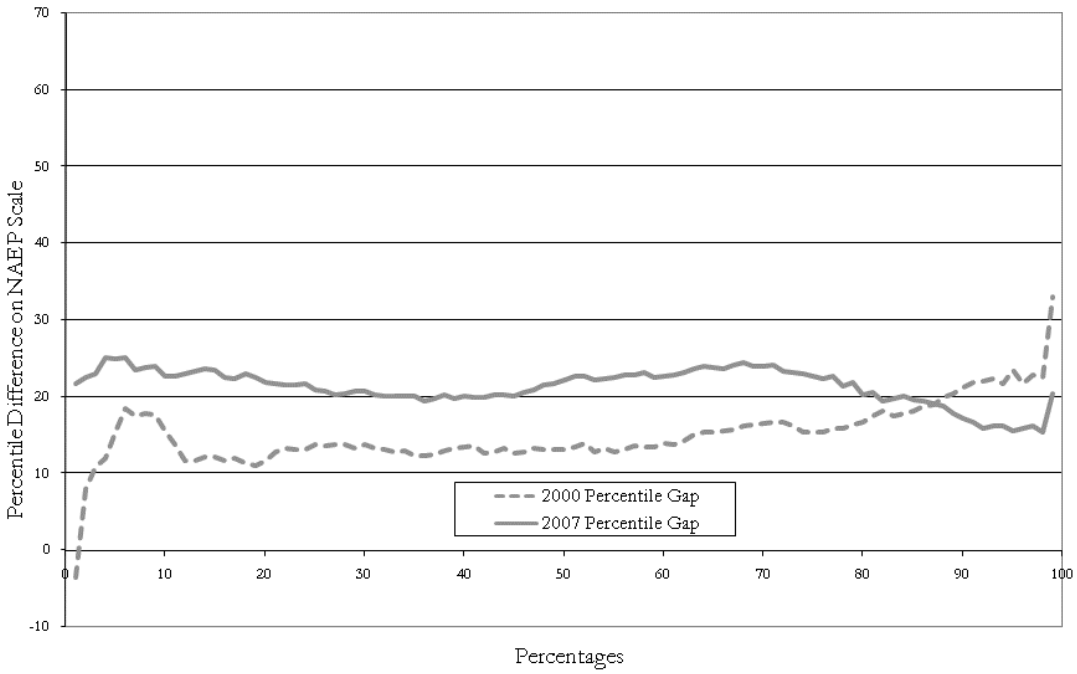

Figure 4. KY: S2: Gap in Percentiles between Whites and Blacks.

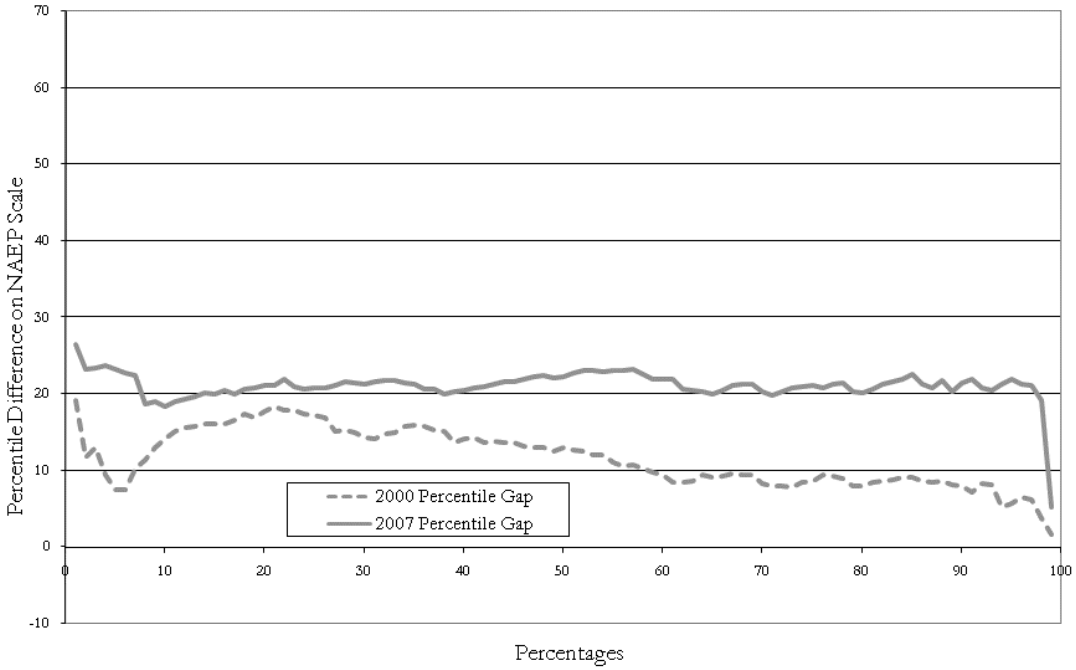

Figure 5. NY: S1: Gap in Percentiles between Whites and Blacks.

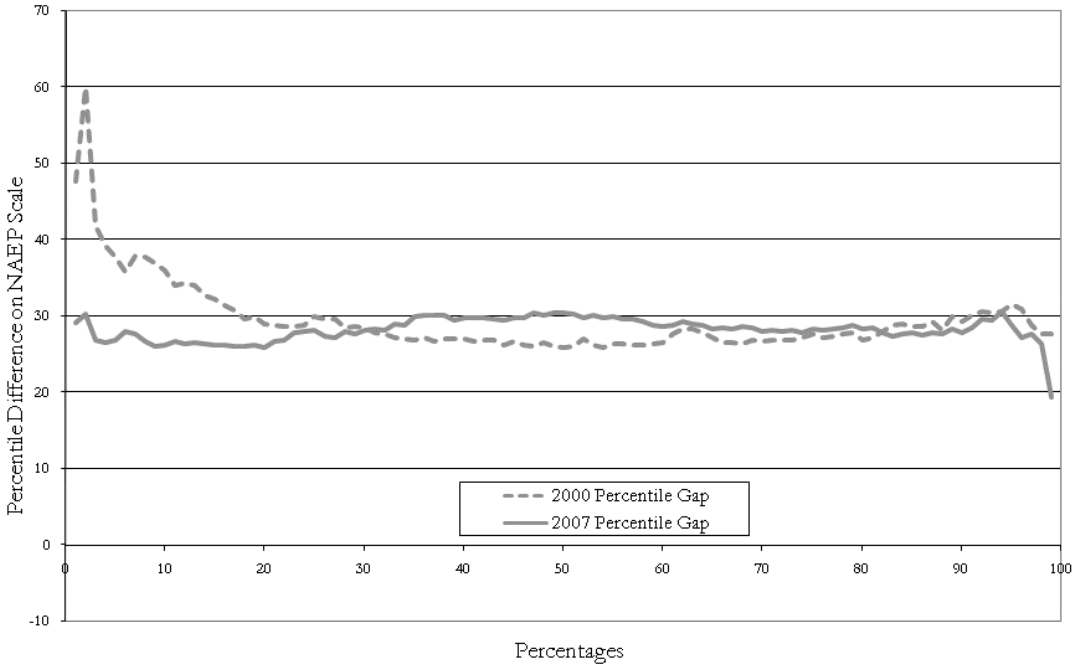

Figure 6. NY: S2: Gap in Percentiles between Whites and Blacks.

Results of HLM Analyses

Changes over the period

Recall that the purpose of the multi-level analyses is to estimate, within each poverty stratum (in each state), the magnitude of the achievement gap after eliminating average differences in achievement among schools within the stratum (in each state). Thus, the results are termed "pooled, within school" estimates of the achievement gaps. There are 20 such estimates, one for each stratum/state combination.

In order to be included in these analyses, a school must have at least one Black student and one White student in the sample of students who took the NAEP assessment. Consequently, some schools in the NAEP school sample cannot contribute data to these analyses. Table 13 presents the counts of schools and students that do contribute data. They are necessarily smaller than those presented in Table 6. In general, the reductions are greater in Stratum 2 than in Stratum 1, attesting to the higher levels of segregation in school populations in higher poverty schools. The reductions are particularly severe in Maryland, Michigan, New York and Tennessee.

Table 14 displays the results for S1. For 2000, the column under "Full School Sample" labeled "Include school effects" contains the estimated gaps (based on the full sample) already presented in Table 11. These should be compared to the gaps in the next column, which are estimated in the same way but are based on the reduced sample. The next two columns display estimates of the adjusted achievement gaps (i.e. eliminating school effects) and estimates of the fully adjusted achievement gaps (i.e. eliminating both school effects and average differences among students on measured covariates). The pattern is repeated for the columns for 2007.
For 2000, the only notable difference between the first two columns occurs for New York, with an increase in the gap from 12 to 23. As noted earlier, the smaller gap appears to be somewhat anomalous. For the other 9 states, the gaps are very similar, and the median gap is the same (28). Turning to the next column,, we observe that the removal of the school effects has only a modest impact, with the median achievement gap remaining unchanged. Adjustment for both school effects and student covariates results in more consistent reductions, with the median achievement gap now 23 points, or 5 points smaller.

The pattern of results for 2007 are similar, with the difference that the removal of school effects yields a reduction in the median achievement gap from 27 to 24 points. Additional adjustment for student covariates yields a further reduction of 5 points. The comparison of the third and seventh gap columns is of particular interest, as it bears directly on the question of whether, from 2000 to 2007, there was a reduction in the achievement gaps in S1, when the gaps are based on pooled, within school estimates. In fact, the reductions ranged from 0 points (Tennessee) to 9 points (California), with a median reduction of 3 points.

Thus, there is evidence of a modest improvement over the period of interest. However, note too, that in both years and for most states the pooled, within school estimates are somewhat larger than the corresponding stratum gaps that were displayed in Table 8. That is, even the "purest" estimates of the achievement gaps are large in magnitude in comparison to other estimates of differences in achievement related to school and student characteristics.

We now turn to Table 15, which displays the results for S2, the higher poverty stratum. The format is identical to that of Table 14. Comparing the first two gap columns, we note that in 2000 for nine of the states (except Kentucky) the estimated gaps based on the reduced sample are smaller than the gaps based on the full sample. The differences from the findings for S1 are due to the greater reduction in the numbers of schools and students, with the likely elimination of highly segregated, more poorly performing schools. Removal of school effects, results in small reductions

for nine states (except Texas), with a further, three point decrease in the median achievement gap to 24 points. Moreover, as displayed in the fourth gap column, adjustment for differences among students in measured covariates results in reductions for 7 states, with another three point decrease in the median gap to 21 points.

Comparing the first two gap columns for 2007, we observe only slight differences in the achievement gaps for the two samples. Removal of school effects has mixed impact and the median gap is only decreased by one point to 22 points. Further adjustment for differences among students results in a four point decrease in the median gap to 18 points. In this case, the comparison of the third and seventh gap columns yields mixed results. Three states (California, Kentucky and New York) experienced increases and the median reduction in the achievement gap over this period was 2.5 points. Thus, there is evidence of only modest progress and the achievement gaps are again large in magnitude in comparison to other gaps related to school and student characteristics.

Table 16, the last table in this section, draws on Tables 14 and 15 to illustrate two different types of comparisons. Columns 1 and 2 display the differences in adjusted achievement gaps between strata for 2000 and 2007, respectively. For example, in 2000, the adjusted achievement gap in California for S1 was 39 points and for S2 it was 17 points. The difference 39-17 = 22 is the entry for California in column 1. We observe that in 2000 the adjusted achievement gaps were typically larger in S1 than in S2 and were also quite variable across states, ranging from -13 points to 22 points.[18] By 2007, the differences between strata had become somewhat smaller with a reduction in the median difference from 7 to 4 points.[19] There was also a dramatic reduction in variability, with entries ranging only from -4 to 9 points. Thus, from 2000 to 2007, with regard to adjusted achievement gaps, strata within states had become more alike and states more homogeneous.

Columns 3 and 4 display the reductions, over the period 2000 to 2007, in adjusted achievement gaps within each stratum. For example, in California S1 the adjusted achievement gap was 39 points in 2000 and 30 points in 2007. The difference 39-30 = 9 is the entry for California in Column 3. We observe that in S1 all states, except Tennessee, experienced a reduction in the adjusted achievement gap over the period. The median reduction of 3 points, however, is rather small and very nearly the same as the median reduction of 2 points in the unadjusted achievement gaps (Table 11). Turning to S2, the median reduction over the period was only 1.5 points and the variability across states increased, ranging from -7 to 11. In fact, three states experienced increases in the adjusted gaps. This stands in contrast to the results for the unadjusted gaps, where the median reduction was 7 points and only one state experienced an increase (Table 11). In general, the patterns and trends for adjusted achievement gaps differ from those for unadjusted achievement gaps. In particular, states' relative rankings for the two strata are more weakly associated than was the case for unadjusted achievement gaps.

[18] This stands in contrast to the results for the unadjusted gaps (Table 11), where the gaps were typically somewhat larger in S2.

[19] The direction here is consistent with the results for the unadjusted gaps (Table 11), where the gaps were typically slightly larger in S1.

Table 13
Numbers of schools and students by state (Reduced school sample).

		2000				2007			
		# of	# of Students			# of	# of Students		
State	stratum	Schools	Total	White	Black	Schools	Total	White	Black
CA	S1	20	500	270	50	70	2100	870	210
	S2	20	300	70	50	80	2200	330	260
KY	S1	40	900	730	130	30	700	560	80
	S2	20	500	410	120	30	700	530	160
MD	S1	70	1700	1080	430	70	1700	950	520
	S2	10	200	30	130	10	300	80	170
MI	S1	20	500	400	40	40	900	690	110
	S2	10	200	40	110	20	300	150	130
NY	S1	20	500	340	60	40	900	650	110
	S2	20	400	100	150	40	1000	230	250
NC	S1	60	1500	1040	400	70	2000	1220	510
	S2	20	500	160	270	60	1600	430	780
SC	S1	50	1200	850	290	50	1100	750	310
	S2	40	1000	450	500	50	1200	550	620
TN	S1	50	1200	950	220	40	1000	760	210
	S2	10	300	150	120	30	700	330	280
TX	S1	40	900	500	150	70	2100	1120	300
	S2	20	400	130	90	70	2800	440	560
VA	S1	80	1900	1360	430	80	2000	1230	440
	S2	10	300	100	160	10	300	80	230

Note: Numbers have been rounded to conform with NCES publication policies.

Table 14

Lower poverty stratum (S1): Black - White achievement gaps from four analyses.

	2000								2007							
	Full School Sample		Minimal Reduced Sample						Full School Sample		Minimal Reduced Sample					
	Include school effects		Include school effects		Remove school effects		Remove school effects and student covariates		Include school effects		Include school effects		Remove school effects		Remove school effects and student covariates	
State	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E
CA	43	(9.3)	40	(8.2)	39	(6.0)	34	(4.7)	29	(3.1)	29	(3.5)	30	(4.7)	26	(4.6)
MD	31	(2.2)	30	(2.4)	29	(2.6)	23	(2.5)	35	(1.9)	35	(2.1)	26	(4.0)	24	(2.2)
TN	26	(4.9)	24	(3.5)	21	(3.3)	20	(3.3)	20	(2.5)	24	(2.5)	21	(3.4)	19	(3.0)
NC	32	(2.0)	33	(2.2)	31	(2.2)	24	(2.1)	26	(2.4)	26	(2.4)	27	(2.8)	20	(2.4)
TX	30	(4.7)	28	(4.5)	28	(3.4)	24	(3.7)	28	(3.2)	27	(3.1)	26	(2.8)	21	(2.0)
MI	29	(3.8)	28	(5.1)	26	(4.9)	20	(4.9)	30	(3.8)	27	(3.2)	23	(4.7)	22	(4.1)
SC	28	(2.3)	28	(2.3)	28	(2.5)	23	(2.5)	25	(2.3)	24	(2.3)	25	(2.7)	18	(3.0)
VA	26	(2.5)	25	(2.5)	25	(2.1)	22	(2.0)	24	(2.1)	24	(2.2)	23	(2.5)	17	(1.9)
NY	12	(7.1)	23	(3.7)	27	(5.0)	20	(4.6)	21	(3.9)	23	(4.5)	20	(4.3)	16	(4.1)
KY	25	(3.0)	26	(3.2)	28	(4.1)	23	(3.8)	28	(4.9)	28	(4.4)	23	(4.8)	19	(4.6)
Median	28		28		28		23		27		27		24		19	

Table 15

Higher poverty stratum (S2): Black-White achievement gaps from four analyses.

	2000								2007							
	Full School Sample		Minimal Reduced Sample						Full School Sample		Minimal Reduced Sample					
	Include School effects		Include School effects		Remove school effects		Remove school effects and student covariates		Include School effects		Include School effects		Remove School effects		Remove school effects and student covariates	
State	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E	Mean	S.E
CA	34	(5.6)	22	(8.3)	17	(8.2)	12	(8.4)	28	(3.1)	23	(3.6)	21	(4.8)	18	(4.8)
MD	30	(11.3)	22	(10.6)	18	(7.9)	18	(8.8)	21	(3.7)	19	(4.9)	17	(6.3)	15	(5.8)
TN	36	(5.2)	37	(9.5)	34	(7.7)	23	(5.4)	26	(3.1)	21	(3.8)	23	(3.0)	24	(3.1)
NC	31	(3.6)	27	(3.7)	24	(3.0)	21	(2.9)	26	(2.8)	23	(2.9)	23	(2.9)	19	(2.6)
TX	34	(3.8)	31	(5.2)	31	(3.1)	29	(5.6)	21	(2.4)	21	(2.9)	23	(2.9)	19	(2.7)
MI	36	(4.0)	30	(7.0)	29	(4.7)	35	(5.9)	29	(4.7)	29	(4.8)	27	(3.9)	24	(4.1)
SC	26	(2.9)	26	(3.1)	24	(2.7)	21	(2.3)	24	(2.1)	24	(2.3)	21	(2.8)	18	(2.7)
VA	33	(5.0)	30	(4.7)	24	(5.4)	21	(4.4)	25	(3.6)	22	(5.5)	19	(6.9)	17	(3.6)
NY	30	(6.2)	26	(7.1)	17	(6.3)	18	(5.7)	28	(3.5)	26	(3.5)	24	(3.4)	17	(3.9)
KY	15	(4.6)	18	(4.4)	13	(6.3)	12	(4.2)	22	(3.4)	22	(3.5)	21	(4.0)	18	(3.2)
Median	32		27		24		21		25		23		22		18	

Table 16
Statistics on Black-White achievement gaps (Reduced school sample with school effects removed).

	Differences in Adjusted Achievement Gaps Between Strata [S1-S2]		Reductions in Adjusted Achievement Gaps Within Strata [2000 to 2007]	
	2000	2007	S1	S2
CA	**22**	**9**	**9**	-4
MD	11	**9**	3	1
TN	**-13**	-2	**0**	**11**
NC	13	4	4	1
TX	-3	3	2	8
MI	-3	**-4**	3	2
SC	4	4	3	3
VA	1	4	2	5
NY	10	**-4**	7	-7
KY	15	2	5	**-8**
Median	7	4	3	1.5

Comparisons with the previous period

In the comparison of the results of the period 1992 to 2000 to the period 2000 to 2007, we will refer to the former as P1 and the latter as P2. In P1, state-level achievement gains ranged from 1.3 to 21.7 points with a median of 9.5 points. The median exclusion rate increased by 4.6 percent. In P2, the achievement gains ranged from 0 to 17 points, with a median of 11.4 points. At the same time, the median exclusion rate increased by only 0.5 percent. Thus, during the later period the typical state experienced a slightly larger increase in overall achievement with no increase in the exclusion rate.

Turning to results at the stratum level, in both periods and for all states, overall achievement increased in both S1 and S2. For S1 the median increase was 11 points in both P1 and P2. For S2, the median increase was 11.5 points in S1 and 15 points in S2. Finally, in P1 the median stratum-level gap increased by two points, while in P2 the median stratum-level gap decreased by three points. Thus, so far our results point to slightly greater progress in the more recent period.

The focus now shifts to results by race. At the state level, in both periods, average achievement increased for both White students and Black students in all ten states. However, in P1 the median reduction in the achievement gap was 0, while in P2 it was five points. At the stratum level, for S1 and S2, average achievement increased for White students and for Black students in all ten states over P1. Over P2, however, the results were not quite so uniform. Nonetheless, the median increases in average achievement were greater over P2 than over P1.

For example, in S1 the median increase for Black students was nine points in P1 and 16 points in P2. In S2, the median increase for Black students in P1 was 11 points and 16 points in P2. There was correspondingly greater progress in reducing the achievement gap. In S1, the median achievement gap decreased by 0.5 points in P1 but by two points in P2. In S2, the median achievement gap decreased by 2.5 points in P1 but by 7 points in P2. Thus, again, our results indicate slightly more progress in reducing the achievement gap in the more recent period.

Turning to consideration of the achievement gaps estimated with the removal of school effects, for S1 the median adjusted achievement gap decreased by 0.5 points in P1 but by three points in P2. By contrast, for S2, the median adjusted achievement gap decreased by 2.5 points in

both periods. Thus, while there was progress for both strata, it was only greater for S1 in the more recent period. In sum, there is little evidence for NCLB having had a substantial impact on the pace of reductions in the achievement gaps.

Linking Results

Recall that a principal goal of the study is to examine the association between states' overall policy rankings (Table 4) and their rankings on selected student outcomes (Tables 10 and 11). It is important to recognize that patterns of association should not be interpreted causally. Rather they may offer useful insights on the nature of the relationships between policy and outcomes. Specifically, they can provide an indication of whether, among the ten states in the study, those states that are judged to be leaders on a number of policy levers are also (relatively) more successful with respect to outcomes. To the extent that is not the case, we would conclude either that the assigned rankings are problematic or that the longitudinal policy profile remains incomplete in important respects. If the latter, then researchers must be more assiduous in ferreting out relevant policy information in order to better understand the links between policy actions and outcomes.

In their review of the literature, Nichols et al. (2006) point out that many earlier investigations of the link between policy and outcomes compared distinct groups of states (e.g. those that had implemented high-stakes accountability and those that had not). However, with the advent of NCLB, all states embarked on policies designed to bring them into compliance with federal law. Accordingly, it is both more appropriate and more convenient to characterize states along a single dimension that summarizes the relevant policy actions and then to correlate states' "scores" on that dimension with their rankings on an outcome measure of interest. In fact, this was the strategy adopted by Braun et al. (2006) and by Nichols et al. (2006). In the former study, rather than computing a rank correlation, the ten states were classified into three ordered categories both on policy and on outcomes. The resulting pattern of association was assessed qualitatively. The rationale was that the accuracy implied by the assignment of numerical ranks on policy would not be commensurate with the nature of the policy data obtained and the subjective judgments employed to obtain an overall policy ranking. Moreover, the qualitative analysis supported by the dual classification of states was adequate for the intended purpose. As noted earlier, we adopted the same strategy for the present study. The linkage pattern for state policy and the outcome "Improving Black student achievement" is displayed in Figure 7. The three rows represent states' classifications with respect to the overall policy composite. Thus, states in the first row were classified as "leaders". Similarly, states in the first column were classified as leaders with respect to the outcome. The figure contains the results for both strata in each state. Perfect alignment would occur if all the state symbols fell in the three cells along the diagonal. Recall from Table 12 that only four states were placed in the same category on this outcome for both S1 and S2. Consequently, there is a limit to the degree of alignment that can be observed between the policy rankings and the dual outcome rankings.

Examination of Figure 7 reveals a modest association, with nine of twenty results located along the diagonal.[20] The association is stronger for lower poverty strata (six of ten) than for higher poverty strata (three of ten). On the positive side, all four states who were leaders on policy had at least one outcome in the leader category. The major outlier was California, with the higher poverty stratum falling in the laggard outcome category. Further, among the three states classified as laggards on policy, two had at least one result in the laggard category on the outcome. Here, North Carolina

[20] The association for this period is somewhat weaker than what was found in the period 1992 -2000. Compare with Figure 6 in Braun et al. (2006).

is the outlier with the lower poverty stratum located in the leader category on the outcome and the higher poverty stratum located in the middle category. One question that arises is whether the modest association is due to the way the categorization on overall policy was carried out. However, the association with each of the policy levers, including "Assessment and Accountability", is weaker than that with overall policy.

Turning to some of the anomalies observed, as noted above California was classified as a laggard on the basis of results for the higher poverty stratum (S2), but a leader for the lower poverty stratum (S1). Interestingly, in both strata the mean scores for Blacks in 2000 were 235. Accordingly, policy makers – and educators – need to investigate why the state's constellation of reforms had greater impact in lower poverty schools. Further, the Holland plots (Figures 1 and 2) show that for S1 the entire score distribution shifted rightward, but for S2 the shift was confined to the lower tail of the distribution. New York was judged to be in the middle category on overall policy but a laggard on results for both S1 and S2. These results were not only weak in relation to the other states but absolutely. That is, even though White students' gains were modest, Black students in S1 lost ground (mean scores for Black students decreased by 3 points) and Black students in S2 gained just about as much as White students. The Holland plots (Figures 5 and 6) show that for S1 there was a uniform shift to the left, while for S2 the improvement was confined to the lower tail of the distribution. New York's weak showing is particularly puzzling as the state was a leader in the period 1992 to 2000.

North Carolina was judged to be a laggard on overall policy based on middle ground and laggard rankings on all four policy levers. At the same time, its outcomes were commendable and just below those of South Carolina, which was judged a leader in overall policy. One possible explanation is that North Carolina was a policy leader during the 1992 to 2000 period and that even though it did not continue to introduce new reforms in the current period, the earlier reforms continued to have an impact as they took hold across the state. [21]

[21] We are indebted to an anonymous referee for a suggestion along these lines.

Overall Policy Rankings	Improving Black Student Achievement-Rankings					
	1		2		3	
1	ca sc tn va	TN		SC VA		CA
2	tx	MD TX	md		ny	NY
3	nc		ky	NC	mi	KY MI

Figure 7. Relationship between overall policy ranking and improving Black student achievement. Cell entries are states: lower case abbreviation denotes the lower poverty stratum and upper case abbreviation denotes the higher poverty stratum.

Note: The states that fall on the diagonal, representing the expected pattern, are in black font, while those states that do not are in red.

Turning to the outcome "Closing the achievement gap", the linkage pattern is displayed in Figure 8. As above, the strength of the alignment with the policy categorization is limited by the fact that only four states were placed in the same category for both S1 and S2. The association is modest, with eight of twenty results located along the diagonal. Again, the association is somewhat stronger for the lower poverty strata. Of the four states classified as leaders on policy, three had at least one outcome in the leader category. South Carolina was the exception, with the higher poverty stratum actually falling in the laggard category. Among the three states in the laggard category on policy, two had at least one outcome in the laggard category. The exception, again, was North Carolina, with the lower poverty stratum falling in the leader category.

One explanation for South Carolina's S2 laggard status is that even though the 18 point gain of Black students over the period was above the ten state median of 16 points, White students gained 16 points, resulting in a reduction of only two points in the achievement gap. On the other hand, the 16 point gain of Black students in North Carolina's lower poverty stratum (S1) was the median value for the ten states, but White students only gained 9 points. The resulting achievement gap reduction was the second largest among the ten states.

These examples point to the difficulty in studying changes in achievement gaps over time. Each change is a difference of differences and, hence, its magnitude is a function of four constituent elements. Ranking the states with respect to these changes adds another level of complexity. Thus, consideration of relative rankings should be supplemented by examination of the absolute gains. Finally, it is difficult to determine the extent to which states' reform policies explicitly addressed the

achievement gap beyond some focus on students of poverty and/or underachieving students, as well as the schools they attend.

Overall Policy Rankings	Closing the Achievement Gap-Rankings					
	1 (Leader)		2		3 (Laggard)	
1 (Leader)	ca tn	TN	sc va	CA VA		SC
2		TX	tx	MD	md ny	NY
3 (Laggard)	nc			MI NC	ky mi	KY

Figure 8. Relationship between overall policy ranking and closing the Black-White achievement gap. Cell entries are states: lower case abbreviation denotes the lower poverty stratum and upper case abbreviation denotes the higher poverty stratum.
Note: Those states that fall on the diagonal, representing the expected pattern, are in black font, while those states that do not are in red.

Discussion

The magnitude and persistence of the achievement gap between majority and minority students is a well-known and much-studied phenomenon. One can argue that since enactment of the first Elementary and Secondary Education Act in 1965, there have ongoing policy efforts to improve the academic performance of all students and, in particular, to eliminate (or at least greatly reduce) the gap among different race/ethnic groups. Although there have been successes over the past 45 years, it is evident from the references cited in the literature review that on many fronts progress has been slow or even stalled altogether. At the same time there is a lack of consensus on what constellation of policies are most likely to be successful, given the constraints under which public schools operate (See, for example, Porter [2005]).

From a methodological perspective, there are a number of challenges in addressing this issue. First, data at the national or even state level can mask the substantial school-level heterogeneity that characterizes American education – a heterogeneity that may contain useful hints for policy makers on "what works" in education reform. With regard to the question of "what works", analyses of data from large-scale cross-sectional assessment surveys such as NAEP can only suggest plausible hypotheses to be subjected to further examination. The difficulties include adequately representing the policy environment in each state over a long period of time and the lack of a randomization basis for causal inference. To these must be added limitations in the data

available with respect to grades, subjects and sample sizes. In particular, despite the large samples now collected by NAEP, once the data are disaggregated by school poverty stratum and race/ethnicity, sample sizes can be rather small, resulting in large estimated standard errors for the estimates of interest.

The present study was undertaken to extend the work reported in Braun et al. (2006) regarding patterns in NAEP performance in 8th grade mathematics from 1992 to 2000, as well as the nature of the relationship between those patterns and the various education initiatives adopted by states. A preliminary goal of the study was to extract, display and interpret trends in the achievement gaps in 8th grade mathematics between Black students and White students (enrolled in public schools) over the period 2000 to 2007. This was done for ten states (educating nearly half of all Black students attending public schools) and at three levels of aggregation: the state, school poverty stratum within state, and school within school poverty stratum within state. Although this goal is a worthwhile one in its own right, it is also a prerequisite to two further aims with more direct relevance to education policy. The first is to compare the results for this period with analogous results for the period 1992 to 2000, thereby providing some evidence regarding the impact of No Child Left Behind on educational achievement. The second is to relate differences in trends among states to differences in their policy efforts over the relevant period. Strong patterns can be suggestive of what combinations of policies are most effective in reducing achievement gaps.

In most respects, the results for the current period are similar to those obtained in the earlier period. For example, in 2007 the range in mean state achievement was about 12 points. At the same time, the stratum gaps within states ranged from 11 points to 34 points, with a median of 21 points. Thus, as before, within state variability was much greater than between state differences. To be fair, in eight of ten states the stratum gaps were smaller in 2007 than in 2000, though only three reductions reached statistical significance. Similarly, considering the twenty stratum/state combinations, White students improved over the period in 19 cases and Black students in 18 cases. There were six states in which the gains for White students and for Black students were statistically significant at the 0.10 level. In the current period, median gains in each stratum for both groups were somewhat larger than the corresponding median gains in the earlier period.

Turning to achievement gaps, the median reduction was 2 points in the lower poverty stratum (0.5 points in the earlier period) and 7 points in the higher poverty stratum (2.5 points in the earlier period). Thus, the pace of improvement was somewhat greater in the current period. There was only modest consistency in states' rankings across the two periods. In the current period, states' rankings by stratum and outcome (increasing Black student achievement and reducing the achievement gap) showed somewhat greater consistency.

The strength of the association between states' policy rankings and states' achievement results in this period appears to be slightly weaker than it was in the previous period. This is not entirely unexpected as the mandates of NCLB have tended to reduce the variability among states on major policy dimensions. As a result, we had to differentiate among states on more nuanced aspects of policy implementation for which it is more difficult to obtain relevant information. As before, the linkage in this period was somewhat stronger for improving Black student achievement than for reducing the achievement gap. Multi-level analyses were also carried out to determine the extent to which systematic differences in average achievement among schools within a stratum may have been confounded with estimates of the stratum level achievement gap. However, the pooled, within school estimates of the achievement gap for a stratum were roughly of the same magnitude as the stratum level estimate. Even further adjustment for a number of measured student characteristics typically resulted in a reduction of only a few points.

Although the ten states certainly differed in their outcomes, the general picture is quite clear: The introduction of high stakes test-based accountability through NCLB has had, at best, a very

modest impact on the rates of improvement for Black students and on the pace of reductions in the achievement gaps between Black students and White students. This is the case at all three levels of aggregation examined in this study. It is most worrisome not only that the magnitudes of these achievement gaps are so large relative to differences between states (and even stratum gaps within states) but also that the annual rate of reduction over a seven year period ranges from less than one-third of a point in S1 to just one point in S2. Linear extrapolation predicts that achievement gaps will remain substantial for the foreseeable future. So we must conclude on the basis of our analysis of the data from 2000 to 2007, as we did on the basis of data from 1992 to 2000, that the achievement gap remains "pervasive, profound, and persistent."

In summary, the association between policy and outcomes was stronger for the overall policy ranking than that for any of the individual policy levers. This suggests that states should adopt a comprehensive reform strategy rather than relying on one that is narrowly focused. In particular, we should not pin our hopes solely on the sort of test-based accountability enshrined in NCLB – it does not appear to be up to the job.

Finally, the modest association between our overall index of states' policy reform efforts and states' results (particularly for schools in the higher poverty stratum) suggest that we are in need of both more powerful theories of reform and a greater capacity to measure the relevant policy dimensions as they play out in schools. Better analytic tools, such as the Holland plots, for the analysis of repeated cross-sectional data, as well as student-level longitudinal data will also contribute to our understanding of the structure of successful school improvement initiatives.

References

Allen, N.L., Johnson, E.J., Mislevy, R.J., & Thomas, N. (1999). Scaling Procedures. In N.L. Allen, J. Carlson, & C.A. Zelenak (Eds.) *The NAEP 1996 Technical Report*, NCES 1999-452 (pp. 75-95). Washington, DC: National Center for Education Statistics.

American Federation of Teacher Research Department. (n.d.). *AFT salary surveys*. Retrieved June 2007 from http://www.aft.org/research/salary/home.htm.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved June 30, 2009 from http://epaa.asa.edu/epaa/v10n18/.

Braden, L., Finn, C.E., Lerner, L.S., Munroe, S., Petrilli, M.J., Raimi, R.A., et al. (2000). *The state of the state standards 2000*. Retrieved September 2007 from http://www.edexcellence.net/detail/news.cfm?news_id=24&id=130.

Braun, H.I. (2004) Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives, 12*(1). Retrieved June 3, 2009 from http://epaa.asu.edu/epaa/v12n1/.

Braun, H. & Qian, J. (2008). *Mapping state standards to the NAEP scale* (Report No. RR-08-57). Princeton, New Jersey: Educational Testing Service.

Braun, H.I., Wang, A., Jenkins, F., & Weinbaum, E. (2006) The Black-White achievement gap: Do state policies matter? *Education Policy Analysis Archives, 14*(8). Retrieved April 20, 2006 from http://epaa.asu.edu/epaa/v14n8/.

Burke, J. & James, J. (1997). *Weighting procedures and variance estimation*. In N.L. Allen et al. (Eds.), *Technical report of the NAEP 1996 state assessment program in mathematics*. Washington, DC: National Center for Education Statistics.

California Commission on Teacher Credentialing. (2001). *Standards of quality and effectiveness for teacher preparation programs for preliminary multiple and single subject teaching credentials*. Retrieved September 2007 from http://www.ctc.ca.gov/educator-prep/standards/AdoptedPreparationStandards.pdf.

California Department of Education. (2004). *Accountability: Fact book 2004*. September 2007 from http://www.cde.ca.gov/re/pn/fb/yr04accountability.asp.

California Department of Education. (2007). *All curriculum frameworks*. Retrieved September 2007 from http://www.cde.ca.gov/ci/cr/cf/allfwks.asp.

Carey, K. (2004). *The funding gap 2004*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/30B3C1B3-3DA6-4809-AFB92DAACF11CF88/0/funding2004.pdf.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" tell us? *Education Researcher, 31*(9), 13-25.

Feuer, M.J. , Holland, P.W., Green, B.F., Bertenthal, M.W., and Hemphill, F.C. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, DC: National Academy Press.

Friedman, B. (2005). *The moral consequences of economic growth*. New York: Knopf.

Hanushek, E.A. & Rivkin, S.G.(2006). *School quality and the Black-White achievement gap. NBER working paper 12651*. Cambridge, MA: National Bureau of Economic Research.

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*, 209-238.

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, 37(6), 351-360.

Hoff, D. (2003). State of the states: Kentucky. *Education Week*, 22 (17), 134.

Holland, P.W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, 27, 3-17.

Ingersoll, R. & Curran, B. (2004). *Out-of-field teaching: The great obstacle to meeting the "highly qualified" teacher challenge.* Retrieved July 2007 from www.nga.org/Files/pdf/0408HQTEACHER.pdf.

Institute for a Competitive Workforce. (2007). *Leaders and laggards: A state-by-state report card on educational effectiveness*. Washington, DC: U.S. Printing Office.

Kentucky Department of Education. (1999). CATS: Commonwealth accountability testing system. Retrieved July 2007 from http://www.kde.state.ky.us/KDE/Administrative+Resources/Testing+and+Reporting+/CATS/Accountability+System/Introduction+to+CATS+PowerPoint.htm.

Kentucky Department of Education. (2007). *Program of study: Revised 2006*. Retrieved July 2007 from http://education.ky.gov/KDE/Instructional+Resources/Curriculum+Documents+and+Resources/Program+of+Studies/default.htm.

Kirsch, I., Braun, H. I., Yamamoto, K. & Sum, A. (2006). *America's Perfect Storm: Three Forces Changing our Nation's Future*. Princeton, NJ: Policy Information Center, Educational Testing Service.

Klein, D., Braams, B.J., Parker, T., Quirk, W., Schmid, W., Wilson,W.S., et al. (2005). *The state of state math standards 2005*. Retrieved September 2007 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=338&pubsubid=1214#1214.

Koretz, D. (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice, 22* (2), 18-26.

Lee, J. (2007). *The testing gap*. Charlotte, NC: Information Age Publishing.

Levin, H.M. (2009). The economic payoff to investing in educational justice. *Educational Researcher, 38*(10), 5-20.

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

Liu, G. (2006). *The funding gap 2006*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/CDEF9403-5A75-437E-93FF-EBF1174181FB/0/FundingGap2006.pdf.

Magnuson, K. & Waldfogel, J. (2008). *Steady gains and stalled progress*. New York: Russell Sage Foundation.

Maryland State Department of Education. (2004). *Maryland teacher professional development standards*. Retrieved October 2007 from http://www.mdk12.org/instruction/professional_development/teachers_standards.html.

Michigan Department of Education. (1996). *Michigan curriculum framework*. Retrieved July 2007 from www.michigan.gov/documents/MichiganCurriculumFramework_8172_7.pdf - 2005-11-17.

Michigan Department of Education. (2004). *Design and validity of the MEAP test*. Retrieved July 2007 from http://www.michigan.gov/mde/0,1607,7-140-22709_31168-94522--,00.html.

Michigan Department of Education. (2007 a). *Curriculum and standards*. Retrieved July 2007 from http://www.michigan.gov/mde/0,1607,7-140-28753---,00.html.

Michigan Department of Education. (2007 b). *Professional development*. Retrieved July 2007 from http://www.michigan.gov/mde/0,1607,7-140-6530_5683_5703---,00.html.

Morris, J.E. & Monroe, C.R. (2009). Why Study the US South? The nexus of race and place in investigating black student achievement. *Educational Researcher, 38* (1), 21-36.

National Center for Education Statistics. (n.d.a). *Early childhood longitudinal program*. Retrieved September 2009 from http://nces.ed.gov/ecls/Kindergarten.asp.

National Center for Educational Statistics. (n.d.b). *National public education financial survey: Fiscal years 1990-2002*. Retrieved June 2007 from: http://nces.ed.gov/ccd/bat/.

National Center for Education Statistics. (2006). *Current expenditures for public elementary and secondary education: School year 2003-2004*. Retrieved June 2007 from: http://nces.ed.gov/pubs2006/expenditures/a1.asp.

National Center for Education Statistics (2007a). *The Nation's Report Card: Mathematics 2007*. (NCES 2007-494).Washington DC: National Center for Education Statistics.

National Center for Education Statistics. (2007b). *Expenditures for public elementary and secondary education: School year 2004-2005*. Retrieved June 2007 from: http://nces.ed.gov/pubs2007/2007356.pdf.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, D.C.: U.S. Department of Education.

New York State Education Department. (n.d.). *Core curriculum*. Retrieved January 2008 from http://www.emsc.nysed.gov/ciai/cores.htm.

New York State Education Department. (2005 a). *Introduction to the grades 3-8 testing program in English language arts and mathematics*. Retrieved December 2007 from http://www.emsc.nysed.gov/3-8/intro.pdf.

Nichols, S.L., Glass, G.V. & Berliner, D.C. (2006). High stakes testing and student achievement: Problems for the No Child Left Behind Act. *Education Policy Analysis Archives*, 14(1). Retrieved April 18, 2010 from http://epaa.asu.edu/epaa/v14n1/.

North Carolina Department of Public Instruction. (2005). *North Carolina standard course of study*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/curriculum/2005ncscos.pdf.

O'Day, J. & Smith, M. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing Coherent Education Policy: Improving the system* (pp. 313-322). San Francisco: Jossey-Bass.

Porter, A.C. (2005). Prospects for school reform in closing the achievement gap. In C.A. Dwyer (Ed.), *Measurement and research in the accountability era* (pp. 59-95). Mahwah, NJ: Lawrence Erlbaum.

Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.

Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html.

Raudenbush, S.W. (2009). The Brown legacy and the O'Connor challenge: Transforming schools in the images of children's potential. *Educational Researcher, 38*(3), 169-180.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S.W., Bryk, A., Chong, Y.F., & Congdon, R. (2004). HLM6: Hierarchical linear and non-linear modeling. Lincolnwood, IL: Scientific Software International Inc.

Skinner, R. (2005). State report cards. *Education Week, 24*, 106-137.

South Carolina Department of Education (2007). *State board*. Retrieved August 2007 from http://ed.sc.gov/.

South Carolina Education Oversight Committee. (2004). *Report on the review of the South Carolina science curriculum standards*. Retrieved August 2007 from https://ed.sc.gov/agency/offices/cso/science/documents/EOCScienceStandardsReviewReport.doc.

South Carolina General Assembly. (1998). *112th session, 1997-1998*. Retrieved August 2007 from http://www.scstatehouse.net/sess112_1997-1998/bills/850.htm.

South Carolina Legislature. (1998). *South Carolina code of laws: Title 59*. Retrieved August 2007 from http://www.scstatehouse.net/code/t59c018.htm.

State of California (1998). *STAR 1998: Standardized testing and reporting program*. Retrieved September 2007 from http://star.cde.ca.gov/star98/.

State of California (2000) *STAR 2000: Standardized testing and reporting program*. Retrieved September 2007 from http://star.cde.ca.gov/star2000f/.

State of California. (2001). *About STAR 2001*. Retrieved September 2007 from http://star.cde.ca.gov/star2001/help/AboutSTAR.html.

Swanson, C. (2008). Grading the states. *Education Week, 27,* 36-57.

Tennessee Department of Education. (n.d. a). *Blueprint for learning*. Retrieved October 2007 from http://www.state.tn.us/education/ci/standards/blueprint/index.shtml.

Tennessee Department of Education. (n.d. b). *TCAP: Frequently asked questions*. Retrieved October 2007 from http://www.tennessee.gov/education/assessment/tsachfaq.shtml.

Tennessee Department of Education. (2006). *Curriculum standards*. Retrieved October 2007 from http://www.state.tn.us/education/ci/standards/.

Virginia Department of Education. (1998). *Routes to licensure in Virginia*. Retrieved November 2007 from http://www.pen.k12.va.us/VDOE/newvdoe/licroute.htm.

Virginia Department of Education. (2004). *High-quality professional development criteria*. Retrieved December 2007 from http://www.pen.k12.va.us/VDOE/nclb/HQPDcriteria4-04.pdf.

About the Authors

Henry Braun
Boston College
Henry Braun is the Boisi Professor of Education and Public Policy in the Lynch School of Education at Boston College. His interests include educational accountability, value-added modeling, large-scale assessments, statistical methodology and assessment design. He can be reached at braunh@bc.edu.

Lauren Chapman
Boston College
Lauren Chapman is a doctoral candidate in the Educational Research, Measurement and Evaluation department in the Lynch School at Boston College. Her research interests include the implementation and evaluation of school-wide reforms, both within the United States, as well as internationally. She can be reached at chapmala@bc.edu.

Sailesh Vezzu
Educational Testing Service
Sailesh Vezzu is a Senior Research Data Analyst in the Research and Development Division at Educational Testing Service in Princeton, NJ. He received Master's degrees in Statistics and Computer Sciences from Kansas State University in December 2004. His interests include statistical computing, sampling, multivariate analysis, hierarchical linear modeling, statistical graphics, and application development. He can be reached at svezzu@ets.org.

education policy analysis archives

Volume 18 Number 21 September 10th, 2010 ISSN 1068–2341

Appendix A

This appendix provides some detail on the variables and coding for the vector **X** employed in model (2) of the multi-level analyses to generate estimates of the fully adjusted achievement gaps. Numbers of respondents presented below are raw counts taken from the 2000 assessment. Variables and coding were essentially the same for both assessments.

The first characteristic is Home Environment. It is a composite constructed from responses to four questions relating to the types of literacy materials at home. We employed four categories based on the number of types: 0-2, 3, 4 and omitted.[22] These generate three indicator functions, with 0-2 types as the base category. The second characteristic is Parental Education. There are five categories: Did not finish high school, graduated high school, some education after high school, graduated college, I don't know or omitted.[23] These generate four indicator functions, with "did not finish high school" as the base category.

The third characteristic is eligibility for free or reduced price lunch. The categories are: Not eligible, eligible and information not available. These generate two indicator functions, with "not eligible" as the base category.[24] The fourth characteristic is number of days absent in the previous month. The categories are: None, 1-4 days, 5 or more days, omitted.[25] These generate three indicator functions, with "none" as the base category.

[22] The number omitted is 113 out of 14,737 responses. The numbers in the other categories are 3100, 4684 and 6840.

[23] The number responding "I don't know" is 1442 and the number omitted is 49.

[24] The number with "information not available" is 1170.

[25] The number omitted is 160.

Appendix B: North Carolina Profile

CONTEXT

Education System

During the 1998-1999 school year, there were 1,254,821 total students enrolled in public school in the state of North Carolina. There were a total of 79,531 teachers during that period, thus yielding a 15.8 average student to teacher ratio. At that time, North Carolina had a total of 180 school districts, with 2,106 total schools (NCES, 2000). By the 2004-2005 school year, the number of total public school students increased to 1,385,754, total teachers increased to 92,550 and the average student to teacher ratio was 15.0. There were also increases in total school districts and total schools, then at 214 and 2,290 respectively (NCES, 2007 c).

GOVERNANCE

After North Carolina's accountability system, the ABC's of Public Education, was implemented in the 1996-1997 school year, the state began a period of increased accountability and standards. "Dramatic gains on national assessments and other achievement measures have moved North Carolina from the bottom on state rankings to about average over the past several years" (Kennedy Manzo, 2001, p. 28). Under the leadership of a governor with a "near-singular focus on education", the state began the 1998-2005 time period with lofty goals for improvements in student achievement (Kennedy Manzo, 2001, p. 28).

The North Carolina State Board of Education (SBE) is responsible for "supervising and administering 'the free public school system and the educational funds provided for its support'" (NC SBE, n.d. a). The board is made up of the state Lieutenant governor, the Treasurer, as well as eleven members appointed by the governor, who serve overlapping eight year terms (NC SBE, n.d. a). Eight of the eleven members represent the eight education districts in North Carolina, while the remaining three positions can be awarded to someone from any area in the state. The governor's appointees must be approved by the North Carolina General Assembly (NC SBE, n.d. a).

The 1998-2005 period in North Carolina began under the leadership of Governor James Baxter Hunt Jr., who served a total of four terms in the gubernatorial position, leaving office in 2001. He was characterized as "a man whose passion and persistence on school policy issues have made him one of the nation's most influential 'education governors'" (Kennedy Manzo, 2001, p. 28). The Democratic Hunt, was succeeded by a fellow Democrat, Governor Michael F. Easley, who remained in office through the rest of the examined time period.

FINANCE

Table B1 illustrates the percent that each state contributed towards their total education revenues. North Carolina was below both the national average and ten state average throughout the examined time period. North Carolina's distribution increased between the 1998-1999 and 2000-2001 period, then decreased from the 2001-2002 through 2004-2005 school years. North Carolina's percent distribution went from 42.2% in 1998-1999 to 43.0% in 2004-2005.

Table B1
Trends in State Proportion of Revenues for Public Elementary and Secondary Education

State or Jurisdiction	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005
Nation	**48.7**	**49.5**	**49.7**	**49.2**	**48.7**	**48.2**	**46.9**
California	59.3	60.3	61.5	59.4	58.9	55.6	59.2
Kentucky	61.8	60.7	59.9	59.6	58.8	57.3	56.7
Maryland	39.5	39.0	37.3	37.2	38.3	38.1	37.7
Michigan	64.7	64.6	64.8	64.6	63.3	61.8	60.1
New York	68.7	67.6	66.3	64.5	63.7	62.9	62.7
North Carolina	**42.2**	**44.8**	**46.2**	**48.2**	**45.6**	**43.2**	**43.0**
South Carolina	52.1	52.8	53.9	51.0	48.1	46.1	45.3
Tennessee	47.2	45.8	44.3	43.6	43.8	42.9	43.2
Texas	42.4	44.2	42.2	40.8	40.9	38.6	35.9
Virginia	33.8	42.6	42.3	40.9	39.6	38.8	40.6
Ten State Average	**51.2**	**52.2**	**51.9**	**51.0**	**50.1**	**48.5**	**48.4**

Source: U.S. Department of Education, National Center for Educational Statistics (n.d.)

Table B2 illustrates the total expenditures per student for the 1998-1999 through 2004-2005 school years. As shown in the table, North Carolina's average total expenditures per pupil were consistently lower than the national average, as well as the ten state average. They continuously increased throughout the 1998-2005 period however, beginning at $5,656 during the 1998-1999 school years, and growing to $6,904 during the 2004-2005 school year.

Table B2
Trends in Total Expenditures per Pupil

State or Jurisdiction	Total Current (Unadjusted U.S. $) 1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005
Nation	**7013**	**7394**	**7904**	**8259**	**8600**	**8310**	**8701**
California	5801	6314	6987	7405	7552	7673	7905
Kentucky	5637	5921	6079	6523	6661	6861	7132
Maryland	7326	7731	8260	8692	9153	9433	10031
Michigan	7432	8110	8278	8653	8781	9094	9340
New York	9344	9846	10716	11218	11961	12638	13703
North Carolina	**5656**	**6045**	**6340**	**6495**	**6562**	**6613**	**6904**
South Carolina	5656	6130	6631	7017	7040	7177	7549
Tennessee	5123	5383	5687	5948	6118	6466	6850
Texas	5685	6288	6539	6771	7136	7151	7246
Virginia	6350	6841	7281	7496	7822	8219	8886
Ten State Average	6401	6861	7280	7622	7879	8133	8555

Sources: National Center for Educational Statistics. (n.d.). *National public education financial survey: Fiscal years 1990-2002*. Retrieved June 2007 from: http://nces.ed.gov/ccd/bat/.
National Center for Educational Statistics. (2007 a). *Expenditures for public elementary and secondary education: School year 2004-2005*. Retrieved June 2007 from: http://nces.ed.gov/pubs2007/2007356.pdf.
National Center for Educational Statistics. (2006 a). *Current expenditures for public elementary and secondary education: School year 2003-2004*. Retrieved June 2007 from: http://nces.ed.gov/pubs2006/expenditures/a1.asp.

Table B3 contains data illustrating the poverty gap over the 1997-2004 school years. For the state of North Carolina in 1997 there was a $464 gap between the highest and lowest-poverty districts, with students in the highest poverty schools receiving less funding than students in the lowest poverty schools. This gap was smaller than the national average of $1,208. In North Carolina, this gap increased between 1997 and 2001 to $751, then decreased between 2001 and 2004, during which the students in highest poverty schools were receiving $622 less per student. The total change in the gap between the 1997 and 2004 period increased by $79 per student, lower than the national average for the change in gap during that same period which increased by $99 per student.

Table B3
State and Local Funding Gaps over Time: 1997-2004[1]

State	Gap Between Highest and Lowest-Poverty Districts 1997	Gap Between Highest and Lowest-Poverty Districts 2001	Gap Between Highest and Lowest-Poverty Districts 2002	Gap Between Highest and Lowest-Poverty Districts 2004	Poverty Gap Change in Dollars 1997-2004
Nation	($1,208)	($1,287)	($1,348)	($1,307)	($99)
California	($205)	($418)	($301)	($259)	($54)
Kentucky	($119)	($143)	($357)	$448	$567
Maryland	($961)	($735)	($772)	($432)	$529
Michigan	($1,407)	($1,099)	($1,085)	($1,072)	$335
New York	($2,938)	($2,264)	($2,615)	($2,927)	$11
North Carolina	($464)	($751)	($622)	($543)	($79)
South Carolina	($370)	($343)	$43	$127	$497
Tennessee	$124	$536	$281	$330	$206
Texas	($437)	($875)	($936)	($757)	($320)
Virginia	($972)	($1,341)	($1,430)	($436)	$536

[1]Cost-adjusted dollars, 40% adjustment for low-income students

Sources: Carey, K. (2004). *The funding gap 2004*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/30B3C1B3-3DA6-4809-AFB92DAACF11CF88/0/funding2004.pdf.

Liu G. (2006). *The funding gap 2006*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/CDEF9403-5A75-437E-93FF-EBF1174181FB/0/FundingGap2006.pdf

There were several court cases during the 1998-2005 period dealing with state funding for low-wealth districts. The original court case was brought about in 1994 by five low-wealth districts alleging that the state was under funding schools with disadvantaged children (Kennedy Manzo, 2002). In April 2002, and again in July 2004, the North Carolina Supreme Court ruled that the state was not "meeting its constitutional duty to provide a sound basic education for students in needy districts" (Kennedy Manzo, 2004, p. 23).

TEACHER QUALITY

Certification

In 2005, *Education Week* rated North Carolina "in the top tier of states this year" in their efforts to improve teacher quality (Skinner, 2005). Teaching licenses in North Carolina were available in the areas of birth through kindergarten, elementary (K-6), middle grades (6-9), secondary grades (9-11), special subjects (K-12), exceptional children (K-12), as well as vocational education (NC SBE, n.d. c). Teacher candidates were required to pass the Praxis II subject tests in order to qualify for certification. As of 2005, North Carolina was only one of 16 states that required middle and high school teacher candidates to take subject knowledge tests in addition to basic-skills and subject-specific pedagogy tests (Skinner, 2005).

Once a teacher candidate met all of the North Carolina certification requirements, they were issued one of two licenses; initial or continuing (also called the Standard professional 2 license). Initial licenses were issued to teachers who were either inexperienced, or from a state that did not have a reciprocal licensing agreement with the state of North Carolina. As of the year 2000, in order for a candidate to be issued an initial license, they first had to complete a bachelor's degree program, as well as receive an endorsement from the Institutes of Higher Education (IHE), who determined whether or not the candidate had satisfied requirements, including score requirements on their teaching examinations, prior to recommending the teacher candidate (NC SBE, 2000b). After 1998, teacher candidates who were issued initial licenses were required to complete a three-year Initial Licensure Program through a North Carolina school district (NC SBE, n.d. c). Continuing licenses were issued to teachers who were considered experienced, having had at least three years of teaching in a public school. Continuing licenses were issued on five-year renewal cycles, and had to be renewed at the end of the cycle in order for the license to remain valid (NC SBE, n.d. c). In accordance with NCLB requirements, beginning in the 2002-2003 school year new teachers in Title I schools had to qualify, both for a full North Carolina state license, but also be considered "highly qualified" in order to teach a core academic subject (NC SBE, 2002a).

Professional Development

In order for teachers to renew their continuing (Standard Professional 2) teaching licenses, they were required to participate in professional development. The Excellent Schools Act, approved in 1997, placed increased emphasis on "meaningful and continued professional development opportunities aligned with State Board goals and directed toward improved student achievement" (NC SBE, n.d. c). The general policies regarding teacher license renewal requirements were revised by the North Carolina SBE during the examined period in the years 1998, 2003, and 2005. According to the most recent revision to these requirements, made in 2005, in order for a teacher to

renew their continuing license, they had to participate in 15 units of renewal credits. In order to obtain these renewal credits, teachers could take classes at a college or university, earn National Board for Professional Teacher Standards certification or its renewal, do an independent study, or participate in other activities that increased the teacher's knowledge in their licensure area (NC SBE, 2005c).

Teacher Compensation

Table B4 presents the trends in average teacher salaries from the 1998-1999 through 2004-2005 school years. As shown in the table, North Carolina's average teacher salary was below the national average in every year. The state's average teacher salary was also consistently below the ten state average. While North Carolina's average teacher salary increased continuously from $36,883 in 1998 to $43,343 in 2005, it still remained below the national average.

Table B4
Trends in Average Teacher Salary (in unadjusted U.S. dollars)

	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005
Nation	**$40,574**	**$41,820**	**$43,250**	**$44,367**	**$45,578**	**$46,565**	**$47,602**
California	46,326	47,680	52,480	54,348	55,673	56,444	57,604
Kentucky	35,383	36,255	36,688	37,951	38,485	40,310	41,075
Maryland	42,545	43,720	45,963	48,251	49,679	50,303	52,330
Michigan	48,711	48,729	50,515	52,497	53,178	52,161	53,959
New York	49,686	51,020	51,020	51,020	53,017	55,181	55,665
North Carolina	**36,883**	**39,404**	**41,496**	**42,118**	**42,411**	**43,211**	**43,343**
South Carolina	34,506	36,081	37,938	39,923	40,362	41,162	42,189
Tennessee	35,490	36,328	37,413	38,515	39,186	40,318	42,076
Texas	34,448	37,567	38,359	39,230	39,972	40,476	41,009
Virginia	37,709	38,992	40,247	41,752	42,677	43,936	45,377
Ten State Average	**40,169**	**41,578**	**43,212**	**44,561**	**45,464**	**46,350**	**47,463**

Source: American Federation of Teacher Research Department (n.d.). Retrieved June 2007 from http://www.aft.org/research/salary/home.htm

Out-of-field Teaching

While the North Carolina SBE did not permit out-of-field teaching, provisional licenses could be issued for teachers' assigned out-of-field (NC SBE, 2002a). These provisional licenses were valid for one year. The end of the 2005-2006 school year marked the last year that teachers were able to obtain provisional licenses for the elementary grades or for middle and high school core academic subjects (NC SBE, 2003).

CURRICULUM AND STANDARDS

Curriculum and Standards

The North Carolina *Standard Course of Study* was the statewide curriculum. It was made up of content standards, and a curriculum that highlighted what students should be able to do as they progressed through the grades, focusing on themes and concepts as opposed to facts (NC Department of Public Instruction, 2005). Subjects featured in the *Standard Course of Study* included English/language arts, mathematics, science, social studies, arts education, second language studies, healthful living, information/computer skills, guidance and workforce development education (NC Department of Public Instruction, 2005). It was originally created in the 1890's, with the most recent total revisions made in 1985. Further revisions had taken place during the 1998-2005 time period, including the information skills curriculum which was revised in 1999, the arts education and healthful living curriculums were revised in 2000, the English language development and social studies curriculums were revised in 2003, and the English language arts and second language curriculums were revised in 2004. The science curriculum was revised in 1999, and again in 2004. The mathematics curriculum was revised in 1998 and again in 2003 (NC Department of Public Instruction, 2005). According to the Thomas B. Fordham Institute's, *The State of the State Standards 2000*, after the first revisions in 1998, the revised math standards were "much like its predecessor, which was one of the best, and though now slightly degraded it is still one of the best" (Braden et al., 2000). However, the 2003 revisions, examined in *The State of the State Math Standards 2005,* were referred to as "a misstep" (Klein, et al., 2005). After the 2003 revisions, the report went on to say that while North Carolina's math standards were still "reasonably clear", the "content coverage is mediocre at all levels, with pervasive shortcomings such as an overemphasis on patterns, data analysis, and probability, and an inappropriate use of technology" (Klein, et al., 2005).

According to the Institute for a Competitive Workforce (ICW), an affiliate of the U.S. Chamber of Commerce, North Carolina overall received "a middling grade for the rigor of its standards" (2007, p. 44). North Carolina's standards were characterized as clear and specific in all subjects in elementary, middle, and high school, except for the social studies/history standards that were only considered to be clear and specific at the high school level (Skinner, 2005).

Table B5 presents data from mapping the individual state proficiency standards in mathematics onto the NAEP score equivalents. Using the 2003 NAEP data, the estimated NAEP score equivalent to the North Carolina state standards was 247, lower than then NAEP *Basic* cut score, which is set at 262. At this level, it was estimated that around 82% of 8th grade math students in North Carolina were meeting this state proficiency standard. In 2005 North Carolina's estimated NAEP score equivalent remained the same at 247. It was estimated that an increased 84% of 8th grade math students in North Carolina were meeting this state proficiency standard in 2005.

Table B5

Results of Mapping State Standards to the Grade 8 NAEP Mathematics Scale

Year	State	Number of schools used in mapping	Estimate of proportion meeting the state proficiency standard	Estimated NAEP score equivalent to the state standard	Estimated standard error of the NAEP score equivalent
2003	Kentucky	112	0.32	291	1.2
	Maryland	95	0.43	286	1.2
	Michigan	105	0.51	278	1.4
	North Carolina	**129**	**0.82**	**247**	**2.1**
	New York	141	0.54	279	1.4
	South Carolina	92	0.20	306	1.5
	Texas	142	0.71	260	1.2
2005	Kentucky	115	0.37	285	1.4
	Maryland	105	0.53	276	1.7
	Michigan	111	0.61	269	1.9
	North Carolina	**133**	**0.84**	**247**	**1.2**
	New York	173	0.56	275	0.9
	South Carolina	104	0.24	305	1.1
	Tennessee	111	0.88	230	1.6
	Texas	270	0.61	273	0.8

Note: NAEP mathematics cut scores at grade 8 are 262 for *Basic* and 299 for *Proficient*. CA and TN (2003) data not available

Source: National Center for Educational Statistics. (2007 b). *Mapping 2005 state proficiency standards onto the NAEP scales: Research and development report*. Washington, DC: U.S. Printing Office.

Textbooks

The textbook adoption process was led by the North Carolina Textbook Commission, a board made up of twenty-three members, all appointed by the Governor. Every year publishers were invited to submit textbooks to be evaluated by Regional Textbook Advisory Committees, which were appointed by the Textbook Commission. These committees evaluated the materials, paying special attention to how close the textbooks aligned with the *Standard Course of Study*. The Textbook Commission used these evaluations from the committees to formulate its own list of textbooks that it presented to the North Carolina State Board of Education. The Board then chose textbooks for state adoption, considering the Textbook Commission's suggestions. The approved state list of textbooks was then sent to local districts to choose the textbooks that best met the needs of their students (NC SBE, 2000d). There were not any major revisions to the North Carolina state textbook adoption process during the years 1998-2005.

ASSESSMENT AND ACCOUNTABILITY

Accountability System

North Carolina's accountability system, known as the ABC's of Public Education, was first implemented in the 1996-1997 school year, and continued to be used throughout the 1998-2005 period. The system was based on three main components: A, strong accountability; B, mastery of basic skills; and C, localized control (NC SBE, 2006b). The model used year-end test scores in reading and mathematics during grades 3 through 8 to determine if a school was meeting growth expectations based on the percentage of students scoring at or above the proficiency standard (NC SBE, 2006b).

In North Carolina, students were given a pre-test in grade 3 and were tested in mathematics and English/language arts every year in grades 3 through 8. The North Carolina High School Comprehensive Test was first given to 10th grade students in April 1998. Due to budget issues the test was abolished in 2001, but was reinstated in 2002-2003 in order to comply with NCLB requirements (NC SBE, 2006b). There was not any state administrated science or social studies/history tests for students in grades 3 through 8 during the 1998-2005 time period. During the 2005-2006 school year, the first online computer skills assessment was given to 8th graders in the state.

End-of-course tests were developed in response to the North Carolina Elementary and Secondary Reform Act of 1984, and were offered for the following subjects; algebra I, algebra II, biology, chemistry, economic, legal and political systems, English I, geometry, physical science, and physics (NC SBE, n.d. b). Administration of end-of-course tests was mandated during the 1998-1999 school year, whereas prior to that point, school systems had the option of administering the tests (NC SBE, 2000c). These assessments were later updated to reflect the updated state curriculum and standards. Based on these revisions, the updated end-of-course biology exam was first administered statewide in the fall of the 2001-2002 school year (NC SBE, 2005b). The updated end-of course algebra I exam was first administered statewide in the spring of 2000-2001 (NC SBE, 2004b). The updated end-of-course civics and economics (adapted from economic, legal and political systems), and U.S. history exams were first administered operationally statewide during the fall of 2005 (NC SBE, 2007a ; NC SBE, 2007b). Beginning in the 2001-2002 school year, the results

from end-of-course tests were required to be calculated into at least 25% of students' final grade in the course (NC SBE, 2004a).

An Alternate Assessment Portfolio (AAP) was added to the assessment system in the spring of 2000. This yearlong assessment allowed students with disabilities to demonstrate what they learned in addition to meeting the goals in their Individualized Education Plan (IEP) (NC SBE, 2004c). Another assessment developed for students with disabilities was the Alternate Assessment Academic Inventory (AAAI). The AAAI was originally piloted in 2000-2001 school year, and then operationally administered in the 2001-2002 school year (NC SBE, 2004c). The AAAI was designed for students with IEP's or Section 504 plans, who are unable to take the state assessments under normal testing conditions. As part of the assessment, students were evaluated by their classroom teachers on the tested objectives three times a year; at the beginning of the school year, midyear and then again at the end of the school year. Teachers used an academic checklist to grade students based on the objectives, determining whether or not the student was performing on grade level, as well as assigned the student an achievement level (NC SBE, 2004c).

The annual North Carolina School Report Cards were first produced in 2001. They were provided to parents both by the districts in the form of printed documents called "Snapshots", as well as available online. The state report cards provided parents, educators and citizens with information regarding academic performance, school climate and safety, as well as teacher quality, on statewide, district and local school levels (NC SBE, 2007c).

The North Carolina ABC's program resulted in rewards and sanctions tied to accountability. These rewards and sanctions were included in the original implementation of the act in 1996, with refinements made in incorporating elements of NCLB. Schools that made higher than expected growth were given recognition based on the percentage of their students that scored at or above a specified achievement level. Teacher incentive rewards were awarded to staff members in schools that met the high growth standard, and smaller incentive rewards were awarded to staff members in schools that met the expected growth standard. Schools that did not meet their expected growth standard and were labeled "low performing" were subject to a variety of sanctions, including notifying the parents of students who attended the schools, and state assistance (NC SBE, 2001).

Student Accountability

In 1999, the ABC accountability system was expanded to include student accountability. Grade promotion requirements were set for students in grades 3, 5, and 8, as well as requirements for high school graduation. These requirements included scoring at least proficient on the end-of-grade reading and math tests, as well as on the 4th and 7th grade writing tests. For high school the requirements included, in addition to meeting local and state graduation requirements, getting a passing score on the 11th grade high school exit exam, and passing the computer skills test given to students while in 8th grade (NC SBE, 2000a). North Carolina's graduation requirements, including multiple courses of study, were first implemented for the incoming 9th graders that entered high school during the 2000-2001 school year (NC SBE, 1999). The courses of study included career preparation, college/technical preparation, college/university preparation and the occupational course of study for students with disabilities. The resulting requirements were different based on students' chosen course of study. According to the ICW, North Carolina "has yet to align its high

school graduation requirements with college and workplace expectations or to enact a rigorous graduation exit exam" (2007, p. 44).

In 2004, the graduation rate for students in all public schools in the state of North Carolina was 66.1%. Between the years 2000 and 2004 the graduation rate within the state decreased by 5.8% (Swanson, 2008).

School Accountability

The ABC's accountability program was a school-based accountability model. In concentrating the accountability at the school, rather than district level, the state was able to give incentives to educators in schools that excelled and provide extra assistance to schools that were lower performing (NC SBE, 2002b). In 2001, new legislation was passed that provided increased assistance to low-performing schools, including reducing class sizes in grades K-3, hiring more instructional support staff, increasing the number of staff development days, and adding signing bonuses for mathematics, science and special education teachers (NC SBE, 2002b).

State Assessment System (mathematics)

When the 1998-2005 period began, students in North Carolina were given mathematics assessments based on the *Standard Course of Study* curriculum that had last been revised in 1989 (grades K-8) and 1992 (9-12). During the 2000-2001 school year, the mathematics assessments were revised to correspond with the *Standard Course of Study* under the 1998 revisions. Along with curricular revisions, changes on the grade 3-pretest, end-of-grade tests (grades 3-8), and end-of-course tests (algebra I and II, geometry) were also made (NC SBE, 1997). In addition, field test items were no longer embedded in the mathematics exams, but instead were done during stand-alone administrations (NC SBE, 2000b). Beginning with the 2005-2006 school year, the end-of-year mathematics tests for grades 3-8 were revised to correspond with the 2003 North Carolina mathematics *Standard Course of Study* (NC SBE, 2005a). Differences included the development of five strands of mathematics for grades K-8, extending the integrated mathematics sequence to a fourth year, as well as adding the new high school mathematics elective, advanced functions and modeling (NC SBE, 2005d).

LOOKING BEYOND

According to *Education Week's 2008 Quality Counts* report, North Carolina received an overall grade of C for its policies through the 2007-2008 school year. For the K-12 Student Achievement Index, which examines status, change, and equity, North Carolina received a grade of D+. During the 2003-2007 period, the NAEP math scale score for fourth grade students decreased by 0.4 points, and for eighth grade students increased by 2.6 points. During this same period, the poverty-gap change in scores on the NAEP mathematics assessment for eighth grade students increased by 0.1 points, showing a slight widening of the gap. For the 2007 NAEP mathematics test, 8.0% of eighth grade students in North Carolina were considered Advanced. This was up by 0.9% from 2003 (Swanson, 2008).

For the category of "Standards, Assessments and Accountability," North Carolina received an overall grade of B+. By the 2007-2008 school year, the state had adopted standards that were considered "clear, specific and grounded" at all three levels (elementary, middle and high school) in the core subjects of English/language arts, mathematics and science, and at the high school level only in social studies/history (Swanson, 2008, p. 45). North Carolina also developed a regular timeline for revising these standards. The type of test items used for the state assessments by the 2007-2008 school year consisted of strictly multiple-choice items, with the exception of the English/language arts assessment which also used extended response items. All of the assessments were aligned to the state standards by this time, except the social studies/history assessment which was only aligned with the state standards at the high school level (Swanson, 2008).

For the category "The Teaching Profession," North Carolina received an overall grade of B. By the 2007-2008 school year, prospective teachers were not required to do any substantial amount of formal coursework in the subject areas that they planned on teaching. While these teachers were required to pass written tests in basic skills, they were not required to pass tests in subject-specific knowledge or subject-specific pedagogy (Swanson, 2008, p. 50). In terms of out-of-field teaching, the state did not require parental notification of out-of-field teaching, and as of the 2007-2008 school year, did not ban or put a cap on the number of out-of-field teachers. The state of North Carolina required all of its teachers to be formally evaluated, and the evaluation of teachers was tied to student achievement, however evaluations did not take place annually (Swanson, 2008). North Carolina had professional development standards and financed professional development for the districts. While the professional development programs were required to be aligned with the local priorities and goals, the state did not require schools/districts to set aside time for their professional development (Swanson, 2008).

References

American Federation of Teacher Research Department. (n.d.). *AFT salary surveys*. Retrieved June 2007 from http://www.aft.org/research/salary/home.htm.

Braden et al. (2000). *The state of the state standards 2000*. Retrieved September 2007 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=24&pubsubid=226#226.

Carey, K. (2004). *The funding gap 2004*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/30B3C1B3-3DA6-4809-AFB92DAACF11CF88/0/funding2004.pdf.

Institute for a Competitive Workforce. (2007). *Leaders and laggards: A state-by-state report card on educational effectiveness*. Washington, DC: U.S. Printing Office.

Kennedy Manzo, K. (2001). Dean of education governors departs. *Education Week*, *20,* 28-29.

Kennedy Manzo, K. (2002). Do more for needy students, N.C. court orders. *Education Week, 21,* 15.

Kennedy Manzo, K. (2004). N.C. high court rules state must spend more on schools. *Education Week, 23*, 23.

Klein, D., et al. (2005). *The state of state math standards 2005*. Retrieved September 2007 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=338&pubsubid=1214#1214.

Liu G. (2006). *The funding gap 2006*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/CDEF9403-5A75-437E-93FF-EBF1174181FB/0/FundingGap2006.pdf

National Center for Educational Statistics. (n.d). National public education financial survey: Fiscal years 1990-2002. Retrieved June 2007 from: http://nces.ed.gov/ccd/bat/.

National Center for Educational Statistics. (2000). CCD data file: Local education agency universe survey: School year 1998-1999. Retrieved June 2007 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000346.

National Center for Educational Statistics. (2006 a). *Current expenditures for public elementary and secondary education: School year 2003-2004*. Retrieved June 2007 from: http://nces.ed.gov/pubs2006/expenditures/a1.asp.

National Center for Educational Statistics. (2006 b). *State education reforms: Table 1.1*. Retrieved June, 2007 from http://nces.ed.gov/programs/statereformsaa_tab1.asp?referrer=tables.

National Center for Educational Statistics. (2006 c). *State education reforms: Table 1.2*. Retrieved June, 2007 from http://nces.ed.gov/programs/statereform/saa_tab2.asp.

National Center for Educational Statistics. (2007 a). *Expenditures for public elementary and secondary education: School year 2004-2005*. Retrieved June 2007 from: http://nces.ed.gov/pubs2007/2007356.pdf.

National Center for Educational Statistics. (2007 b). *Mapping 2005 state proficiency standards onto the NAEP scales: Research and development report*. Washington, DC: U.S. Printing Office.

National Center for Educational Statistics. (2007 c). *Public elementary and secondary school student enrollment, high school completions, and staff from the common core of data: School year 2005-2006*. Retrieved June 2007 from http://nces.ed.gov/pubs2007/2007352.pdf.

North Carolina Department of Public Instruction. (2005). *North Carolina standard course of study*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/curriculum/2005ncscos.pdf.

North Carolina State Board of Education (n.d. a). *About the state board of education*. Retrieved August 2007 from http://www.ncpublicschools.org/state_board/AboutSBE.html.

North Carolina State Board of Education (n.d. b). *North Carolina end-of-state tests*. Retrieved September 2007 from http://www.ncpublicschools.org/accountability/testing/eoc/.

North Carolina State Board of Education (n.d. c). *North Carolina licensure*. Retrieved August 2007 from http://www.ncpublicschools.org/licensure/app_packet/question_answers.html.

North Carolina State Board of Education (n.d. d). *What the excellent schools act means*. Retrieved August 2007 from http://www.ncpublicschools.org/esareports/esa.html.

North Carolina State Board of Education. (1997). *Assessment brief*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/eog/asb_assmth.pdf.

North Carolina State Board of Education. (1999). *Graduation requirements*. Retrieved August 2007 from http://www.ncpublicschools.org/curriculum/graduation.

North Carolina State Board of Education. (2000a). *1999-2000 annual report for the North Carolina state board of education*. Retrieved August 2007 from http://www.ncpublicschools.org/state_board/annualrpt/99-00/annualreport.pdf.

North Carolina State Board of Education. (2000b). *Assessment brief*. Retrieved August 2007 from https://mail.bc.edu/Redirect/www.ncpublicschools.org/docs/accountability/testing/briefs/MathematicsScs/MathematicsScs.pdf.

North Carolina State Board of Education. (2000c). *North Carolina state board of education policy manual*. Retrieved August 2007 from http://sbepolicy.dpi.state.nc.us/.

North Carolina State Board of Education. (2000d). *Textbook adoption process*. Retrieved August 2007 from http://www.ncpublicschools.org/textbook/process/.

North Carolina State Board of Education. (2000e). *The 1998-1999 North Carolina state testing results*. Retrieved August 2007 from http://www.ncpublicschools.org/accountability/testing/reports/green/1998-1999.

North Carolina State Board of Education. (2001). *ABC's awards and recognition categories*. Retrieved August 2007 from http://www.ncpublicschools.org/accountability/reporting/abc/2001-02/refinement.

North Carolina State Board of Education. (2002a). *Licensure update: Winter 2002-2003*. Retrieved August 2007 from http://www.ncpublicschools.org/licensure/Licensure.ppt.

North Carolina State Board of Education. (2002b). *Taking the next step to superior schools*. Retrieved August 2007 from http://www.ncpublicschools.org/state_board/annualrpt/01-02/biennialreport.pdf.

North Carolina State Board of Education. (2003). *Executive summary: Proposal to add facilitate add-on licensing of teachers*. Retrieved August 2007 from http://www.ncpublicschools.org/sbe_meetings/0308/0308_QP04.pdf.

North Carolina State Board of Education. (2004a). *Assessment brief*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/briefs/eocassessmentbrief200304draft(v2.2).pdf.

North Carolina State Board of Education. (2004b). *North Carolina end of course tests: Algebra I*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/eoc/Algebra1/RevisedAlg1TIS.pdf.

North Carolina State Board of Education. (2004c). *North Carolina state testing results 2002-2003*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/reports/green/gbfinal02-03.pdf.

North Carolina State Board of Education. (2005a). *End-of-grade mathematics tests*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/eog/matheogsummary_03.pdf.

North Carolina State Board of Education. (2005b). *North Carolina end of course tests: Biology*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/eoc/Biology/biology_tis_7_21_05.pdf.

North Carolina State Board of Education. (2005c). *North Carolina state board of education policy manual*. Retrieved August 2007 from http://sbepolicy.dpi.state.nc.us/.

North Carolina State Board of Education. (2005d). *The mathematics program.* Retrieved August 2007 from http://www.ncpublicschools.org/curriculum/mathematics/scos/2003/intro/09mathprogram.

North Carolina State Board of Education. (2006a). *Education acronyms*. Retrieved August 2007 from http://www.dpi.state.nc.us/docs/acronyms/1106acronyms.pdf.

North Carolina State Board of Education. (2006b). *The North Carolina testing program*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/NORTHCgeneralpolicies.pdf.

North Carolina State Board of Education. (2007a). *North Carolina end-of-course test of civics and economics*. Retrieved August 2007 http://www.ncpublicschools.org/docs/accountability/testing/2007civicseconomicstis.pdf.

North Carolina State Board of Education. (2007b). *North Carolina end-of-course test of U.S. History*. Retrieved August 2007 from http://www.ncpublicschools.org/docs/accountability/testing/2007ushistorytis..pdf.

North Carolina State Board of Education. (2007c). *2006 North Carolina school report cards launch today*. Retrieved August 2007 from http://www.ncpublicschools.org/newsroom/news/2006-07/20070130-01.

Skinner, R. (2005). State report cards. *Education Week, 24*, 106-137.

Swanson, C. (2008). Grading the states. *Education Week, 27,* 36-57.

Appendix C

FINANCE RANKINGS

Table C1
Trends in Total Expenditures per Pupil

	Total Current (Unadjusted U.S. $)							
State or Jurisdiction	**1998-1999**	**1999-2000**	**2000-2001**	**2001-2002**	**2002-2003**	**2003-2004**	**2004-2005**	**Difference Between 1998 and 2005**
Nation	7013	7394	7904	8259	8600	8310	8701	1688
New York	9344	9846	10716	11218	11961	12638	13703	4359
Maryland	7326	7731	8260	8692	9153	9433	10031	2705
Virginia	6350	6841	7281	7496	7822	8219	8886	2536
California	5801	6314	6987	7405	7552	7673	7905	2104
Michigan	7432	8110	8278	8653	8781	9094	9340	1908
South Carolina	5656	6130	6631	7017	7040	7177	7549	1893
Tennessee	5123	5383	5687	5948	6118	6466	6850	1727
Texas	5685	6288	6539	6771	7136	7151	7246	1561
Kentucky	5637	5921	6079	6523	6661	6861	7132	1495
North Carolina	5656	6045	6340	6495	6562	6613	6904	1248
Ten State Average	6401	6861	7280	7622	7879	8133	8555	2154

Sources: National Center for Educational Statistics. (n.d.). *National public education financial survey: Fiscal years 1990-2002*. Retrieved June 2007 from: http://nces.ed.gov/ccd/bat/.
National Center for Educational Statistics. (2007). *Expenditures for public elementary and secondary education: School year 2004-2005*. Retrieved June 2007 from: http://nces.ed.gov/pubs2007/2007356.pdf.
National Center for Educational Statistics. (2006). *Current expenditures for public elementary and secondary education: School year 2003-2004*. Retrieved June 2007 from: http://nces.ed.gov/pubs2006/expenditures/a1.asp.

Table C2
Financial Equity

	Examination of Poverty Gap					Quality Counts Equity Grade	
State or Jurisdiction	**Gap Between Highest and Lowest-Poverty Districts 1997**	**Gap Between Highest and Lowest-Poverty Districts 2001**	**Gap Between Highest and Lowest-Poverty Districts 2002**	**Gap Between Highest and Lowest-Poverty Districts 2004**	**Poverty Gap Change in Dollars 1997-2004**	**2000**	**2005**
California	($205)	($418)	($301)	($259)	($54)	C+	D+
Kentucky	($119)	($143)	($357)	$448	$567	C+	C
Maryland	($961)	($735)	($772)	($432)	$529	D-	B
Michigan	($1,407)	($1,099)	($1,085)	($1,072)	$335	C-	B-
New York	($2,938)	($2,264)	($2,615)	($2,927)	$11	D+	B+
North Carolina	($464)	($751)	($622)	($543)	($79)	C+	C-
South Carolina	($370)	($343)	$43	$127	$497	C-	C
Tennessee	$124	$536	$281	$330	$206	D+	C-
Texas	($437)	($875)	($936)	($757)	($320)	B-	C-
Virginia	($972)	($1,341)	($1,430)	($436)	$536	C-	C+
Nation	($1,208)	($1,287)	($1,348)	($1,307)	($99)	N/A	N/A

[1]Examination of Poverty Gap: Cost-adjusted dollars, 40% adjustment for low-income students

Sources: Carey, K. (2004). *The funding gap 2004*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/30B3C1B3-3DA6-4809-AFB92DAACF11CF88/0/funding2004.pdf.
Liu G. (2006). *The funding gap 2006*. Retrieved June 2007 from http://www2.edtrust.org/NR/rdonlyres/CDEF9403-5A75-437E-93FF-EBF1174181FB/0/FundingGap2006.pdf
Quality Counts. (2003). If I can't learn from you: Ensuring a highly qualified teacher for every classroom. *Education Week, 22*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC03full.pdf.
Quality Counts. (2008). Tapping into teaching: Unlocking the key to student success. *Education Week, 27*(17). Available: http://www.edweek.org/ew/toc/2008/01/10/index.html.

Ranking	State	Rationale
Leaders	New York	New York was placed in the top spot due to a combination of both dramatically increasing annual per pupil expenditures during the 1998-2005 period, as well as making strides in improving equity between low and high poverty districts. It is important to note that NY did have the largest discrepancy between high and low poverty districts throughout the entire period. However, the state's increase in per pupil expenditures between 1998 and 2005 was over double that of either the ten state or national averages. Also, according to the grade assigned by *Education Week's Quality Counts* report, the state's overall grade in school finance equity improved by two letter grades between 2000 and 2005. This was the second highest increase behind Maryland, out of the ten states.
	Maryland	Maryland is also considered a leader in finance, mostly because of its efforts in improving school finance equity. Out of the ten states, Maryland was the most improved during the 2000 through 2005 period according the *Quality Counts* reports, who gave the state a score of D- in 2000, up to a B in 2005, the biggest score increase of all ten states. Looking at Table 2, the difference in the gap between the state's lowest and highest poverty districts decreased by $529, the third largest gap decrease in the study, and well better than the national average, which showed an increase in the gap by $99. Also notable about Maryland was its increase in annual per pupil expenditures. While the state had the second highest increase during the 1998-2005 period, it was still just over half of the amount that NY had increased during that same period.
	Virginia	Out of the ten states, Virginia had the third highest increase in per pupil expenditures between 1998 and 2005. In addition their equity grade increased almost a letter grade during this period (from a C- to a C+), and the discrepancy between high poverty and low poverty districts closed by $536, the second largest gap decrease in the sample.
Middle Ground	South Carolina	With the state's per pupil expenditure difference between 1998 and 2005 below the ten state average though above the national average, South Carolina finds itself in the middle ground category. The state was one of only three in the study that managed to flip the gap between high and low poverty districts, with the higher poverty districts receiving more funding per student than their

Ranking	State	Rationale
		lower poverty counterparts. While their *Quality Counts* equity grade increased during that period, it was slight (from C- to C).
	Michigan	With a per pupil expenditure difference between 1998 to 2005 very close to that of South Carolina, and still below the ten state average, Michigan also finds itself in the middle ground category Michigan was the final state in the top five in terms of growth in per pupil expenditures, however, with an increase of $1,908 between 2000 and 2005, the state still fell below the ten state average of $2,154. The state decreased the poverty gap between highest and lowest poverty districts by $335, and the state's *Quality Counts* equity grade increased by a full letter grade from C- in 2000 to B- in 2005.
	Tennessee	Although Tennessee's per pupil expenditures between 1998 and 2005 were $427 below the ten state average, the state's difference in expenditures did remain above the national average. Also, the state did see a slight increase in its *Quality Counts* equity grade. In addition, TN was the only state to have a reverse spending gap for the entire period, and increased that gap in favor of the higher poverty districts further between 1997 and 2004. The combination of these factors, in comparison to the improvements made by the other states, earned Tennessee a middling rank.
	California	While the state of California had a moderate increase in the amount of per pupil expenditures during the 1998 to 2005 period, placing it at just below the ten state average, the state also saw an increase in the gap between high and low poverty school districts within the state ($54), as well as an overall decrease in their *Quality Counts* equity grade. The combination of these two factors contributed to the state earning their place in the middle section of the rankings.
	Kentucky	During the 1998 to 2005 period, the per pupil expenditures in the state of Kentucky only increased by a total of $1,495-less than both the ten state and national averages. The state did have the largest decrease in the poverty gap among the ten states, moving from higher poverty students receiving $119 less than their lower poverty counterparts in 1997, to higher poverty students receiving $448 more in 2005. The state's *Quality Counts* equity grade saw a slight decrease over the period.
Laggards	North Carolina	Out of the ten states that are examined in this study, North Carolina's difference in per pupil expenditures

Ranking	State	Rationale
		between 1998 and 2005 was the lowest. The state also had an overall decrease in their *Quality Counts* equity grade during the 2000 to 2005 period. In addition, the gap between highest and lowest poverty districts increased by $79 between 1997 and 2004.
	Texas	Texas earned its place at the bottom spot in the finance rankings for two reasons. First, the difference in the state's per pupil expenditures was one of the bottom three, falling below both the ten state and national averages. Second, the poverty gap between the state's highest and lowest poverty districts saw an overall $320 widening of the gap between 1997 and 2004. Texas' *Quality Counts* equity score decreased by a whole letter grade from a B- in 2000 to a C- in 2005, which was the largest score decrease among the ten states.

Appendix D

CURRICULUM AND STANDARDS

Table D1
State Standards

	State has standards that are clear, specific, and grounded in content in the subject of mathematics		State has a regular timeline for revising standards		Thomas B. Fordham Foundation State of the Math Standards Grade		Institute for a Competitive Workforce: Rigor of Standards Rating
	2000	**2005**	**2001**	**2005**	**2000**	**2005**	**2007**
California	ES MS HS	ES MS HS			A	A	Excellent
Kentucky	ES MS HS	ES MS HS			B	C	Average
Maryland	ES MS HS	ES MS HS	X	X	C	C	Average
Michigan	ES MS HS	ES MS HS			F	C	Average
New York	ES MS HS	ES MS HS	X		B	C	Excellent
North Carolina	ES MS HS	ES MS HS	X	X	A	C	Middling
South Carolina	ES MS HS	ES MS HS	X	X	B	D	Solid
Tennessee		ES MS HS	X		F	D	Modest
Texas	ES MS HS	ES MS HS		X	B	C	Solid
Virginia	ES MS HS	ES MS HS	X	X	B	C	Higher than Average

Sources: Klein, D., et al. (2005). The state of state math standards 2005. Retrieved September 2007 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=338&pubsubid=1214#1214.
Institute for a Competitive Workforce. (2007). *Leaders and laggards: A state-by-state report card on educational effectiveness*. Washington, DC: U.S. Printing Office.
Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.
Quality Counts. (2002). Building blocks for success: State efforts in early childhood education. *Education Week, 21*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC02full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html.

Table D2
State Textbook Policies

State textbook selection policies: 2000			
	State policy	**Use state content standards**	**State revisions in textbook policies between 2000 and 2005**
California	Select	Yes	Yes
Kentucky	Yes	Yes	Yes
Maryland	No		No
Michigan	No		No
New York	No		No
North Carolina	Recommend	Yes	No
South Carolina	Select	Yes	No
Tennessee	Select	Yes	No
Texas	Select	Yes	No
Virginia	Recommend	Yes	Yes

Note: State officials were asked whether their state has a policy by which the state selects textbooks or curriculum materials or recommends textbooks or curriculum materials. Select=state SELECTS texts/materials; Recommend=state RECOMMENDS text/materials

Source: National Center for Educational Statistics. (2003). Overview and inventory of state education reforms: 1999 to 2000. Retrieved April 2008 from http://nces.ed.gov/pubs2003/2003020.pdf.

Ranking	State	Rationale
Leaders	California	The state of California was given a consistent grade of A from the Fordham Foundation for their mathematics state standards, and was the only one of the ten states to receive a grade of A for their standards from the Foundation in 2005. The state was considered to have standards that were "clear, specific, and grounded" in mathematics at all three levels, however, as of 2005, they did not have a time line for revising their standards. The state also received a grade of "excellent" from the Institute for a Competitive Workforce (ICW) for the rigor of their standards. The state selects text/materials for their districts to choose from, and has revised their textbook policies since 2000. (From profile) The state's reading/language arts framework was adopted by the California SBE in 1998. The content standards were adopted the year prior to the framework in 1997. The mathematics content standards were also adopted in 1997. The corresponding framework was adopted in 1998, with revisions in 2005. The science framework was adopted in 2001, with content standards adopted prior to the framework in 1998. The history/social science frameworks were originally published in 1987, with content standards added in 1998, and further updates in 2001 and 2005.
	Tennessee	Tennessee is the only one of the ten states that did not have standards that were "clear, specific, and grounded" in mathematics at any level in 2000. However, by 2005 they had developed standards that fit that description in mathematics at all three levels. While the state had a regular timeline through which to revise their standards in 2001, they no longer did by 2005. Over the 2000 to 2005 period, according to the Fordham Foundation, the math standards in the state of Tennessee went from having a failing grade to passing with a D. The ICW gave them a "modest" rating for the rigor of their standards. The state textbook policy involved the state selecting textbooks for districts, and these textbooks were aligned with the state standards. The state did not have any textbook policy revisions between 2000 and 2005. (From profile) The Tennessee Curriculum Standards were developed for English/language arts, mathematics and science in 2001. The social studies curriculum standards were developed in 2002. The high school mathematics standards were developed in 1998 and revised in 2004.
Middle Ground	Virginia	Virginia saw a decrease in the quality of their standards according to the Fordham Foundation, with a grade that

Ranking	State	Rationale
		went from B in 2000, to C in 2005. The state consistently had standards in mathematics that were considered "clear, specific, and grounded" at all three levels of schooling as well as a revision process of the standards throughout the examined period. In 2007, the ICW referred to the rigor of the Virginia mathematics standards as "higher than average." Virginia's textbook policy involved the state recommending text/materials. The state did have textbook policy revisions during the 2000 to 2005 period, specifically in 2002, when the state's textbook policies were adapted to reflect changes in the state's standards. (From profile) These standards were revised, as well as new standards were developed for the following subjects in the identified years: Fine Arts (2000), Foreign Language (2000), History/Social Sciences (2001), Mathematics (2001), Physical Education (2001), Health (2001), English (2002), and Science (2003).
	New York	The state of New York also received a lower grade on their mathematics standards from the Fordham Foundation in 2005 (grade of C) than it did in 2000 (grade of B), and did not enact a regular revision process at all during the 2000 to 2005 period. However, the state had mathematics standards that were characterized as "clear, specific, and grounded" at all three levels throughout the period, and in 2007 the ICW called the rigor of the New York state mathematics standards "excellent." The state did not have any formal textbook policies at all during this time period, with the New York State Education Department declaring that New York was "not a textbook adoption state." (From profile) The state's learning standards were unchanged throughout the 1998-2005 time period. However, core curricula were developed in the following subjects; English/language arts-1998, revised in 2005, Mathematics-1999, revised in 2005, Social studies-1999.
	Texas	Like New York, the grade of Texas' mathematics standards was also lowered from a B in 2000 to a C in 2005. However, according to the ICW, the rigor of the state's mathematics standards in 2007 were "solid" and the mathematics standards were considered "clear, specific and grounded" throughout the period. The state also devised a regular timeline for revising their standards by 2005. While the state began the examined time period by selecting text/materials that aligned with the state standards, there were not any new textbook policies enacted between 2000 and 2005. (From profile) The Texas Essential Knowledge and Skills (TEKS) were implemented in 1998, and districts

Ranking	State	Rationale
		were required to provide instruction based on curriculum that was in accordance with the TEKS in the appropriate corresponding grade level, by the 2003-2004 school year.
	Maryland	In terms of state education policies on curriculum, standards, and textbooks/materials, the state of Maryland, between the years 1998 and 2005, remained consistent. The state began the period with mathematic standards that were considered to be "clear, specific, and grounded" at all three levels, and remained that way throughout, as well as a revision process for the standards from 2000 through 2005. The state's mathematics standards received a grade of C from the Fordham Foundation in both 2000 as well as 2005, and in 2007 the ICW qualified the rigor of the standards as being "average". The state did not have any policies regarding state adoption of textbooks, and did not enact any during the examined period. (From profile) The Maryland Content Standards were developed for the subjects English/language arts, mathematics, science and social studies in 2001. The Core Learning Goals (CLG) were standards that corresponded with the High School Assessments (HSA); developed in English before 1998 (but with updates in 2000 and 2004), in algebra and data analysis in 2001 and in government and biology in 2002.
	South Carolina	The state of South Carolina also saw little change over the course of the 1998 to 2005 period, with the exception of the Fordham Foundation grade of their mathematics standards which decreased from a B in 2000 to a D in 2005. The state had standards that were considered "clear, specific and grounded" at all three levels throughout the period, as well as a regular timeline for revising the standards. In 2007, the ICW characterized the rigor of the South Carolina standards as "solid". While, in 2000, the state had in place textbook policies in which the state selected text/materials that were aligned with the state standards, there were no further revisions during the examined period. (From profile) The South Carolina mathematics, science and social studies standards were first introduced in 2000. The state's English/language arts standards were developed in 2002.
Laggards	Michigan	While the state of Michigan, according to the Fordham Foundation, had standards that improved from failing with an F in 2000 to a C in 2005, the state did not introduce many policy initiatives during this time. They had standards in mathematics that were considered to be "clear, specific and grounded" at all three levels of schooling in both 2000 as well as 2005 but never

Ranking	State	Rationale
		developed a timeline for revising these standards. Their standards also received a grade of "average" for their rigor according to the ICW. The state did not have a textbook policy at any point during the examined time period. (From profile) The state's original curriculum was published in 1996. The mathematics curriculum framework including teaching and learning activities was added in 1998.
	Kentucky	Kentucky's standards received a lower grade from the Fordham Foundation in 2005 (grade of C), than they did in 2000 (grade of B), though the state did have mathematics standards that were considered "clear, specific and grounded" at all three levels throughout the time period. Also, the ICW referred to the rigor of the Kentucky state standards as "average". The state did not have any regular revision process for their standards during this period. The state did however, have textbook policy changes, such as in 2003, when legislation was passed requiring all textbook publishers to furnish Kentucky schools with electronic versions of their texts. (From profile) The state's standards were first implemented before 1998. However, Kentucky's *Program of Studies for Grades Primary – 12*, were added in 1998 to clarify the standards.
	North Carolina	While North Carolina began the examined time period with mathematics standards that had received a grade of A from the Fordham Foundation, by 2005, the grade given to their standards had fallen to a C. The state did however, have math standards that were considered "clear, specific and grounded" as well as a timeline for revising the standards throughout the period. In 2007 the ICW gave the rigor of the North Carolina standards a "middling" rating. The state began the period recommending text/materials that aligned with the state standards, and did not enact any further textbook revisions between 2000 and 2005. (From profile) The state standards were developed before the examined time period. Revisions did take place during the 1998-2005 time period, including social studies curriculum in 2003, the English language arts and second language curriculums in 2004, science curriculum in 1999, and again in 2004, and the mathematics curriculum in 1998 and again in 2003.

Appendix E

TEACHER QUALITY

Table E1
Percentage of Public School Teachers without a Full Certificate in the Field Taught By State (1999-2000)

	Number of out-of-field teachers (1999-2000)		State has a ban or cap on the number of out-of-field teachers	
	Elementary	Secondary Math	2003-2004	2005-2006
Nation	6.56	28.62	N/A	N/A
California	13.14	41.53	X	X
Kentucky	9.88	26.39	X	X
Maryland	6.32	20.89		
Michigan	16.49	38.81		X
New York	10.88	33.66		
North Carolina	3.06	28.57		
South Carolina	3.20	13.55	X	X
Tennessee	5.56	24.07		
Texas	8.18	36.14		
Virginia	8.97	26.29		X
Ten state average	8.57	28.99	N/A	N/A

Sources: Ingersoll, R., and Curran, B. (2004). *Out-of field teaching: The great obstacle to meeting the "highly qualified" teacher challenge*. NGA Center for Best Practices. Available from http://www.nga.org/Files/pdf/0408HQTEACHER.pdf.
Quality Counts. (2004). Count me in: Special education in an era of standards. *Education Week*, *27*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC04full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html.

Table E2
Trends in Average Teacher Salary (in unadjusted U.S. dollars)

	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	Difference Between 1998 and 2005
Nation	$40,574	$41,820	$43,250	$44,367	$45,578	$46,565	$47,602	$7,028
California	46,326	47,680	52,480	54,348	55,673	56,444	57,604	11,278
Maryland	42,545	43,720	45,963	48,251	49,679	50,303	52,330	9,785
South Carolina	34,506	36,081	37,938	39,923	40,362	41,162	42,189	7,683
Virginia	37,709	38,992	40,247	41,752	42,677	43,936	45,377	7,668
Tennessee	35,490	36,328	37,413	38,515	39,186	40,318	42,076	6,586
Texas	34,448	37,567	38,359	39,230	39,972	40,476	41,009	6,561
North Carolina	36,883	39,404	41,496	42,118	42,411	43,211	43,343	6,460
New York	49,686	51,020	51,020	51,020	53,017	55,181	55,665	5,979
Kentucky	35,383	36,255	36,688	37,951	38,485	40,310	41,075	5,692
Michigan	48,711	48,729	50,515	52,497	53,178	52,161	53,959	5,248
Ten State Average	40,169	41,578	43,212	44,561	45,464	46,350	47,463	7,294

Source: American Federation of Teacher Research Department. (n.d.). AFT salary surveys. Retrieved June 2007 from http://www.aft.org/research/salary/home.htm

Table E3
Teacher Assessment

	State Requires Written Test for Beginning Teacher License (2000)			Prospective Teachers Must Pass Written Tests (2007-2008)		
	Basic Skills	Subject Knowledge	Subject Specific Pedagogy	Basic Skills	Subject Specific Knowledge	Subject Specific Pedagogy
California	**X**	**X**	**X**	**X**		
Kentucky	**X**	**X**	**X**		**X**	
Maryland	**X**	**X**	**X**	**X**	**X**	**X**
Michigan	**X**	**X**		**X**	**X**	
New York	**X**	**X**	**X**	**X**	**X**	**X**
North Carolina	**X**	**X**	**X**	**X**		
South Carolina	**X**	**X**	**X**	**X**	**X**	
Tennessee	**X**	**X**	**X**	**X**	**X**	
Texas		**X**	**X**	**X**	**X**	**X**
Virginia	**X**	**X**		**X**	**X**	

Source:
Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.
Quality Counts. (2008). Tapping into teaching: Unlocking the key to student success. *Education Week, 27*(17). Available: http://www.edweek.org/ew/toc/2008/01/10/index.html.

Table E4.
Teacher Qualifications

State	Minimum degree/coursework required for initial secondary teacher license in Middle School		State requires performance assessment for second stage of certification				Number of National Board Certified Teachers	
			Portfolio		Classroom observation			
	2000	2005	2000	2005	2000	2005	2000	2005
California	major[1]						785	3,377
Kentucky				X		X	79	899
Maryland							70	660
Michigan	minor (major)[1]						89	189
New York	major						104	588
North Carolina	N/A	major[2]	X		X		2,377	9,815
South Carolina				X		X	371	4,445
Tennessee	major[1]			X		X	35	173
Texas	N/A						35	231
Virginia	minor (major)[1]	minor					144	905
Ten State Average							408.9	2,128

1: State requires a non-education major, but certification need not be in that subject area.
2:: State requires teacher-candidates to demonstrate subject matter competency either by obtaining a major in the subject taught or by passing a content test.
Sources:
Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html

Table E5
Professional Support

	State finances professional development		Time set aside for professional development		State has written professional-development standards		Quality Counts Teacher Quality Grade	
	2000	**2005**	**2000**	**2005**	**2003-2004**	**2005-2006**	**2000**	**2005**
California	X	X	X		X	X	C-	B-
Kentucky	X	X	X	X	X	X	C+	B
Maryland	X	X				X	C	C+
Michigan	X		X	X	X	X	C-	D
New York	X	X			X	X	C	B-
North Carolina	X	X			X	X	B+	B
South Carolina	X	X	X	X	X	X	B-	A
Tennessee	X	X	X	X	X	X	C	C+
Texas	X						D	C-
Virginia	X	X				X	C	B+

Sources: Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.
Quality Counts. (2004). Count me in: Special education in an era of standards. *Education Week*, *27*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC04full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html.

Ranking	State	Rationale
Leaders	Virginia	While starting the examined time period with a percentage of out-of-field elementary and secondary mathematics teachers that were at or slightly above the national and ten-state averages, by the 2005-2006 school year the state had placed a cap on the number of out-of-field teachers within the state. The annual teacher salary in the state between the 1998-1999 and 2004-2005 school years increased more than both the ten-state and national averages. The state did not require its teachers to pass a subject-specific pedagogy test for licensure, but did require at least a specific minor for an initial secondary teaching license in middle school. The state also did not require its teachers to complete a portfolio or be subject to a classroom observation, and the number of National Board Certified teachers in the state was far lower than in other states. However, the state did finance professional development throughout the period, and by the 2005-2006 school year had developed professional development standards. Over the 2000-2005 time period, the state's *Quality Counts* grade increased from a C to a B+, giving it the second highest score in 2005, behind South Carolina. (From profile) The Virginia teacher licensure requirements were revised in 1998. In 2004, the DOE mandated that the state's professional development was to align with the SOL.
	South Carolina	South Carolina began the examined time period with a percentage of teachers that were out-of-field that was much lower than that of the national and ten-state average for both elementary school as well as secondary math. The state also had a ban on the number of out-of-field teachers in place from 2003 through 2006. The annual teacher salary in the state between the 1998-1999 and 2004-2005 school years increased more than the ten-state and national averages. While by the 2007-2008 school year, the state no longer required teachers to take subject specific pedagogy tests, and there was not any minimum degree/coursework required for teachers to get a secondary license in middle school, the state did implement performance assessments for certification in the forms of both a portfolio and classroom observation by 2005, as well as dramatically increased the number of National Board Certified teachers in the state. Throughout the period, the state had professional development standards, as well as provided time and funding for professional development. The state's *Quality Counts* score increased overall during the period, and was the only one of the ten states to receive a grade of A in 2005. (From profile) Professional development was one of the objectives in the Education Accountability Act of 1998, which specified that teachers in low performing schools who

Ranking	State	Rationale
		participated in professional development activities that addressed needs in their school's improvement plan could earn credits towards recertification. It also called for the development of a professional development accountability system.
	California	The state of California began the examined period with a percentage of out-of-field teachers at the elementary and secondary mathematics levels that were much higher than both the national and ten-state averages, though the state had a cap on the number of out-of-field teachers from 2003 through 2006. The annual teacher salary in the state between the 1998-1999 and 2004-2005 school years increased the most out of the ten states during that period. The state decreased the number of tests that teacher candidates were required to pass for licensure, getting rid of the subject knowledge and subject-specific pedagogy tests by the 2007-2008 school year. The state also stopped requiring a major in a specific subject as a requirement for an initial secondary teacher's license in middle school. The state did however, have more teachers that held National Board Certification than the ten state average. While the state had professional development standards, as well as financed professional development during the 2000-2005 period, the state stopped requiring districts to set aside time for professional development by 2005. The state's *Quality Counts* grade improved from a C- to a B- during this period. (From profile) The state's teacher preparation program standards during the examined period were based on Senate Bill 2042, passed in 1998. This legislation created multiple routes to teacher certification. The legislation also established a two-tiered teacher credential structure that included teacher preparation and professional teacher induction.
Middle Ground	North Carolina	While at the elementary level, North Carolina's out-of field teaching percentage was below national and state-averages, at the secondary mathematics level their percentage was in line with the national and ten-state averages, and the state failed to place a cap/ban on the number of out-of-field teachers in the state at any point during the examined period. From 1998-2005, the average teacher salary in North Carolina increased by $6,460, the difference falling below national and the ten-state averages. Between 2000 and 2007, the state did away with two of its three state tests, requiring teacher candidates to only pass a basic skills test in order to be certified in 2007-2008. While the state did require teacher candidates to demonstrate subject knowledge through obtaining a major by 2005 (or by passing a content test), the state abolished both of their performance assessments by that same year. The state did, however, have the

Ranking	State	Rationale
		highest number of National Board certified teachers out of the ten states. While the state had written professional development standards during the 2000-2005 period and funded PD activities, the state did not require districts to set aside time for PD. Overall, North Carolina's *Quality Counts* score decreased from B+ in 2000 to B in 2005. (From profile) As of the year 2000, in order for a candidate to be issued an initial license, they first had to complete a bachelor's degree program, as well as receive an endorsement from the Institutes of Higher Education. After 1998, teacher candidates who were issued initial licenses were required to complete a three-year Initial Licensure Program through one of the state's school districts.
	New York	The state of New York had a higher than average out-of-field teaching percentage in both the elementary and secondary math levels, as well as failed to place a ban/cap on the number of out-of-field teachers within the state during the examined period. The state's average teacher salary increased by $5,979 from 1998-2005, which was below the national and ten-state averages. While the state held consistent in the tests required for licensure, it stopped requiring a specific major by the 2005 school year, and did not require any performance assessments for certification throughout the period. Also, the number of National Board certified teachers in the state was well below the ten-state average. While the state did have written professional development standards and funded PD throughout the 2000-2005 period, the state did not require districts to set aside time for PD activities. The state's overall *Quality Counts* grade increased from a C to a B- between 2000 and 2005. (From profile) The state changed their teacher certificates in 2004, doing away with permanent certification. In addition teacher candidates were required to pass a Content Specialty Test (CST) in the area of their teaching certificate. Also in 2004, teachers who were issued professional certificates were required to complete 175 hours of professional development within a five year period in order to be reissued another teaching certificate.
	Tennessee	Tennessee began the examined period with a percentage of out-of-field teachers that was below the national and ten-state averages at both the elementary and secondary mathematics levels, but failed to place a ban/cap on the number of out-of-field teachers in the state. The state's average teacher salary increase was just below both the national and ten-state averages. By the 2007-2008 school year, the state did away with the subject-specific pedagogy test as a certification requirement, as well as no longer requiring a specific major for licensure by

Ranking	State	Rationale
		2005. However, by that same year the state did require that teacher candidates complete a performance assessment including both a portfolio and classroom observation. The state's number of National Board certified teachers was well under the ten-state average. But, the state did consistently support professional development with written standards, as well as funding and requiring districts to allot time for PD activities. Overall, the state's *Quality Counts* grade increased from a C in 2000 to a C+ in 2005. (From profile) The state's standards for teacher licensure were revised in 1997 with compliance to the updated standards by 2001. The licensure requirements for secondary education (grades 7-12), including English, mathematics, sciences, social studies and foreign languages were implemented in 2001. The state had Teacher Licensure Standards for different areas of endorsement. These performance standards were first implemented in 2003. There were also updates to the state's teacher licensure examinations during the examined period.
	Texas	The number of out-of-field teachers in Texas during the 1998-1999 school year were at (Elementary) and above (Secondary) the ten-state averages, and the state failed to place a cap on the number of out-of-field teachers within the state during any point of the examined time period. The average teacher salary from 1998 to 2005 increased by $6,561, which was lower than the national and ten-state averages. Texas did add a testing requirement to their certification process by 2007, with the inclusion of a basic skills test, however the state did not require any specified degree/coursework, nor did it require any performance assessments for certification. The state also had very low numbers of teachers that held National Board certification throughout the time period. While the state did finance professional development in 2000, it no longer did by 2005, and the state also did not have any written PD standards, nor did it require districts to set aside time for PD activities. The state's *Quality Counts* grade slightly increased from a D in 2000 to a C- in 2005. (From profile) Starting in 1999, the state's standard certificate was issued to new Texas teachers for five year increments. Beginning with the 2002-2003 school year, the standard certificate was given in two main categories, generalists and those who were specialized in a particular subject.
Laggards	Kentucky	Kentucky began the examined period in 1998-1999 with an out-of-field teaching percentage that was around both the national and ten-state averages, and from 2003-2006 had a cap in effect for the number of out-of-field teachers in the state. The annual teacher salary in the state increased by $5,692,

Ranking	State	Rationale
		which was below both the ten-state and national averages. The state did away with the basic skills and subject specific pedagogy tests between 2000 and 2007, requiring teachers to only pass a subject knowledge test for a beginning license. However, by 2005, the state did require performance assessments in the forms of a portfolio and classroom observation for certification. But, the state did have less teachers holding National Board certification than the ten-state average. Throughout the period of 2000 to 2005, the state was dedicated to professional development, having standards, as well as funding PD and requiring districts to set aside time for PD activities. The state's *Quality Counts* grade increased from a C+ in 2000 to a B in 2005. (From profile) There weren't any major changes to the state's teacher certification policies during the examined period. However, as of 2004, Kentucky was the only state to ban out-of-field teaching (Ingersoll and Curran, 2004). But, due to Kentucky's shortage of teachers in some subjects (including special education, mathematics, science and foreign languages) the state did employ teachers on emergency certificates.
	Maryland	While the state of Maryland started the time period with a percentage of out-of-field teachers that was below the ten-state and national averages, throughout the entire period the state failed to place a ban/cap on the number of out-of-field teachers permitted in the state. The state's annual teacher salary increased by $9,785, which was the second highest increase in the group (behind California) and was above the national and ten-state averages. The state had consistent testing procedures throughout the time period, requiring teachers to pass basic skills, subject knowledge, and subject specific pedagogy. The state did not, however require teachers to complete any minimum coursework or pass a performance assessment in order to obtain a license, and also had fewer National Board certified teachers than the ten-state average. The state did finance professional development throughout the 2000-2005 period, and by 2005 had developed professional development standards. The state's *Quality Counts* grade increased from a C in 2000 to a C+ in 2005. (From profile) There appear to have been changes in the types of certification during the examined period, but documents found on the state DOE's website did not indicate exactly when these changes took place.
	Michigan	Michigan began the examined time period with percentages of out-of-field teaching at both levels that were above ten-state and national averages, and did not place a ban/cap on the number of out-of-field teachers throughout the examined period. The state had an average teacher salary increase

Ranking	State	Rationale
		between 1998 and 2005 of $5,248, the smallest increase of any of the ten states. The state's teacher testing program stayed consistent during the 2000 to 2007 period, requiring teachers to pass both a basic skills and subject knowledge test, but not a subject specific pedagogy test. While the state did require a non-education major in terms of teacher coursework in 2000, they did away with this requirement by 2005, and did not require any performance assessments of teacher candidates throughout the period. The state also had very low numbers of teachers that held National Board certification. While the state did have written professional development standards and required districts to set aside time for PD activities throughout the entire period, the state stopped funding PD by 2005. The state's *Quality Counts* grade decreased from a C- in 2000 to a D in 2005. (From profile) There were not any major policy changes in terms of teacher certification during the 1998-2005 time period. The state did update their teacher professional development standards in 2003.

Appendix F

ASSESSMENT AND ACCOUNTABILITY

Table F1
Assessments and Accountability Grade

	State has mathematics assessments aligned with the state standards		State standards-based tests have undergone an external alignment review by 2001	State standards-based tests have undergone an external alignment review since 2001 (2005-2005)	Quality Counts: Overall grade for standards and accountability	
	2000-2001	**2005-2006**			**2001**	**2005**
California	ES MS HS	ES MS HS	X	X	B	B+
Kentucky	ES MS HS	ES MS HS		in process	A	B+
Maryland	ES MS HS	ES MS HS	X	X	A	A-
Michigan	ES HS	ES MS HS	X	X	C	B
New York	ES MS HS	ES MS HS			A	A
North Carolina	ES MS HS	ES MS HS	X		B	B
South Carolina	ES MS	ES MS HS		X	B+	A
Tennessee	HS	ES MS HS	X		C+	B
Texas	ES MS HS	ES MS HS			B-	B-
Virginia	ES MS HS	ES MS HS		X	B	B

Sources:
Quality Counts. (2001). A better balance: Standards, tests and the tools to succeed. *Education Week*, *20*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC01full.pdf.
Quality Counts. (2002). Building blocks for success: State efforts in early childhood education. *Education Week, 21*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC02full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html.

Table F2
AYP

	State holds schools accountable for performance				Sanctions the state has authority to use for persistently low performing schools					
	2001-2002		2005-2006		2001-2002			2005-2006		
	Assistance	Sanctions	Assistance	Sanctions	School closure	Reconstitution	Permit student transfers	School closure	Reconstitution	Permit student transfers
California	X		X	X				X	X	X
Kentucky	X	X	X	X			X		X	X
Maryland	X	X	X	X	X	X		X	X	
Michigan			X	X				X	X	X
New York	X	X	X	X	X	X		X	X	X
North Carolina	X	X	X	X		X			X	
South Carolina	X	X	X	X	X	X			X	
Tennessee	X	X	X	X		X	X		X	X
Texas	X	X	X	X	X	X	X	X	X	X
Virginia	X		X							

Sources: Quality Counts. (2002). Building blocks for success: State efforts in early childhood education. *Education Week, 21*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC02full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html

****Michigan was the only state that did not produce school report cards in the year 2001. By 2005-2006 all ten states produced school report cards.**

Table F3
Graduation Rates

	Graduation contingent on performance on statewide exit or end-of-course exams		Graduation Rates in percents (All students, all public schools)	
	2001	2005	2004	Change 2000 to 2004
Nation			69.9	3.1
California		X	70.7	2.1
Kentucky			70.0	6.3
Maryland	X	class of 2009	74.7	2.0
Michigan			69.1	-3.8
New York	X	X	65.0	4.5
North Carolina	X	X	66.1	5.8
South Carolina	X	X	53.8	5.4
Tennessee	X	X	72.2	14.7
Texas	X	X	63.7	4.4
Virginia		X	73.1	-4.2
Ten state average			67.8	3.7

Sources:
Quality Counts. (2002). Building blocks for success: State efforts in early childhood education. *Education Week, 21*(17). Available: http://www.edweek.org/media/ew/qc/archives/QC02full.pdf.
Quality Counts. (2006). Quality counts at 10: A decade of standards based education. *Education Week, 25*(17). Available: http://www.edweek.org/ew/toc/2006/01/05/index.html
Quality Counts. (2008). Tapping into teaching: Unlocking the key to student success. *Education Week, 27*(17). Available: http://www.edweek.org/ew/toc/2008/01/10/index.html.

Ranking	State	Rationale
Leaders	South Carolina	South Carolina earned a spot in the leader category for implementing a number of educational policies regarding assessment and accountability during the 1998-2005 period. While the state began the period with mathematics assessments aligned with the state standards at the elementary and middle school levels, by the 2005-2006 school year, a high school mathematics assessment aligned with the state standards had also been developed. As of 2001 the state tests had not undergone an external alignment review, but such a review had taken place by the 2005-2006 school year. The state held schools accountable for performance and offered assistance as well as imposed sanctions from 2001 through 2005, however the state ceased to use school closure as a sanction by the 2005-2006 school year, leaving reconstitution as an option. The state provided rewards for high performing schools throughout the period. South Carolina used an exam on which graduation was contingent throughout, and increased their graduation rate by 5.4%, to 53.8% in 2004. While the increase was larger than both the national and ten state-averages, the percent of students that graduated were below both. The state's overall *Quality Counts* grade in the category of standards and accountability increased from a B+ in 2001 to an A in 2005. (From profile) In 1998, the state established a performance-based accountability system titled the Education Accountability Act (EAA). The Palmetto Achievement Challenge Tests (PACT) were developed in accordance with the EAA, and assessed students in the four core academic subjects. The mathematics and English/Language Arts assessments began in 1999, science in 2001 and social studies in 2002.
	Tennessee	Similarly to South Carolina, Tennessee also made the top category through implementing a variety of educational policies regarding the state's assessment and accountability programs during the 1998-2005 period. While the state began the period with mathematics assessments aligned with the state standards at the high school level, by the 2005-2006 school year, elementary and middle school mathematics assessments that aligned with the state standards had also been developed. As of 2001 the state tests had undergone an external alignment review, but such a review did not take place again between then and the 2005-2006 school year. The state held schools accountable for performance and offered assistance as well as imposed sanctions throughout the entire 2001 through 2006 period, including the sanctions of reconstitution as well as permitting students to transfer schools. The state did not provide rewards for high performing schools at all during the examined period. Tennessee used an exam on

Ranking	State	Rationale
		which graduation was contingent throughout the period, and increased their graduation rate by 14.7.%, to 72.2%. This increase was larger than that of any of the other ten states, and the overall student graduation percentage rate was above both the ten-state and national averages. The state's overall *Quality Counts* grade in the category of standards and accountability increased from a C+ in 2001 to a B in 2005. (From profile) While the Tennessee Comprehensive Assessment Program (TCAP) was in place before the examined time period, the TCAP Secondary Assessments were developed in 1998. These assessments were designated as the end of course examinations for ten high school courses, thereby extending the state's accountability system into secondary schools.
	California	The state of California had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. As of 2001 the state tests had undergone an external alignment review, and they were reviewed again between then and the 2005-2006 school year. The state held schools accountable for performance and offered assistance during the 2001-2002 school year, and by the 2005-2006 school year the state also imposed sanctions, including school closure, reconstitution, and permitting students to transfer schools. The state did provide rewards for high performing schools during the 2001-2002 school year, but stopped providing rewards by 2005-2006. While California did not have an exam on which graduation was contingent in 2001, they had developed one by 2005. The state saw an increase in their graduation rate by 2.1.%, an increase that was smaller than both the ten state as well as national averages. The state's overall Quality Counts grade in the category of standards and accountability increased from a B in 2001 to a B+ in 2005. (From profile) The state enacted the Public Schools Accountability Act (PSAA) of 1999, which included the Accountability Performance Index (API), used to measure school performance on an index that ranged from 200 to 1000. While the Standardized Testing and Reporting (STAR) program was authorized in 1997, the Stanford 9 was designated as the state test for administration beginning in 1998. The test was given to students in grades 2 through 8 in reading, mathematics, written expression, and spelling. Secondary students (grades 9 through 11) were tested in reading, writing, mathematics, science, and history/social science. Other changes in accountability included the edition of an assessment for native Spanish speakers, the Spanish Assessment of Basic Education, 2nd Edition (SABE/2), in 1999. Also during this period, the California Content Standards tests were developed. These assessments were comprised only of California specific items,

Ranking	State	Rationale
		unlike earlier assessments that consisted of a combination of Stanford 9 questions and items created specifically for the California tests.
Middle Ground	North Carolina	North Carolina had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests underwent an external alignment review by 2001, but did not undergo another between 2001 and 2005-2006. The state held schools accountable for performance and offered assistance and sanctions during the entire period, allowing the use of reconstitution throughout. The state provided rewards for high performing schools during the entire examined period. North Carolina used an exam on which graduation was contingent throughout the period, and increased their graduation rate by 5.8%, to 66.1%. The state's overall *Quality Counts* grade in the category of standards and accountability was consistent at a B both in 2001 and 2005. (From profile) Beginning in 1998, students were given a High School Comprehensive Test. The North Carolina High School Comprehensive Test was first given to 10th grade students in 1998, the test was then abolished in 2001 because of budget issues, then reinstated in 2002 to comply with NCLB requirements. There were not any state administrated science or social studies/history tests for students in grades 3 through 8 during the 1998-2005 time period.
	Kentucky	The state of Kentucky had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests did not undergo an external alignment review at all during this period, but were in the process of a review by the 2005-2006 school year. The state held schools accountable for performance and offered assistance and sanctions during the entire period, including permitting students to transfer in 2001-2002, with the addition of reconstitution by 2005-2006. The state did provide rewards for high performing schools during the 2001-2002 school year, but stopped providing rewards by 2005-2006. Kentucky did not use an exam on which graduation was contingent at all during the examined period, but did increase their graduation rate by 6.3%, to 70.0%. The state's overall *Quality Counts* grade in the category of standards and accountability decreased from an A in 2001, to a B+ in 2005. (From profile) In 1998, the state developed a new assessment system, named the Commonwealth Accountability Testing System (CATS). Assessments from this system were first given to students in 1999. These tests, referred to as the Kentucky Core Content Tests, were standards based , and consisted of open response and multiple choice items, on-demand writing, portfolios, as well as alternate portfolios for

Ranking	State	Rationale
		special needs students.
	Virginia	Virginia had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests had not underwent an external alignment review by 2001, but did undergo a review between 2001 and 2005. The state held schools accountable for performance and offered assistance during the entire period, but did not have the authority to impose any sanctions on persistently low performing schools. The state did not provide rewards for high performing schools during the entire examined period. Virginia did not have an exam on which graduation was contingent in 2001, but developed one by 2005. Overall the state's graduation rate decreased by 4.2%, to 73.1%. The state's overall *Quality Counts* grade in the category of standards and accountability was consistent at a B both in 2001 and 2005. (From profile) The Virginia Standards of Learning Assessments were first implemented in 1998, and were designed to test all of the content in the Virginia Standards of Learning (SOL). When first implemented, these assessments were given to students in grades 3, 5, 8 and high school, but were extended to all grades 3 through 8 in 2005.
	Texas	The state of Texas had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests did not undergo an external alignment review at all during the examined time period. The state held schools accountable for performance and offered assistance and sanctions during the entire period, including school closure and reconstitution and permitting students to transfer, throughout. The state did provide rewards for high performing schools during the 2001-2002 school year, but stopped providing rewards by 2005-2006. Texas used an exam on which graduation was contingent throughout the period, and increased their graduation rate by 4.4%, to 63.7%. The state's overall *Quality Counts* grade in the category of standards and accountability was consistent at B- both in 2001 and 2005. (From profile) The state revised their assessments to align with the state standards. These assessments were called the Texas Assessment of Knowledge and Skills (TAKS), and began administration in 2003. The TAKS tested students in reading in grades 3-9, writing in grades 4 and 7, language arts in grades 10 and 11, mathematics in grades 3-11, science in grades 5, 10, and 11, and social studies in grades 8, 10, and 11.
	Maryland	Maryland had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests had undergone an external alignment review by 2001, as well as an additional review

Ranking	State	Rationale
		between 2001 and 2005. The state held schools accountable for performance and offered assistance and imposed sanctions during the entire period, including school closure and reconstitution. The state did provide rewards for high performing schools during the entire examined period. Maryland did not have an exam on which graduation was contingent in 2001. Overall the state's graduation rate increased by 2.0%, a difference lower than both the national and ten state average, to 74.7%. The state's overall *Quality Counts* grade in the category of standards and accountability decreased from an A in 2001 to an A- in 2005. (From profile) Maryland's accountability system, Maryland School Performance Assessment Program (MSPAP), was developed prior to the examined time period, and discontinued in 2002 in order to meet the requirements of NCLB. The Maryland School Assessment (MSA) was developed to assess students' knowledge of the Maryland content standards in reading and mathematics for students in grades 3 through 8. The High School Assessment (HSA) was first administered to students who entered ninth grade in 2001.
Laggards	New York	During the examined time period, the state of New York had a consistent, and high quality accountability system, however the state enacted very few changes in their accountability system. The state had mathematics assessments aligned with the state standards at all three levels, elementary, middle and high school, throughout the entire period. The state tests did not undergo an external alignment review at all during that time. The state held schools accountable for performance and offered assistance and sanctions during the entire period, including school closure and reconstitution in 2001-2002, in addition to permitting students to transfer schools by 2005-2006. The state did not provide rewards for high performing schools at all during the time period. New York used an exam on which graduation was contingent throughout the period, and increased their graduation rate by 4.5%, to 65.0%. The state's overall *Quality Counts* grade in the category of standards and accountability was consistent at an A both in 2001 and 2005. (From profile) In 2005, New York began testing all students in grades 3 through 8 in mathematics and English/language arts , as per the NCLB requirement. Prior to this, only students in grades 4 and 8 were assessed in those subjects.
	Michigan	Like New York, the state of Michigan implemented few changes regarding their state assessment and accountability policies during the 1998-2005 period, thus landing them in the laggard category. In addition, their overall graduation rate decreased by 3.8% during the 2000 through 2004 period, to an overall graduation rate of 69.1% which is just above the ten-state average, and just

Ranking	State	Rationale
		below the national average. While the state began the period with mathematics assessments aligned with the state standards at the elementary and high school levels, by the 2005-2006 school year, the middle school mathematics assessments aligned with the state standards had also been developed. As of 2001 the state tests had undergone an external alignment review, as well as another review that took place again between then and the 2005-2006 school year. The state did not hold schools responsible for performance, not offering any assistance or sanctions in 2001, but the state did hold schools accountable for performance and offered assistance as well as imposed sanctions by 2005, including school closure, reconstitution as well as permitting students to transfer schools. The state did provide rewards for high performing schools during the 2001-2002 school year, but stopped providing rewards by 2005-2006. It was the only one of the ten states that did not produce a school report card in 2001, but had developed report cards by the 2005-2006 school year. The state's overall *Quality Counts* grade in the category of standards and accountability increased from a C in 2001 to a B in 2005. (From profile) The Michigan Educational Assessment Program (MEAP) was first administered during the 1969-1970 school year, and continued to be the primary method of student accountability through the examined time period. However, the system did undergo some changes in the fall of 2005 in order to comply with NCLB requirements including redesigning the assessments and complying with testing schedules in content areas and grade levels.

education policy analysis archives