
education policy analysis archives

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



Arizona State University

Volume 20 Number 20

July 20, 2012

ISSN 1068-2341

High-stakes Testing and Student Achievement: Updated Analyses with NAEP Data

Sharon L. Nichols

University of Texas at San Antonio

Gene V Glass

University of Colorado Boulder

David C. Berliner

Arizona State University
USA

Citation: Nichols, S. L., Glass, G. V., Berliner, D.C. (2012) High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives*, 20 (20) Retrieved [date], from <http://epaa.asu.edu/ojs/article/view/1048>

Abstract: The present research is a follow-up study of earlier published analyses that looked at the relationship between high-stakes testing pressure and student achievement in 25 states. Using the previously derived Accountability Pressure Index (APR) as a measure of state-level policy pressure for performance on standardized tests, a series of correlation analyses was conducted to explore relationships between high-stakes testing accountability pressure and student achievement as measured by the National Assessment for Education Progress (NAEP) in reading and math. Consistent with earlier work, stronger positive correlations between the pressure index and NAEP performance in fourth grade math and weaker connections between pressure and fourth and eighth grade reading performance were found. Policy implications and future directions for research are discussed.

Keywords: high-stakes testing; NAEP; Accountability Pressure Index (APR).

Journal website: <http://epaa.asu.edu/ojs/>
Facebook: /EPAAAAPE
Twitter: @epaa_aape

Manuscript received: 11/18/2011
Revisions received: 02/20/2012
Accepted: 03/20/2012

Pruebas de consecuencias severas y rendimiento de los estudiantes: análisis actualizado con datos de NAEP

Resumen: Este estudio da seguimiento a investigaciones ya publicadas que analizaron la relación entre las presiones de las pruebas de consecuencias severas y el rendimiento de estudiantes en 25 estados. Utilizando el Índice de Presión y Responsabilidad (APR por su sigla en inglés) como medida de presión política a nivel estatal sobre el rendimiento en los exámenes estandarizados, se realizaron una serie de análisis de correlación para explorar las relaciones entre pruebas de consecuencias severas, rendición de cuentas y rendimiento de los estudiantes, medidos por los resultados de el Instituto Nacional Evaluación para el Progreso de la Educación (NAEP por su sigla en inglés) en lectura y matemáticas. Consistente con resultados de trabajos anteriores, las correlaciones positivas más fuertes fueron entre el índice de presión y el rendimiento de la NAEP en matemáticas en cuarto grado y las conexiones más débiles entre la presión y el rendimiento de lectura en cuarto y octavo grado. El trabajo concluye con implicaciones para las políticas educativas y sugerencias para continuar investigando este tema.

Palabras clave: pruebas de consecuencias severas; NAEP; Índice de Presión y Responsabilidad

Provas de consequências severas e desempenho do aluno: análise atualizada de dados NAEP

Resumo: Este estudo prossegue uma investigação anteriormente publicada que examinou a relação entre as provas de consequências severas e o desempenho de alunos em 25 estados. Usando o Índice de Pressão e Responsabilidade (APR – sigla em Inglês) como medida de pressão política a nível estadual sobre o desempenho em testes padronizados foram realizados uma série de correlações para explorar as relações entre provas de consequências severas, prestação de contas e desempenho dos alunos, medidos pelos resultados da Avaliação Nacional do Progresso educacional (NAEP – sigla em Inglês) em leitura e matemática. Consistente com os resultados de estudos anteriores, as correlações positivas foram mais fortes entre o índice de pressão e desempenho em NAEP em matemática na quarta série e ligações mais fracas entre pressão e desempenho de leitura na quarta e oitava séries. O artigo conclui com implicações para as políticas educacionais e sugestões para futuras pesquisas nesse tema.

Palavras-chave: provas de consequências graves; NAEP; Índice de Pressão e Responsabilidade.

Introduction

The present study adds to a growing literature on the relationship of high-stakes testing accountability and student achievement. The major goal of federal and state high-stakes testing policies is to improve schools. The theory of action undergirding this approach suggests that by tying negative consequences (e.g., public exposure, external takeover) to standardized test performance, teachers and students in low performing schools will work harder and more effectively, thereby increasing what students learn. Although the practice of high-stakes testing dates back several decades in various districts (Chicago public schools) and states (Texas, New York, Florida), the passage of the No Child Left Behind Act in 2002 mandated high-stakes testing nationwide and at many more grade levels than was customary. The extant literature on high-stakes testing and student achievement can be organized into three types. In the first type, researchers use two-group designs to compare achievement patterns in states with accountability practices versus those without such practices, or in states with a long history of accountability versus those with shorter histories (Amrein & Berliner, 2002a; Amrein-Beardsley & Berliner, 2003; Braun, 2004; Dee & Jacob, 2009). A second approach has analysts ranking states according to some measure of

accountability and then using correlation or regression techniques to ascertain the form and significance of the relationship between accountability measures and student achievement (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005). A third type of research focuses on specific aspects of high-stakes testing practice and impact as they affect particular districts, regions, or states (Clarke, Haney, & Madaus, 2000; Jacob, 2001; Winters, Trivitt, & Greene, 2010).

Each approach has limitations of methods, making it difficult to determine with confidence the effects of high-stakes testing. Still, a pattern seems to have emerged that suggests that high-stakes testing has little or no relationship to reading achievement, and a weak to moderate relationship to math, especially in fourth grade but only for certain student groups (Braun, Wang, Jenkins, & Weinbaum, 2006; Braun, Chapman, & Vezzu, 2010; Figlio & Ladd, 2008; Nichols, Glass, & Berliner, 2006). This particular pattern of results (only affecting fourth grade math) raises serious questions about whether high-stakes testing increases learning or merely more vigorous test preparation practices (i.e., teaching to the test).

This study is a follow-up to our earlier work in which we used an empirically-derived measure of state level high-stakes testing policy to examine the relationship between accountability policy implementation and student achievement (as measured by the National Assessment for Education Progress, NAEP). In contrast to other research that measures high-stakes testing accountability according to the number of laws passed that are associated with accountability (Clarke *et al.*, 2003; Pedulla *et al.*, 2003), or by estimating the acceptance of accountability based on state level variables such as funding and student demographic characteristics (Braun *et al.*, 2006; Carnoy & Loeb, 2002), our measure was derived from both legislative efforts as well as “on-the-ground” implementation, response, and reaction (see Nichols *et al.* 2006). In our earlier analyses, we used the unique measure of accountability pressures that we created with NAEP 4th and 8th grade data from 1992-2003. The purpose of this follow-up study is to examine state policy accountability (measured by our state level accountability pressure index) as it relates to more recent (e.g., 2005, 2007, and 2009) NAEP data available for 4th and 8th grade math and reading.

Review of Literature

High-stakes testing is the process of attaching significant consequences to standardized test performance with the goal of incentivizing teacher effectiveness and student achievement (Herman & Haertel, 2005; Ryan, 2004). The rationale is that by attaching significant rewards or serious threats to changes in student test scores, teachers and their students will inevitably be prompted to work harder, better, and learn more. Although most tests students take are arguably “high-stakes” to them (i.e., failing a teacher-made test could result in failing a class or not passing to the next grade), “high-stakes” here refers to standardized tests developed specifically for the purpose of evaluating teachers and students. Performance on these tests may result in important consequences to schools, administrators, teachers, and students. Passing could bring rewards to teachers (bonuses) and schools (positive reviews in local newspapers), whereas failure could bring severe penalties to teachers and principals (termination), schools (closure or “take-over”), and to students (denied diploma or retained in grade).

Although the practice of high-stakes testing gained a prominent position in educational reform with the passage of the No Child Left Behind Act (NCLB) of 2002, its use as a lever for school change preceded NCLB. Tests have been used to distribute rewards and sanctions to teachers in urban schools since the mid 1800s (Tyack, 1974) and for most schools throughout the United States since at least the 1970s (Haertel & Herman, 2005). New York state in particular has led the United States in test-based accountability efforts, “implementing state-developed (1965) and

mandated minimal competency testing (MCT) before most other states (1978) and disseminating information to the media about local district performance on the state assessments before it became routinely popular (1985)” (Allington & McGill-Franzen, 1992, p. 398).

Standardized achievement test invention, development, and use paralleled these reform efforts (Giordano, 2005). The evolution of valid and reliable measurement techniques influenced views of how one might gauge educational quality (McDonnell, 2005). The passage of NCLB in 2002, mandated the most intrusive use of tests for influencing how and what teachers would teach and how and what students would learn. In spite of a growing literature indicating that high-stakes testing has had deleterious effects on teaching practices and student motivation, policymakers continue to argue for its effectiveness in increasing student learning as evidenced in newer proposals (e.g., U.S. Department of Education, 2009) and recommendations for the reauthorization of NCLB (U.S. Department of Education, 2010).

High-Stakes Testing and Student Outcomes

The gradual adoption of accountability-based practices in the years leading up to NCLB provided a context in which to study their effects on achievement at the state level. By the late 1990s, increasing numbers of states had begun adopting test-based accountability plans; however, their form and function varied widely. Some states had developed criterion-based standardized tests, others were just starting the process. Some states were actively using tests to hold teachers and students accountable; others were in the process of developing such mechanisms. Around this time it was discovered that scores on virtually all of these tests went up, probably as a function of test preparation and familiarity with the tests, rather than because students were learning much more (Linn, Graue, & Sanders, 1990; Shepard, 1990). This should have been a lesson for those who designed NCLB accountability, but this pervasive finding and explanations of it were ignored.

Most of the research conducted around the time of NCLB provides scant support for the effectiveness of high-stakes tests in increasing student achievement (Amrein & Berliner, 2002a, b; Braun, 2004; Rosenshine, 2003) or graduation rates (Haney *et al.*, 2004; Heubert & Hauser, 1999; Marchant & Paulson, 2005). Since our initial study, no data have emerged to contradict the findings that accountability pressure has some relationship to fourth grade math, virtually no influence on reading (Dee & Jacob, 2009), and only negative influence on student graduation rates (Holme, Richards, Jimerson, & Cohen, 2010; Orfield, Losen, Wald, & Swanson, 2004). Studies focusing on both high- and low-stakes exit exams repeatedly reveal that these types of incentives/threats have little to no impact on student achievement over time (e.g., Bishop, Mane, Bishop, & Moriarty, 2001; Grodsky, Warren, & Kalogrides, 2009; Reardon, Arshan, Atteberry, & Kurlaender, 2008; Reardon, Atteberry, Arshan, & Kurlaender, 2009). In addition, the reduction of the achievement gap between income groups and between racial and ethnic groups, a major goal of the high-stakes accountability movement, either did not occur or was only marginally effective in the years these policies have been in place (Reardon, 2011; Timar & Maxwell-Jolly, 2012).

The Initial Study

Our initial study (Nichols *et al.*, 2006) was prompted by our view that existing approaches to the *measurement* of test-based accountability policies and practice at the state level were largely inadequate because of their reliance on inspection of state level legislation, as opposed to actual practices. That is, most researchers measured testing “pressure” by examining the number of laws that states had passed prior to or up to the enactment of NCLB. Although reliable, the validity of such approaches for capturing the on-the-ground feeling of pressure was doubtful. In our initial study, we addressed this problem by spending considerable time and effort conceptualizing and

building a measure that would more closely represent high-stakes testing policy implementation and which we labeled the Accountability Pressure Rating (APR).

Our method for deriving APR values for our 25¹ study states was guided by the method of “comparative judgments” used for ordering complex and abstract psychological data (Torgerson, 1960). This approach seemed ideal for our purposes since our goal was to transform complex qualitative data (state level policy legislation enactment and implementation) into a quantitative indicator that can be used in subsequent analyses. Our work involved three steps. First, we created state-level portfolios that included a range of legislative documentation, state-generated accountability reports, and newspaper articles documenting the range of ways policy changes both impacted and were viewed by the public (see Nichols *et al.*, 2006 for a complete description of portfolio contents). One unique aspect of our approach was the inclusion of newspaper references (both leading stories as well as editorials) that were used to capture the on-the-ground effects of and reactions to local and statewide test-based accountability practices. In contrast to other studies that relied on quantitative estimations of policies (e.g., Braun *et al.*, 2006; Carnoy & Loeb, 2002) our measure included evidence that described how policies played out in local school systems.

Next, we asked 300 graduate students each to view two states’ portfolios and to make two judgments—which state exerted more pressure and by about how much (on a scale of 1-7²). Last, we took the ratings provided by our students and applied the least-squares solution for uni-dimensional scale values due to Mosteller (as outlined in Torgerson, 1960, pp. 170–173). The result was a scale ranging from .54 to 4.78. This rating served as our measure of state-level testing pressure as of 2004 (See Table 1). As can be seen in Table 1, Kentucky’s policies and practices were consistently rated below other states in terms of test-based pressure (APR = .54), whereas Texas’s policies and practices were consistently viewed as having the highest test-based accountability pressure (APR= 4.78).

Table 1
Accountability Pressure Rating (APR, 2004)

| State | APR | State | APR | State | APR |
|--------------|------|---------------|------|-------------|------|
| Kentucky | 0.54 | Utah | 2.80 | Georgia | 3.44 |
| Wyoming | 1.00 | Maryland | 2.82 | Tennessee | 3.50 |
| Connecticut | 1.60 | Alabama | 3.06 | Louisiana | 3.72 |
| Hawaii | 1.76 | Virginia | 3.08 | Mississippi | 3.82 |
| Maine | 1.78 | West Virginia | 3.08 | New York | 4.08 |
| Rhode Island | 1.90 | Massachusetts | 3.18 | NC | 4.14 |
| Missouri | 2.14 | SC | 3.20 | Texas | 4.78 |
| California | 2.56 | New Mexico | 3.28 | | |
| Arkansas | 2.60 | Arizona | 3.36 | | |

¹ NAEP began disaggregating student achievement by state in 1990. Eighteen states participated in this assessment schedule since its inception and therefore have available a complete set of NAEP data on fourth- and eighth-grade students in math and reading. These are Alabama, Arizona, Arkansas, California, Connecticut, Georgia, Hawaii, Kentucky, Louisiana, Maryland, New Mexico, New York, North Carolina, Rhode Island, Texas, Virginia, West Virginia, and Wyoming. Seven states are missing one assessment—the eighth-grade math test from 1990. These are South Carolina, Massachusetts, Maine, Mississippi, Missouri, Tennessee, and Utah. All 25 states are the focus of this study.

² We felt a 7-point scale would provide enough variability in raters’ responses while at the same time minimize reliability issues emanating from wider scales.

Using our APR, we performed correlation and regression analyses to examine patterns in the relationships between our APR and fourth and eighth grade reading and math NAEP through 2003. Our findings revealed that APR was connected most consistently with gains in fourth grade math performance, only slightly connected to gains in eighth grade math, and not correlated with gains in reading at either fourth or eighth grade levels (Nichols *et al.*, 2006).

Study Goals and Research Questions

To date there is no consistent and therefore no convincing evidence that high-stakes testing works to increase student achievement, except weakly in certain areas of the math curriculum. Thus, in spite of the claims of some (Raymond & Hanushek, 2003) who argue that the benefits of high-stakes testing are well established, it appears that most research fails to support the contention that high-stakes testing increases student learning. Further, the continued emphasis on test-based accountability as a panacea for school reform (e.g., the 2010 Race to the Top initiative) prompts us to reconsider the relationship of high-stakes testing policies with student achievement. Therefore, our goal in this study is to re-examine the relationship between high-stakes testing pressure using our APR measure as it relates to NAEP data emanating from later years of NCLB enactment.

The primary research question guiding this study is: What is the relationship between state-level high-stakes testing pressure and student achievement? More specifically, we want to know: What is the pattern of correlations between APR and fourth and eighth grade NAEP scores in reading and math from 2005-2009:

- over time;
- when disaggregated by student ethnicity;
- when disaggregated by student socioeconomic status;

Additionally, what is the relationship between APR and four-year NAEP gains (both cohort and non-cohort) in math and reading?

Method

Our analyses are organized into two parts. First, we use descriptive statistics to analyze fourth and eighth grade NAEP data during the period 2000-2009 in reading and math.³ Next, we conduct a series of partial, part, and simple bivariate correlations to examine relationships among state level demographic characteristics, APR, and NAEP indicators.

Data

For the achievement data, we used state-level NAEP (scale scores) in fourth and eighth grade math and reading for all students and disaggregated by student socioeconomic status⁴ and ethnicity.⁵ State-level demographic data were drawn from a variety of online databases and include characteristics of students in the state (e.g., percent who are African American), percent of state

³ Although the focus of this study is with NAEP 2005, 2007, 2009, we include in these analyses earlier data available (2000 and 2002) in order to look at trends over time from before NCLB and in the year just as it was passed. These years are also important benchmarks since our APR was derived around the time NCLB was just getting started (i.e., 2004).

⁴ Students were characterized according to two categories: Eligible for free and reduced lunch (Low SES) and Not Eligible for free and reduced lunch (High SES).

⁵ Students self identified according to choices of: African American, Hispanic, White, Asian/Pacific Islander, and Other. We focus on African American, Hispanic, and White student subgroups in this study.

population living in poverty, school enrollment characteristics, and total per student revenues.⁶ This study focuses on 25 study states, some of which did not have large enough African American or Hispanic student populations to generate NAEP data for each group. Thus, throughout our analysis, we report specific sample sizes when data involve NAEP data disaggregated by student ethnicity.

Results

Part I: Descriptive Analysis. Fourth and eighth grade math.

Means and standard deviations in fourth and eighth grade math across time and disaggregated by student ethnicity and socioeconomic status are presented in Table 2. All subgroup averages fall below the level of “proficiency” set by NAEP. Interestingly, except for Hispanic students in 2000 and 2005, fourth graders’ math performance demonstrated less variability than eighth graders. Average NAEP performance across time and disaggregated by student ethnicity and socioeconomic status are also displayed in Figures 1 and 2. Across each administration of NAEP, high SES students scored on average consistently higher than any other subgroup. By contrast, African American students consistently posted lowest average NAEP scores in both fourth and eighth grade math. Fourth and eighth grade math average scores rose more dramatically from 2000 to 2003 (pre-NCLB) than from 2003 to 2009 (post-NCLB).

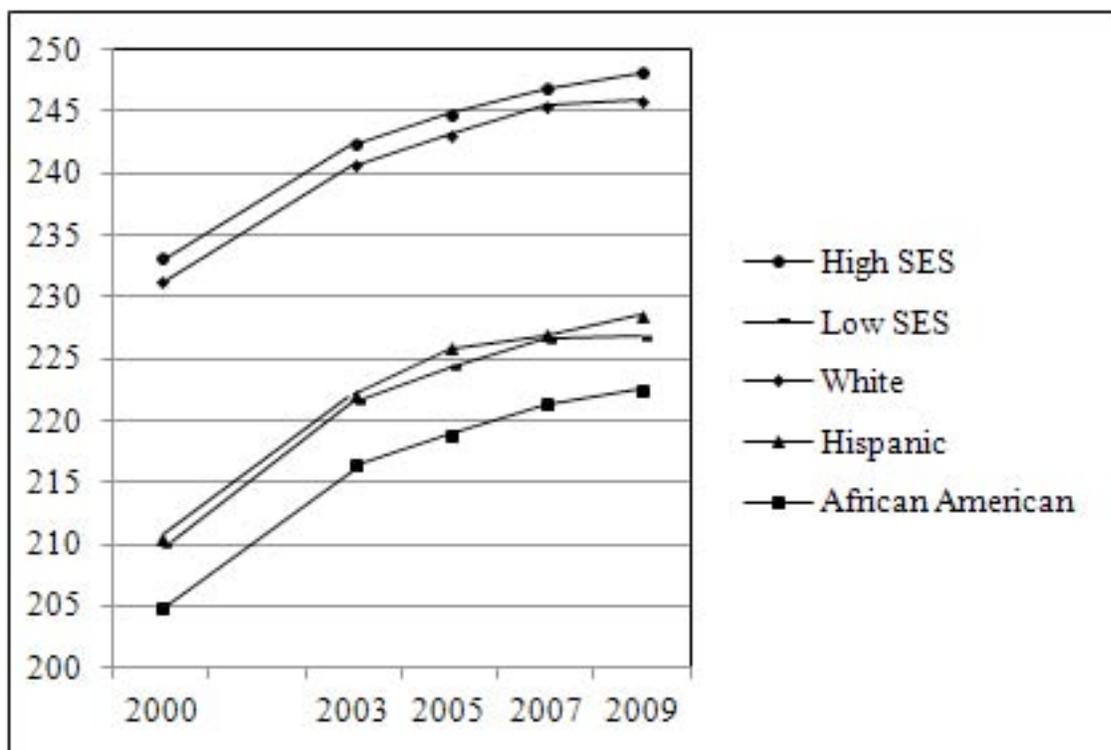


Figure 1. Fourth Grade Math NAEP State-Level Standard Score Averages: 2000-2009

⁶ NAEP data downloaded from the National Center for Education Statistics website, nces.ed.gov. Demographic and revenue data both come from the National Center for Education Statistics website through their Common Core of Data, nces.ed.gov; Enrollment figures downloaded from US Census Bureau website, <http://www.census.gov>

Table 2

Fourth and Eighth Grade Math NAEP: Means and Standard Deviations Disaggregated by Student Ethnicity and Student Socioeconomic Status (25 Study States: 2000-2009)

| | | Fourth Grade | | | Eighth Grade | | |
|-------------------|--------------------------|--------------|---------------|-------------|--------------|---------------|-------------|
| | | N | M | SD | N | M | SD |
| All Students | 2009 | 25 | 237.64 | 6.31 | 25 | 279.65 | 7.00 |
| | 2007 | 25 | 237.24 | 6.02 | 25 | 278.11 | 8.09 |
| | 2005 | 25 | 235.08 | 6.05 | 25 | 274.86 | 7.58 |
| | 2003 | 25 | 232.40 | 6.12 | 25 | 274.88 | 7.64 |
| | 2000 | 25 | 222.68 | 6.71 | 25 | 268.38 | 7.59 |
| | <i>All Years Average</i> | 25 | <i>233.01</i> | <i>6.25</i> | 25 | <i>275.17</i> | <i>7.78</i> |
| Low SES Students | 2009 | 25 | 226.86 | 5.27 | 25 | 265.60 | 5.87 |
| | 2007 | 25 | 226.56 | 5.46 | 25 | 264.29 | 6.73 |
| | 2005 | 25 | 224.36 | 5.27 | 25 | 260.53 | 6.85 |
| | 2003 | 25 | 221.56 | 5.32 | 25 | 258.43 | 6.24 |
| | 2000 | 25 | 210.08 | 6.12 | 25 | 251.04 | 7.01 |
| | <i>All Years Average</i> | 25 | <i>221.84</i> | <i>5.49</i> | 25 | <i>259.98</i> | <i>6.54</i> |
| High SES Students | 2009 | 25 | 248.15 | 4.60 | 25 | 290.71 | 6.47 |
| | 2007 | 25 | 246.88 | 4.24 | 25 | 288.29 | 6.74 |
| | 2005 | 25 | 244.84 | 4.61 | 25 | 285.21 | 6.13 |
| | 2003 | 25 | 242.32 | 4.69 | 25 | 283.98 | 6.08 |
| | 2000 | 25 | 233.20 | 4.62 | 25 | 278.28 | 6.39 |
| | <i>All Years Average</i> | 25 | <i>243.08</i> | <i>4.55</i> | 25 | <i>285.29</i> | <i>6.36</i> |
| African American | 2009 | 24 | 222.56 | 6.09 | 23 | 260.65 | 7.06 |
| | 2007 | 23 | 221.39 | 4.97 | 21 | 258.57 | 7.08 |
| | 2005 | 22 | 218.90 | 5.23 | 21 | 253.97 | 7.32 |
| | 2003 | 22 | 216.44 | 5.65 | 21 | 251.74 | 6.81 |
| | 2000 | 21 | 204.79 | 6.93 | 20 | 243.86 | 7.55 |
| | <i>All Years Average</i> | 20 | <i>216.82</i> | <i>5.77</i> | <i>21.2</i> | <i>253.76</i> | <i>7.16</i> |
| Hispanic | 2009 | 22 | 228.54 | 6.13 | 21 | 268.34 | 7.43 |
| | 2007 | 22 | 226.91 | 6.27 | 20 | 264.32 | 8.45 |
| | 2005 | 19 | 225.81 | 6.97 | 17 | 260.71 | 7.09 |
| | 2003 | 19 | 222.04 | 6.90 | 16 | 258.19 | 7.23 |
| | 2000 | 14 | 210.53 | 8.65 | 12 | 251.29 | 8.13 |
| | <i>All Years Average</i> | 18 | <i>222.77</i> | <i>6.98</i> | <i>17.2</i> | <i>260.57</i> | <i>7.67</i> |
| White | 2009 | 25 | 245.94 | 5.78 | 25 | 289.07 | 7.94 |
| | 2007 | 25 | 245.44 | 5.34 | 25 | 287.60 | 7.75 |
| | 2005 | 25 | 243.10 | 5.86 | 25 | 284.46 | 7.22 |
| | 2003 | 25 | 240.69 | 5.58 | 25 | 283.41 | 6.89 |
| | 2000 | 25 | 231.17 | 5.61 | 25 | 277.90 | 6.45 |
| | <i>All Years Average</i> | 25 | <i>241.27</i> | <i>5.63</i> | 25 | <i>284.49</i> | <i>7.26</i> |

NOTE: SES=Student response based on eligibility for free/reduced lunch program at school, where Low SES=Eligible, High SES=Not Eligible.

Proficiency Levels: 4th Grade: Basic 214-248; Proficient 250-281; Advanced 282+

8th Grade: Basic 262-298; Proficient 299-332; Advanced 333+

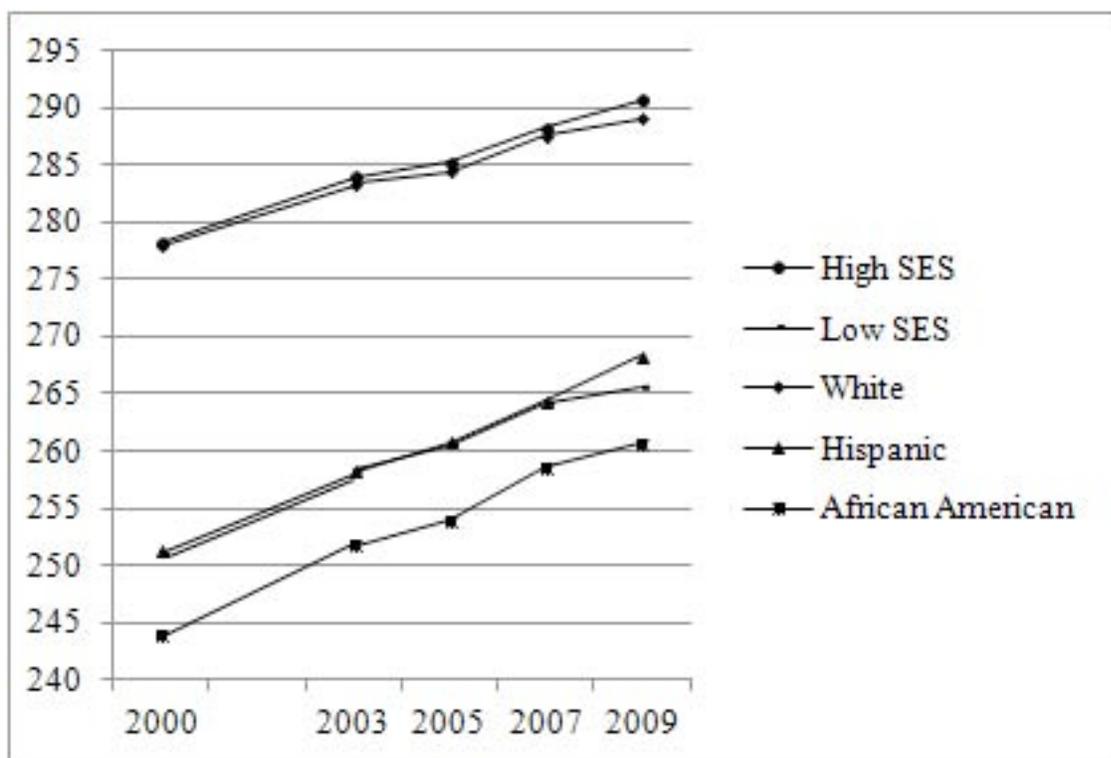


Figure 2. Eighth Grade Math NAEP State-Level Standard Score Averages: 2000-2009

We were also interested in average achievement gap patterns over time and between White and Black (WB: calculated as White subgroup average score in state i – Black subgroup average score in state i), White and Hispanic (WH: calculated as White subgroup average score in state i – Hispanic subgroup average score in state i), and Hispanic and Black (HB: calculated as Hispanic subgroup average score in state i – Black subgroup average score in state i) student sub groups for both fourth and eighth grade math. Average NAEP standard score differences and standard deviations for each subgroup are displayed in Table 3 and in Figures 3 (fourth grade) and 4 (eighth grade). Figure 3 suggests that the WB and WH achievement gaps dropped relatively more steeply in the period 2000-2003, while levelling off but still declining slowly from 2003 to 2009. By contrast, the HB gap seemed to increase over time. An examination of averages in fourth grade math suggests that consistently, Hispanic students outperform African American students.

Table 3
Average Math NAEP Achievement Gap: 2000-2009

| | Fourth Grade | | | Eighth Grade | | |
|---------------------------|--------------|-------|------|--------------|-------|------|
| | N | M | SD | N | M | SD |
| White-Black Gap | | | | | | |
| 2009 | 24 | 23.44 | 4.94 | 23 | 28.39 | 7.38 |
| 2007 | 23 | 24.09 | 4.35 | 21 | 29.47 | 5.30 |
| 2005 | 22 | 24.27 | 5.40 | 21 | 31.07 | 5.83 |
| 2003 | 22 | 24.37 | 5.22 | 21 | 32.05 | 4.65 |
| 2000 | 21 | 26.73 | 4.78 | 20 | 34.34 | 6.47 |
| White-Hispanic Gap | | | | | | |
| 2009 | 22 | 18.25 | 5.85 | 21 | 22.44 | 8.82 |
| 2007 | 22 | 19.31 | 5.47 | 20 | 25.08 | 6.52 |
| 2005 | 19 | 19.35 | 5.70 | 17 | 26.78 | 6.08 |
| 2003 | 19 | 20.45 | 4.99 | 16 | 27.60 | 7.26 |
| 2000 | 14 | 23.63 | 7.79 | 12 | 30.18 | 7.41 |
| Hispanic-Black Gap | | | | | | |
| 2009 | 21 | 5.91 | 6.33 | 19 | 7.54 | 7.22 |
| 2007 | 20 | 5.62 | 5.98 | 17 | 4.68 | 5.17 |
| 2005 | 17 | 6.35 | 5.33 | 14 | 4.35 | 6.56 |
| 2003 | 17 | 4.75 | 5.01 | 13 | 4.36 | 4.24 |
| 2000 | 11 | 3.31 | 7.85 | 9 | 4.76 | 8.75 |

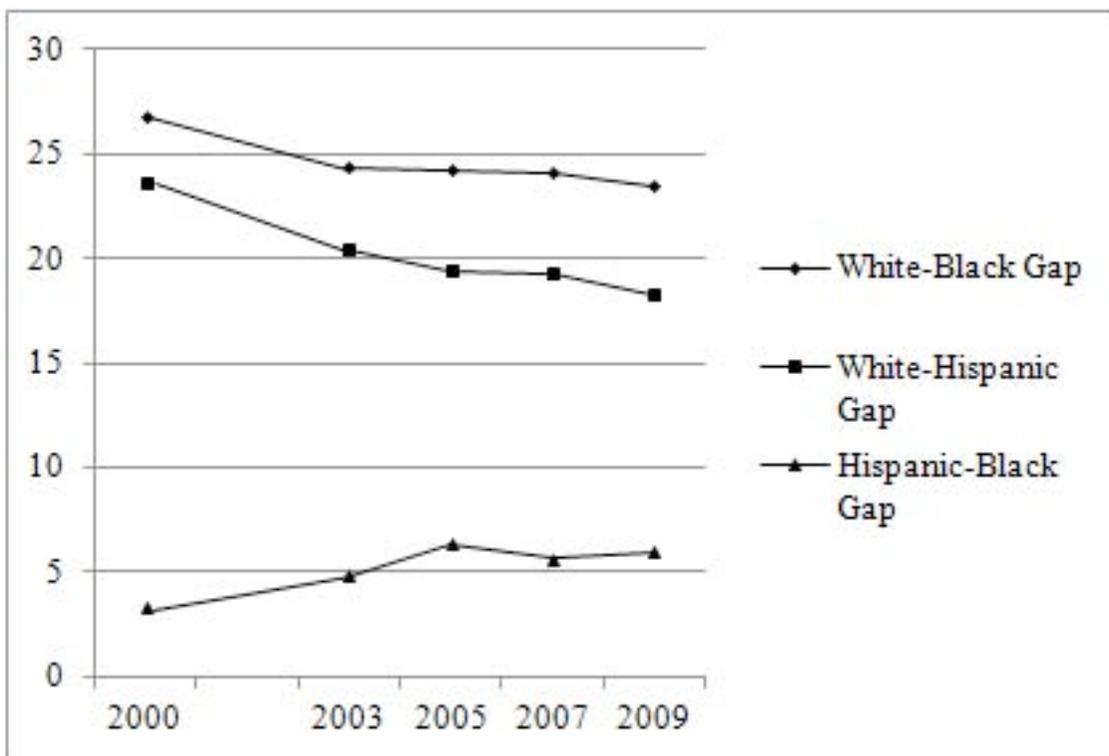


Figure 3. Fourth grade average math NAEP scale score differences: 2000-2009

In eighth grade, average math achievement gaps between WB and WH narrowed relatively steadily over time (Figure 4). By contrast, the HB gap stayed flat over time, followed by a steep increase in the period 2007-2009. As with fourth grade, in eighth grade, Hispanic students on average consistently score better than their Black peers.

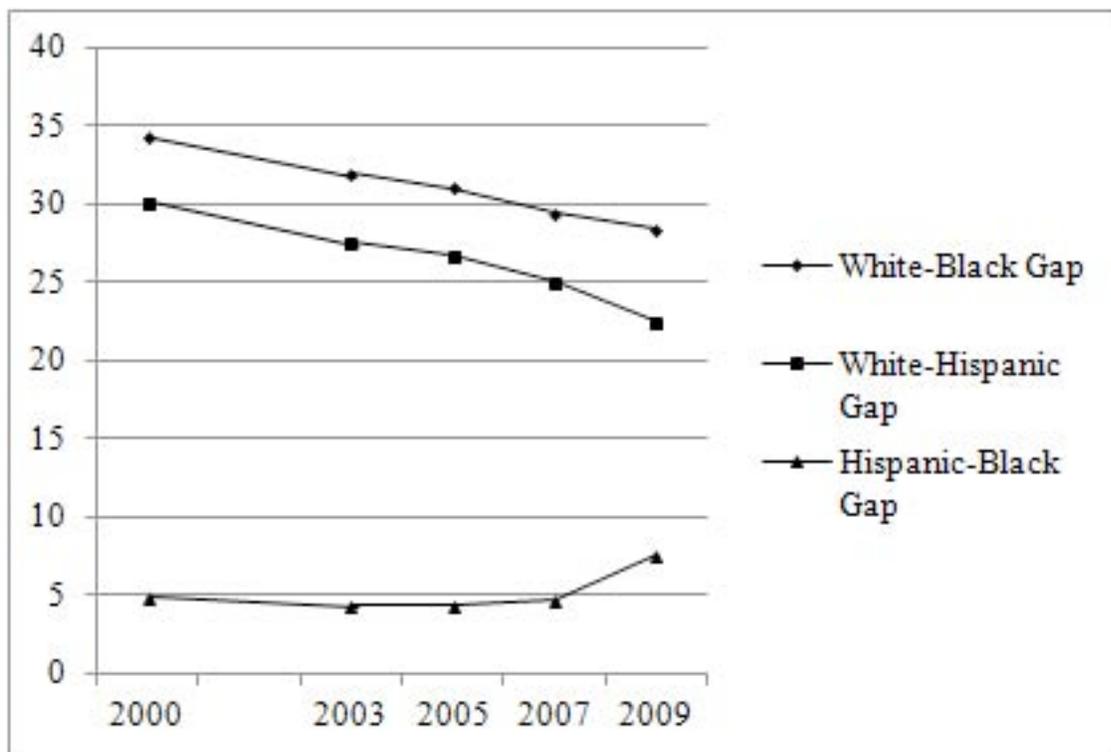


Figure 4. Eighth grade average math NAEP scale score differences: 2000-2009

Fourth and eighth grade reading.

NAEP means and standard deviations for all subgroups over time in fourth and eighth grade reading are displayed in Table 4. An examination of the pattern of standard deviations suggests that eighth graders showed less variability on average than fourth graders.

Table 4

Fourth and Eighth Grade Reading NAEP: Means and Standard Deviations Disaggregated by Student Ethnicity and Student Socioeconomic Status (25 Study States: 2002-2009)

| | | Fourth Grade | | | Eighth Grade | | |
|-------------------|--------------------------|--------------|---------------|-------------|--------------|---------------|-------------|
| | | N | M | SD | N | M | SD |
| All Students | 2009 | 25 | 218.88 | 7.06 | 25 | 261.16 | 6.38 |
| | 2007 | 25 | 218.68 | 6.99 | 25 | 259.52 | 6.40 |
| | 2005 | 25 | 216.52 | 7.20 | 25 | 259.40 | 6.71 |
| | 2003 | 25 | 215.80 | 7.36 | 25 | 260.44 | 6.18 |
| | 2002 | 25 | 216.48 | 7.99 | 25 | 261.24 | 5.83 |
| | <i>All Years Average</i> | 25 | <i>217.27</i> | <i>7.32</i> | 25 | <i>254.35</i> | <i>6.30</i> |
| Low SES Students | 2009 | 25 | 206.24 | 5.39 | 25 | 248.84 | 4.69 |
| | 2007 | 25 | 205.44 | 5.35 | 25 | 247.44 | 5.47 |
| | 2005 | 25 | 203.24 | 6.67 | 25 | 247.16 | 6.50 |
| | 2003 | 25 | 202.56 | 6.48 | 25 | 247.24 | 5.62 |
| | 2002 | 25 | 203.76 | 6.97 | 25 | 248.88 | 5.62 |
| | <i>All Years Average</i> | 25 | <i>204.25</i> | <i>6.17</i> | 25 | <i>247.91</i> | <i>5.58</i> |
| High SES Students | 2009 | 25 | 230.60 | 5.27 | 25 | 270.84 | 4.65 |
| | 2007 | 25 | 230.36 | 4.95 | 25 | 268.76 | 4.79 |
| | 2005 | 25 | 228.44 | 4.40 | 25 | 268.28 | 5.00 |
| | 2003 | 25 | 228.00 | 4.95 | 25 | 269.08 | 4.88 |
| | 2002 | 25 | 228.08 | 5.51 | 25 | 269.48 | 4.48 |
| | <i>All Years Average</i> | 25 | <i>229.10</i> | <i>5.01</i> | 25 | <i>269.29</i> | <i>4.76</i> |
| African American | 2009 | 24 | 204.17 | 5.17 | 23 | 245.74 | 5.45 |
| | 2007 | 22 | 202.82 | 5.95 | 22 | 244.14 | 5.63 |
| | 2005 | 22 | 199.64 | 5.98 | 20 | 241.90 | 4.73 |
| | 2003 | 22 | 198.78 | 6.47 | 21 | 243.57 | 4.64 |
| | 2002 | 21 | 199.43 | 6.65 | 21 | 244.43 | 4.86 |
| | <i>All Years Average</i> | 22.2 | <i>200.97</i> | <i>6.04</i> | 21.4 | <i>243.96</i> | <i>5.06</i> |
| Hispanic | 2009 | 23 | 206.83 | 7.11 | 21 | 251.43 | 6.35 |
| | 2007 | 21 | 205.91 | 6.13 | 20 | 247.20 | 5.96 |
| | 2005 | 19 | 204.47 | 7.59 | 17 | 247.71 | 6.51 |
| | 2003 | 19 | 204.05 | 7.16 | 16 | 247.06 | 7.57 |
| | 2002 | 17 | 203.06 | 8.56 | 15 | 246.27 | 6.57 |
| | <i>All Years Average</i> | 19.8 | <i>204.86</i> | <i>7.31</i> | 17.8 | <i>248.53</i> | <i>6.59</i> |
| White | 2009 | 25 | 227.80 | 5.80 | 25 | 269.24 | 5.31 |
| | 2007 | 25 | 227.80 | 5.53 | 25 | 268.28 | 5.24 |
| | 2005 | 25 | 226.04 | 4.98 | 25 | 267.56 | 4.95 |
| | 2003 | 25 | 225.56 | 5.15 | 25 | 268.68 | 4.69 |
| | 2002 | 25 | 225.64 | 6.04 | 25 | 269.24 | 4.38 |
| | <i>All Years Average</i> | 25 | <i>226.57</i> | <i>5.50</i> | 25 | <i>268.60</i> | <i>4.91</i> |

Note: SES=Student response based on eligibility for free/reduced lunch program at school.

Low SES=Eligible, High SES=Not Eligible.

Proficiency Levels: 4th Grade: Basic 208-237; Proficient 238-267; Advanced 268+

8th Grade: Basic 243-280; Proficient 281-322; Advanced 323+

Average reading performance over time and disaggregated by student ethnicity and socioeconomic status are also displayed in Figures 5 (fourth grade) and 6 (eighth grade). At both the fourth and eighth grade levels, high SES students and White students outperformed low SES,

African American, and Hispanic student groups consistently over time. Average fourth grade reading NAEP scores for all student subgroups (except for Hispanics) were relatively flat in the period 2002-2005, followed by a rise in the period 2005-2007. High SES and White students leveled off in the period 2007-2009 whereas Hispanic, low SES and African American students' performance continued to rise moderately. In eighth grade, NAEP reading performance for all subgroups of students (except low SES) dropped steadily in the period 2002-2005 following by a steady rise in the period 2005-2009 (Figure 6).

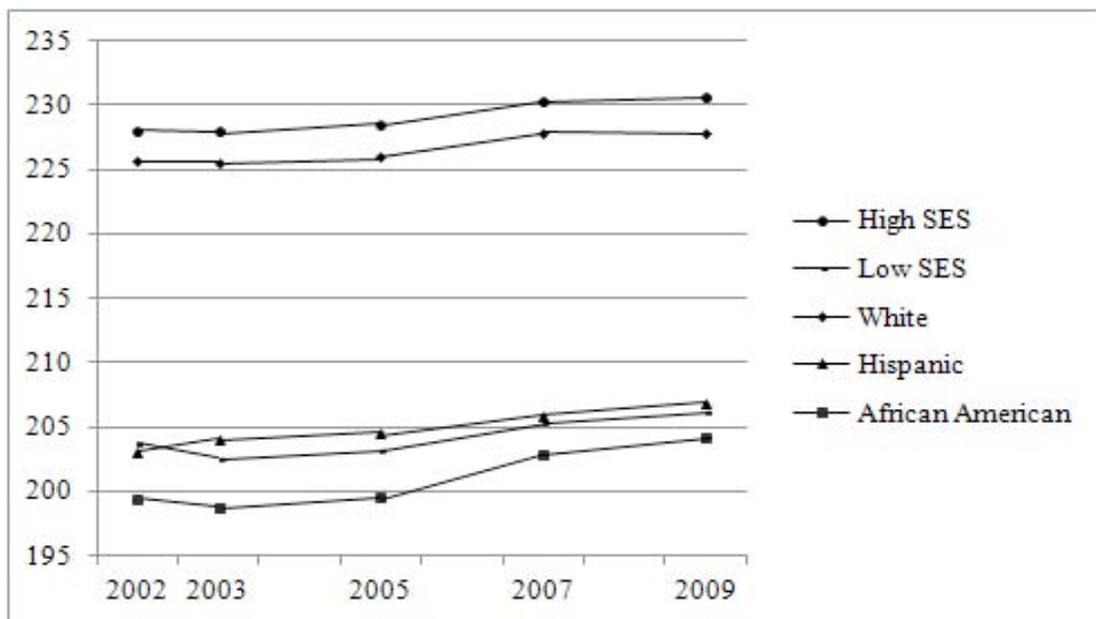


Figure 5. Fourth Grade Reading NAEP State-Level Standard Score Averages: 2002-2009

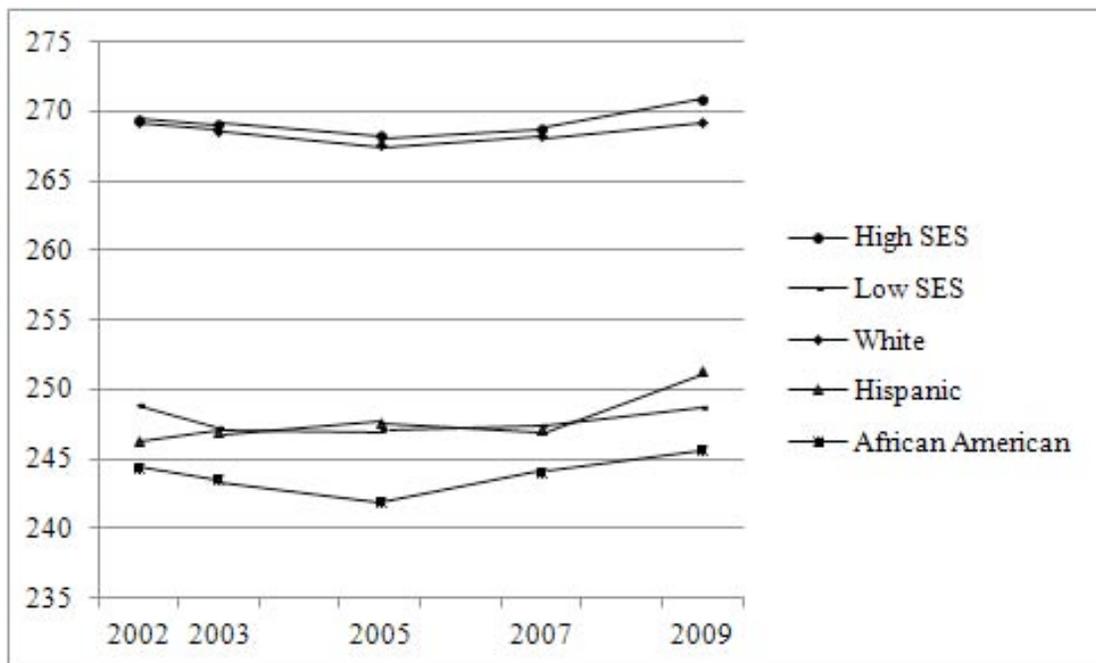


Figure 6. Eighth-Grade Reading NAEP State-Level Standard Score Averages: 2002-2009

Average NAEP standard score differences and standard deviations for each subgroup are displayed in Table 5. Average NAEP reading achievement gap trends are displayed in Figure 7 (fourth grade) and Figure 8 (eighth grade). The WB achievement gap in fourth grade reading increases at first but then drops slowly in the period 2003-2009. The WH gap stays relatively flat in the period 2002-2009.

Table 5
Average Reading NAEP Achievement Gap: 2000-2009

| | Fourth Grade | | | Eighth Grade | | |
|--------------------|--------------|-------|------|--------------|-------|------|
| | N | M | SD | N | M | SD |
| White-Black Gap | | | | | | |
| 2009 | 24 | 23.79 | 3.79 | 23 | 23.48 | 6.59 |
| 2007 | 22 | 25.14 | 5.24 | 22 | 24.14 | 5.73 |
| 2005 | 22 | 26.41 | 4.96 | 20 | 26.05 | 3.63 |
| 2003 | 22 | 27.05 | 6.21 | 21 | 25.57 | 4.30 |
| 2002 | 21 | 26.52 | 5.71 | 21 | 25.14 | 5.73 |
| White-Hispanic Gap | | | | | | |
| 2009 | 23 | 21.65 | 7.02 | 21 | 19.20 | 7.20 |
| 2007 | 21 | 22.95 | 6.91 | 20 | 22.25 | 6.30 |
| 2005 | 19 | 22.21 | 7.22 | 17 | 21.35 | 5.89 |
| 2003 | 19 | 22.84 | 7.03 | 16 | 22.75 | 7.90 |
| 2002 | 17 | 24.35 | 7.08 | 15 | 24.33 | 6.22 |
| Hispanic-Black Gap | | | | | | |
| 2009 | 22 | 1.79 | 8.72 | 19 | 5.63 | 6.32 |
| 2007 | 19 | 4.12 | 7.24 | 18 | 2.72 | 6.51 |
| 2005 | 17 | 4.00 | 7.61 | 13 | 5.00 | 7.15 |
| 2003 | 17 | 2.68 | 7.62 | 13 | 2.21 | 8.93 |
| 2002 | 14 | 2.14 | 7.09 | 12 | 0.17 | 5.59 |

In eighth grade, the gap trends vary. The WH gap trends downward over time, whereas the WB and BH gap increases in the period 2002-2005 followed by declines in the period 2005-2007 and then another increase for the HB gap during the period 2007 to 2009.

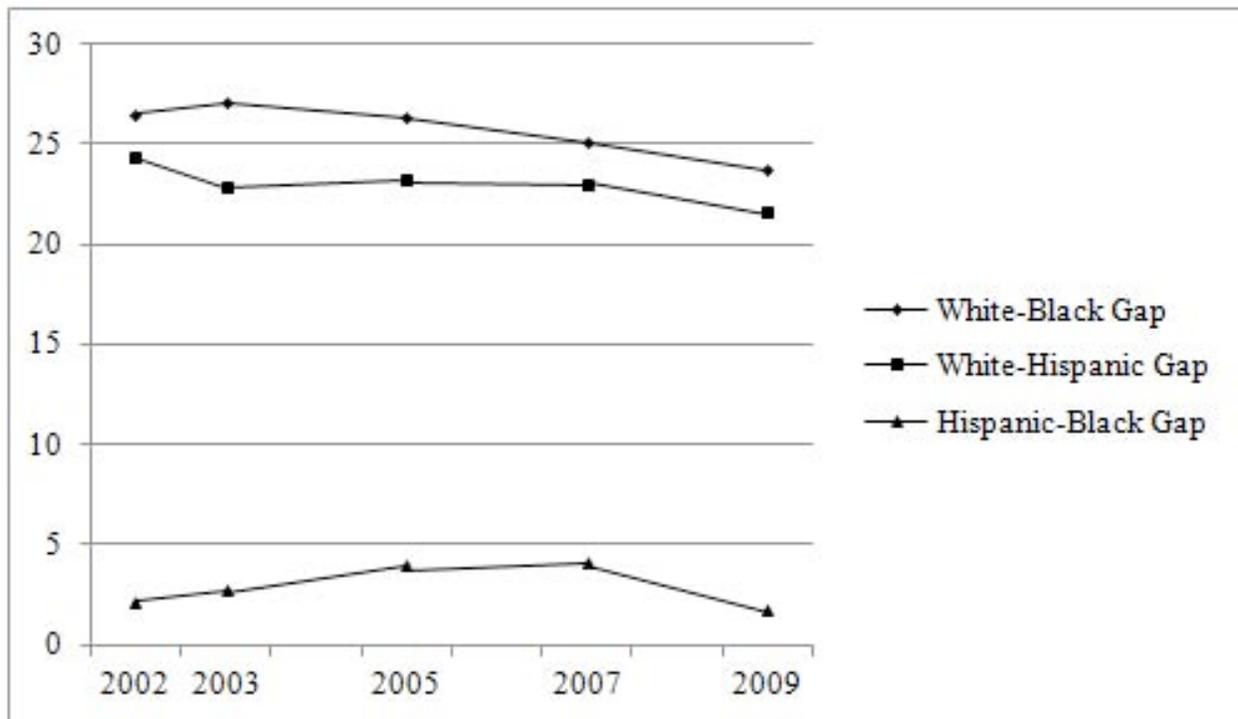


Figure 7. Fourth grade reading average NAEP scale score differences: 2002-2009

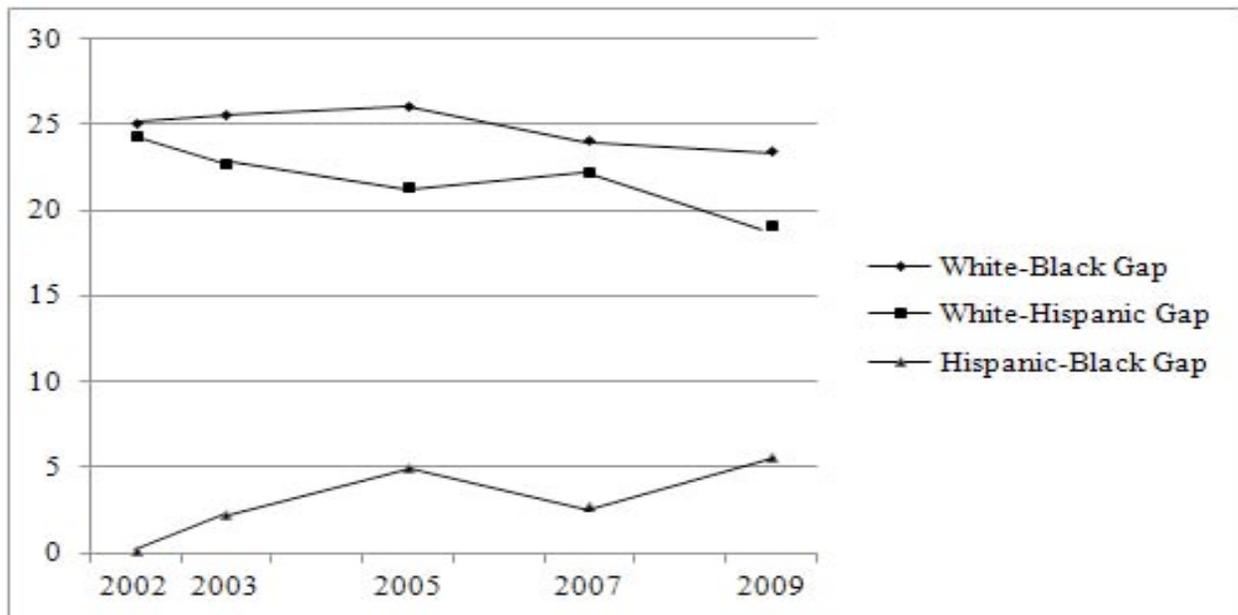


Figure 8. Eighth grade reading average NAEP scale score differences: 2002-2009

Part II: Correlation Analysis

In this section, bivariate, part, and partial correlation coefficients are reported in an examination of the relationships between our APR indicator and NAEP achievement. We began by running correlations to see what state level demographic variables are associated with our APR. Extant research has consistently shown that state poverty rates and racial composition of students are associated with state accountability practices in that poorer states and those with greater

numbers of students of color tend to adopt more punitive accountability policies than states with largely white and more affluent student populations (e.g. Carnoy & Loeb, 2002; Nichols *et al.*, 2006). We wanted to see whether APR shared variance with these and/or other state demographic variables and to see how these relationships may change over time. Because our APR was constructed in 2004, we were also interested to see if correlations with state poverty and other characteristics remained stable over time. As shown in Table 6, APR is consistently associated with percent poverty levels as well as percent of students who are African American in the state. APR is not associated with the percent of Hispanic students or state-level revenue/expenditure patterns.

Table 6
Correlations of APR and State Level Variables over Time

| | % Poverty in state | Expenditures per student | Revenues per student | % Black students | % Hispanic students |
|------|--------------------------|-----------------------------|----------------------------|---------------------|---------------------------|
| 2001 | .399* | -.085 | -.210 | .426* | .256 |
| 2002 | .422* | -.139 | -.282 | .423* | .258 |
| 2003 | .461* | -.168 | -.285 | .423* | .260 |
| 2004 | .452* | -.227 | -.279 | .423* | .253 |
| 2005 | .441* | -.260 | -.320 | .423* | .265 |
| 2006 | .427* | -.272 | -.309 | .478* | .275 |
| 2007 | .419* | -.326 | -.368 | .466* | .281 |
| 2008 | .391 | + | + | + | + |

Note: * = $p < .05$, $n = 25$, + = data not available at the time of analysis

Fourth and eighth grade math.

In Table 7, we correlate APR and fourth and eighth grade math NAEP disaggregated by student SES. There is no significant correlation between APR and student performance when divided into high and low socioeconomic status for either grade level.

Table 7
Correlations APR and math NAEP disaggregated by student SES

| | Low SES | | High SES | | |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-------|
| | 4 th Grade | 8 th Grade | 4 th Grade | 8 th Grade | |
| 2003 | 0.021 | -0.207 | 2003 | 0.167 | 0.015 |
| 2005 | 0.021 | 0.023 | 2005 | 0.275 | 0.197 |
| 2007 | 0.06 | 0.023 | 2007 | 0.13 | 0.125 |
| 2009 | -0.091 | -0.051 | 2009 | -0.04 | 0.105 |

Note: * = $p < .05$, $n = 25$ all cells

In the next analysis (and a similar one conducted for reading data, Tables 17-20), we used three slightly different approaches to examine the relationship between APR and math NAEP disaggregated by student ethnicity. In the first column of results in Table 8 (Column 1), we report bivariate correlations between APR and NAEP scale scores. Next, we remove the effects of state poverty from APR using regression techniques to generate standardized residuals of APR from state poverty data as of 2004 and use those residuals in correlation with NAEP scale scores. Lastly, we use partial correlation techniques to look at the relationships of APR residuals and NAEP while at the same time partialing out the effects of exclusion rates.

As displayed in Table 8, correlations of APR and NAEP suggest that the correlation of accountability pressure and math achievement ranges from a low of .038 (African American, 2009, 4th grade) to a high of .463 (Hispanic, 2005, 8th grade). For some groups in some years, the relationship is barely evident, whereas for other groups in other years, there seems to be a relatively strong correlation. A closer look at the pattern of correlations shows that the within group strength of the APR-NAEP relationship for Hispanic students increases 2003-2005/2007 followed by a significant decrease in 2009 in both fourth and eighth grade (the same pattern is true for African American students in eighth grade). However, when state poverty is removed from APR, the pattern of results slightly changes. Although correlations still seem to decrease overtime, absolute values across the board are greater when state poverty is removed. An examination of Column 2 suggests that pressure and NAEP are connected more strongly for 4th and 8th grade White and Black students but less connected for Hispanic students. This pattern changes little when exclusion rates are removed (Column 3).

Table 8

Bivariate, Part, and Partial Correlations: APR and Math NAEP Disaggregated by Student Ethnicity

| | | Column 1 | Column 2 | Column 3 |
|------------------|------|-----------------|------------------------------|--|
| | | APR and NAEP | APR Residuals and NAEP | APR Residuals and NAEP partial Exclusion Rates |
| | | n/df | r | Partial r |
| Fourth Grade | | | | |
| African American | 2003 | 22/19 | .282 | .461* |
| | 2005 | 22/19 | .390 | .534* |
| | 2007 | 23/20 | .191 | .451* |
| | 2009 | 24/21 | .038 | .308 |
| Hispanic | 2003 | 19/16 | .291 | .367 |
| | 2005 | 19/16 | .362 | .375 |
| | 2007 | 22/19 | .399 | .443* |
| | 2009 | 22/19 | .162 | .237 |
| White | 2003 | 25/22 | .308 | .545** |
| | 2005 | 25/22 | .297 | .560** |
| | 2007 | 25/22 | .240 | .543** |
| | 2009 | 25/22 | .120 | .431* |
| Eighth Grade | | | | |
| African American | 2003 | 21/18 | .290 | .482* |
| | 2005 | 21/18 | .358 | .501* |
| | 2007 | 21/18 | .410 | .542* |
| | 2009 | 23/20 | .152 | .411 |
| Hispanic | 2003 | 16/13 | .259 | .385 |
| | 2005 | 17/14 | .463 | .468 |
| | 2007 | 20/17 | .312 | .441 |
| | 2009 | 21/18 | .063 | .105 |
| White | 2003 | 25/22 | .254 | .489* |
| | 2005 | 25/22 | .294 | .536** |
| | 2007 | 25/22 | .225 | .486* |
| | 2009 | 25/22 | .211 | .486* |

Note: *= $p < .05$, **= $p < .01$ (two-tailed); n=state sample size for bivariate correlations; df=degrees of freedom for partial correlation.

Fourth and eighth grade math: Gain and Cohort Analysis.

We were particularly interested in the relationship between APR and NAEP gains across 2003-2007 (Table 9) and 2005-2009 (Table 10). These correlations reflect the degree to which test-related pressures in various states, as measured in 2004, are related to subsequent NAEP gains. To address regression effects associated with gain score analyses, we used linear regression techniques to generate standardized residuals of NAEP 2007 (2009) from NAEP 2003 (2005) as the estimation of NAEP gain for this and all subsequent analyses involving NAEP standard score gains over time. As displayed in Tables 9 and 10, we estimated the relationship between our APR indicator and NAEP gain in two ways. First, we correlated APR and NAEP gain measured as a regression residual. Second, we again accounted for the shared variance of APR and state poverty by using APR residuals as the correlate with NAEP gain residuals. Lastly, we used partial correlation techniques to examine these relationships while partialing out the exclusion rates.

Table 9

Correlations of APR and 2003-2007 Math NAEP Gain Scores Disaggregated by Student Ethnicity

| | APR and NAEP Gain Residuals | | APR Residuals and NAEP Gain Residuals | |
|------------------|-----------------------------|------------|---------------------------------------|------------|
| | Fourth Grade (2003-2007) | | | |
| | r | Partial r | r | Partial r |
| All Students | -.057 (25) | -.145 (22) | .040 (25) | -.065 (22) |
| African American | -.081 (22) | -.139 (19) | .148 (22) | .059 (19) |
| Hispanic | .200 (19) | .031 (16) | .239 (19) | .086 (16) |
| White | -.089 (25) | -.219 (22) | .116 (25) | -.018 (22) |
| | Eighth Grade (2003-2007) | | | |
| All Students | .236 (25) | .280 (22) | .247 (25) | .201 (22) |
| African American | .302 (21) | .344 (18) | .256 (21) | .238 (18) |
| Hispanic | .348 (16) | .177 (13) | .375 (16) | .071 (13) |
| White | -.009 (25) | -.042 (22) | .105 (25) | .014 (22) |

Note: Partial correlation between APR indicator and NAEP gain indicator, holding 2007 exclusion rates constant (for fourth and eighth grade respectively)

When it comes to relationships between our APR and 2003-2007 NAEP gains, bivariate correlations suggest moderate to low connections for all student groups in fourth grade but stronger connections for African American and Hispanic eighth graders. When poverty is removed, this pattern changes little; however the connection between high stakes testing pressure and both fourth and eighth grade gains diminishes significantly once exclusion rates are partialled out of the question (last column, Table 9). With the exception of African American eighth grade achievement, the relationship between high-stakes testing pressure (APR) and NAEP gains in math is relatively absent in both fourth and eighth grades. When it comes to 2005-2009 NAEP gains (Table 10), the pattern reverses significantly.

As shown in Table 10, our APR is negatively associated with NAEP gains for virtually all student groups in both fourth and eighth grade. Greater pressure in 2004 is associated with decreasing NAEP gains in both fourth and eighth grade math for the 2005-2009 years.

We wanted to see if high-stakes testing pressure was related to changes in achievement among cohorts of students (i.e., “cohort” analyses follow the achievement trends of students as

they progress from fourth to eighth grade⁷). For these, and all subsequent cohort analyses, cohort NAEP gains are calculated as: [eighth-grade achievement year i] – [fourth-grade achievement year ($i - 4$)]. To account for regression effects in gain analysis, we generate standardized residuals of 2009 (2007) eighth grade math achievement from 2005 (2003) fourth grade math achievement. Using these residuals, we correlate with APR and with APR residuals (removing state poverty). Results displayed in Table 11 suggest that when it comes to the 2003-2007 cohort, pressure is related to achievement for African American students, but not for White or Hispanic students. When it comes to the 2005-2009 cohort, however, all relationships disappear and for African Americans, the relationship inverts such that greater pressure is associated with declines in cohort NAEP performance.

Table 10

Correlations of APR and 2005-2009 Math NAEP Gain Scores Disaggregated by Student Ethnicity in Fourth Grade

| | Fourth Grade (2005-2009) | | | |
|------------------|-----------------------------|--------------|---------------------------------------|-------------|
| | APR and NAEP Gain Residuals | | APR Residuals and NAEP Gain Residuals | |
| | r | Partial r | r | Partial r |
| All Students | -.485* (25) | -.528** (22) | -.329 (25) | -.421* (22) |
| African American | -.284 (22) | -.270 (19) | -.029 (22) | -.058 (19) |
| Hispanic | -.120 (19) | -.176 (16) | -.021 (19) | -.122 (16) |
| White | -.319 (22) | -.348 (22) | -.152 (25) | -.226 (22) |
| | Eighth Grade (2005-2009) | | | |
| All Students | -.303 (25) | -.362 (22) | -.142 (25) | -.240 (22) |
| African American | -.078 (21) | -.077 (18) | .109 (21) | .074 (18) |
| Hispanic | -.103 (17) | -.207 (14) | .070 (17) | .071 (14) |
| White | -.190 (25) | -.252 (22) | -.052 (25) | -.155 (22) |

Note: Partial correlation between APR and NAEP gain, holding 2009 exclusion rates constant (for fourth and eighth grade respectively).

⁷ In this analysis, a “cohort” is not a true cohort in the sense that we follow the same students from fourth to eighth grade. Rather, it is a proxy of a true cohort—following the achievement trends of two different random samples of students as they progress through the intermediary grades from fourth to eighth grade.

Table 11
Math Cohort Analysis: Correlations and Partial Correlations

| | Cohort 2003-2007 | | | |
|------------------|--|------------|--|------------|
| | APR and Cohort Standardized Residual Gain Scores | | APR Residuals and Cohort Standardized Residual Gain Scores | |
| | r | Partial r | r | Partial r |
| All Students | -.027 (25) | -.096 (22) | .022 (25) | -.164 (22) |
| African American | .223 (21) | .278 (18) | .270 (21) | .252 (18) |
| Hispanic | .212 (19) | -.020 (16) | .243 (19) | -.109 (16) |
| White | -.066 (25) | -.182 (22) | .045 (25) | -.151 (22) |
| | Cohort 2005-2009 | | | |
| All Students | -.242 (25) | -.364 (22) | -.117 (25) | -.303 (22) |
| African American | -.228 (22) | -.245 (19) | .010 (22) | -.040 (19) |
| Hispanic | -.051 (19) | -.141 (16) | .014 (19) | -.119 (16) |
| White | -.150 (25) | -.246 (22) | -.066 (25) | -.228 (22) |

Note: Partial correlations represent the association of APR and NAEP cohort gain indicator while holding 2007 (2009) eighth grade math exclusion rates constant.

Fourth and eighth grade reading

Correlations of APR and NAEP for fourth grade reading and disaggregated by SES are displayed in Table 12. APR and reading achievement are most strongly and negatively related for low-income eighth graders (2003, 2005, 2009).

Table 12
Fourth and Eighth Grade Reading NAEP Disaggregated by SES

| | | 4 th | 8 th |
|----------|------|-----------------|-----------------|
| | | Grade | Grade |
| Low SES | 2002 | -.176 | .290 |
| | 2003 | -.282 | -.370 |
| | 2005 | -.279 | -.336 |
| | 2007 | -.234 | -.245 |
| | 2009 | -.214 | .386 |
| High SES | 2002 | -.026 | .094 |
| | 2003 | -.048 | -.128 |
| | 2005 | .015 | -.077 |
| | 2007 | -.111 | .030 |
| | 2009 | -.122 | .125 |

Note: *= $p < .05$, **= $p < .01$ (two-tailed): $n=25$ all cells

Table 13

Correlations: Fourth and Eighth Grade Reading Disaggregated by Student Ethnicity, 2003-2009

| | | n/df | APR and NAEP | APR Residuals and NAEP | APR Residuals and NAEP partial Exclusion Rates |
|------------------|------|-------|--------------|------------------------|--|
| | | | r | r | Partial r |
| Fourth Grade | | | | | |
| African American | 2003 | 22/19 | -.169 | .108 | .075 |
| | 2005 | 22/09 | -.028 | .209 | .185 |
| | 2007 | 22/19 | -.017 | .280 | .254 |
| | 2009 | 24/21 | .091 | .357 | .364 |
| Hispanic | 2003 | 19/16 | -.075 | -.013 | -.130 |
| | 2005 | 19/16 | .098 | .178 | -.014 |
| | 2007 | 21/18 | .190 | .294 | .191 |
| | 2009 | 23/20 | -.177 | -.091 | -.108 |
| White | 2003 | 25/22 | .140 | .405* | .385 |
| | 2005 | 25/22 | .098 | .406* | .384 |
| | 2007 | 25/22 | .005 | .317 | .279 |
| | 2009 | 25/22 | .003 | .295 | .293 |
| Eighth Grade | | | | | |
| African American | 2003 | 16/13 | -.059 | .035 | -.132 |
| | 2005 | 17/14 | .004 | .070 | -.065 |
| | 2007 | 20/17 | .242 | .359 | .105 |
| | 2009 | 21/18 | .324 | .357 | .356 |
| Hispanic | 2003 | 21/18 | .170 | .356 | .378 |
| | 2005 | 20/17 | -.120 | .120 | .128 |
| | 2007 | 22/19 | -.077 | .202 | .147 |
| | 2009 | 23/20 | .217 | .220 | .220 |
| White | 2003 | 25/22 | .143 | .313 | .298 |
| | 2005 | 25/22 | .030 | .279 | .272 |
| | 2007 | 25/22 | .149 | .419* | .384 |
| | 2009 | 25/22 | -.021 | -.025 | -.028 |

Note: *= $p < .05$, **= $p < .01$ (two-tailed).

Similar to the math analysis, we looked at APR and reading achievement disaggregated by student ethnicity in three ways. As displayed in Table 13, bivariate correlations of APR and NAEP reading show no relationship in fourth or eighth grade. However, when poverty is removed from APR, the pattern of associations shifts such that APR and reading achievement in fourth (2003, 2005) and eighth grade (2007) are positively linked for White students. When exclusion rates are partialled out of the relationships, many of the correlations subsequently diminish except for White students in fourth and eighth (2003-2007) and African American students in fourth grade (2009), and Hispanic students in eighth grade (2003).

Fourth and eighth grade reading: Gain and Cohort Analysis.

As shown in Tables 14 and 15, APR and NAEP gains in reading are positive for Hispanic students in fourth (2003-2007) and eighth (both 2003-2007 and 2005-2009) grades. By contrast APR and NAEP gains for Hispanic students in fourth grade 2005-2009 are negative.

Table 14
Correlations and Partial Correlations of APR and 2003-2007 Reading NAEP GAIN Scores Disaggregated by Student Ethnicity

| | Fourth Grade (2003-2007) | | | |
|------------------|-----------------------------|------------|---------------------------------------|------------|
| | APR and NAEP Gain Residuals | | APR Residuals and NAEP Gain Residuals | |
| | r | Partial r | r | Partial r |
| All Students | -.045 (25) | -.152 (22) | .086 (25) | -.014 (22) |
| African American | .207 (22) | .180 (19) | .332 (22) | .302 (19) |
| Hispanic | .330 (19) | .039 (16) | .454* (19) | .294 (19) |
| White | -.184 (25) | -.303 (25) | -.018 (25) | -.126 (22) |
| | Eighth Grade (2003-2007) | | | |
| All Students | .124 (25) | .068 (22) | .318 (25) | .269 (22) |
| African American | -.142 (21) | -.234 (18) | .000 (21) | -.123 (18) |
| Hispanic | .451* (16) | .246 (13) | .577* (16) | .430 (13) |
| White | .049 (25) | .026 (22) | .303 (25) | .282 (22) |

Note: Partial correlations represent the association of APR and NAEP cohort gain indicator while holding 2007 eighth grade reading exclusion rates constant. *= $p < .05$, **= $p < .01$ (two-tailed).

Table 15
Correlations and Partial Correlations of APR and 2005-2009 Reading NAEP GAIN Scores Disaggregated by Student Ethnicity

| | Fourth Grade (2005-2009) | | | |
|------------------|-----------------------------|------------|---------------------------------------|------------|
| | APR and NAEP Gain Residuals | | APR Residuals and NAEP Gain Residuals | |
| | r | Partial r | r | Partial r |
| All Students | -.190 (25) | -.199 (22) | -.167 (25) | -.196 (22) |
| African American | .068 (22) | .073 (19) | .270 (24) | .271 (19) |
| Hispanic | -.228 (19) | -.325 (16) | -.056 (19) | -.157 (16) |
| White | -.176 (25) | -.183 (22) | -.133 (25) | -.155 (22) |
| | Eighth Grade (2005-2009) | | | |
| All Students | -.304 (25) | -.315 (22) | -.100 (25) | -.112 (22) |
| African American | .294 (18) | .288 (15) | .268 (18) | .265 (15) |
| Hispanic | .381 (13) | .305 (10) | .352 (18) | .250 (10) |
| White | -.022 (25) | -.019 (22) | -.032 (25) | -.035 (22) |

*= $p < .05$, **= $p < .01$ (two-tailed).

Correlation analysis of NAEP cohort gains in reading and APR suggest there are few meaningful relationships for the 2003-2007 time span; however, APR and Cohort gains among African American and Hispanic students in reading are slightly stronger for the 2005-2009 time frame (Table 16).

Table 16
Correlations and Partial Correlations—Cohort Analysis for Reading

| Cohort 2003-2007 | | | | |
|------------------|--|---------------|--|---------------|
| | APR and Cohort Standardized Residual Gain Scores | | APR Residuals and Cohort Standardized Residual Gain Scores | |
| | r(n) | Partial r(df) | r(n) | Partial r(df) |
| All Students | .036 (25) | -.066 (22) | .105 (25) | -.009 (22) |
| African American | .098 (22) | .040 (19) | .194 (22) | .125 (19) |
| Hispanic | .342 (19) | -.101 (16) | .465* (19) | .155 (16) |
| White | .058 (25) | -.029 (22) | .142 (25) | .047 (22) |
| Cohort 2005-2009 | | | | |
| All Students | -.394 (25) | -.418 (22) | -.311 (25) | -.344 (22) |
| African American | .318 (20) | .314 (17) | .285 (20) | .282 (17) |
| Hispanic | .356 (15) | .287 (17) | .360 (15) | .287 (12) |
| White | -.015 (25) | -.012 (22) | -.001 (25) | -.004 (22) |

Note: Partial correlations represent the association of APR and NAEP cohort gain indicator while holding 2007 (2009) eighth grade reading exclusion rates constant. *= $p < .05$, **= $p < .01$ (two-tailed).

Discussion

In this study, we used correlational techniques to look at the relationship of high-stakes testing pressure and student achievement in 25 states. Using our empirically derived measure of state-level high-stakes testing pressure, the Accountability Pressure Rating (APR) developed in an earlier study (Nichols *et al.*, 2006), we looked at the ways in which state level pressure was associated with state level achievement as measured by NAEP in reading and math in fourth and eighth grades since the inception of NCLB. The data tell a very familiar story.

Descriptives

Math and reading NAEP data reveal a few interesting patterns. In math, pre-NCLB achievement gains were greater than post-NCLB gains. Thus, students were progressing in math at a much faster rate before the national high-stakes testing movement spawned by NCLB. By comparison, fourth and eighth grade reading achievement remained relatively stable over time, with the exception of small increases for fourth graders (2005-2007) and small decreases for eighth graders (2003-2005) after NCLB. When it comes to NAEP achievement from 2002 to 2009, the institution of the NCLB was followed by varied achievement patterns in fourth and eighth grade math.

When disaggregated by ethnicity and SES, White students consistently outscored African American and Hispanic students and richer students consistently outperformed poorer students. Based on these descriptive data, it appears as if the achievement gaps are narrowing, although very slowly. Elsewhere, Braun *et al.* (2010), using more sophisticated analytic techniques in an attempt to isolate the effects of NCLB on these trends in 10 states, conclude the following regarding the achievement gap problem,

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 and White students (pp. 41-42).

Our data here cannot explain the nature of these gap trends over time; however, our analysis seems to reiterate the point that achievement gaps have insignificantly changed as a result of the policies emanating from NCLB.

Correlation Analysis

Our correlation analysis revealed several notable patterns. For example, our data suggest that test related pressure is significantly and positively correlated with state poverty index (percent poverty in state). That is, states with greater number of individuals living in poverty also tended to employ test-related practices that exerted greater amounts of pressure. The nation's poorest children, and the teachers who teach them, tend to feel more pressure when it comes to high-stakes tests than their more privileged contemporaries. When disaggregated by SES and race, data suggest that the relationship between APR and NAEP performance is mixed. In terms of SES, high-stakes testing pressure has no connection to NAEP performance in math. By contrast, APR is more strongly and negatively connected with NAEP performance in reading, especially for low-income students. Thus, high-stakes testing pressure seems to have no measurable connection to NAEP math performance for either rich or poor students, but pressure is deleterious for poor students' NAEP reading performance.

Our data also show that APR is positively correlated with fourth and eighth grade math performance among all groups of students at different points in time (Table 8). Notably, when exclusion rates are removed from the relationship, the APR-NAEP connection diminishes for Hispanic students, raising the question of the role exclusion rate practices play in facilitating the connection between pressure and test performance. By contrast, when it comes to reading, the only substantive connection between pressure and NAEP performance emerges for White students in both fourth and eighth grades (and in particular, in the earlier years of implementation, 2003 and 2005, see Table 13). Overall, these correlations suggest that test-related pressure connects more strongly with increases in math performance than in reading (in both 4th and 8th grades), a pattern that seems more prevalent for White students than for African American or Hispanic students.

We looked at the relationship between APR and NAEP gain scores across two time periods (2003-2007 and 2005-2009) for math and reading. Starting with math, an interesting pattern emerged. When it comes to the 2003-2007 gain years, pressure emerged as a more positive correlate with student math gains—especially among eighth graders (Table 9). By contrast, when it came to the later 2005-2009 span, these correlations virtually all transformed into negative relationships. Thus, as time goes by, it seemed as if earlier levels of pressure in state policy enactments led to later decreases in math gain achievement. For reading, a mixed picture emerges. Pressure is positively related to Hispanic student performance in both fourth and eighth grade and for both time spans examined. However, the relationships are more mixed for other groups: APR is weakly connected to White student reading performance across both time spans and inconsistently related to African American performance (i.e., sometimes positive, sometimes negative, sometimes strongly, sometimes faintly; see Tables 14 and 15).

In terms of cohort achievements in math, APR is positively related to African American cohort achievement 2003-2007 and negatively related to African American cohort achievement 2005-2009. Overtime, pressure has diminishing returns for African American students. When it comes to reading performance, APR has a positive connection with Hispanic reading performance in both 2003-2007 and 2005-2009.

From these results, it is very difficult to come to any simple conclusion regarding the relationship of pressure and student achievement. In some cases there are positive connections,

whereas in other cases, there are negative connections. In our earlier study, we rank ordered all of our correlation coefficients to try to ascertain a meaningful pattern. In that study, our pattern of correlations revealed that the strongest positive associations between APR and NAEP gain was in fourth grade math (Nichols et al., 2006). In Table 17, we rank order by absolute value all correlations emerging from analyses where student achievement was disaggregated by ethnicity. These 48 correlations reveal an interesting pattern. Among the first 24 (or half) of these correlations, 19 come from math and 5 from reading. By contrast, among the bottom 24 correlations, there are 19 in reading and 5 in math. Positive relationships between pressure and NAEP performance exist primarily in math across all subgroups. In contrast to our previous work in which we found significant relationships between APR and NAEP in fourth grade math only, these correlations seem evenly spread among fourth and eighth grade performance.

Table 17

Rank ordering of APR-NAEP correlations disaggregated by student ethnicity

| Student Subgroup | APR/NAEP Correlation | Grade | Subject | Year |
|-------------------------|-----------------------------|--------------|----------------|-------------|
| Hispanic | .463 | 8 | Math | 2005 |
| African American | .410 | 8 | Math | 2007 |
| Hispanic | .399 | 4 | Math | 2007 |
| African American | .390 | 4 | Math | 2005 |
| Hispanic | .362 | 4 | Math | 2005 |
| African American | .358 | 8 | Math | 2005 |
| African American | .324 | 8 | Reading | 2009 |
| Hispanic | .312 | 8 | Math | 2007 |
| White | .308 | 4 | Math | 2003 |
| White | .297 | 4 | Math | 2005 |
| White | .294 | 8 | Math | 2005 |
| Hispanic | .291 | 4 | Math | 2003 |
| African American | .290 | 8 | Math | 2003 |
| African American | .282 | 4 | Math | 2003 |
| Hispanic | .259 | 8 | Math | 2003 |
| White | .254 | 8 | Math | 2003 |
| African American | .242 | 8 | Reading | 2007 |
| White | .240 | 4 | Math | 2007 |
| Hispanic | .217 | 8 | Reading | 2009 |
| White | .211 | 8 | Math | 2009 |
| African American | .191 | 4 | Math | 2007 |
| Hispanic | .190 | 4 | Reading | 2007 |
| Hispanic | .170 | 8 | Reading | 2003 |
| Hispanic | .162 | 4 | Math | 2009 |

Table 17 (continued)

Rank ordering of APR-NAEP correlations disaggregated by student ethnicity

| Student Subgroup | APR/NAEP | | | |
|------------------|-------------|-------|---------|------|
| | Correlation | Grade | Subject | Year |
| African American | .152 | 8 | Math | 2009 |
| White | .149 | 8 | Reading | 2007 |
| White | .143 | 8 | Reading | 2003 |
| White | .140 | 4 | Reading | 2003 |
| White | .120 | 4 | Math | 2009 |
| Hispanic | .098 | 4 | Reading | 2005 |
| White | .098 | 4 | Reading | 2005 |
| African American | .091 | 4 | Reading | 2009 |
| Hispanic | .063 | 8 | Math | 2009 |
| African American | .038 | 4 | Math | 2009 |
| White | .030 | 8 | Reading | 2005 |
| White | .005 | 4 | Reading | 2007 |
| African American | .004 | 8 | Reading | 2005 |
| White | .003 | 4 | Reading | 2009 |
| African American | -.017 | 4 | Reading | 2007 |
| White | -.021 | 8 | Reading | 2009 |
| African American | -.028 | 4 | Reading | 2005 |
| African American | -.059 | 8 | Reading | 2003 |
| Hispanic | -.075 | 4 | Reading | 2003 |
| Hispanic | -.077 | 8 | Reading | 2007 |
| Hispanic | -.120 | 8 | Reading | 2005 |
| African American | -.169 | 4 | Reading | 2003 |
| Hispanic | -.177 | 4 | Reading | 2009 |

Implications

The research on the impact of accountability-based policies and student achievement is varied, limited, and relatively inconclusive. One explanation for this state of affairs is that it is very difficult to isolate cause-effect relationships between complex policy implementation and subsequent student achievement. Still, our data here and elsewhere, as well as work by others reiterate a familiar story: Increased testing pressure is related to increases in achievement in math more consistently than in reading. Differences in the nature of the mathematics and reading curriculum, and /or differences in the ways one can prepare for assessments in these two areas may have something to do with the fact that state level pressure to perform well on high-stakes tests is more strongly and positively related to math achievement and negatively related to reading achievement.

Although our overall correlations reveal that pressure is more connected with math achievement than with reading, our gain and cohort analyses tell a slightly different story. When it comes to math, pressure has no relationship to NAEP changes over time (for either cohorts of students or cross sectional groups of students). By contrast, pressure is positively associated with some student group gain scores in reading. This reversal is perplexing and difficult to interpret. If our APR holds up over time (which is questionable, see the limitations section next), then these data suggest that pressure has diminishing returns for math achievement over time, but slightly positive returns for reading achievement over time (but only for some students groups). Some of these

trends may be explained by the fact that correlations in both math and reading diminish when exclusion rates are partialled out—schools may be excluding lower scoring students at greater rates in later years.

We still contend from these data, as we did in our earlier study (Nichols *et al.*, 2006), that the overall pattern of correlations (math more strongly connected to pressure than reading), points to the likelihood that under pressure, teachers grow more efficient at training students for the test. The math curriculum (versus the reading/language arts curriculum) is structured in such a way as to make it much more amenable for teachers to teach to the test. However, as our data suggest, as time passes, pressure seems to play more of a role in increasing reading scores raising the question of how increasing pressure translates into practice in reading classrooms. What are teachers and students doing differently when it comes to preparing for reading assessments as a result of this increasing pressure? Although more difficult, it is possible that as time progresses, teachers become more skilled at deconstructing the reading curriculum to help students prepare for test questions. Of course, as NCLB persists, it becomes increasingly important but more difficult to disentangle the effects of pressure on student ability to take tests from pressure that genuinely affects student learning. This pattern of results is too varied to make any definitive claims regarding how test-based practices in the classroom may connect with these achievement outcomes.

Limitations of Study

There are a few limitations to this study. First of all, correlation data reveal nothing about the causal nature of relationships. Therefore, although we detect certain consistent patterns in the relationship of APR and NAEP, we cannot make claims regarding casual direction. Further, we also recognize that as time passes, it becomes more difficult to ascertain the meaningfulness of correlations between our APR derived in 2004 and subsequent NAEP data. We acknowledge that as time passes, states' test-based accountability practices have likely changed such that the validity of our 2004-derived index may be less relevant. Since we have no measure of accountability practices beyond 2004, we have no way of knowing the relative accuracy of APR for capturing these changes over time. Still, we believe that since there have been no new federal laws that have mandated sweeping accountability changes until 2010 when the Race to the Top was first signed into law, we believe that states' test-based policies as they were in 2004 have likely changed very little over the course of our study years (2005-2009). Although we have no reason to think states have changed so dramatically that their pressure-based ranking may have altered in any significant way (e.g., Guisbond, Neill, & Schaeffer, 2012), the question left unexamined is to what degree states' policies under NCLB have evened out over time

Future Directions

In light of the fact that high-stakes testing mandates are not going away anytime soon, it seems important that more research is done to understand how the pressures of testing influence classroom-based teaching practices in different curriculum areas. The areas most affected by cultural practices at home, reading, for example, may require a different policy approach than that required by school subjects more related to in-school learning, such as mathematics or science. It is clear that the policy positions of the Obama administration support test-based accountability and the government is pleased with the pressure those policies exert on the functioning of our schools. Because of that, it seems incumbent upon policy researchers to continue work that sheds light on the ways in which test-based instructional practices affected by accountability pressures impact students' motivation, development, and achievement.

References

- Allington, R. L. & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy*, 6, 397-414. doi:10.1177/0895904892006004003
- Amrein, A. L. & Berliner, D. C. (2002a). *The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP Test results in states with high school graduation exams*. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved from <http://www.asu.edu/educ/eps1/EPRU/documents/EP1-0211-126-EPRU.pdf>
- Amrein, A. L. & Berliner, D. C. (2002b). High-Stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved from <http://epaa.asu.edu/epaa/v10n18/>
- Amrein-Beardsley, A. & Berliner, D. (2003, August). Re-analysis of NAEP Math and Reading scores in states with and without high-stakes tests: Response to Rosenshine. *Education Policy Analysis Archives*, 11(25). Retrieved February 5, 2005 from <http://epaa.asu.edu/epaa/v11n25/>
- Bishop, J. H., Mane, F., Bishop, M., & Moriarty, J. (2001). *The role of end-of-course exams and minimum competency exams in standards-based reforms*. *Brookings Papers on Educational Policy*, 4, 267-345.
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives*, 12(1), 1-40. Retrieved from <http://epaa.asu.edu/epaa/v12n1/>
- 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 from <http://epaa.asu.edu/epaa/v14n8/>.
- Braun, H., Chapman, L., & Vezzu, S. (2010). The Black-White achievement gap revisited. *Education Policy Analysis Archives*, 18(21). Retrieved from <http://epaa.asu.edu/ojs/article/view/772>
- Carnoy, M. & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Clarke, M., Haney, W., & Madaus, G. (2000). *High Stakes Testing and High School Completion*. Boston, MA: Boston College, Lynch School of Education, National Board on Educational Testing and Public Policy.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Boston, MA: Boston College, National Board on Educational Testing and Public Policy. Retrieved from <http://www.bc.edu/research/nbetpp/statements/nbr1.pdf>
- Dee, T. & Jacob, B. (2009, November). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Figlio, D., N. & Ladd, H. F. (2008). School accountability and student achievement. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 166-182). NY: Routledge.
- Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessment*. New York: Peter Lang.
- Grodsky, E. S., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971-2004. *Educational Policy*, 23, 589-614. doi: 10.1177/0395909808320678

- Guisbond, L., Neill, M., & Schaeffer, B. (2012, January). *NCLB's lost decade for educational progress: What can we learn from this policy failure?* Jamaica Plain, MA: Fairtest. Retrieved from, http://fairtest.org/sites/default/files/NCLB_Report_Final_Layout.pdf
- Haertel, E. H. & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement*. The 104th Yearbook of the National Society for the Study of Education (part 2, pp. 1-34). Malden, MA: Blackwell. doi: 10.1111/j.1744-7984.2005.00023.x
- Haney, W., Madaus, G., Abrams, L., Wheelock, A., Miao, J., & Gruia, I. (2004, January). The education pipeline in the United States 1970-2000. National Board on Educational Testing and Public Policy. Chestnut Hill, MA: Boston College.
- Hanushek, E. & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Herman, J. L. & Haertel, E. H. (Eds.) (2005). *Uses and misuses of data for educational accountability and improvement*. The 104th Yearbook of the National Society for the Study of Education (part 2). Malden, MA: Blackwell.
- Heubert, J. P. & Hauser, R. M., (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research*, 80(4), 476-526. doi:10.3102/0034654310383147
- Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99-121.
- Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and practice*, 9(3), 5-14. doi: 10.1111/j.1745-3992.1990.tb00372.x
- Marchant, G. J. & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives*, 13(6). Retrieved from <http://epaa.asu.edu/epaa/v13n6/>.
- McDonnell, L. (2005). Assessment and accountability from the policymaker's perspective. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement*. The 104th Yearbook of the National Society for the Study of Education (part 2, pp. 35-54). Malden, MA: Blackwell.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved July 20, 2009, from <http://epaa.asu.edu/epaa/v14n1/>.
- Orfield, G., Losen, D., Wald, J., & Swanson, C. B. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston, MA: Boston College, National Board on Educational Testing and Public Policy. Retrieved from <http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>
- Raymond, M. E. & Hanushek, E. A. (2003, Summer). High-stakes research. *Education Next*, 3(3), 48-55. Retrieved from <http://www.educationnext.org/>
- Reardon, S. F. (2011). The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations In Richard Murnane & Greg Duncan

- (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. New York: Russell Sage Foundation.
- Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2008). *High stakes, no effects: Effects of failing the California High School Exit Exam* (Working Paper 2008-10). Stanford, CA: Stanford University, Institute for Research on Education Policy & Practice.
- Reardon, S. F., Atteberry, A., Arshan, N., & Kurlaender, M. (2009, April 21). *Effects of the California High School Exit Exam on Student Persistence, Achievement and Graduation* (Working Paper 2009-12). Stanford, CA: Stanford University, Institute for Research on Education Policy & Practice.
- Rosenshine, B. (2003). High-Stakes testing: Another analysis. *Education Policy Analysis Archives*, 11(24). Retrieved from <http://epaa.asu.edu/epaa/v11n24/>
- Ryan, J. E. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review*, 79, 932-989. doi: 10.2139/ssrn.476463
- Shepard, L. A. (1990). Inflated test scores gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Timar, T. B. & Maxwell-Jolly, J. (Eds) (2012). *Narrowing the achievement gap: Perspectives and strategies for challenging times*. Cambridge, Massachusetts: Harvard Education Press.
- Torgerson, W. S. (1960). *Theory and Methods of Scaling*. New York: John Wiley.
- Tyack, D. B. (1974). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.
- U.S. Department of Education (November 2009). *Race to the Top: Executive Summary*. Washington, DC: US Department of Education. Retrieved from <http://ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education (March 2010). *A blueprint for reform. The reauthorization of the Elementary and Secondary School Act*. Washington, DC: US Department of Education. Retrieved from <http://www2.ed.gov/policy/elsec/leg/blueprint/>.
- Winters, M. A., Trivitt, J. R., & Greene, J. P. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review*, 29, 138-146. doi: 10.1016/j.econedurev.2009.07.004

About the Authors

Sharon L. Nichols
University of Texas at San Antonio

Email: sharon.nichols@utsa.edu

Sharon L. Nichols is an Associate Professor of Educational Psychology at the University of Texas at San Antonio. She is the past chair of the Adolescence and Youth Development Special Interest Group of AERA and the current treasurer for Division 15 (Educational Psychology) of APA. She is coauthor of two books including *Collateral damage: How high-stakes testing corrupts America's schools* (with D. C. Berliner, Harvard Education Press, 2007) and *America's teenagers—myths and realities: Media images, schooling and the social costs of careless indifference* (with T. L. Good, Erlbaum, 2004). Her current work focuses on the impact of test-based accountability on teacher practice and student motivation, learning, and development.

Gene V Glass
University of Colorado Boulder

Email: gene@gvglass.net

Gene V Glass is currently a Research Professor in the School of Education at the University of Colorado Boulder and a Regents' Professor Emeritus from Arizona State University. Trained originally in statistics, his interests broadened to include psychotherapy research, evaluation methodology, and policy analysis. His work on meta-analysis of psychotherapy outcomes (with M.L. Smith) was named as one of the Forty Studies that Changed Psychology in the book of the same name by Roger R. Hock (1999). He is the founding editor of *Education Policy Analysis Archives* and *Education Review/Reseñas Educativas*. His Ph.D. was awarded in 1965 by the University of Wisconsin, Madison, in educational psychology with a minor in statistics.

David C. Berliner
Arizona State University

Email: berliner@asu.edu

David C. Berliner is Regents' Professor Emeritus in the Mary Lou Fulton Teachers College at Arizona State University. His interests are in the study of teaching, teacher education, and educational policy. Related to this publication are two of his books: *The Manufactured Crisis* (with B. J. Biddle) and *Collateral Damage* (with S. L. Nichols). His recent publications examine how high-stakes testing has narrowed school curriculum for the poor and works against the development of creativity; the many faults that occur when interpreting PISA tests; the effects of inequality and poverty on school achievement; and the inadequacy of value-added models of teacher evaluation.

education policy analysis archives

Volume 20 Number 20

July 20, 2012

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, Socolar (China).

Please contribute commentaries at <http://epaa.info/wordpress/> and send errata notes to Gustavo E. Fischman fischman@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **David R. Garcia** (Arizona State University), **Stephen Lawton** (Arizona State University)

Rick Mintrop, (University of California, Berkeley) **Jeanne M. Powers** (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi* University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman* Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jaekyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

* Members of the New Scholars Board

archivos analíticos de políticas educativas
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel FLACSO, Argentina

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Economicas – UNAM, México

Maria Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin UNESCO, Francia

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Universidad de Toronto-Ontario Institute of Studies in Education, Canadá

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil
Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
Antônio Teodoro Universidade Lusófona, Portugal
Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto [Fundação Carlos Chagas](#), Brasil
Manuela Terrasêca Universidade do Porto, Portugal
Robert Verhine Universidade Federal da Bahia, Brasil
Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil