Patterns and Trends in Racial Academic Achievement Gaps Among States, 1999-2011

Sean F. Reardon
Rachel A. Valentino
Demetra Kalogrides
Kenneth A. Shores
Erica H. Greenberg

Stanford University

August, 2013

Author Note: Please direct correspondence to Sean F. Reardon (sean.reardon@stanford.edu), Stanford University Graduate School of Education, 520 Galvez Mall, #526, Stanford, CA 94305-3084; Office Phone: 650-736-8517; Office Fax: 650-723-9931. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305D110018 to Stanford University. Research support for Valentino, Shores, and Greenberg was also provided by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B090016 to Stanford University. We particularly thank Ross Santy for his assistance providing data from EdFacts; Ximena Portilla and Natassia Rodriguez for their comments and feedback on earlier versions of this paper; and the following research assistants for their help collecting data and contacting state departments of education: Donna Saadati-Soto, Ricky Vargas, Austin Block, Leah Weiser, Walter Herring, Sarah Quartey, and James Chu. The findings, conclusions, and opinions here are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Patterns and Trends in Racial Academic Achievement Gaps Among States, 1999-2011

Abstract

In this paper we combine National Assessment of Educational Progress and state accountability test data to examine variation among states in achievement gap levels and trends. Although national trends in gaps have been well studied, little research has examined variation in gaps across states or the extent to which differences in state demographics or policies account for these differences. We find that white-black and white-Hispanic achievement gaps are large and have been closing very slowly over the last two decades, but the sizes and rates of change of these gaps vary considerably across states. Much of the variation in gap levels, but none of the variation in gap trends, can be explained by differences across states in the socioeconomic status of minority children and school racial segregation.

Racial differences in average academic achievement are large and persistent. Data from the National Assessment of Educational Progress (NAEP) indicate that black and Hispanic students score, on average, roughly three-quarters of a standard deviation lower than white students in both math and reading – the equivalent of about four years of learning in middle or high school (Reardon, 2011). Although these gaps are substantially smaller than they were 40-50 years ago, they remain stubbornly large (Grissmer, Flanagan, & Williamson, 1998; Hedges & Nowell, 1998; Neal, 2006; Reardon, 2011; Vanneman, Hamilton, Baldwin Anderson, & Rahman, 2009; Rampey, Dion, & Donahue, 2009). Moreover, although white-black and white-Hispanic gaps are similar in magnitude at kindergarten entry, they follow different trajectories as children age. White-black gaps increase during the first six years of schooling in both math and reading, while white-Hispanic gaps decrease during this period (Fryer & Levitt, 2004; 2006; Reardon & Galindo, 2006; Reardon & Robinson, 2008).

NAEP data indicate that racial achievement gaps vary in size substantially across states, ranging from 0.50 standard deviations or less to over 1.25 standard deviations (Vanneman, et al., 2009). Moreover, these gaps are narrowing or widening at different rates across states (Hemphill & Vanneman, 2010). Because of NAEP's modest sample sizes, however, estimates of state-specific rates of change are relatively imprecise. Given this imprecision, some researchers have attempted to use state accountability test data to estimate within-state achievement gap trends, a solution that is appealing because the number of students tested is substantially larger in state tests than in NAEP (e.g., Kober, Chudowsky, & Chudowsky, 2010). Unfortunately, many reports of achievement gaps based on state test data measure achievement gaps as differences in the proportions of students scoring at or above state-specific definitions of "proficiency." Because the size and trends in achievement gaps defined this way are very sensitive to how states define "proficiency," such measures are ill-suited for comparisons of gaps between states or over time (Ho, 2008). As a result, we have relatively little detailed information about cross-state variation in achievement gap patterns and trends.

There are several reasons to expect that achievement gaps and their trends may vary among states. First, there are large racial differences in family socioeconomic conditions, and these differences account for 50 to 85% of white-black and white-Hispanic achievement gaps among young children (Fryer & Levitt, 2006; Duncan & Magnuson, 2005). Black and Hispanic children are approximately three times more likely to live in poverty than white children; black and Hispanic median family incomes are less than two-thirds of white median family income (Economic Policy Institute, 2012); and the percentage of parents of black and Hispanic children who have a Bachelor's degree or higher (21% and 12.8%, respectively) is substantially lower than that of white children (38.5%; Aud, Fox, & Kewal Ramani, 2010). To the extent that these racial socioeconomic differences vary among states, they may lead to between-state differences in achievement gaps and their trends. While we know that racial differences in socioeconomic status explains much of the achievement gap at the individual level, we have no clear evidence as to whether they can account for between-state differences in achievement gaps and their trends.

We also might expect larger racial achievement gaps in states where there are large racial disparities in school quality than in states where there is little or no correlation between a student's race and the quality of the school she attends. Because segregation is a necessary—but not sufficient—condition for racial disparities in access to high-quality schools, measures of school segregation serve as a crude proxy for differential exposure to high-quality schooling. That is, if segregation is high, and school quality varies across states, then student race and socioeconomic status are likely to be correlated with educational resource allocation and, ultimately, school quality (Aud et al., 2010). Further, school desegregation in the late 1960s and 1970s resulted in significant increases in educational attainment and decreases in the dropout rates of blacks, yet had no effect on these outcomes for whites (Johnson, 2011; Guryan, 2004). These findings suggest that states with lower (and declining) levels of school segregation might also demonstrate smaller (and narrowing) achievement gaps.

In analyses that follow, we first describe average within-state white-black and white-Hispanic achievement gap levels and trends, both over time and across grades. To do so, we use test score data from both NAEP and state testing programs. We then describe the between-state variation in achievement gap levels and trends, identifying the states with the largest and smallest gaps and those where gaps are widening or narrowing more and less rapidly. Using state accountability test data allows us to obtain very reliable estimates of achievement gaps and their trends because states test nearly every student in grades 3-8 annually. Finally, we identify the extent to which variation among states is explained by state-level economic and demographic characteristics. Understanding the variation among states in their achievement gap patterns and trends is a critical first step toward learning *why* some states have been able to narrow these gaps more quickly than other states, and what policies might be most effective at ultimately eliminating gaps.

Data

We use data from two sources to estimate state-specific achievement gaps: State NAEP and state accountability tests. NAEP math and reading test score data are available for 4^{th}- and 8^{th}-graders between 1999 and 2011, though the tests were typically administered only every other year. Data from annual state accountability tests, in the form of counts of students scoring in categorical proficiency levels, are available from 1999 through 2011 for grades two through eight, though only a few states test students in second grade. Although not all states reported or administered state accountability tests as far back as 1999, we have state test data for over half the states (26) starting beginning in 2002 and virtually all states beginning in 2006. We do not analyze data from secondary grades, as states vary in the specific content covered in such tests and the ages of students tested, and because often students in the same grade take different tests, depending on what courses they are enrolled in, rendering the computation of achievement gaps complicated or impossible.

We obtained NAEP and state test data from the U.S. Department of Education, state Departments of Education, and the Center for Education Policy (CEP, 2009). For a given year and grade, we have as many as 400 achievement gap estimates (one for each state-subject-source-group combination: 50 states, in math and reading, from NAEP and state tests, between blacks and whites and between Hispanics and whites). The dataset we use here includes 1,231 and 4,504 white-black gap estimates, as well as 1,241 and 4,492 white-Hispanic gap estimates, from NAEP and state tests, respectively. Gaps are measured so that positive gaps indicate that white students' scores are higher, on average, than black or Hispanic students' scores; negative gaps indicate the opposite.

Estimating Achievement Gaps

We measure achievement gaps using the V-statistic (Ho & Reardon, 2012) for each state-grade-year-subject-assessment cell. We use V rather than racial differences in proficiency rates because proficiency differences are very sensitive to differences among states in their definitions of proficiency, while V is not (Ho, 2008; Ho & Reardon, 2012). The V-statistic is similar to Cohen's d (the difference in means, divided by the pooled standard deviation), but relies only on the ordered nature of test scores rather than on an assumption that the scores reflect an interval scale. Indeed, the V statistic is equivalent to Cohen's d when the distributions are both normal, but unlike d, is unaffected by non-linear monotonic distortions of a test metric that leave the ordering of individuals unchanged.

We choose to use V for two reasons. First, we rely on data from many different tests—NAEP and state tests, tests in different states, years, and grades—and scores on these tests may not all have a defensible interval scale. V permits valid comparison of gaps across test scores reported on different scales, assuming they measure the same underlying constructs. Shores, Valentino, & Reardon (2013) show that estimates of achievement gaps from NAEP and state tests display similar within-state trends and magnitudes, on average. Moreover, state-specific gap magnitudes are highly correlated between NAEP and state assessments, and for white-black gaps state-specific gap trends are also highly correlated. This implies that comparisons of gaps across data sources are largely valid. The second reason we use V is that it can be easily and reliably (typically with errors less than 0.01-0.02 standard deviations of the true value; Ho & Reardon, 2012) computed from both student-level data that is available from NAEP and from publicly reported counts of students of who score in each of a small number of ordered categories (e.g. the percent of students of each race scoring “Below Basic”, “Basic”, “Proficient,” and “Advanced”) from state assessments. Because all states report counts of this type, and only some report race-specific means and standard deviations, we can readily estimate V for all states.

One important difference between NAEP test scores and state test scores is that the NAEP scoring procedure implicitly removes some of the measurement error—the item-level measurement error—from the test scores, which has the effect of partially disattenuating achievement gaps. State test scores are not similarly adjusted. To make the NAEP and state test gap estimates more comparable, we disattenuate gaps estimated from state assessment data by dividing them by the square root of the estimated reliability of the test used. We use a reliability estimate of 0.90 for all tests, based on evidence from Reardon and Ho (2013).

State-level characteristics

We consider how much of the between-state variation in achievement gaps can be accounted for by two types of state-level time-varying and time-invariant factors: 1) socioeconomic differences between racial groups (including white-black or white-Hispanic household income ratios, household poverty ratios, unemployment rate ratios, and differences in average parental years of completed schooling), and 2) white-black or white-Hispanic school segregation, understood as a proxy for potential differences in access to school resources and quality, as discussed above. We also control for the proportion of the school-aged population that is black or Hispanic when we include school segregation in the model. We construct measures of these factors using data from two main sources: the Current Population Survey (CPS), which we use to compute measures of the relative economic position of minorities and whites, and the Common Core of Data (CCD), which we use to compute the racial composition and segregation of schools.

Using the CPS, we restrict the data to records pertaining to children ages birth to 14 years old. For each state-year-age-race combination, we compute the average income, poverty status, unemployment status, and parental education level of children's families, using the sampling weights to ensure our variables are representative of the average child's household in each cell. Exposure to unemployment is measured by whether any adult in the child's household is unemployed. Parental education is measured using the highest number of years of schooling completed among adults in the household. We then compute white-black and white-Hispanic ratios (or differences, in the case of education) of these measures within each cell.

We use the CCD data to measure the proportion of public school students that are black or Hispanic, as well as the between-school racial segregation for each state-year-grade combination. We compute segregation using the information theory index (H) (Theil & Finezza, 1971; Reardon & Firebaugh, 2002).

We include three versions of each of the CPS-derived variables and two versions of the CCD-derived variables in our models. First, we include average values of each variable within a state over all the years/cohorts we have available. The state-level versions of the covariates, denoted $\mathbf{X}_s$ in our models below, may explain between-state variation in average levels of achievement gaps, but not their trends. Second, we include state-by-cohort specific measures of the CPS-derived socioeconomic variables. For each student cohort, we compute the average of the value in the years the cohort was between the age of 0 and 5. These variables, denoted $\mathbf{W}_{sc}$, reflect the average socioeconomic inequality experienced by students in a given state and cohort in early childhood; they may explain within-state, between-cohort trends in achievement gaps. Third, we include state-by-cohort-by-grade measures of the cumulative exposure a cohort has had to each of the covariates from kindergarten through that specific grade. These variables, denoted $\mathbf{Z}_{scg}$, indicate the within-cohort time-varying cumulative exposure to socioecononomic inequality or segregation during the school years; they may explain within-state and -cohort changes in achievement gaps across grades.

Methods

We estimate the average levels and trends in achievement gaps, as well as the extent to which these vary among states, using a set of precision-weighted random coefficient models. We fit these models separately for white-black and white-Hispanic achievement gaps, pooling across data sources and subjects for each model. Although there are some modest differences between the math and reading gap patterns, and between the gaps estimated from NAEP and state assessments, we construct the models so that the coefficients of interest describe the averages of the gaps and trends in both math and reading, and from both NAEP and state assessments. As robustness checks, we also estimate the models separately by source (i.e. NAEP and state test data), and by subject (results available upon request). Pooling the data and constructing the models this way enables us to provide parsimonious summaries of the trends in achievement gaps across both subjects and assessments.

Equation (1) below describes the model we use. Here s indexes states, c indexes student cohorts, g indexes grades, t indexes test subjects, and a indexes assessments (NAEP or state assessments). We model the estimated achievement gap $\hat{G}_{scgta}$ as a function of cohort (the year a cohort entered kindergarten, centered at 2002), grade (centered at grade 4), test subject (a binary indicator variable for math or reading, centered at ½), and assessment type (a binary indicator for

NAEP or state, centered at ½). Based on preliminary analysis of the data, we allow average cohort trends to differ between the math and reading test gaps, and we allow the cohort, grade, and test subject coefficients to differ between the state and NAEP tests by including interactions between the assessment type variable and the cohort, grade, test, and test-by-cohort subject variables. Because we center the test subject and assessment type variables at ½ prior to interacting them with the other variables, the intercept is interpreted as the average of the within-state math and reading achievement gaps, estimated from both NAEP and state assessment data, in 4th grade in Spring 2007 (corresponding to the cohort of students who entered kindergarten in 2002). The cohort and grade coefficients in the model are interpreted as the average of the within-state math and reading cohort and grade trends, as estimated from both NAEP and state assessments. In some models we include vectors of state, state-by-cohort, and state-by-cohort-by-grade covariates (described above, and denoted $\mathbf{X}_s$, $\mathbf{W}_{sc}$, and $\mathbf{Z}_{scg}$, respectively). We fit the models using precision-weighted random coefficients models, weighting each gap estimate by the inverse of its sampling variance and allowing the intercepts, cohort trends, math-reading differences, grade slopes, and NAEP-state differences in intercepts and slopes to vary across states. The full model is:

$$
\begin{aligned}
\hat{G}_{scgta} &= \gamma_{0s} + \gamma_{1s} cohort_{csgta} + \gamma_{2s} grade_{csgta} + \gamma_{3s} math_{csgta} + \gamma_4 cohort * math_{scgta} \\
&\quad + \left(\gamma_{5s} + \gamma_{6s} cohort_{csgta} + \gamma_7 grade_{csgta} + \gamma_8 math_{csgta} + \gamma_9 cohort * math_{scgta}\right)(NAEP_a) \\
&\quad + \mathbf{X}_s \mathbf{A} + \mathbf{W}_{sc} \mathbf{B} + \mathbf{Z}_{scg} \mathbf{\Gamma} + e_{scgta} + \varepsilon_{scgta}
\end{aligned}
$$

$$
e_{scgta} \sim N[0, \sigma^2]
$$

$$
\varepsilon_{scgta} \sim N\left[0, \omega^2_{scgta}\right] = N\left[0, var\left(\hat{G}_{scgta}\right)\right]
$$

$$
\begin{bmatrix} \gamma_{0s} \\ \gamma_{1s} \\ \gamma_{2s} \\ \gamma_{3s} \\ \gamma_{5s} \\ \gamma_{6s} \end{bmatrix} \sim N \left[\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_5 \\ \gamma_6 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} & \tau_{02} & \tau_{03} & \tau_{05} & \tau_{06} \\ \tau_{10} & \tau_{11} & \tau_{12} & \tau_{13} & \tau_{15} & \tau_{16} \\ \tau_{20} & \tau_{21} & \tau_{22} & \tau_{23} & \tau_{25} & \tau_{26} \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_{33} & \tau_{35} & \tau_{36} \\ \tau_{50} & \tau_{51} & \tau_{52} & \tau_{53} & \tau_{55} & \tau_{56} \\ \tau_{60} & \tau_{61} & \tau_{62} & \tau_{63} & \tau_{65} & \tau_{66} \end{pmatrix} \right]
$$

(1)

Several coefficients in the model are of particular interest. The coefficient γ_0 indicates the average achievement gap (neither subject- nor test-specific) in grade 4 in 2007. Similarly, γ_1 represents the average achievement gap trend across student cohorts and γ_2 is the average rate at which the gaps change as cohorts progress through school. In addition to these average effects, the variance components τ_{00}, τ_{11}, and τ_{22} indicate the variances of the gap sizes, cohort trends, and grade trends, respectively, among states. In addition to these parameters, the model also yields estimates of the reliabilities with which we can estimate state-specific gap levels and trends, and provides Empirical Bayes (EB) estimates of the state-specific levels and trends.

We begin by fitting the model above without the covariates in order to estimate the unconditional average intercepts and slopes, their variances, and the state-specific Empirical Bayes estimates. We then fit models designed to assess the extent to which the variances of the intercepts are reduced by the inclusion of the state-level covariates $\mathbf{X}_s$, and the extent to which the variances of the cohort and grade slopes are reduced by the state-cohort and state-cohort-grade covariates $\mathbf{W}_{sc}$ and $\mathbf{Z}_{scg}$, respectively. More specifically, we fit three different models including different combinations of covariates: one including only the socioeconomic covariates; one including only the school segregation and racial composition measures; and one including both sets. In each model we include the state, state-by-cohort, and state-by-cohort-by grade versions of the covariates.

Results

Average Achievement Gap Magnitudes and Trends

Figure 1 describes the average trends in within-state achievement gaps. These trends are estimated separately from NAEP and from state assessment data, using a version of Model 1 that includes a non-parametric cohort trend rather than a linear trend. Although the gaps estimated from NAEP are slightly larger than those estimated from state assessments, both data sources indicate that the white-black and white-Hispanic gaps have declined modestly over the last two decades. This provides some justification for pooling the estimates from the two sources in the same model (see also Shores et al., 2013).

[INSERT FIGURE 1 ABOUT HERE]

Table 1 reports the parametric estimates of the average size and trends in achievement gaps, based on the pooled model described above. We first focus attention on the models without covariates (Models B1 for white-black and H1 for white-Hispanic gaps). The base white-black model (B1) indicates that the average achievement gap for the 2002 cohort in grade 4 was 0.78 standard deviations (SD). Across cohorts, these gaps have been narrowing over time, at the very slow rate of 0.006 SD per year. At this rate, the average white-black achievement will take over a century to be eliminated. White-black math gaps are, on average, 0.13 SD larger than reading gaps (given the centering in the model, this implies that the average reading gap is 0.71 SD and the average math gap is 0.84 SD). Moreover, math gaps are narrowing at a somewhat faster rate than reading gaps (reading gaps are narrowing at an estimated -0.0045 SD/year; math gaps at -0.0075 SD/year), but still would take over a century to reach 0. On average, the white black-gaps do not change significantly as children progress from second to eighth grade.

[INSERT TABLE 1 ABOUT HERE]

Model H1 reports the comparable estimates for white-Hispanic gap patterns. The average white-Hispanic gap is marginally smaller than the average white-black gap for 4^{th} graders in 2002 (0.64 SD). White-Hispanic gaps are closing slightly faster, on average, than white-black gaps (0.009 SD/ year). The white-Hispanic math gaps are very slightly larger than reading gaps, by 0.02 SD, and again, math gaps are narrowing faster (-0.011 SD/year) than reading gaps (-0.007 SD/year). These rates are roughly 50% faster than the rate of change of white-black gaps, but it would still take over 50 years for the white-Hispanic gaps to be eliminated at this rate. Finally, although the white-Hispanic achievement gap does narrow, on average, as children progress through school, it does so very slowly (-0.007 SD/grade); this implies that the average gap in 8th grade is 0.035 SD smaller than the gap in 3rd grade, a trivial difference in comparison to the size of the gap.

Between-State Variation in Gap Magnitudes and Trends

We are interested not only in the average levels and trends in achievement gaps, but also in how these vary among states. The bottom section of Table 1 presents the estimated standard deviation of the state-specific intercepts, cohort slopes, grade slopes, and math effects among states. Of particular interest is the variation of the intercepts and cohort slopes. It is also worth noting that the reliabilities of the estimated state-specific intercepts, cohort trends, and grade trends are very high in most cases—above 0.90, except for the reliability of the white-Hispanic cohort trends, which is 0.73. These high reliabilities mean that we can distinguish states from one another in terms of their gap levels and trends.

The unconditional white-black gap model (B1) indicates that there is considerable variation among states in achievement gap magnitudes (the standard deviation of the intercepts is 0.167), which implies that 95% of states have achievement gaps between 0.45 SD and 1.10 SD. The cohort trends vary considerably as well (SD = 0.013 SD/year). Given that the mean trend is 0.006 SD/year, this implies that the white-black achievement gap is widening in a number of states. The variation in the size and trends in white-black achievement gaps is displayed in the left side of Figure 2, which plots the EB estimates of the state-specific intercepts and trends in the gaps. White-black levels and trends are inversely correlated (r = -0.55): on average, achievement gaps are closing the fastest in states where gap levels are the largest, implying that the variance in achievement gaps across states has been narrowing in recent years. In Illinois, for example, the achievement gap level in grade 4 of 2007 is greater than 1 SD, and thus larger than the average across all states, but also is closing by more than 0.02 SD per year, a much faster rate than the average rate of gap closure across states. Finally, the grade slopes also vary significantly between states, indicating that in some states white-black gaps are narrowing as children progress through school, and in others, these gaps are widening.

[INSERT FIGURE 2 ABOUT HERE]

White-Hispanic gaps, like white-black gaps, vary considerably among states. There is, however, somewhat less variation between states in trends in white-Hispanic achievement gaps than in white-black gap trends. These patterns can be seen in the right side of Figure 2 as well as in the standard deviations reported in Table 1, Model H1. As with white-black gaps, there is a negative relationship between state achievement gap levels and rates of change in gaps ($r = -0.34$), but this relationship is much weaker for white-Hispanic gaps than it was for white-black gaps. Finally, the rate at which the white-Hispanic gap changes across grades also varies substantially across states, again suggesting that gaps are widening across grades in some states while they are narrowing in others.

Explaining Between-State Variation in Gap Magnitudes and Trends

The first part of our analysis suggested that the size of gaps and the rates at which they change over time and across grades vary considerably across states. We now consider how much of this variation can be accounted for by two categories of state-level factors: racial differences in socioeconomic circumstances and racial segregation. Table 1 reports the estimates from three additional models, each including variables measuring one or both of these two sets of factors. We do not show the coefficients on the individual covariates because we are primarily interested in their collective power to explain the variation among states than in their specific partial associations with achievement gaps, none of which can be interpreted causally and some of which are confounded by high collinearity with other variables in the models. In each model, we compute the proportion of the unconditional variances of the intercepts, cohort trends, and grade trends (from Models B1 or H1) that is accounted for when the covariates are added to the models.

Models B2 and B3 show that the socioeconomic disparity and segregation/demographic covariates each explain about three-fifths (61% and 57%, respectively) of the between state variance in achievement gap levels. Combined (Model B4), these covariates explain three quarters of the total variation in achievement gap levels. In other words, differences between states in the degree of racial socioeconomic disparities and racial segregation account for the lion's share of between-state differences in achievement gaps. Surprisingly, however, none of the variance between states in rates of gap closure across time or grades is accounted for by these variables, despite our inclusion of time- and grade-varying controls.

The patterns are relatively similar with respect to variation in white-Hispanic achievement gaps. Over three-quarters of the between-state variance in gap magnitudes can be explained by white-Hispanic differences in socioeconomic conditions and white-Hispanic segregation. Interestingly, however, the socioeconomic disparities between white and Hispanic families appear to account for all of this result; including the segregation and ethnic composition variables in Model H4 explains none of the remaining variance after the socioeconomic variables are included in Model H2. As with the white-black gap, none of the variance in trends across time or grades is explained by covariates (the apparent negative variance "explained" in models H3 and H4 is a result of the fact that inclusion of the covariates makes the estimation of variance more precise).

Figure 3 illustrates these patterns by plotting the covariate-adjusted EB estimates of the intercepts and cohort trends from models B4 and H4. Compared to the unconditional estimates plotted in Figure 2, Figure 3 shows much less variation in gap levels, but no reduction in the variation of gap trends. The inverse relationship between the levels of within-state achievement gap levels and trends disappears and is no longer significant.

[INSERT FIGURE 3 ABOUT HERE]

Why does our short list of covariates explain so much of the variation between states in achievement gaps levels, but so little of the variation in achievement gaps across cohorts and across grades? One possibility is that the trends in the covariates do not vary much across states, which would suggest that there is not enough variation in the covariate trends to explain much of the variation in achievement gap trends. We investigate this by estimating simple models of within-state levels and trends in achievement gaps (see Table 2). These models show sizable and significant variation across states in the average levels of these covariates and in their trends. Indeed, for most of the socioeconomic variables, the amount of variation in trends relative to the amount of variation in the intercepts is larger for the covariates than in the achievement gap models B1 and H1. This suggests that the failure of the models to explain any of the between-state variation in achievement gaps trends is not due to insufficient variation across states in covariate trends. Bolstering this argument is the fact that the coefficients (not shown here) on the state covariates ($\mathbf{X}_s$) are generally substantially larger than the corresponding coefficients on the state-by-cohort covariates ($\mathbf{W}_{sc}$). This suggests that it is not socioeconomic disparities *per se* that drive the wide variation between states in the size of achievement gaps, but rather some correlated factors that do not vary much over time or that affect gaps too slowly to register in our models.

[TABLE 2 ABOUT HERE]

Discussion

Across all 50 states, white-black and white-Hispanic gaps remain quite large. Although these gaps have been closing, on average, over the last two decades, the rate of change is, in most states, extremely slow. In addition, there is substantial variation across states in gap magnitudes and rates of change: achievement gaps have been widening in some states and narrowing in others. Finally, while much of the between-state variation in gap levels is associated with racial socioeconomic disparities and racial segregation levels, none of the variation in trends can be accounted for by variation in changes in these factors.

This last puzzle requires explanation. Understanding what factors lead to narrowing achievement gaps could help educators and policymakers craft effective strategies to speed up the reduction in gaps. One plausible explanation is that state education policy differences account for some of the variation in the rates at which the gaps are narrowing. In particular, state policies around school accountability, early childhood education (ECE), and equalization of school funding could plausibly affect achievement gaps.

State school accountability policies might affect achievement gaps to the extent that they include provisions for subgroup-specific accountability, as No Child Left Behind (NCLB) was intended to do. However, the extent to which policies like NCLB put pressure on schools to improve minority students' test scores may vary across states depending on features of the state and the design of the policy. If this is the case, some states may close achievement gaps more quickly than others (see Reardon, Greenberg, Kalogrides, Shores, & Valentino, 2013).

Second, publically funded ECE programs have been expanding across the country—more in some states than others. There is some suggestive evidence that the effects of preschool are largest for disadvantaged children, and that the effects of high-quality preschool can have long-lasting effects on student achievement (Garces, Thomas, & Currie, 2002; Magnuson, Ruhm, & Waldfogel, 2007; Barnett, 1998). Because black and Hispanic children are disproportionately low-income relative to whites, the expansion of preschool programs targeted to low-income children in some states may have led to reductions in achievement gaps in those states more than in states without such programs. The expansion of universal preschool programs with access for all children may have also led to narrowing gaps, if such expansions differentially benefited low-income children.

Finally, between-district inequalities in per pupil spending also varies across states. If spending is related to increases in achievement and decreases in dropouts among disadvantaged students (a contested issue, but see Downes & Figlio, 1997; Hoxby, 2001), then states that either increased allocations to their poorest districts, or that otherwise equalized spending across districts within their states over time, might also be those that experienced faster rates of achievement gap closure relative to states that did not.

Although changes in these three policy areas are plausible explanations for why we observe the variation in achievement gap trends presented here, they are certainly not the only possibilities. There are other differences in state contexts, such as differences in non-educational social policies. that could explain some of the between-state differences in rates of change in achievement gaps.

The bad news in our results is that achievement gaps, on average, remain large and are closing at glacial pace. The good news, perhaps, is that there is a lot of variation across states in how quickly achievement gaps are closing, and little of this variation seems related to variation in trends in racial socioeconomic disparities. If we can understand what has led to the more rapid reduction in gaps in some states—like Pennsylvania, Mississippi, Illinois, and New York have with respect to white-black gaps, and New York, Louisiana, and South Dakota have with respect to white-Hispanic gaps—then perhaps we can apply this knowledge to closing gaps in all states.

References

Aud, S., Fox, M. A., & Kewal Ramani, A. (2010). Status and trends in the education of racial and ethnic minorities. *Washington, DC: National Center for Education Statistics*.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). The Condition of Education 2011. *Washington, DC: National Center for Education Statistics.*

Barnett, W.S. (1998). Long-term effects on cognitive development and school success. In W.S. Barnett & S.S. Boocock (Eds.), *Early care and education for children in poverty: Promises, programs, and long-term results*. (11-44). Albany, NY: State University of New York Press.

Center for Education Policy (2009). *State testing data and student achievement: State profiles & data.* Retrieved February 24, 2012, from http://www.cep-dc.org/index.cfm?DocumentSubTopicID=22 for publicly .

Downes, T. A., & Figlio, D. N. (1997). *School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?* Institute for Research on Poverty, University of Wisconsin--Madison.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps?. *The future of children*, *15*(1), 35-54.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447-464.

Fryer, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, *8*(2), 249-281.

Garces, D. Thomas, J. Currie. (2002). Longer term effects of Head Start. *The American Economic Review*, *92*(4), 999-1012.

Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the Black-White score gap narrow in the 1970s and 1980s. *The Black-White test score gap*, 182-228.

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review, 94*, 2004, 919-943.

Hedges, L. V., & Nowell, A. (1998). Black-White test score convergence since 1965. *The Black-White test score gap*, 149-181.

Hemphill, F.C., & Vanneman, A. (2011). Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress. NCES 2011-459. *National Center for Education Statistics*.

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

Ho, A. D., & Reardon, S. F. (2012). Estimating Achievement Gaps From Test Scores Reported in Ordinal "Proficiency" Categories. *Journal of Educational and Behavioral Statistics*, *37*(4), 489-517.

Hoxby, C. M. (2001). All school finance equalizations are not created equal. *The Quarterly Journal of Economics*, *116*(4), 1189-1231.

Johnson, R. C. (2011). *Long-run impacts of school desegregation & school quality on adult attainments* (No. w16664). National Bureau of Economic Research.

Kober, N., Chudowsky, N., & Chudowsky, V. (2010). State Test Score Trends through 2008-09, Part 2: Slow and Uneven Progress in Narrowing Gaps. *Center on Education Policy*.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance?. *Economics of Education Review*, *26*(1), 33-51.

Neal, D. (2006). Why has black–white skill convergence stopped?. *Handbook of the Economics of Education*, *1*, 511-576.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). NAEP 2008: Trends in Academic Progress. NCES 2009-479. *National Center for Education Statistics*.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, 91-116.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological methodology*, *32*(1), 33-67.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853-891.

Reardon, S.F., Greenberg, E.H., Kalogrides, D., Shores, K.A., & Valentino, R.A. (2013). "Left behind? The effect of no child left behind on academic achievement gaps." Presented at the *Annual Meeting of the American Education Research Association*. San Francisco, CA.

Reardon, S.F., Ho, A., & Kalogrides, D. (2013). "Addressing measurement error and sampling variability in nonparametric gap estimation." Presented at the *Annual Meeting of the National Council on Measurement in Education*. San Francisco, CA.

Reardon, S. F., & Robinson, J. P. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. *Handbook of research in education finance and policy*, 497-516.

Shores, K.A., Valentino, R.A., and Reardon, S.F. (2013). "Trends in nonparametric achievement gaps in the NCLB era." Presented at the *Annual Meeting of the National Council on Measurement in Education*. San Francisco, CA.

Theil, H., & Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts†.

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2009-455. *National Center for Education Statistics*.

Table 1. Achievement Gap Regression Model Parameter Estimates

	White-Black				White-Hispanic			
	Model B1	Model B2	Model B3	Model B4	Model H1	Model H2	Model H3	Model H4
Coefficient Estimates								
Intercept	0.776 ***	0.774 ***	0.783 ***	0.780 ***	0.637 ***	0.636 ***	0.640	0.639 ***
	(0.024)	(0.015)	(0.019)	(0.062)	(0.026)	(0.024)	(1.187)	(0.190)
Cohort	-0.006 ***	-0.007 ***	-0.006 ***	-0.006	-0.009 ***	-0.009 ***	-0.007	-0.007
	(0.002)	(0.002)	(0.002)	(0.005)	(0.001)	(0.002)	(0.027)	(0.007)
Grade	-0.001	0.000	-0.005	-0.004	-0.008 ***	-0.007 *	-0.008	-0.007
	(0.002)	(0.003)	(0.007)	(0.008)	(0.002)	(0.003)	(0.050)	(0.025)
Math (vs.Reading)	0.134 ***	0.134 ***	0.134 ***	0.134 ***	0.028 ***	0.028 ***	0.028	0.028
	(0.007)	(0.007)	(0.007)	(0.014)	(0.008)	(0.008)	(0.055)	(0.041)
Math*Cohort	-0.003 **	-0.003 **	-0.003 **	-0.003 **	-0.004 **	-0.004 **	-0.004 **	-0.004 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Covariates Included								
Socioeconomic Disparities		X		X		X		X
Segregation and Composition			X	X			X	X
Random Coefficient Parameters								
Intercept								
Reliability	0.986	0.967	0.969	0.951	0.966	0.921	0.936	0.901
Standard Deviation	0.167 ***	0.106 ***	0.109 ***	0.086 ***	0.181 ***	0.077 ***	0.151 ***	0.078 ***
Proportion of Variance Explained		61%	57%	75%		82%	30%	82%
Cohort Trend								
Reliability	0.908	0.908	0.908	0.908	0.734	0.714	0.724	0.701
Standard Deviation	0.013 ***	0.013 ***	0.013 ***	0.013 ***	0.009 ***	0.009 ***	0.009 ***	0.009 ***
Proportion of Variance Explained		0%	0%	0%		0%	0%	0%
Grade								
Reliability	0.929	0.929	0.927	0.926	0.915	0.915	0.934	0.932
Standard Deviation	0.020 ***	0.020 ***	0.019 ***	0.019 ***	0.018 ***	0.018 ***	0.021 ***	0.020 ***
Proportion of Variance Explained		0%	3%	4%		1%	-35%	-30%
SD(Math)	0.009 ***	0.009 ***	0.009 ***	0.008 ***	0.013 ***	0.013 ***	0.013 ***	0.013 ***
Random Coefficient Correlations								
Corr(Intercept and Cohort)	-0.556	-0.251	0.010	0.331	-0.364	-0.217	-0.238	-0.219
Corr(Intercept and Grade)	-0.247	-0.149	-0.056	0.077	-0.077	-0.314	-0.044	-0.396
Total Obs.	5,736	5,736	5,736	5,736	5,733	5,733	5,733	5,733

Table 2: Estimated Trends in State-Level Covariates

	Income Ratio		Poverty Ratio		Unemploy-ment Ratio		Education Difference		Segregation		Percent Black/Hisp	
	White-Black											
Intercept	0.588	***	3.706	***	2.771	***	-1.229	***	0.387	***	14.54	***
	(0.017)		(0.180)		(0.170)		(0.095)		(0.020)		(1.831)	
Cohort	-0.005		0.008		-0.018		-0.033	*	0.002	**	0.029	
	(0.003)		(0.027)		(0.018)		(0.015)		(0.001)		(0.026)	
SD(Intercept)	0.101		1.141		0.989		0.628		0.140		13.068	
SD(Cohort)	0.020		0.160		0.040		0.095		0.004		0.151	
Corr(Intercept and Cohort)	-0.458		0.093		-0.931		0.165		-0.460		-0.128	
Reliability of Intercept	0.681		0.789		0.653		0.851		0.999		0.998	
Reliability of Cohort Slope	0.743		0.709		0.092		0.812		0.956		0.668	
Total Obs.	850		850		850		850		850		850	
	White-Hispanic											
Intercept	0.626	***	3.473	***	2.122	***	-2.252	***	0.327	***	12.459	***
	(0.018)		(0.179)		(0.128)		(0.116)		(0.015)		(1.811)	
Cohort	-0.009	*	0.009		0.026		-0.026	+	0.003	***	0.526	***
	(0.005)		(0.018)		(0.016)		(0.015)		(0.001)		(0.048)	
SD(Intercept)	0.102		1.145		0.741		0.778		0.109		12.937	
SD(Cohort)	0.030		0.085		0.070		0.096		0.004		0.337	
Corr(Intercept and Cohort)	-0.566		-0.263		0.120		0.189		-0.499		0.700	
Reliability of Intercept	0.628		0.799		0.652		0.884		0.996		1	
Reliability of Cohort Slope	0.824		0.417		0.358		0.793		0.909		0.986	
Total Obs.	850		850		850		850		850		850	

*Cohorts 1991 to 2007 are included. The maximum number of observations would be 16 cohorts*50 states =

Figure 1. Within-state achievement gap trends over time, by data source

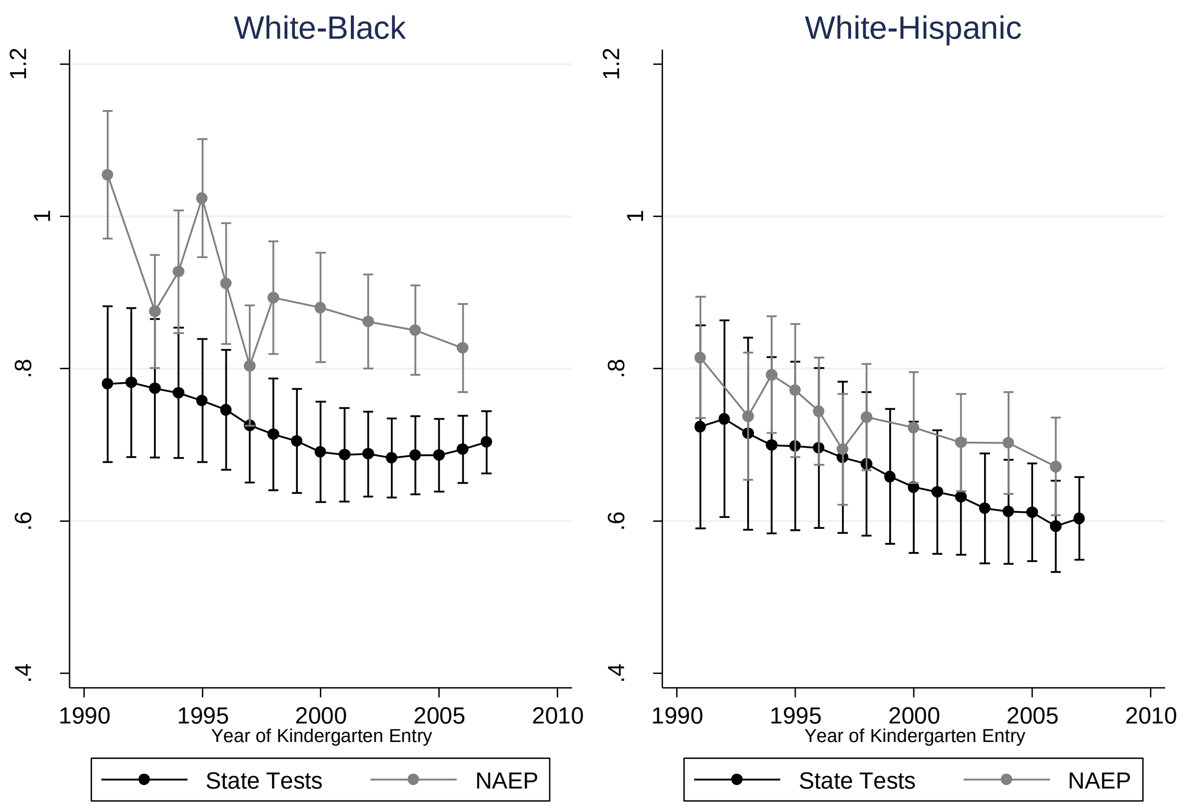

Figure 2. Achievement gap sizes and trends, by state, unadjusted

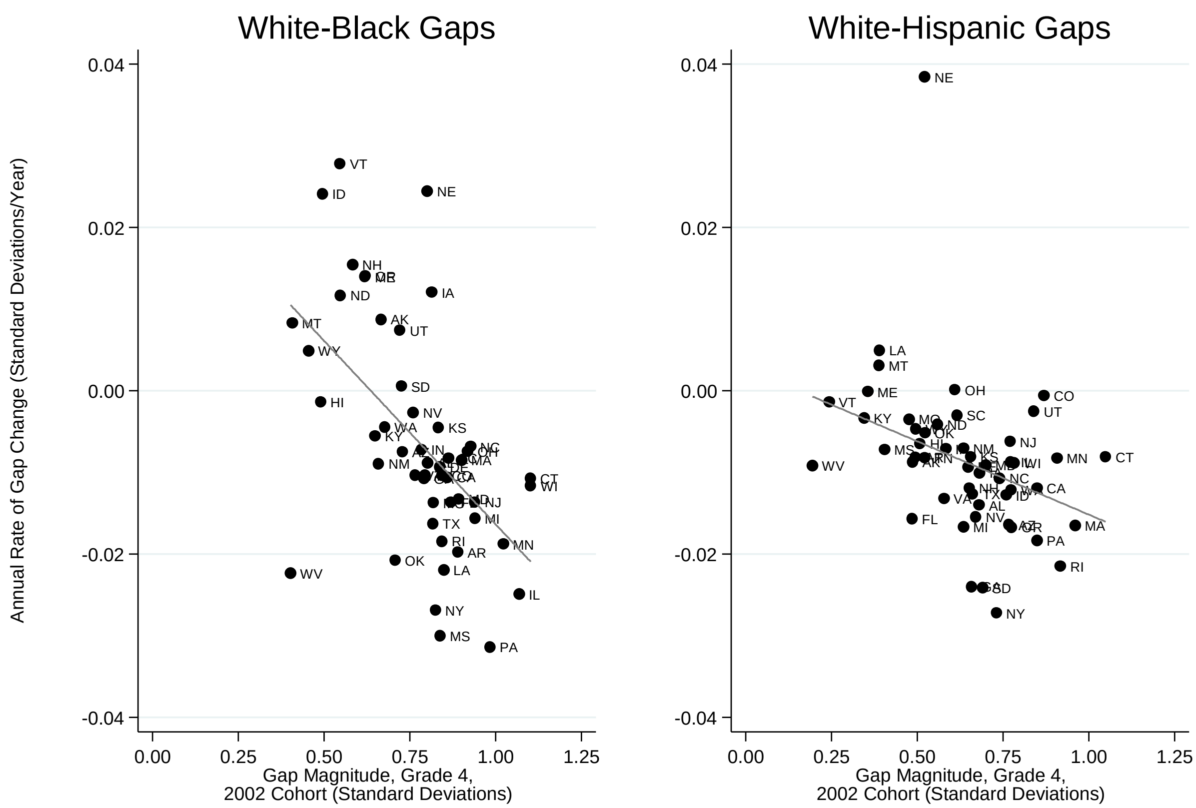

Figure 3. Achievement gap sizes and trends, by state, adjusted