

Peltzman, Sam

Working Paper

Political Economy of Public Education: Non-college-bound Students

Working Paper, No. 108

Provided in Cooperation with:

George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business

Suggested Citation: Peltzman, Sam (1994) : Political Economy of Public Education: Non-college-bound Students, Working Paper, No. 108, The University of Chicago, Center for the Study of the Economy and the State, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/262510>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Political Economy of Public Education: Non-College-Bound Students

Sam Peltzman

Journal of Law and Economics, Vol. 39, No. 1 (Apr., 1996), 73-120.

Stable URL:

<http://links.jstor.org/sici?sici=0022-2186%28199604%2939%3A1%3C73%3APEOPEN%3E2.0.CO%3B2-3>

Journal of Law and Economics is currently published by The University of Chicago Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.



POLITICAL ECONOMY OF PUBLIC EDUCATION: NON-COLLEGE-BOUND STUDENTS*

SAM PELTZMAN
University of Chicago

ABSTRACT

My previous research showed that two important changes in the political environment of public schools—growing teacher unionization and a shift of funding responsibility to state governments—adversely affected the performance of college-bound students. Here I show similar impacts for public school students who do not go to college. These effects are found in analyses of 1971–91 changes in a school performance measure derived from individual scores on the Armed Forces Qualifying Test. Comparative analysis of performance trends in different areas of the same state suggests that the adverse performance effects of teacher unionization and spending centralization stem from their impact on state educational policy rather than on the direct operation of schools. These adverse effects are also found for students in the lower tail of achievement and for black students. They are not plausibly related to broader political and social changes.

I. INTRODUCTION

THE recent performance of American public education has been disappointing. Over the last 3 decades real spending per pupil has almost tripled, but available measures of student achievement have deteriorated. The most often cited measure is the Scholastic Aptitude Test (SAT). The national average on this widely used college entrance test declined about .4 standard deviations from 1965 to 1980 and has essentially remained at this lower level since then. Scores on the other widely used college entrance test, the American College Testing Program's ACT Assessment, follow a similar pattern. While there is room for quibbling about details, it is reasonably clear that trends on the SAT and ACT are broadly repre-

* I am grateful to the Defense Manpower Data Center for providing the data on military applicants that provides the basis for this study and for providing data on zip code characteristics from their DORIS database. I also want to thank Derek Neal for making available to me his data from the National Longitudinal Survey of Youth and for his help in understanding these data and for his comments on an earlier draft. Kevin Murphy, Eric Hanushek, Sherwin Rosen, and William Fischel provided valuable comments and suggestions. Jamie Johnson and David Greeley provided invaluable research assistance. Research support from grants to the George J. Stigler Center for the Study of the Economy and the State, University of Chicago, by the Sarah Scaife Foundation and the Lynde and Harry Bradley Foundation is acknowledged with gratitude.

[*Journal of Law and Economics*, vol. XXXIX (April 1996)]

© 1996 by The University of Chicago. All rights reserved. 0022-2186/96/3901-0003\$01.50

sentative, at least for the eventful 1965–80 period. The basic literacy and numeracy of public school students waned even as the resources devoted to the enterprise waxed.¹

In a previous article,² I argued that analysis of the performance of American elementary and secondary schools should not ignore the mainly public character of that enterprise³ and the resulting political context within which most school resources are allocated. Evidence in that article suggested that changes in the political economy of public education—specifically, the growth of teacher unionization, the shift of responsibility for school finance from local school boards to state governments, and the increased reliance of politically influential employers on colleges rather than public schools for employee education—may have contributed to declining school performance. The nature of this evidence was that in states where these trends went furthest performance on the SAT and ACT tended to decline most.

This article extends the analysis to students who do not go on to college. It finds that trends in their school performance have been at least as powerfully affected by changes in the political background as college-bound students.

There are two important reasons for studying the performance of non-college-bound students. First, ignoring them misses much of the story. Fully half of the relevant age cohort typically never enters college. Second, recent trends in the performance of these students seem substantially better than for the SAT and ACT population.⁴ This divergence, which began around 1980 when SAT and ACT scores reached a nadir, suggests caution about drawing broad conclusions entirely from the performance of the SAT and ACT population, at least for the last decade or so.

¹ The evidence is summarized in Congressional Budget Office, *Trends in Educational Achievement* (1986), and analyzed further in Congressional Budget Office, *Educational Achievement: Explanations and Implications of Recent Trends* (1987).

² Sam Peltzman, *The Political Economy of the Decline of American Public Education*, 36 *J. Law & Econ.* 331 (1993).

³ Public schools enroll about 90 percent of all students.

⁴ Evidence on this point is summarized by Charles Murray & R. J. Herrnstein, *What's Really behind the SAT-Score Decline?* *Public Interest*, Winter 1992, at 32. In general, results for broad samples of high school seniors and juniors improved after 1980 while SAT and ACT scores have remained flat. For example, results from the Preliminary SAT, which is administered irregularly to a nationally representative sample, show the same decline as the SAT and ACT from 1960 to 1974 but an improvement over the next decade. The Iowa Test of Educational Development which has been administered to virtually all Iowa high school students since the 1950s shows a roughly similar pattern: a substantial decline from the early 1960s to around 1980 just like the SAT, but an equally substantial rise in the next decade.

The next section of the article reviews briefly the empirical strategy I employ to study the effect of political forces on public school performance, and it summarizes the political forces to be studied. This is followed by a section describing my use of a voluminous data set on applicants to the armed forces to measure trends in school performance for non-college-bound students. This measure is then subject to an analysis of the impact of political changes on the performance of these students which mainly parallels my previous analysis of SAT and ACT scores. By so doing comparisons of the relative importance of political forces on the two groups can be made. The implications of these comparisons as well as other findings on intrastate and racial differences are discussed in a concluding section.

II. ANALYTICAL FRAMEWORK AND PREVIOUS RESULTS

The few available time series of school performance exhibit considerable inertia. For example, mean SAT scores declined in every one of the years between 1963 and 1980. The recent upturn among non-college-bound students appears to be similarly persistent.⁵ Trends like these are essentially single events which preclude conventional statistical analysis of their correlates. In effect, there are too few degrees of freedom with which to "fit" a plausible explanation.

A more feasible, if more modest, strategy is permitted by the constitutional devolution of responsibility for public education to the states. We can ask why schools in some states did better or worse than others. More specifically, to gain insight into the broad trends we should focus on cross-state differences in the *changes* in school performance over time. This strategy is pursued here as it was in the predecessor study.

Some restraint in pursuing this strategy is needed, because so little is known about the sources of the broad trends. Even a plausible model of what went wrong in the 1960s and 1970s and what went right thereafter is lacking.⁶ Accordingly, there is little to bound the domain of inquiry. Two considerations provide such bounds here:

1. If an analysis of cross-state differences in school performance trends is to shed light on the national trends, the candidate explanations should follow broadly comparable trends. For example, the Congressional Budget Office's 1987 study⁷ showed that television watching peaked well

⁵ Scores on the Iowa Test (see note 4 *supra*) declined in only 1 year between 1978 and 1990.

⁶ See Congressional Budget Office, *supra* note 1.

⁷*Id.* at 70.

before the achievement decline began. It used this timing mismatch to cast doubt on the "common view" that television contributed substantially to declining school performance. The example suggests that, in general, we should look for candidate explanations among those things that changed considerably in the 1960s and 1970s and either stopped changing or reversed course thereafter.⁸

2. I focus mainly on a few changes in the political economy of public education without pretending to provide a complete model of changes in school performance. This narrow focus is meant to redress a considerable neglect of politics in the existing literature on school performance and to facilitate comparison with my previous research on college-bound students. With the notable exception of a few comparisons of private and public schools,⁹ the literature essentially ignores the mainly public character of American elementary and secondary education.¹⁰ However, my previous research uncovered important effects of changes in the political background of public education on the performance of college-bound students. Those results motivate extending the inquiry to non-college-bound students.

In sum, to investigate why public school performance *declined* (and subsequently rebounded), I analyze *changes* in school performance across states over the same time period that these broad trends were operative. I try to relate changes in a state's school performance to changes in some political background variables in the state (plus a few controls). These variables are chosen because they generally follow a time pattern broadly congruent with that of achievement—changing much prior to 1980¹¹ and either changing little or reversing course thereafter.

⁸ This is no hard and fast rule. Models with changing marginal effects of monotonically trended variables could produce the observed time pattern of school performance.

⁹ For example, James Coleman, Thomas Hoffer, & Sally Kilgore, *High School Achievement: Public, Catholic and Private Schools Compared* (1982); and John Chubb & Terry Moe, *Politics, Markets and America's Schools* (1990). More recent work includes Derek Neal, *The Effect of Catholic Secondary Schooling on Educational Attainment* (Working Paper No. 95, Univ. Chicago, Center for the Study of the Economy and the State 1994); and Caroline Minter Hoxby, *Do Private Schools Provide Competition for Public Schools?* (unpublished manuscript, Mass. Inst. Tech., Dep't Economics 1993).

¹⁰ In economics, the dominant paradigm has been the educational production function literature which views schools, analogously to firms, as producing an output (achievement) with inputs like teacher time and training. This literature, which commonly finds that measured inputs contribute little to achievement, is reviewed in Eric Hanushek, *The Economics of Schooling: Production and Efficiency in Public Schools*, 24 *J. Econ. Literature* 1141, 1148–70 (1986).

¹¹ Some allowance for lags is necessary. High school achievement in 1980 is affected by education received as far back as 1968.

I provide here only a brief summary of the political background variables to be studied (more extensive discussion and references can be found in my 1993 article):

1. Teacher unionization was insignificant prior to 1961 when the American Federation of Teachers (AFT) succeeded in organizing the New York City teachers. This breakthrough induced a reluctant professional association, the National Education Association (NEA) to 'abandon long-standing opposition to unionization and to compete head-on with the AFT. At the same time, the legal background was growing generally more favorable to public sector unions. By the late 1960s around half of all teachers were unionized, with an additional quarter organized in the subsequent decade. Union density among teachers today is at roughly the 1980 level, with a bit over half of all teachers belonging to NEA affiliates and around a fifth belonging to AFT locals. In general, the AFT's main success occurred in or near the big cities of the Northeast and Midwest, while the NEA was more successful in smaller cities and suburbs. Much of the South remains unorganized. Teacher unions and their political action arms have become formidable forces in state capitals and have exerted influence over a wide variety of nonwage issues such as class-size, curriculum, textbooks, and teacher promotion.

My 1993 article analyzed changes in SAT and ACT scores for 1971–89 in a cross section of 44 states.¹² Some of the detailed results are reproduced later in this article (see Table 5). They show generally unfavorable effects of the growth of teacher unionization on SAT and ACT score changes, with the largest impact coming from growth of the initially more militant AFT.¹³

2. From the end of World War II to around 1960, local school boards raised about 60 percent of the funds for U.S. public schools. State governments provided most of the remainder. This 60-40 local-state split eroded to 50-50 in the next 20 years.¹⁴ My 1993 article found that states which went furthest in centralizing their school finances tended to have steeper declines in SAT and ACT scores.

Political and legal opposition to interdistrict spending inequality contributed to the push for centralizing school finance. Thus the negative

¹² The SAT and ACT scores were combined and adjusted for effects of changes in the number of test takers and their racial composition.

¹³ The plausible mechanism underlying these generally unfavorable effects on student performance is surely not that teacher organizations pursued that objective directly. Instead, it is that they became able to assert claims on limited resources (money, policy making effort, and so forth) in the pursuit of objectives competing with student achievement (for example, shifting emphasis from teacher "merit" to seniority in promotion decisions).

¹⁴ The federal role was minor throughout the period.

impact on SAT and ACT scores may simply reflect a reallocation of school resources away from wealthier districts with many college-bound students. This article is able to confront that possibility and finds it wanting.

3. The spread of college education has reduced the public schools' share in human capital formation. This loss of "market share" was especially pronounced in the 1960–80 period when the share of adults with a college degree more than doubled. Users of public education can generally be expected to press for higher quality output per dollar of resources than suppliers of inputs to public education like teachers. However, growing use of college-educated human capital should weaken the incentive to exert this pressure. The reason is that college-educated labor is mobile while American public education policy and politics are local. This makes local employer groups, who are education users with ability to exert political pressure, less eager to do so as their reliance on entirely homegrown human capital wanes.

In my 1993 article, I identified an industry's political importance with its share of state employment and its (non)reliance on public schools with the share of its employees who are college graduates.¹⁵ I found that in states where politically important industries employed more college graduates or where industries which employed many college graduates grew in political importance, the decline in SAT and ACT scores was atypically large.

There is surely more to the politics of the supply of public school performance than the growth of teacher unions and the weakened power of local school boards. Similarly, the spread of college education cannot

¹⁵ Specifically the underlying model is that achievement (A) is influenced by employer groups according to

$$A = \sum W_i U_i + \text{other variables,}$$

where

- W = political weight or influence of group i , and
- U = strength of i 's preference for student achievement (which depends on competing uses of the group's political capital, its demand for public school provided human capital, and so forth).

Since I am focusing on the change in achievement, I want to estimate

$$\frac{dA}{dt} = \sum \left[W_i \frac{dU_i}{dt} + U_i \frac{dW_i}{dt} \right] + \dots$$

Empirical implementation of this model is described in note 37 *infra*.

be the only relevant demand variable.¹⁶ The purpose of recounting these developments is primarily methodological: these forces changed substantially in the crucial 1960–80 period when SAT and ACT scores declined substantially. And taken together they accounted for a significant part (nearly half) of the cross-state variation in this decline up to 1980. The end of the national decline is accompanied by a weakening both of these forces and of their ability to explain cross-state differences. Thus crude measures of politics and performance give hints of a link between the two that deserve to be pursued.

III. MEASURING THE PERFORMANCE OF NON-COLLEGE-BOUND STUDENTS

Each year up to 1 million applicants to the U.S. military take an SAT-like test of ability and achievement, the Armed Forces Qualifying Test (AFQT). Most of these are young people who have recently left their public school system, and the vast majority do not go on to college. I have the AFQT percentile score for every applicant to the military for 1976–91 and for every inductee for 1971–75.¹⁷ These 10 million plus individual records provide the raw material for my measure of public school performance.

My unit of analysis is the state. This is the level of government responsible for basic policy on public education, and it is today the most important source of funds. Thus finer levels of aggregation would risk missing the proverbial forest for the trees. (I do analyze some substate aggregates but this reinforces the importance of state policy.) The AFQT percentile scores are periodically normed to yield a mean of 50 for the military applicant pool. Thus they cannot shed light on broad national trends in achievement of this pool. However, scores in a state with improving relative performance will tend to move up in the percentile distribution over time. So, we can measure the states' changes in relative achievement, and this is the measure I analyze.

Scores on the AFQT are affected by more than the quality of schooling. I attempt to remove the influence of these nonschooling effects as follows: each year for 17–20-year-olds I estimate the general relation

$$\text{AFQT}_{ijt} = a_r X_{ijt} + U_{ijt}, \quad (1)$$

¹⁶ For example, in about half the states, both supply and demand conditions were significantly influenced by pressure for desegregation after the 1964 Civil Rights Act. In my 1993 article (Peltzman, *supra* note 2), I found that desegregation had, if anything, a beneficial effect on school performance.

¹⁷ I am grateful to the Defense Manpower Data Center for providing the data.

where

- AFQT_{*ijt*} = the score of individual *i* in state *j* in year *t*,
 X_{ijt} = vector of measurable “background” characteristics
 assumed unrelated to school quality,
 a_t = vector of coefficients which may vary over time but not
 across states, and
 U_{ijt} = error term capturing unmeasured nonschooling effects and
 the quality of *i*’s education.

Then I average the U_{ijt} across all 17–20-year-old test takers in state *j*. The result, \bar{U}_{jt} , is my estimate of the relative standing of *j*’s public school system in year *t*. (I also analyze a measure based on the distribution of U_{ijt} —the fraction of the U_{it} in *j* above the 25th percentile in the national distribution. This measures the ability of a school system to avoid the worst outcomes.) Finally, I estimate

$$\bar{U}_{jt} = \text{constant} + b_j \text{TIME} \quad (2)$$

for each state.

The coefficient b_j is the focus of the analysis. It measures the average annual change in the relative performance of *j*’s school system.

Some problems with the approach summarized by (1) and (2) follow:

1. My procedure assumes that the effects of school quality are unrelated to a student’s background (as measured by the X ’s). This raises substantive and econometric concerns.

Background and school quality effects are not always distinguishable. For example, I include years of schooling in the X vector. The reason is that even the best school system will have dropouts, and they will score below those who stay on. But a bad system may also encourage more dropouts. Thus, to some degree, adjusting AFQT scores for years of education overstates a poor system’s performance.¹⁸

Even if student background and school quality are separable, they may be related. For example, school systems might add more value in states where students have higher average potential. In that case, my procedure of deducting this potential (as measured by the X ’s) from the AFQT score and using the resulting residual as a measure of the level of school quality is biased.¹⁹ However, this particular bias can be removed by adding state

¹⁸ The problem is mitigated, but not eliminated, by constraining the adjustment to be the same across all states. This puts especially poor performance by a state’s dropouts in the performance measure. Evidence that removing dropouts from a school quality measure imparts a conservative bias to my results is presented in note 40 *infra*.

¹⁹ It will tend to understate the true differences in school quality across states.

fixed effects to (1) and estimating

$$\text{AFQT}_{ijt} = a' X_{ijt} + \sum_j k_{jt} D_j + U'_{ijt}, \quad (1')$$

where

D_j = a vector of state dummies in year t , and
 k_{jt} = coefficients of these dummies.

Equation (1') says that the student's expected AFQT score in any year is the sum of background and state-specific effects (the k_{jt}) which include school quality and which can change over time. The ordinary least squares estimator of k_{jt} would be an unbiased estimator of nonbackground effects common to all students in a state in a given year if the X 's fully capture student background. Then, instead of regressing a state-average residual on a time index as in (2), we would estimate

$$k_{jt} = \text{constant} + b'_j \text{TIME} \quad (2')$$

for each state and use b'_j instead of b_j as the measure of the average annual change in school performance.

Unfortunately, we never have a complete set of background variables. And if the omitted background variables are correlated with school performance, both of the alternative inputs to the estimate of trends in school performance—the residuals used in (2) and the state fixed effects used in (2')—will be biased. Neither will be obviously closer to the truth. In the circumstances, I adopt an eclectic approach. Mainly I use the residuals from (1) to estimate performance trends because they are more convenient computationally. But I also use estimates based on state fixed effects to check the sensitivity of the key results. Finally, it needs to be emphasized that, because I am analyzing *changes* in performance over time, biases in estimating performance *levels* are troublesome only if the biases themselves change systematically overtime.

2. As shown in Table 1, the AFQT sample is unrepresentative in a number of ways. It is almost entirely devoid of college students, especially in the 17–20 age-group on which I will focus. This is a compelling virtue of the sample for the present purpose. On the debit side, AFQT takers are also dominantly male, and blacks are overrepresented.²⁰ However, these imbalances do not appear serious. I adjust scores for broad

²⁰ In part, because they are less likely to go on to college. In 1980, the probability that a 17-year-old would graduate high school and go on to college was .37 for whites and .26 for blacks.

TABLE 1
CHARACTERISTICS OF AFQT TAKERS AND U.S. 17-20-YEAR-OLDS, 1980

CHARACTERISTIC	UNITED STATES, 17-20-YEAR-OLDS	AFQT SAMPLE	
		17-20-Year-Olds	Total
1. Female (%)	49.6	16.3	18.0
2. Black (%)	13.9	23.6	24.4
3. Reside in metropolitan area (%)	75.1*	75.2	...
4. Some college education (%):†			
a. Undergraduates 18-21/population 17-20	31.3	2.2	7.1
b. Percentage of 20- and 21-year-olds in school	31.0		
c. Percentage of 17-year-olds enrolling in college within 1 year of high school graduation	35.8		
5. State share of sample/state share of the United States:‡			
a. Mean	99
b. SD:			
48 states	20
32 largest§	16
c. Number of states within .8 < ratio < 1.2:			
48 states		...	32
32 largest		...	25
6. Region: share of sample/share of the United States:‡			
Northeast			.99
North central			.98
South			1.06
West			.92

NOTE.—AFQT = Armed Forces Qualifying Test.

* Total U.S. population.

† Data were obtained from U.S. Department of Education, Digest of Education Statistics (various years).

‡ The denominator is based on 17-20-year-olds. Results are essentially identical if total population share is used. Based on 48 states.

§ Over 1 percent of population ages 17-20.

trends in racial differences, and there is no reason to suspect dramatically different trends by gender.²¹ Moreover, Table 1 shows that the AFQT sample is geographically well-balanced. It mirrors the largely urban character of the population, and it is more or less evenly spread across states: in 1980 (a typical year) the state sample sizes were within 80-120 percent of a random sample for 80 percent of the larger states. The slight tendency

²¹ The broad decline in SAT and ACT scores exempted neither gender. The main difference was a greater decline in female verbal scores.

for the South to be overrepresented and the West underrepresented is due to the racial and educational makeup of the sample.

3. Takers of the AFQT are self-selected.²² For my purpose, it is the dynamic aspects of the selection process that are potentially worrisome. For example, improvement in my performance measure (a rising \bar{U}_{jt}) could result from better (unmeasured) characteristics of test takers rather than from better schools. However, several considerations suggest that such selection bias is not driving subsequent results:

a. The measurable characteristics of the sample do not change in any way obviously symptomatic of a dynamic bias. The number of test takers ranges between roughly 500,000 and 1 million per year with no trend. The broad geographic spread and the degree of overrepresentation of males and blacks shown in Table 1 do not change systematically over time.

b. I found that state average scores and test-taker densities were uncorrelated. A (positive or negative) correlation would hint that those selecting into or out of the sample were systematically different (better or worse, respectively) from the rest. Specifically, when I added state test-taker densities²³ to the right-hand side of (1), the coefficients ranged around zero and yielded nugatory effects. I also regressed the change in \bar{U}_{jt} on the change in relative densities²⁴ for each of the 48 states in my sample. The coefficients again clustered about zero, the number significant being no greater than expected by chance.²⁵ Finally, the current and lagged state unemployment rates are included on the right-hand side of (1) to control for changes in the potential applicant pool.²⁶

c. For a single year, 1980, I can compare results from my sample to a random sample²⁷ of youth aged 15–23 who were given the AFQT and whose subsequent education was tracked. I estimated equation (1) for the 17–20-year-old subset of this random sample with no subsequent college

²² This is literally true after 1972 when the draft ended, but draftees were only 10 percent of inductees in 1971 and 1972.

²³ Test takers in a state divided by the 17–20-year-old population.

²⁴ The state's density minus the national density.

²⁵ These results are broadly consistent with work indicating that military recruits come from both a high- and low-quality margin. The former have good civilian alternatives including further schooling, while the latter have poor civilian alternatives. See David T. Ellwood & David A. Wise, *Uncle Sam Wants You—Sometimes: Military Enlistments and the Youth Labor Market in Public Sector Payrolls* (David Wise ed. 1987). Extra test takers can be either above or below average. This contrasts with the SAT and ACT, where extra test takers usually drag down the average (see Peltzman, *supra* note 2).

²⁶ The coefficients were usually significant and opposite in sign, but the pattern was inconsistent. Thus tighter civilian labor markets can raise or lower average applicant quality as discussed in note 25 *supra*. The numerical impact of these unemployment variables was always trivial.

²⁷ The National Longitudinal Survey of Youth. I am indebted to Derek Neal for making the data available.

education and correlated average state residuals with those from the 1980 military applicants. The results (in Appendix D) suggest that the state-specific components of the two sets of residuals are identical but are much more precisely measured for the much larger military sample. This result is hardly dispositive, because it applies to the level rather than the change in residuals. But it adds confidence that selection bias is generally unimportant in these data.

d. Because my focus is on trends in performance over 1 or 2 decades, any biases from, for example, short-run changes in the labor market are arguably unimportant. However, gradual changes in unmeasured attributes of a state's population would be more serious, because they can induce trends unrelated to school performance. One potential source of such gradual change is mobility. About 10 percent of a typical state's population lived in another state 5 years before, and, more important, there is substantial persistence in mobility propensities across states.²⁸ If new residents usually differ substantially from long-term residents, this could induce a trend in my performance measure.²⁹ And the induced

²⁸ The correlation of the 1980 and 1990 shares of a state's population living in another state 5 years earlier is .9.

²⁹ More precisely, *changes* in the quality of new residents or in their number could induce spurious trends. To see this, suppose there are only two types of test takers in a state. One group (average score U_s) received all its education in that state, and another (whose share is p and average score is U_m) was educated entirely in another state. In any year, the state average score (U) would be

$$U = U_s + p(U_m - U_s).$$

The performance measure I analyze is the time derivative of U (\dot{U}), but the correct measure is \dot{U}_s . The two are related by

$$\dot{U} = \dot{U}_s + p(\dot{U}_m - \dot{U}_s) + (U_m - U_s)\dot{p}.$$

This reveals two ways in which \dot{U} could seriously misrepresent \dot{U}_s : (1) The in-migrants are different than they used to be, and in a different direction from \dot{U}_s . For example, $(\dot{U}_m - \dot{U}_s)$ could be sufficiently positive to make \dot{U} positive even if \dot{U}_s is negative. (2) The in-migrants are systematically different from stayers ($(U_m - U_s) \neq 0$), and their share is changing ($\dot{p} \neq 0$).

Because of the persistence in p (see note 28 *supra*), the first possibility is the more trenchant. However, it requires a possibly unreasonable combination of a high p with a strong negative correlation between \dot{U}_m and \dot{U}_s . For example, with equal variances for \dot{U}_m and \dot{U}_s and $\dot{p} = 0$, the cross-section relation between \dot{U} and \dot{U}_s becomes

$$\dot{U} = \dot{U}_s[1 + p(r - 1)],$$

where r = correlation of \dot{U}_m with \dot{U}_s . Even if $r = -1$, the correlation between \dot{U} and \dot{U}_s remains positive for $p < .5$. Note that if the in-migrants are just a random draw from the rest of the country ($\dot{U}_m = 0$), then $\dot{U} (= (1 - p)\dot{U}_s)$ is biased toward zero but has the same sign as \dot{U}_s .

There are of course further potential complications if \dot{U}_s is systematically unrepresentative of the change in performance of out-migrants who were educated in the state.

trend would be larger absolutely where new residents are more prevalent. However, I found no such correlation in my data.³⁰

4. For 1971–75 we have data only for inductees. Moreover, military induction standards were being progressively tightened in this period as the draft was being wound down. So sample truncation is more severe in 1975 than in 1971. These limitations will tend to make bad states look better over 1971–75, because their worst cases increasingly do not appear in the sample. I tried to remove this bias rather than discard the pre-1976 information. Details are in Appendix C. The basic idea was to estimate residuals from an artificially truncated 1976 sample and use the differences (by zip code) between the residuals from the full and truncated samples to correct the pre-1976 residuals. Appendix C shows that this adjustment produces residuals with seemingly reasonable time-series properties. However, I always report results for the more homogenous post-1975 samples as well.

The important patterns in the sample are summarized in Tables 2 and 3. (The regressions from which these Tables and the subsequently analyzed residuals come are described in Appendix B).³¹ Table 2 gives data for 1980, the chronological midpoint of the sample period. They are broadly representative of the whole period: there is a large racial differential that grows with schooling. High school dropouts (anybody with 2 or fewer years of high school and the 19- and 20-year-olds without a diploma) score roughly 10–15 percentile points below high school graduates. And there are similar-sized penalties if the applicant's parents are blue-collar and poor.³² With the caveat that the education and racial effects may be proxies for otherwise unmeasured family background effects, these results are broadly consistent with much previous work.³³

Table 3 shows the most important change in these effects—the dramatic narrowing in racial differences—and compares it to similarly scaled (but unadjusted) differences from the SAT. The comparison reveals the

³⁰ The correlations between the average of the two new resident shares described in note 29 *supra* and the absolute values of my estimates of the 1971–91 and 1976–91 change in school performance are .11 and .10, respectively. Both are insignificant.

³¹ Regressions with state dummy variables yield substantively identical results to those shown in Tables 2 and 3.

³² The gain (6.7 points) for blacks from female-headed households offsets the loss from poverty (6 points). Since the two effects are frequently conjoined for blacks, the results imply that marginal effects of these variables are confined to whites.

³³ See the summary in Hanushek, *supra* note 10. On the correlation between racial and family background effects, see Chubb & Moe, *supra* note 9, at 126.

narrowing of racial differences to be pervasive and suggests that it is more substantial for the non-college-bound students.³⁴ While the racial differences remain large, the clear message of Table 3 is that they are not immutable.³⁵

IV. PERFORMANCE TRENDS ACROSS STATES

A. *Average Scores*

National achievement trends in the 1980s and 1970s differ greatly. So I begin analysis of cross-state trend differences by dividing the sample period in half. Table 4 shows correlation coefficients for the two subperiods for trends in AFQT and college entrance exam scores. The interesting results are mainly negative: The low (insignificant) negative autocorrelations ($-.11$, $-.12$) suggest that the 2 decades are indeed different. The low correlations across the tests mean that the two samples are providing independent information about school performance. Systems which serve one group well do not systematically do so for the other group. The only sign of a commonality among the two groups occurs in the last line of the table. The mild positive correlations between AFQT trends in the 1980s and SAT/ACT trends in both decades³⁶ hint at effects which "trickle down" from college-bound to non-college-bound students. Subsequent results are consistent with this hint.

My main goal is to see if the political economy of public education, especially those aspects of it susceptible to influence by organized interests, affects the performance of non-college-bound students. Because my previous research uncovered such effects for college-bound students, a general reference to those results might have sufficed had the intergroup correlations in Table 4 been much stronger. The weak correlations are not dispositive either. They could betoken a complete absence of common influences, different weights on them, different timing patterns (as sug-

³⁴ Results from the pre-1976 inductee samples imply that the narrowing goes back at least to 1971. For 1971-75, for example, the racial difference for high school graduates declined 6 points in this sample.

The black-white differences in Table 3 are evaluated at the black applicants' mean probability of coming from a household headed by a female and by an executive, technical, or professional employee. Nearly identical trends, but higher average differentials (around 5 points), are obtained by evaluating black-white differentials at the mean of these characteristics for the whole sample.

³⁵ Blacks comprise a trendless 20-25 percent of the AFQT sample throughout the sample period. This suggests that the declining racial gap is not due to improved preselection of black test takers.

³⁶ The multiple correlation coefficient is .25, which is significant at .10.

TABLE 2
SOCIOECONOMIC CHARACTERISTICS AND AFQT SCORES, 17-20-YEAR-OLDS, 1980 SAMPLE
A. AGE, EDUCATION, RACE

	High School Graduates	3 or 4 Years of High School	≤2 Years of High School
I. Race: mean AFQT scores for 17-year-olds: [*]			
1. Whites	56.3	53.5	39.8
2. Blacks	39.0	38.6	29.7
3. Difference	17.3	14.9	10.1
II. Age: deviation from 17-year-olds:			
1. 18-year-olds	-.8	-5.2	...†
2. 19-year-olds	-1.8	-9.0	...†
3. 20-year-olds	-1.5	-9.0	...†

B. HOUSEHOLD/NEIGHBORHOOD CHARACTERISTICS‡

Mean effect on AFQT score if applicant:		
1. Lives in all-white neighborhood		+3.1
2. Household has income 1 SD above mean		+1.9
3. Parent's occupation is "executive, professional, technical"		+10.5
4. Is from female-headed household		-1.8 (+6.7 for blacks)
5. Is from poverty household		-6.0

NOTE.—AFQT = Armed Forces Qualifying Test.

* Holds other characteristics at sample mean for blacks.

† The same as for high school graduates.

‡ See Appendix A for construction of household characteristics.

gested in Table 4), and so on. Table 5 begins to sort these possibilities by reproducing the main results for college-bound students and replicating them for the AFQT sample.

All the regressions in Table 5 and subsequent tables are estimates of

$$\Delta \text{PERF}_j = f(\Delta \text{POLECON}_j, \dots), \quad (3)$$

where

ΔPERF = the rate of change of some school performance measure over time in state j —usually this is the slope b_j from equation (2)—and

$\Delta \text{DPOLECON}$ = a vector of variables characterizing the change in the political economy of state j over the relevant period.

TABLE 3
RACIAL DIFFERENCES IN AFQT SCORES AND SAT SCORES, 1976-91

YEAR	WHITE-BLACK SCORES ON:			
	SAT ÷ 10*	AFQT†		
		High School Graduates	3 or 4 Years of High School	≤2 Years of High School
1976	26	22	17	9
1977	25	21	16	8
1978	25	19	15	8
1979	24	20	16	9
1980	23	17	15	10
1981	23	19	15	7
1982	22	17	13	6
1983	22	15	11	5
1984	22	15	11	5
1985	22	15	10	4
1986	N.A.	13	9	3
1987	21	13	9	2
1988	20	13	8	0
1989	20	14	9	2
1990	20	11	8	2
1991	19	11	7	3

NOTE.—AFQT = Armed Forces Qualifying Test. SAT = Scholastic Aptitude Test. N.A. = not available.

* Division puts SAT on scale comparable to AFQT (standard deviation of SAT ~ 200; of AFQT residuals ~ 20). Source of SAT: U.S. Department of Education, Digest of Education Statistics (various years).

† Household and neighborhood characteristics held constant at means for blacks over sample period.

TABLE 4
CORRELATION OF PERFORMANCE TRENDS: 1970s AND 1980

	SAT/ACT 1970s	SAT/ACT 1980s	AFQT 1970s
SAT/ACT 1980s	-.11		
AFQT 1970s	.03	-.02	
AFQT 1980s	.17	.16	-.12

NOTE.—SAT/ACT = annual change in average state score on college entrance exams, 1972-81 and 1981-89. See Sam Peltzman, *The Political Economy of the Decline of American Public Education*, 36 *J. Law & Econ.* 331 (1993). AFQT = annual change in average state residual on AFQT, 1971-81 and 1981-91. See the text. AFQT sample is 48 states (excludes Alaska and Hawaii). SAT/ACT sample is 44 (excludes Alaska, Hawaii, Wisconsin, Missouri, Iowa, and Idaho).

The ellipses stand for control variables peripheral to the main inquiry. For example, every regression includes the growth of school inputs.

The focus in (3) on changes over time, rather than levels, of performance is mainly driven by the search for clues to the national decline in school performance. However, by eschewing the more conventional analysis of levels, I can also avoid confounding school performance with unmeasured persistent geographic differences in student background.

The Δ POLECON vector tries to quantify three previously discussed aspects of the changed political economy of public education: the centralization of funding (and presumably of power) in state governments, the growth of teacher unions, and the changed interests of those employers with presumably above-average political influence in a state. The changes are measured over periods which recognize that education is cumulative. Thus the change in performance of high school graduates from, say, the early 1970s to date would be affected by background changes from around 1960 to date.

The variables included in the Δ POLECON vector are changes in

1. the share of public school revenues coming from state governments,
2. the share of teachers belonging to unions affiliated with the NEA or AFT,
3. an index of bargaining rights granted to teacher unions by the state legislature (denoted LAW in subsequent tables—higher values mean that unions have more bargaining rights), and
4. an “industry pressure index” based on employment by politically influential industries of human capital produced outside of public schools.³⁷

Because teacher unions were essentially nonexistent prior to 1960, I use post-1960 levels of the union density and LAW variables where appropriate.

³⁷ The index uses the (approximately two-digit) industry as the unit of analysis. Each industry is presumed to have a political weight (W) based on the industry’s share of state employment and a preference (U) based on the share of its labor force with college degrees (S). The index tries to approximate the change in $W \cdot U(S)$ over time, or

$$W \frac{dU(S)}{dt} + U(S) \frac{dW}{dt}.$$

Each of the two terms is approximated by averages of two alternative expressions. For example, the two alternatives for the first term in the derivative are (1) the average change in S for industries with atypically high employment shares (for example, autos in Michigan) and (2) the average employment share for industries with the largest changes in S . The industry pressure index is the equal-weighted average of the four resulting indexes. The changes are measured over 2 decades beginning 1960 or 1970. See Peltzman, *supra* note 2, for further details.

D. Teacher bargaining right

index:§									
Change 1972-84									
1972	-.033	1.1	.006	1.4	.017	.3	.0054	1.1	
E. Industry pressure index:¶									
1960-80	-1.10	3.3	-.085	1.7	.028	.7	.0031	.8	
\bar{R}^2		.44		.21		.11		.12	
SEE		1.22		.185		1.39		.143	

Sources.—Public school finances from U.S. Department of Education, Digest of Education Statistics (various years). Union density for 1967 provided by Hirschel Kasper. The 1982 data are from U.S. Bureau of Census, 3 1982 Census of Governments, n. 3 (1984) [Education employees belonging to collective bargaining units] and NEA Organizing Unit, AFT Membership by States and Locals for Membership Years 1979-80 (1980). The latter gives AFT densities for 1980 which I assume are the same as 1982. The AFT density is subtracted from the total to estimate the 1982 NEA density. Teacher bargaining rights for 1972 are a count of provisions in State law from Thomas Kochan, Correlates of State Public Employee Bargaining Laws, 12 Indus. Rel. 322, 324-329 (1973). The 1972-84 change comes from similar data supplied by Robert Valletta as described in app. B of Richard B. Freeman & Casey Ichniowski, When Public Sector Workers Unionize (1988).

NOTE.—The dependent variable in all regressions is the average annual change in state average test scores. Columns (1) and (3) are reproduced from Sam Peltzman, The Political Economy of the Decline of American Public Education, 36 J. Law & Econ 331 (1993), table 6, col. (1), and table 7, line 1, respectively. These cover 1972-81 and 1981-89, respectively. Columns (2) and (4) are based on residuals from AFQT test score regressions as described in the text and Appendix B. The years covered by the AFQT data are 1971-81 and 1981-91.

* Annual change in log of public school expenditures per pupil minus log change of salary per teacher.

† Public school revenues from state governments as share of total revenues.

‡ Share of teachers covered by collective bargaining agreements and belonging to locals affiliated with NEA or AFT.

§ Count of rights accorded to teacher organizations by state legislature.

¶ See the text and Peltzman, *supra*, for a general description.

Columns (1) and (3) repeat the important results from my 1993 article for college-bound students (denoted SAT/ACT because their performance is measured by a concatenation of SAT and ACT scores): strong negative effects in the 1970s from early success by the AFT which continue into the next decade, positive early effects from NEA organization which wear off as that organization becomes a mature union, modest and persistent negative effects from early moves to centralize school finance, and a strong negative effect (confined to the 1970s) from an industry pressure index based on employment of college graduates. Any comparison of these results with those for the non-college-bound sample in columns (2) and (4) has to be somewhat tentative because of the plethora of weak results. Nevertheless, there is one reasonably clear common pattern: the persistently negative effects of mature teacher unionization in the 1980s. By contrast, none of the early effects of unionization—positive or negative—on the SAT/ACT sample is discernible in AFQT scores. Roughly comparable³⁸ effects of industry pressure and centralization of school finance are however visible in the 1970s. Put together, Tables 4 and 5 suggest that the effects of changes in the political economy of education occur less promptly and are more drawn out for the non-college-bound students.³⁹ I act on that suggestion by combining the two AFQT subsamples in subsequent work.

Table 6 summarizes the main results for non-college-bound students when the 2 decades are combined. The first regression in each of the first two panels has the same set of variables as Table 5, with the union and bargaining rights variables set at early 1980s levels. These regressions tell a straightforward tale about non-college-bound youth: centralization of school finances and the growth of teacher unions are associated with declining school performance. The negative union effects come from successful organization efforts, not from success in securing legal rights from the legislature. Indeed, the latter is associated with improved school performance. Recent work by Hoxby suggests that negative effects of spending centralization can also be found before my sample period.⁴⁰

³⁸ The scaling of the AFQT and SAT/ACT variables differs by a factor of ~ 10 .

³⁹ Some of this may be due to measurement. The AFQT sample includes individuals whose last school experience occurred in years prior to the test. An extreme (and uncommon) example would be a 20-year-old who dropped out of school at age 15.

⁴⁰ See Caroline M. Hoxby, *Is There an Equity-Efficiency Tradeoff in School Finance?* (unpublished manuscript, Harvard Univ., Dep't Economics, 1995). She uses decade changes in high school completion rates to measure performance trends and finds that these are negatively related to the change in the state spending share over every decade from 1940 to 1980. Her results seem to reinforce mine, because I have removed changes in high school completion from my performance measure.

TABLE 6
AFQT PERFORMANCE TRENDS, 1971-91 AND 1976-91

REGRESSION	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION DENSITY 1982				LAW 1984	LAW ²	INDUSTRY PRESSURE		CHANGE IN FEMALE LABOR FORCE PARTICIPATION		\bar{R}^2 /SEE
			NEA	AFT	1970s	1980s			1960-80	1980-90			
A. 1971-91:													
1. Coefficient	371	-.33	-17	-19	.95			-.95	-.34			.30	
<i>t</i>	.9	3.1	2.3	2.0	3.2			.3	.1			.092	
2. Coefficient	287	-.24	-16	-26	.95						-88	-.38	
<i>t</i>	.8	2.2	2.2	2.7	3.3						2.2	.087	
3. Coefficient	130	-.23	-22	-31	4.71		-.088				-.95	.46	
<i>t</i>	.4	2.2	3.2	3.5	3.4		2.8				2.5	.080	
B. 1976-91:													
1. Coefficient	218	-.39	-28	-32	1.06			2.50	-1.29			.21	
<i>t</i>	.4	2.2	3.0	2.4	2.6			.7	.3			.122	
2. Coefficient	1.8	-.31	-29	-35	.96						-55	.23	
<i>t</i>	.1	1.7	3.1	2.6	2.4						1.0	.120	
3. Coefficient	-.258	-.25	-38	-42	6.26		-.124				-.66	.34	
<i>t</i>	.6	1.5	4.1	3.4	3.2		2.8				1.3	.111	
C. State dummies:													
1. 1971-91:													
Coefficient	-.337	-.21	-24	-23	.64						-15	.30	
<i>t</i>	1.0	2.1	3.6	2.6	2.4						.4	.079	
2. 1971-91:													
Coefficient	-.513	-.19	-31	-29	4.87		-.099				-24	.46	
<i>t</i>	1.8	2.2	5.1	3.7	4.0		3.6				.7	.070	
3. 1976-91:													
Coefficient	-.614	-.48	-35	-41	.72						33	.34	
<i>t</i>	1.3	2.6	3.6	3.1	1.8						.6	.122	

NOTE.—See Table 5 for sources and definition of expenditure growth, change in state revenue share union density, and teacher bargaining rights (LAW); see text and notes 15 and 37 for description of industry pressure index. Change in female labor participation is from the Census of Population. Initial year for expenditure growth and state revenue share is 1961 for regressions in panel A; 1966 for panel B. The industry pressure indexes are changes over 1960-80 (1970s) and 1970-89 (1980s). LAW variable is the 1984 level of the index constructed from data supplied by Robert Valletta. The dependent variable in panels A and B is the average annual change (from regression on a time index) of the average adjusted residual AFQT score in a state. For panel C a similarly estimated annual change in the coefficients of state dummy variables is the dependent variable. See text and Appendix A for a description of other variables held constant in estimation of both the average adjusted residual and the state fixed effect in any year. $N = 48$ states; SEE = standard error of estimate. Absolute t -ratios are below coefficients. Coefficients are times 100.

The industry pressure variables prove uninformative for this sample and are dropped subsequently.

The second regression in panels A and B is motivated by the insignificant expenditure effects and recent work by Flyer and Rosen⁴¹ showing that increased female labor force participation has raised demand for school inputs. They argue that school inputs are a substitute for home production of human capital. If so, rising expenditures might only be offsetting the drag on home production of human capital from growing female labor force participation.⁴² Some hint of this drag is found in the second regression, which adds the change in female labor force participation, but the negative effect is somewhat sensitive to the estimation period and seems to have lost its force after 1980. Taking account of female labor force participation does not however disinter any positive spending effects.

The third regression in each of the first two panels attempts to illuminate the surprising positive effect of the LAW variable by adding a quadratic term. The significantly negative coefficients of this term imply maxima at about 1 standard deviation above the mean of LAW. This says that neither complete legislative accommodation nor hard-line resistance to teacher unions is optimal. This result needs to be put in context. The LAW variable is essentially a product of a broad post-1960 thrust for public sector union rights and membership.⁴³ Undoubtedly, success in organizing and in securing legal rights was mutually reinforcing.⁴⁴ So the negative effects from legal hostility to union rights could be offset, according to the regression, if this stunted the growth of unions. But the negative effects of complete accommodation have no similar offset.

Panel C shows representative results when the dependent variable is based on state fixed effects rather than state average residuals. Specifically, the dependent variable in each of these regressions is the trend of the dummy variable coefficients for a state as estimated by b'_j in equation (2'). The key unionization and state spending effects tend to be a bit bigger and a bit stronger here than their counterparts in the preceding panels. This suggests that any bias from use of state average residuals is toward understatement of these effects. The other notable results in panel

⁴¹ Frederick Flyer & Sherwin Rosen, *The New Economics of Teachers and Education* (Working Paper No. 94, Univ. Chicago, Center for the Study of the Economy and the State 1994).

⁴² Growing female labor force participation may also signal broader job opportunities for women, thereby lowering the quality of (predominantly female) teachers.

⁴³ Ninety percent of the variance of LAW is contributed by post-1960 changes.

⁴⁴ The simple correlations of LAW and union density are .62 for NEA, .34 for AFT, and .75 for the sum of the two. All are significant.

C are the suggestive negative spending effects and the unimportance of growing female labor force participation.⁴⁵

Several other variants of the basic regression were tried in attempts to tease out nonlinearities in expenditure effects and in the long run adjustment process. These sometimes produced interesting results,⁴⁶ but none changed the central results. So I spare the reader from the details.

I also found that the key results were insensitive to alternative weighting schemes for the dependent variable which attempted to correct for overrepresentation of blacks and for intrastate population shifts.⁴⁷

B. The Distribution of Outcomes

A comparison of the magnitudes of some of the regression coefficients in Tables 5 and 6 has implications for the distribution of educational outcomes. For example, these tables show that spending centralization has hurt the performance of both the wealthier college-bound and the poorer non-college-bound students. But did centralization contribute to a narrowing difference between them? To answer, suppose the state share of spending rose 20 percentage points between 1960 and 1980 (about 1 standard deviation above the mean increase). I use columns (1) and (3) of Table 5 and regression A2 of Table 6 to estimate the impact of this change on the two groups and divide by the respective sample standard

⁴⁵ This triad of results—stronger political economy effects, more consistently negative spending effects, and essentially no role for female labor force participation—characterizes every replication of panel A or B regressions on the trend of state dummy coefficients, including those not shown in panel C.

The correlation between the alternative performance trends is high but not overwhelming: .88 for 1976–91 and .75 for 1971–91.

⁴⁶ I added an interaction of expenditure growth with the initial level of expenditures to allow for diminishing returns. This yielded significantly positive expenditure growth effects where the initial level was around 1½ SD below the mean (and significantly negative effects at correspondingly above average initial levels). Moreover, the implied diminishing returns to spending growth are substantial: the difference in performance from accelerating spending growth by 1 SD in a low rather than a high spending state is around 1 SD.

Adding the initial level of performance yields a significantly negative coefficient. This means that performance differences across states narrowed over the sample period.

⁴⁷ Two alternative dependent variables were constructed, each based on fixed weighted averages of zip code average residuals (ZAR) rather than the statewide average residual. One alternative weighted each ZAR in each year by 1980 zip code population. The other refined this further by computing black and white ZARs in each year and using the 1980 racial composition of the zip code's population as weights to compute the ZAR. These ZAR were then averaged as above, using 1980 zip code population weights.

The proximate reason that the key results were robust to these alternatives is that the weighting schemes essentially reproduced ($r > .95$) the original dependent variable. Thus that variable is unaffected either by shifts of population to poorer or better schools within a state or any biases stemming from the overrepresentation of blacks in the AFQT sample.

deviations of the performance measure to adjust for their different scales. The result is substantially the same decline (.39 S.D. for the SAT/ACT sample and .44 S.D. for the AFQT sample) for each group.⁴⁸

A similar comparison can be done for the union effects. Consider a typical state which achieved 25 percent unionization by 1967 and 50 percent by 1982. If all unionized teachers belong to the AFT, the result would be similar declines in SAT/ACT and AFQT scores (1.4 SD and 1.2 SD, respectively). For an all NEA state, the decline would be larger for the AFQT sample (.73 SD vs. .05 SD). Since the NEA is around three times as large as the AFT, the last calculation should perhaps be given greatest weight. The safest conclusion is that unionization, like centralization of school finance, has not contributed to narrowing the difference between the top and bottom of the achievement distribution.

The data also allow analysis of differences within the AFQT sample. This is pursued in Table 7, where the dependent variables are based on the share of the state's test takers with residuals above the 25th percentile of the national distribution ($SHARE > 25$). A high value implies success in avoiding very bad outcomes.

The first panel of Table 7 uses the change in $SHARE > 25$ as the dependent variable and yields sign and significance patterns similar to those in Table 6. This means that changes in performance are not limited to one part of the distribution.⁴⁹ The second panel focuses more sharply on the lower tail by removing effects of changes in the mean. The dependent variable here is the trend in $SHARE > 25$ minus what would be expected from the trend in the state average.⁵⁰ The only robust result is

⁴⁸ This broad result is consistent with evidence from California that equalized expenditures financed by the state following the *Serrano* decision did not reduce achievement differences between rich and poor school districts in that state. See Thomas A. Downes, *Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case*, 45 Nat'l Tax J. 405 (1992).

⁴⁹ The simple correlation between the change in $SHARE > 25$ and the change in the average residual is .96 for 1971-91 and .68 for 1976-91.

⁵⁰ For example, suppose the distribution of residuals in the nation and a state is uniform over $(-50, 50)$, so the 25th percentile score is -25 . If the state distribution shifts by $+1$ to $(-49, 51)$, 76 percent of its sample would exceed -25 . So $SHARE > 25$ would be expected to move point for point with the average if all students experience the average change. The dependent variable in panel B just nets out this expected change from the actual change in $SHARE > 25$.

The actual distribution of residuals is not uniform. So the expected change is derived empirically from the national distribution. Each year, I estimate the derivative, $\Delta SHARE > 25 / \Delta AVERAGE$, for this distribution as follows: (1) locate R_{25} , the 25th percentile residual; (2) find the two percentiles associated with $R_{25} \pm 1$; (3) estimate the derivative as the difference between the two percentiles divided by 2; (4) average the derivative over the sample period; and (5) multiply the trend in the state average by this national average derivative to get the expected effect on the trend of $SHARE > 25$ from the trend in the state average. The result of step 4 is 1.65 (SD = .12) for 1971-91 and 1.61 (SD = .09) for

that increased female labor force participation seems to have reduced the gap between the worst and the average performances. The impression in panel A that "rising tides lift all boats" seems strengthened by the lack of consistent results in panel B for the other, more immediately relevant variables.

V. INTRASTATE EFFECTS

Do the forces so far uncovered operate differently in, say, rural areas than in cities and suburbs? The answer is of more than intrinsic interest. It may help illuminate what has so far been a "black box": the mechanism by which these forces exert their effects, specifically, whether they work directly on the operation of schools or more broadly on the policy-making process (where state governments are of primary importance). For example, teacher unions influence what goes on within schools by negotiating pay differentials, promotion rules, and so on. They have also become powerful lobbying organizations in state capitals. Which of the two roles is plausibly responsible for the negative correlation of unionization and performance? The uneven growth of teacher unions may provide a clue. Teacher unions developed first in large cities and still tend to be overrepresented in and near them. So, for example, larger unionization effects on urban performance would suggest predominance of the unions' "within-school" role. Centralization of school finance also has both specific, partly geographical, and broader effects. It seeks to allocate resources away from wealthy suburbs, and it arguably strengthens the hand of state policy makers (and those able to influence them).

Information on each test taker's residence was used to construct four state subsamples: nonmetropolitan ("rural"), small metropolitan (standard metropolitan statistical areas [SMSAs] < 1 million), and large metropolitan, which is subdivided into central city and suburban populations. Some rural and small SMSA residents are found in every state, but fewer than half the states are in the large SMSA subsamples.

An interesting by-product of this classification was the absence of any divergence among the four groups at the national aggregate level. Annual national averages of the within group AFQT residuals stay within a narrow bound of about ± 1 point over the entire 1971–91 period. Thus, for example, concerns that big city school systems have done especially badly seem misplaced at least for this period.

The meaningful patterns in these data occur at the state rather than the

1976–91. The dependent variable in panel B is therefore the trend in $\text{SHARE} > 25$ less 1.65, or 1.61 times the trend in the average residual.

TABLE 7
TREND IN SHARE OF AFQT TAKERS ABOVE 25TH PERCENTILE, 1971-91 AND 1976-91

REGRESSION	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION SHARE		LAW	CHANGE IN FEMALE LABOR FORCE PARTICIPATION		\bar{R}^2/SEE
			NEA	AFT		1960-80	1980-90	
A. Annual change in share:								
1. 1971-91:								
Coefficient	283	-.46	-33	-55	1.8	-87	-17	.38
t	.5	2.5	2.8	3.5	3.8	1.3	.2	.142
2. 1976-91:								
Coefficient	201	-.44	-16	-40	1.6	83	194	.09
t	.2	1.3	.9	1.6	2.1	.8	1.2	.231
B. Annual change: residual of share vs. mean:								
1. 1971-91:								
Coefficient	-189	-.06	-7	-12	.23	57	-6	.26
t	1.1	1.1	1.8	2.6	1.6	2.8	.2	.044
2. 1976-91:								
Coefficient	179	.06	31	15	.08	171	40	.44
t	.4	.3	2.8	1.0	.2	2.8	.4	.140

NOTE.—Dependent variable is estimated annual change in percentage of state's Armed Forces Qualifying Test (AFQT) takers with residual scores above the 25th percentile of national distribution. In panel B, the effect of changes in the state's mean on the change in share is removed (see text and note 49 for elaboration). Annual change is estimated from regression on time index. See Tables 5 and 6 for an explanation and sources of independent variables. Coefficients are times 100. SEE = standard error of estimate.

TABLE 8
CORRELATION OF AFQT TRENDS BY RESIDENCE,
1971-91 AND 1976-91

	Rural	Small SMSA	Large SMSA: Central City
Small SMSA:			
1971-91	.70		
1976-91	.77		
Large SMSA:			
Central city:			
1971-91	.77	.61	
1976-91	.81	.71	
Suburbs:			
1971-91	.83	.71	.87
1976-91	.83	.69	.81

NOTE.—Data are simple correlations across states between the annual change in average Armed Forces Qualifying Test (AFQT) residuals in the indicated substate areas. Large standard metropolitan statistical areas (SMSAs) are Class A (generally over 1 million population) metropolitan areas as of 1980. Sample sizes: 19 states for central city, large SMSA; 22 for their suburbs; 48 for rural and small SMSA. All correlation coefficients are significant at $P(.05)$.

national level. I computed trends in the average AFQT residual for each available subsample in each state. Table 8 contains the simple correlations of these trends across the state subsamples. Without exception, these correlations are significantly positive and substantial in magnitude. They tend to cluster in the middle of a range from $+.6$ to $+.9$. These results mean that students in all parts of a state tend to share a common performance trend. These shared performance trends hint at a shared impact of state education policy. However, the high correlations might reflect only broader regional similarities not confined to a state. For example, they could result from a tendency for students in the North to perform more like other Northerners than like Southerners and vice versa. But further tests clearly reject this sort of possibility.⁵¹ The high

⁵¹ One test involved pairing each observation (for example, rural Wisconsin) with its counterpart in a contiguous state (rural Minnesota). This yields four pairings where the populations are similar (both rural, for example) but live in different states. For 1971-91 data, the average (range) of the four correlations was $.31$ ($0, .51$) compared to the $.6-.9$ range in Table 8. So, at least with respect to school performance, nearby rural (or urban) areas have less in common with each other than with different areas in the same state.

When different areas are paired, as in Table 8, but across state lines (for example, rural Wisconsin and suburban Minnesota), the average (range) of the 12 correlations corresponding to those in Table 8 is $.15$ ($-.30, .56$), again, well below the within-state correlations.

I also estimated regressions of the form: $y_0 = a + by_N + Cx_0$, where the y , x are trends in substate regions and O, N denotes own and nearby state, respectively. The results when x is defined as the state's rural trend are typical. For 1971-91 data, the y_0 , x_0 partial correlations are $(.71, .74, .79)$, where y_0 = small SMSA, large SMSA central city, and

positive correlations in Table 8 are indeed being driven almost exclusively by common intrastate trends. In this sense, students in the Bronx have more in common with those in a rural New York hamlet than with students in Newark or Philadelphia.

The question of what lies behind the strong link between apparently dissimilar students from the same state is explored in Table 9. Here the same regression is run four times. Each dependent variable is the performance trend for a specific part of a state (for example, rural areas). The independent variables are exactly the same as in the second regression in Table 6; they are all statewide averages.⁵² So, in effect, Table 9 asks, "How are students in rural areas of a state affected by union density in the whole state, by the share of the state's total spending on schools provided by state government, and so forth?" The broad answer emerging from Table 9 is "about the same as students in other parts of the state." That is, a glance down the columns of Table 9 reveals broadly similar signs and magnitudes of the key unionization and spending coefficients. This similarity adds weight to the notion that the main channel through which these variables operate is through their effect on state policy, not through direct effects on the operation of schools. The unionization effects in the nonmetropolitan regressions probably make the point most clearly. They are virtually the same as for the full state sample in Table 6 even though these regressions typically contain only 1/4 of the full sample and even though the statewide union densities grossly mismeasure rural unionization. The AFT, for example, has virtually no rural membership outside New York. One way that unions can affect the performance of unorganized rural schools is by influencing state policies which apply to school systems in all parts of a state.

Some indication of the importance of the link between a state's political economy and its policy can be gleaned by combining the data in Tables 8 and 9. Specifically, we can ask, "How much of the correlation between

large SMSA suburbs, respectively. These are essentially the same as the corresponding simple correlations in Table 8. The associated cross-state (y_0, y_N) partial correlations were (- .10, .44, .29). Clearly, the common within-state component is dominating these regressions.

Finally, I replaced y_N in the above regression with a vector of regional dummies for the rural, small SMSA pair (the only one with enough degrees of freedom to yield meaningful results). The partial correlation of the within-state variables here is .67, again, roughly equal to the simple correlation (.70) in Table 8. None of the regional dummies was significant, though the hypothesis that all are zero could be rejected at about 5 percent.

The safest conclusion from these exercises is that there may be some common regional or locational elements in these data. But these are overshadowed by the within state commonality.

⁵² It would, of course, be interesting to add substate detail on the key unionization and state spending variables, but these are unavailable to me.

TABLE 9

AFQT PERFORMANCE TRENDS WITHIN AND OUTSIDE SMSAs, 1971-91 AND 1976-91

PERIOD AND REGION	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION SHARE		LAW	CHANGE IN FEMALE LABOR FORCE PARTICIPATION		\bar{R}^2 /SEE
			NEA	AFT		1960-80	1980-90	
A. 1971-91:								
1. Nonmetropolitan:								
Coefficient	431	-.16	-17	-31	.74	-85	36	.34
<i>t</i>	1.2	1.4	2.3	3.2	2.5	2.1	.6	.088
2. Small SMSA:								
Coefficient	-.50	-.20	-.22	-.28	1.06	-.75	63	.16
<i>t</i>	.1	1.3	2.1	2.1	2.6	1.3	.7	.123
3. Central city of large SMSA:								
Coefficient	755	-.72	5.7	-4.6	1.31	-.76	-338	.49
<i>t</i>	.6	2.2	.3	.2	2.4	.6	1.8	.115
4. Suburbs of large SMSA:								
Coefficient	843	-.45	-11	-29	1.60	-.39	-321	.48
<i>t</i>	1.0	1.7	.7	1.8	3.3	.5	2.3	.101
B. 1976-91:								
1. Nonmetropolitan:								
Coefficient	443	-.26	-30	-31	1.15	-41	87	.26
<i>t</i>	1.0	1.5	3.2	2.4	2.9	.8	1.0	.119
2. Small SMSA:								
Coefficient	-477	-.36	-.29	-.32	.72	-.78	167	.15
<i>t</i>	.8	1.5	2.4	1.9	1.4	1.1	1.5	.154
3. Central city of large SMSA:								
Coefficient	322	-.42	-.37	-.37	1.54	20	-143	.15
<i>t</i>	.2	.9	1.4	1.5	2.1	.1	.6	.156
4. Suburbs of large SMSA:								
Coefficient	-275	-.19	-.47	-.58	1.64	-.82	-.98	.44
<i>t</i>	.3	.6	2.6	3.3	3.0	.9	.6	.115

NOTE.—See Tables 5 and 6 for a description of independent variables. Dependent variable in each regression is the average annual change in average Armed Forces Qualifying Test (AFQT) residuals for test-takers residing in the indicated portion of the state. Large standard metropolitan statistical areas (SMSAs) are those with population over 1 million. Sample sizes: 48 for nonmetropolitan and small SMSA; 19 for large SMSA, central city; and 22 for suburbs of the large SMSAs. Coefficients are times 100.

the performance of, say, urban and rural students in the same state (in Table 8) is attributable to the similar response of these trends to the same statewide forces (in Table 9)?” To see how this question can be answered, it is best to work through a specific case. The first entry in Table 8 says that the correlation (r) between 1971–91 performance trends (Y) of students in rural areas and in small SMSAs is $+.70$ ($r(Y_R, Y_S) = +.70$). The first two regressions in Table 9 use an identical set of statewide variables to “explain” these rural and small SMSA performance trends. Each of the two regressions generates fitted values (\hat{Y}_R, \hat{Y}_S) and residuals (U_R, U_S) for each state as well as the conventional R^2 “goodness-of-fit” measure.

Because the Y 's respond similarly to unionization, state spending, and so on, the fitted values will be positively correlated ($r(\hat{Y}_R, \hat{Y}_S) > 0$), and this correlation can account for some of the $.70$ correlation between the two performance trends. Just how much can be obtained from the first term on the right-hand side of the relation

$$r(Y_R, Y_S) = r(\hat{Y}_R, \hat{Y}_S) \cdot R_R \cdot R_S + r(U_R, U_S) \cdot [(1 - R_R^2)(1 - R_S^2)]^{1/2} \quad (4)$$

This says that the share of the overall correlation coming from a similar response to the same forces depends on (a) how similar the response is (given by $r(\hat{Y}_R, \hat{Y}_S)$) and (b) how well these forces explain both rural and small SMSA performance trends (as given by the multiple correlation coefficients, R_R and R_S). In this case, the answer is that 47 percent of the $.70$ correlation is explained by the similar response to unionization, and so on. The remaining 53 percent is due to the fact that the residuals from the first two regressions in Table 9 are correlated and to the less-than-perfect fit of these regressions.

When the same method is applied to all 12 pairings in Table 8, the average (range) of the first term on the right-hand side of (4) is 60 (35, 81) percent of the overall correlation of performance trends. This means that something over half of the typical correlation of within-state performance trends is coming from similar responses to variables like unionization and centralized school finance. So this exercise confirms the importance of the particular forces I have focused on. But there is a broader message. It is that a state's political economy plays a large, even dominant role in producing the striking similarity of student performance trends within a typical state.

VI. EXTENSIONS

A. *Nonparametric Measures*

So far, I have analyzed trends in average student performance. One risk with using estimates of trends is that errors in measurement at the beginning or end will be misconstrued as a long-run “trend.”

To hedge against this risk and to check the robustness of the results, I constructed two nonparametric state performance measures which use zip codes as the basic unit of analysis. The sample period was divided into 10 (1971–91) or 8 (1976–91) 2-year subperiods.⁵³

For zip codes with at least 10 observations in adjacent subperiods, I calculated the change in the zip code's average AFQT residual between subperiods. Then, a binary variable, IMP, was defined as +1 if the change is positive, 0 otherwise. The measures derived from IMP are

1. AVGIMP = percentage of a state's zip codes with improving average AFQT residuals over adjacent subperiods averaged over all subperiods (= the average value of IMP over the sample period) and
2. AVGIMP > .5 = percentage of a state's zip codes improving in more than half the subperiods (that is, zip codes where IMP = 1 at least 5 (4) times from 1971 to 1991 (1976–91)).⁵⁴

Neither measure is especially sensitive to the timing or magnitude of measurement errors. Both provide some geographic breadth in that more or less arbitrarily defined chunks of a state (zip codes, with varying populations) rather than individuals are given equal weight. And the AVGIMP > .5 measure adds a persistence dimension by discarding IMP which is not sustained.

These features are not, of course, costless: the magnitude of improvement as well as its existence matters; we would prefer, all else the same, that the more populous zip codes improve; stretching improvement out is not obviously better than having it occur all at once. In sum, the nonparametric measures hide information about how much improvement occurred.

Table 10 replicates familiar regressions on the nonparametric measures and shows how the various performance measures are related. Two broad conclusions emerge: (1) All the performance measures are highly correlated. That is, improved school performance in a state tends to be pervasive. It is not limited to one part of the state, as indeed was already evident in the preceding substate analysis. Nor is improvement confined to particular subperiods. Thus, end-period effects cannot have been driving previous results. (2) The sign patterns of the coefficients in Table 10 are the same as in Table 6. So, the same forces driving average performance work in much the same way on the nonparametric measures. But the effects here tend to be weaker statistically. That could mean either

⁵³ Because 1971–91 has an odd number of years, 1971–73 were grouped in one subperiod.

⁵⁴ Based on zip codes that are in the sample in all subperiods.

TABLE 10
NONPARAMETRIC MEASURES OF CHANGE IN PERFORMANCE, 1971-91

PERIOD AND DEPENDENT VARIABLE	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION SHARE		LAW	CHANGE FEMALE LFP		\bar{R}^2 /SEE	CORRELATIONS WITH:	
			NEA	AFT		1960-80	1980-90		State Average Trend	AVGIMP
A. 1971-91:										
1. AVGIMP:*										
Coefficient	99	-.025	-3.3	-3.3	.16	-15	8	.22	.85	...
t	1.2	.9	2.0	1.5	2.4	1.5	.5	2.03		
2. AVGIMP > .5:†										
Coefficient	494	-.089	-14	-10	.57	-26	6	.14	.75	.86
t	1.4	.8	1.9	1.0	1.9	.6	.1	8.86		
B. 1976-91:										
1. AVGIMP:*										
Coefficient	28	-.027	-5.2	-4.2	.18	-16	20	.15	.84	...
t	.3	.7	2.6	1.5	2.1	1.4	1.1	2.54		
2. AVGIMP > .5:†										
Coefficient	21	-.107	-17	-15	.45	-50	16	.15	.83	.88
t	.6	.8	2.4	1.5	1.5	1.2	.2	9.04		

NOTE.—See Tables 5 and 6 for independent variables. Coefficients are times 100.

* Percentage of zip codes in a state with increased average AFQT residuals from one 2-year subperiod to next, averaged over all subperiods.

† Percentage of zip codes with increased average AFQT residuals in more than half the subperiods.

that the precision of the previously estimated unionization and school finance effects was overstated or that these variables affect the size as well as the direction of performance trends. The uncertainty here reflects the inherent trade-offs in the two kinds of performance measures.

B. Black-White Differences

Detailed analysis of the narrowing of black-white differences (see Table 3) deserves a separate inquiry. Here, I address a more limited question: Have the political forces affecting overall performance plausibly contributed to the narrowing racial gap? In part, sheer timing motivates the question. The racial gap narrowed in the wake of growing unionization and centralization of school finance. The negative findings on the effects of these variables on lower-tail performance (Table 7) provide further motivation. The narrowed racial gap is perhaps the most dramatic specific change in low-end performance in our sample period. And the history of the period suggests that racial differences were politically relevant. The first decade of the period saw the working out of the legal and political pressure for desegregation which followed the Civil Rights Act of 1964. All this implies that political pressures for educational equality also had a specific racial focus.

Table 11 contains regressions of the 1976–81 trend in black-white differences in AFQT residuals on the variables used previously⁵⁵ plus measures of the change in school integration. The latter are based on the number of black students attending integrated schools (those with less than half minority enrollment) in 1968 and 1980. Dividing this number by total black enrollment in a state gives the probability that a black attends an integrated school. Dividing the same number by white enrollment gives a noisy proxy for the probability that a white student attends an integrated school.⁵⁶ The 1968–80 change in these “integration probabilities” is used in the regressions. These changes are positive in most states but tend to be much higher in the South than North. This reflects the persistence of de facto segregation in the North and its substantial erosion in the South.

The opposite signs on the coefficients of the integration variables in

⁵⁵ The female labor force participation change variables are dropped, because there is no reason for expecting this overall measure to affect the difference between two performance trends. (Indeed, the variables have insignificant coefficients when added to the regressions in the table.) The sample period is dictated by lack of accurate pre-1976 black-white differences.

⁵⁶ By the definition, an integrated school has $k > 1$ white students per black student. So, the true probability is k times my proxy, and k is a variable.

TABLE 11
BLACK-WHITE PERFORMANCE TREND DIFFERENCES, 1976-91

REGRESSION	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION SHARE		LAW	INTEGRATION RATIO		STATE AVERAGE TREND*	\bar{R}^2/SEE
			NEA	AFT		Blacks	Whites		
1. Coefficient	1,204	-0	12	16	.54	38	-25	.38	
†	2.0	.0	1.1	1.1	1.2	2.4	.7	.106	
2. Coefficient	1,113	-.05	10	14	.59	37†		.40	
†	2.0	.2	1.1	1.0	1.3	2.5		.104	
3. Coefficient	1,314	.20	21	29	.06	41†		.56	
†	2.8	1.1	2.4	2.3	.1	3.1		.090	

NOTE.—Dependent variable is the difference between annual change in a state's black and white average AFQT residual for 1976-91. Integration ratios = 1968-80 change in ratio of black students attending integrated schools to all black and white students, respectively. See note to Table 6 for definition of other independent variables. All coefficients are times 100.

* Difference between black and white ratio.

† Dependent variable in panel B, Table 6.

the first regression suggest a redistributive impact of integration, though the negative coefficient for whites is imprecise. The second regression uses the roughly equal coefficients to simplify by netting the white from the black integration variable.⁵⁷ Both regressions show significantly positive effects of total expenditure growth and hint at positive unionization effects. These union effects are under half the corresponding negative effects on total performance in panel B of Table 6. So they mean that unionization has smaller negative effects on blacks than whites.

A similar point applies to the third regression of Table 11, which reveals that blacks disproportionately benefited from good overall school system performance: each extra point of overall performance raises black scores by over 1.44 points.⁵⁸ The unionization coefficients in this regression are double those in the preceding regression. But these are now partial, not total, effects. To evaluate the full union effect on the black-white differential (BWD), any negative union effects on the average have to be included because the differential is affected by the average. That is, we want

$$\frac{dBWD}{dUNION} = \frac{\partial BWD}{\partial UNION} + \frac{\partial BWD}{\partial AVERAGE} \cdot \frac{dAverage}{dUnion}.$$

Estimating this, using the coefficients from regression B2 of Table 6 for $dAverage/dUnion$ and the appropriate coefficients from the third regression in Table 11 as the partial derivatives, yields total union effects of .09 and .15 for the NEA and AFT, respectively, or just about the same as in the previous regression.

The qualifications understood, the regressions in Table 11 do suggest some role for unionization in contributing to narrower black-white differentials. The lack of any role for state spending might be surprising in light of its ostensibly egalitarian purpose and the significant total spending effect. Apparently, any beneficial resource allocation effects of state spending are offset by other negative effects.⁵⁹

⁵⁷ Substantively identical results are obtained by dropping the white variable.

⁵⁸ The .44 coefficient understates the total effect for two reasons: (1) Black performance is part of the average performance, so there are feedback effects from any initial increase in black performance. These add around .05 points given that 25 percent of the sample is black. (2) Statistically, errors in the average are dominated by errors in white performance, and this imparts a negative bias to the coefficient. For example, a +1 error in white performance raises the average by .75 and lowers the black-white difference by 1. If black and white performance were purely random with the same variance, the coefficient on the average would be -.8.

⁵⁹ Or state spending does not shift resources toward blacks.

C. *Competition from Private Schools*

From the end of World War II to date, the nonpublic school share of enrollment has fluctuated in a narrow range between 10 and 15 percent. However, over the last 30 years the regional distribution of this share has changed markedly. Up to the mid-1960s, Catholic schools, which are concentrated in the Northeast and Midwest, accounted for around 90 percent of private school enrollment. That share has declined to under half today.⁶⁰ The offsetting growth of non-Catholic private schools was concentrated in the South following the integration of Southern public schools.⁶¹ These divergent regional patterns raise the question of whether the decline in private schools in the North and their growth in the South had corresponding effects on public school performance trends. My data suggest not: the coefficient of the 1966–89 change in a state's private school share is insignificant when added to the basic regression. Attempts to account for potential feedback from public school performance to the private school share yielded the same negative result.⁶² All the previous results were essentially unchanged by these experiments.

D. *The Social Background*

Schools operate in a social as well as political context. Their need to grapple with growing social adversity is sometimes thought to be a primary source of declining performance. Two kinds of evidence support this view. First, the educational production function literature clearly shows the importance of family background. For example, superior performance of children from stable two-parent households is commonly found. Second, several important social indicators began deteriorating significantly around the mid-1960s when school performance began declining. These include divorce rates, the share of households headed by

⁶⁰ From 1960 to 1991, Catholic school enrollment fell by more than half, from 5.3 million to 2.4 million. At the same time, non-Catholic private school enrollment more than quadrupled, from 600,000 to 2.8 million.

⁶¹ The simple correlation between the 1966–89 change in private school share and the previously described white integration ratio is .6.

⁶² A two-stage least squares model was estimated in which the white integration ratio and the 1966 private (that is, Catholic) school share were the exogenous determinants of the change in the private school share. No connection between the change in private share and the change in performance is discernible in either structural equation. The initial Catholic school share and exposure of whites to integration together account for around 80 percent of the interstate variance in the change in private school share.

The only hint of an effect from private school competition occurs when the change in private share is added to regressions with the 1976–91 trend in state fixed effects as the dependent variable. The coefficient is positive and significant at $P(.08)$. No similar result is found for the 1971–91 trend in state fixed effects.

one parent, and crime rates. And some of these—for example, crime and divorce—peaked or flattened around 1980, just when the school performance decline ended.

The political context also has a broader dimension that I have so far ignored. The growth of teacher unions and the push for centralizing school finance began in a climate generally favoring “liberal” social policies like the War on Poverty, Medicare, and widespread unionization of public employees among others. Moreover, growth of teacher unions and the state spending share went further in more liberal states.⁶³

This history arouses suspicion that my results, based on a narrow focus on the politics of education, may be obscuring effects of broader social and political forces. The suspicion, however, receives little statistical support. When I added two measures of social disintegration—the growth of crime and of children from broken homes⁶⁴—and a measure of the liberalism of a state’s voters⁶⁵ to the regressions in Table 6, none yielded significant effects no matter what permutations were used,⁶⁶ and all the other results remained intact. There is, however, a significant effect in the expected direction (partial correlation = $-.45$) between the residual measure of left-tail performance (see Table 7, panel B) and the broken-home variable. These results suggest some appropriately tentative conclusions. One is that the systemic effects of social disintegration may be overrated. Broken homes are indeed not good for school performance. But the result on left-tail performance implies that broken homes mainly hurt those who live in them, not the entire school system.⁶⁷ The results also suggest that the details of education politics really matter; they are not proxies for some more general aspect of a state’s politics. This point

⁶³ A commonly used indicator of a state’s bedrock liberalism is George McGovern’s share of the vote in the 1972 presidential election. This is correlated .27, .47, and .39 with the state spending, NEA, and AFT variables, respectively. All are significant at $P < .10$.

⁶⁴ More specifically, (1) the average of the change in logs of the murder, rape, assault, burglary, robbery, and auto theft rates and (2) the change in share of children under 18 not living with both parents. The changes are measured from 1960–80 to allow for lags in response.

⁶⁵ The aforementioned McGovern vote share.

⁶⁶ For example adding all three yielded partial correlations of $-.14$, $-.08$, and $+.20$ (1971–91 regression) or $-.09$, $-.03$, and $+.13$ (1976–91 regression) for crime, broken homes, and liberalism, respectively. Similar results are obtained from regressions with trends in state fixed effects as the dependent variable except that the negative effect of the crime variable becomes significant at $P(.07)$ in the 1971–91 regression.

⁶⁷ Recall that the performance measure is purged of at least some student-specific disadvantages from broken homes, poverty, residential segregation, and so on. See discussion surrounding Table 2 and Appendix B. If the student-specific effect of these variables has indeed been removed, only system-wide effects (for example, the need to divert resources to remedial work) would remain to be revealed by the tests in this section.

is clear from previous results: granting legal rights to teacher unions helps school performance, while actual organization hurts. Yet liberal states tend to have more of both.⁶⁸

E. A Comprehensive Measure of School Performance

One motive to this research was to gain insight into the forces impinging on public school performance generally, not just the performance of college-bound or non-college-bound students. That insight can be sharpened by combining the available data on the two groups.

I calculated a comprehensive index of the change in state school performance from the two (1971–91 and 1976–91) measures for non-college-bound students analyzed in here and the two college-bound student measures (for 1972–81 and 1981–89) used in my 1993 paper. Each of the four was converted to a standardized (0, 1) normal deviate. The four deviates were averaged, and the resulting average was restandardized to yield a comprehensive measure which gives equal weight to each group and to each time period. It is an index (U.S. average = 0) characterizing the relative change in state school performance over the 2 decades of available data.

Table 12 shows that the comprehensive measure as well as the two components respond similarly to the same set of variables. However, the coefficients in the regression on the comprehensive measure seem closer to those for the non-college-bound than the college-bound regression. Indeed, only the AFT unionization coefficient attains significance in the college-bound regression. Nevertheless, there is a substantial and significant positive correlation (.65) between the fitted values of the regressions on the two components. This correlation accounts for all of the weak (.22) positive correlation between the two components. Moreover, the data fail to reject the restriction, implicit in the first regression, that the coefficients on the components are the same for the two groups. Thus these two more or less independent components are telling much the same story. And, by averaging over them, the comprehensive measure puts that story in sharper focus: complete unionization lowers overall performance by 1.6 to over 3 standard deviations; moving from -2 to

⁶⁸ The following correlation matrix provides a concise summary:

	AFT	NEA	LAW
NEA share 1982	-.12
LAW	.34	.62	. . .
McGovern share	.39	.47	.59

By 1982, the NEA and AFT had little geographic overlap, but both tended to gravitate toward politically hospitable states.

TABLE 12
OVERALL SCHOOL PERFORMANCE TRENDS AND COMPONENTS, 1971-91

REGRESSION	EXPENDITURE GROWTH	CHANGE STATE SHARE	UNION SHARE		LAW	CHANGE IN FEMALE LABOR FORCE PARTICIPATION		\bar{R}^2/SEE
			NEA	AFT		1960-80	1980-90	
I. Overall index:								
Coefficient	6.9	-.02	-1.6	-3.4	.06	-8.6	.1	.42
<i>t</i>	.2	2.1	2.6	4.1	2.3	2.4	.0	.76
II. Components:								
A. SAT/ACT:								
Coefficient	2.1	-.01	-.7	-2.9	.01	-7.9	-3.3	.21
<i>t</i>	.1	.7	.9	3.0	.3	1.9	.5	.89
B. AFQT:								
Coefficient	12.5	-.02	-1.8	-2.4	.08	-5.5	3.5	.32
<i>t</i>	.4	2.2	2.7	2.7	3.0	1.4	.6	.82

NOTE.—All dependent variables are standardized (0, 1) normal deviates. Overall index is standardized average of SAT/ACT and AFQT indexes. SAT/ACT is standardized average of standardized trends for 1972-81 and 1981-89. AFQT is standardized average of standardized trends in 1971-91 and 1976-91.

+2 SD on the state spending share (+55 points) costs 1.1 SD, and a similar shift in the bargaining rights index adds 1.6 SD. The regression on the overall index also confirms the absence of total spending effects and the probable negative effect of growing female labor force participation (a 1.4 SD performance loss for a -2 to $+2$ SD change).⁶⁹

VII. SUMMARY AND CONCLUSIONS

While there are subtle differences in timing, two results consistently cut through the data: teacher unionization and increased reliance on state government funding of schools are associated with declining student achievement. The negative effects seem to show up first in the performance of college-bound students then spread to the non-college-bound. My results also yield a clue about the path by which these negative effects are transmitted. They imply that the primary effects are exerted in state capitals on the policy-making process rather than inside schools.

Two kinds of evidence suggest the primacy of these broad policy effects. The most dramatic, perhaps, is my finding of substantial within-state performance commonalities across disparate groups. As just one example of a pervasive tendency, consider the 19 states with large (>1 million) SMSAs, and compare students in the rural areas of those states with students in the inner city of the SMSAs. In 16 of the 19 states—a proportion too large to have arisen by chance—the two seemingly dissimilar groups share either above- or below-average performance trends. Moreover, I found that a major source of such patterns was common responses to the key variables. For example, the negative unionization effect is as strong in rural areas as in cities, though unions came later and remain weaker in rural areas. One plausible path for such common effects would be through state educational policy, which both groups share.

Further evidence consistent with this interpretation is found in the performance of two groups with below average performance: students in the lowest quartile and black students. Neither group did better in states which most centralized school financing, though one ostensible purpose of centralization was to equalize educational resources. At the same time, I find that for blacks (but for no other group studied) total system resources do matter. Thus, the failure of spending centralization to help blacks implies either a failure to equalize resources or confirmation of the negative effects of centralization.

My results imply that education outcomes are affected importantly by state policies which are shaped by interest groups organized at the state

⁶⁹ Adding the business pressure indexes to any of the regressions proved unavailing.

level. However, the mechanisms providing these linkages remain to be uncovered by future research. The methodological implication of my results is that economists may have overinvested in the study of educational production—the direct link between inputs and outputs. The political context in which the inputs are used deserves more attention than it has received.

My results also have implications for some of the broad trends in student achievement. They suggest that the rise of teacher unions and the shift of financial responsibility to the states which began in the early 1960s set off forces which contributed to the decline in achievement which began shortly thereafter. Similarly, the abatement of those forces may have contributed to the end of declining performance around 1980. (The growth of unionization slowed markedly in the 1980s and the state spending share stopped rising.) My results imply that these forces operated more or less equally on college-bound and non-college-bound students. So, I make no progress on understanding the recent narrowing of the performance gap between the two. There are, however, insights into the narrowing racial differential. My results suggest that the approximate tripling of real per-student spending since 1960, desegregation of schools in the South, and, to some extent, the growth of teacher unions each contributed. But recent trends suggest more modest contributions from all these sources in the future.

Finally, some negative results deserve emphasis. The notion that a broader process of social disintegration is responsible for the decline of public education receives no support from my data. Two measures of social change—growth of crime rates and of children living in single-parent households—had scant marginal explanatory power. I was also able to dismiss the possibility that the unionization and spending effects are proxies for broader political forces. These two forces did tend to make more headway in “liberal” states. But adding a direct measure of a state’s political liberalism to the regressions proved uninformative.

These negative results add weight to the conclusion that the politics of public education has important effects on the way public schools perform.

APPENDIX A

AFQT SCORES AND BACKGROUND VARIABLES

The Defense Manpower Data Center (DMDC) provided an extract from their COHORT database. For every individual inducted from 1971 to 1975 and every applicant to the military from 1976 to 1991, I have the individual’s

1. AFQT percentile score (1–99);
2. zip code of residence (plus state and county IDs);
3. age;

4. sex;
5. race/ethnic category: Black, White, Hispanic, and Other; Hispanic is a subset of White;
6. marital status (not used, married applicants are rare); and
7. years of education (in 12 categories).

The DMDC also provided an extract from its DORIS database, which contains 1980 census data (and sometimes 1980–90 changes) by zip code on a wide variety of socioeconomic characteristics. The DORIS and COHORT data were merged so that we have a set of characteristics of the zip code of residence of each AFQT taker. The specific characteristics ultimately used were the zip code's

1. racial composition and 1980–90 change;
2. average household income separately for blacks, whites, and Hispanics;
3. 1980–90 growth of per capita income;
4. households in poverty;
5. households headed by females; and
6. workers in "executive, professional, technical" occupations.

Item 2 was matched to the individual's race/ethnic identifier to yield an estimate of each test taker's household income. Items 4, 5, and 6 were converted into estimated probabilities for each test taker from regression estimates of

$$y_{ij} = a_j + b_j X_{ij} + u_{ij},$$

where

- y_{ij} = characteristic of zip code i in state j (for example, the percentage of households in poverty),
- X_{ij} = percentage of population black in zip code i ,
- a_j, b_j = state-specific parameters, and
- u_{ij} = zip code error term.

(There are over 30,000 zip codes, so there are on average around 600 observations in each state regression.)

Then the probability that an individual from i has the characteristic (for example, comes from a poverty household) is estimated as $(a_j + b_j + u_{ij})$ for blacks and $(a_j + u_{ij})$ for whites. The estimates are truncated to lie within $(0, 1)$.

Finally, the Bureau of Labor Statistics provided estimates of unemployment rates for each state for 1976 on and for groups of states for prior years. I used the 1971–76 annual changes in the group to estimate pre-1976 state unemployment rates.

APPENDIX B

ESTIMATION OF AFQT RESIDUALS AND STATE FIXED EFFECTS

From the full database I extracted the records of all 17–20-year-olds. This group typically comprises 70+ percent of the full sample. For each year, 1971–91, I use the national sample (excluding observations from Alaska, Hawaii, and the District of Columbia) to regress the individual AFQT scores on a set of background characteristics. The set of characteristics on the right-hand side of these

regressions is the same in each year. These are (with white, non-Hispanic, male, 17-year-old, high school graduates as the reference group)

1. AGE: dummies = +1 for 18-, 19-, or 20-year-olds;
2. SEX = +1 for females (always under 20 percent of the sample);
3. EDUCATION: attainment dummies = +1 for fewer than 3 years of high school (HS), 3 or 4 years HS, some college (always under 5 percent of this sample), or uncoded;
4. RACE/ETHNIC: dummies = +1 for black, Hispanic, or other;
5. PERCENT WHITE in individual's zip code, 1980, and the CHANGE in this percent from 1980 to 1990;
6. UNEMPLOYMENT rates in individual's state in current and preceding year;
7. estimated probabilities that individual's family is (a) in poverty, (b) female headed, and (c) has workers classed as "executive, professional, technical";
8. log of estimated household income, 1980;
9. change in log of zip code per capita income, 1980–90; and
10. interactions between
 - A. RACE and EDUCATION:
 - 1) $HS < 3 \times \text{black, Hispanic}$
 - 2) $HS = 3 \text{ or } 4 \times \text{black}$
 - 3) $\text{uncoded} \times \text{black}$
 - B. AGE and EDUCATION:
 - 1) $HS = 3 \text{ or } 4 \times 18, 19, 20$
 - 2) $\text{uncoded} \times 19, 20$
 - C. RACE and zip code characteristics:
 - 1) $\text{black} \times \text{executive, female headed}$
 - 2) $\text{Hispanic} \times \text{female headed}$
 (the characteristics are zip code averages).

The interesting regularities from the regressions are discussed in the text. Some further general points:

1. The sample sizes for the regressions are on the order of $500,000 \pm 200,000$. Accordingly, *T*-ratios well over 10 are common. The list of variables and interactions ultimately retained in the regressions came from a preliminary analysis which discarded those without consistently important effects.

2. The interactions serve largely to correct for some unreasonable estimates from a purely additive model. For example, consider a black high school dropout who lives in an area with few professionals and many female-headed households. Simply adding the separate negative effects of these attributes could produce estimated scores below zero. AGE \times EDUCATION mainly adjusts for the different meaning of "3 or 4 years of high school" for 17-year-olds (who may shortly graduate) and those older (who likely have dropped out).

3. The INCOME GROWTH term acts to revise the 1980 level. Its coefficient tends to be negative in pre-1980 regressions (when actual income would be below the 1980 level where income is rising) and positive after 1980.

4. AFQT tests for 1976–80 were misnormed so that assigned scores were initially too high and then corrected. This episode is however essentially irrelevant here where the goal is to obtain a relative ranking of test takers across states.

The ranking of test takers is invariant to the scale used, so a wrong scale merely shifts the constant term in the regression.

Each residual from these regressions impounds unmeasured individual and locational characteristics (none of the variables or coefficients are location-specific). The location-specific characteristics include the quality of the school system attended by the individual (assumed to be in the state of residence at time of test). Two location specific measures are calculated from the residuals each year:

1. The average residual for the state.
2. The percent of residuals in each state which lie above the 25th percentile of the whole sample. Holding the average residual constant, this measures the avoidance of very bad outcomes.

Another set of 1971–91 regressions was estimated with 47 state dummy variables (California was the excluded state) added to the list of independent variables. The coefficients of these dummy variables are used instead of the state average residual as a school quality measure in some of the analysis. These state fixed-effect models yielded substantively identical sociodemographic patterns to those described above and in the related text.

APPENDIX C

ADJUSTMENTS TO RESIDUALS

The military uses AFQT scores to select inductees. For 1971–75, when the DMDC sample contains only inductees, straightforward calculation of state average residuals or state fixed effects makes the worst state look less bad. The large negative residuals in those states will not be in the inductee sample, so their potential negative impact on the state's average will be removed. This bias gets worse over the period. As the military phased out draftees, it raised minimum AFQT scores. Accordingly, the share of inductees with AFQTs below 30 fell from 21 percent in 1971 to 6 percent in 1975. As a result, the variance of the average residuals from these samples tends to decline over the period and their year-to-year correlations tend to be lower than for post-1975 samples.

To avoid loss of potentially useful information, I adjusted the pre-1975 residuals instead of discarding them:

1. For 1971–74 I randomly discarded observations from the lower tail of the score distribution to reproduce the restrictive 1975 distribution. I did the same for 1976, the first year which has information for all applicants.

2. Regressions were then estimated on these commonly truncated samples. Let

\hat{U}_{it} = estimated residual for i in year t from such a sample, and
 $U_{i,76}$ = a residual from the full 1976 sample.

3. For individuals in both the full 1976 sample and the artificially truncated one, I computed an estimated bias ($b_{i,76}$) due to truncation:

$$b_{i,76} = U_{i,76} - \hat{U}_{i,76}.$$

I then computed the average bias in each zip code ($b_{z,76}$).

4. I then add $b_{z,76}$ to every \hat{U}_{it} in the same zip code in each year prior to 1976.

In effect, the bias is assumed to be constant over time within a zip code. (Gaps are filled with a statewide average bias.) These adjusted residuals are used in all time-series analysis.

The procedure is crude and ad hoc. But it does produce residuals with reasonable time-series properties.

A summary is provided in the following regression of the correlation coefficient (CORR) between average state residuals from different years on the interval between the residuals (L2 = +1 if $t, t - 2$ residuals are used, L3 = +1 if $t, t - 3$, and so on) and a dummy = +1 if one or both sets of residuals are from the PRE76 period:

$$\text{CORR} = .89 - .07\text{L2} - .11\text{L3} - .13\text{L4} - .15\text{L5} - .009\text{PRE76},$$

(64.0) (3.8) (5.7) (6.5) (7.5) (0.2)

where $\bar{R}^2 = .43$, $\text{SEE} = .06$, $N = 90$ (t -ratios are in parentheses).

The sample includes every correlation for lags 1–5 over the 1972–91 period. The regression shows that they decline from an average of .89 for lag 1 to .74 for lag 5. The insignificant coefficient on PRE76 means that correlations that use the adjusted residuals are no weaker than those that do not. (Similar negative results were obtained when PRE76 was divided into a dummy = +1 when both sets of residuals are from the pre-1976 period and another = +1 if only one set is from this period.) By contrast, if unadjusted residuals from 1971–75 are used to compute the correlations, the coefficient on PRE76 becomes a highly significant $-.14$. Thus at the very least, the adjustment succeeds in eliminating atypically erratic year-to-year movements in the residuals.

For the state fixed-effects models, I first estimated regressions on the full and truncated 1976 samples. This yielded two vectors of 1976 state fixed effects. I used the difference between the coefficients of a state dummy variable from the full and truncated 1976 samples as an estimate of the bias due to sample truncation for that state. I then added this estimate back to the coefficient of the state's dummy variable in every pre-1976 regression on the truncated samples to obtain adjusted state fixed effects for 1971–75. This procedure is substantively the same as steps 3 and 4 above, except that a single statewide adjustment is used instead of zip-code-specific adjustments.

A less important adjustment is necessitated by the DMDC's failure to code nearly all Hispanic applicants for the years 1980–86. (They are about 4 percent of the sample in other years.) For the 2 years on either side of this period (1978–79 and 1987–88), I estimated regressions with and without Hispanic dummies and associated interactions. I then computed for each zip code the difference (D_{it}) in the zip code's average residual with and without the information on Hispanic applicants. The two average differences, $\bar{D}_{i,78,79}$ and $\bar{D}_{i,87,88}$ were then linearly interpolated to yield estimates of D for each year 1980–86. These estimates were then added to each residual from the zip code in those years. This adjustment turned out to be meaningful only for New Mexico, where Hispanic applicants typically exceed 20 percent of the total. Because Hispanics tend to do less well than other whites, the New Mexico residual is "too negative" in 1980–86, and the adjustment corrected this.

APPENDIX D

COMPARISON WITH THE NATIONAL LONGITUDINAL SURVEY OF YOUTH

The National Longitudinal Survey of Youth (NLSY) administered the AFQT to a random sample of 15–23-year-olds in 1980 (this was drawn from a larger sample which overrepresents Blacks and Hispanics). In addition, we have information on each youth's job and education experience over the next 10 years and the 1980 county of residence. This sample can be used to ask: Is the state specific component in the 1980 DMDC sample the same as that from a random sample of non-college-bound youth?

To address this, I subsampled from the NLSY all 17–20-year-olds who did not go on to college in the next decade. This group numbers 1,757, or only about 35 per state (compared to around 12,000 per state for the 1980 DMDC sample for 17–20-year-olds). I then estimated a regression of AFQT scores on background variables for this NLSY group. This had the same right-hand-side variables as the DMDC regression, with the following exceptions:

1. County rather than zip code information had to be used to construct the income, poverty, female-headed household, occupation, and neighborhood racial characteristic variables.
2. The educational attainment dummies were for fewer than 2 years and 2–3 years (versus fewer than 3 and 3–4 years for DMDC).
3. NLSY information on years of schooling subsequent to the test is included (and interacted with age dummies).

The sign pattern and magnitudes of the coefficients from the NLSY regression follow closely those in the DMDC regression. In both, the bulk of the explanatory power lies in the racial/ethnic variables and associated interactions. The point estimates of racial/ethnic differences are larger in the NLSY regression, but never by more than 2 standard errors. The tendency for these differences to widen with education is common to both samples.

Meaningful average residuals could be computed for the 28 states which had at least 15 individuals in the NLSY sample. They were compared to the DMDC state average residuals as follows:

The underlying model is that any individual's residual has both a state-specific and an idiosyncratic component. Or

$$N_{ij} = X_j + u_{ij} \quad (D1)$$

for NLSY and

$$D_{ij} = Z_j + v_{ij} \quad (D2)$$

for DMDC, where

N, D = residuals from the NLSY or DMDC regressions for individual i in state j ,

X, Z = state-specific component (including school system quality), and

u, v = individual-specific component.

We want to know if the same process is generating the residuals in both samples, so that $X_j = Z_j$. These are not, however, directly observable. We have instead the state-average residuals (N_j, D_j). But we know that their variances are

$$\sigma_{N_j}^2 = \sigma_X^2 + \frac{\sigma_u^2}{n_N} \quad (D3)$$

and

$$\sigma_{D_j}^2 = \sigma_Z^2 + \frac{\sigma_v^2}{n_D}, \quad (D4)$$

where the second right-hand-side terms are variances of the mean idiosyncratic component in a state (n_N, n_D = number of observations in a state). From the within-state variances of the residuals from both the NLSY and DMDC regressions, we also know that σ_u^2 and σ_v^2 are on the order of 500. (In a random sample from a rectangular distribution on (1, 99)—that is, no racial, state-specific, or so on, effects—the variance would be $28.6^2 = 817$). The value n_D averages 12,000 and $\sigma_{D_j}^2 = 2.7$, so we have

$$\sigma_{D_j}^2 = \sigma_Z^2 + \frac{500}{12,000} = 2.7 \Rightarrow \sigma_Z^2 \sim 2.7. \quad (D4')$$

That is, the idiosyncratic component of the variance of the mean disappears in a sample of 12,000. Further, if N_j were generated by the same process as D_j , $\sigma_X^2 = \sigma_Z^2$, and we would expect

$$\sigma_{N_j}^2 \sim 2.7 + \frac{500}{60} = 11.0 \quad (D3')$$

given roughly 60 NLSY observations per state in the 28-state sample. The actual $\sigma_{N_j}^2$ in this sample is 11.4, or almost exactly as expected if the two processes are the same.

Further, if the state-specific components were the same ($X_j = Z_j$), $\text{cov } XZ = \sigma_X^2 = \sigma_Z^2$, and the expected regression coefficient (b) when D is regressed on N would be

$$b = \frac{\sigma_X^2}{\sigma_N^2} = \frac{2.7}{11.0} = .245.$$

The actual value (.274, SE = .080) is indistinguishable from the expected value. Or, regressing N on D would yield an expected coefficient ≈ 1.0 ; the actual coefficient (SE) is 1.16 (.34). Or the expected correlation coefficient would be $\sigma_X/\sigma_N = .495$. The actual value of .54 is insignificantly different.

In sum for this year, the NLSY and DMDC samples are yielding estimates of the same underlying reality with respect to state-specific effects. The only difference is that the NLSY estimates are much noisier.

BIBLIOGRAPHY

- Chubb, John, and Moe, Terry. *Politics, Markets and America's Schools*. Washington, D.C.: Brookings Institution, 1990.
- Coleman, James; Hoffer, Thomas; and Kilgore, Sally. *High School Achievement:*

- Public, Catholic and Private Schools Compared*. New York: Basic Books, 1982.
- Downes, Thomas. "Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case." *National Tax Journal* 45 (December 1992): 405–20.
- Ellwood, David T., and Wise, David A. "Uncle Sam Wants You—Sometimes: Military Enlistments and the Youth Labor Market." Ch. 4 in *Public Sector Payrolls*, edited by David Wise. Chicago: University of Chicago Press, 1987.
- Flyer, Frederick, and Rosen, Sherwin. "The New Economics of Teachers and Education." Working Paper No. 94. Chicago: University of Chicago, Center for the Study of the Economy and the State, 1994.
- Freeman, Richard B., and Ichniowski, Casey, eds. *When Public Sector Workers Unionize*. Chicago: University of Chicago Press, 1988.
- Hanushek, Eric. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (September 1986): 1141–77.
- Hoxby, Caroline Minter. "Do Private Schools Provide Competition for Public Schools?" Unpublished manuscript. Cambridge: Massachusetts Institute of Technology, Department of Economics, 1993.
- Hoxby, Caroline Minter. "Is There an Equity-Efficiency Trade-Off in School Finance? Tiebout and a Theory of the Local Public Goods Producer." Unpublished manuscript. Cambridge, Mass.: Harvard University, Department of Economics, 1995.
- Kochan, Thomas. "Correlates of State Public Employee Bargaining Laws." *Industrial Relations* 12 (1973): 322–27.
- Murray, Charles, and Herrnstein, R. J. "What's Really Behind the SAT-Score Decline." *Public Interest* (Winter 1992): 32–55.
- Neal, Derek. "The Effect of Catholic Secondary Schooling on Educational Attainment." Working Paper No. 95. Chicago: University of Chicago, Center for the Study of the Economy and the State, 1994.
- Peltzman, Sam. "The Political Economy of the Decline of American Public Education." *Journal of Law and Economics* 36 (April 1993, Pt. 2): 331–70.
- U.S. Congressional Budget Office. *Trends in Educational Achievement*. Washington, D.C.: Government Printing Office, 1986.
- U.S. Congressional Budget Office. *Educational Achievement: Explanations and Implications of Recent Trends*. Washington, D.C.: Government Printing Office, 1987.