THE EFFECT OF MEASURED SCHOOL INPUTS ON ACADEMIC ACHIEVEMENT: EVIDENCE FROM THE 1920s, 1930s AND 1940s BIRTH COHORTS

Susanna Loeb and John Bound*

Abstract—The study presented here uses data from the NORC General Social Surveys to explore the effects of measurable school characteristics on student achievement. What separates this study from many others is the use of aggregate data on older cohorts, usually associated with research on the influence of school inputs on earnings. We find substantively large effects, similar in size to those found in many earnings-focused studies. Our results point to the importance of aggregation in modeling the relationship between school inputs and student outcomes, bringing into question causal interpretations of the results of studies using aggregate data to assess school input effects.

I. Introduction

SCHOOLS differ markedly in quality and their ability to promote student achievement. Yet, simple measures of school inputs, such as student–teacher ratios, generally have not been found to capture school quality differences that influence student outcomes. The 1966 Coleman Report, the result of an extensive nationally representative study of schools, showed little school quality effect on student test scores once family background characteristics were accounted for. These results were bolstered by Hanushek's (1986) summary of research on the impact of school characteristics on achievement in which he found no measurable characteristic of schools that consistently contributed to student achievement. A recent meta-analysis of the same studies that Hanushek used indicated positive relationships between some school inputs and student outputs, undetected with Hanushek's simpler summary methods (Hedges, Laine, and Greenwald (1994)). Yet, even meta-analyses and studies at the level of micro data in general that have found significant effects of school inputs on achievement have tended not to find substantively large and consistent effects (Childs and Shakeshaft (1986), Glass and Smith (1979)). In light of this history, the corresponding literature on the impact of school characteristics on later earnings seems surprising. A series of studies using aggregate school input data at the state or district level have concluded that school inputs can substantially increase returns to education, as well as educational attainment (e.g., Akin and Garfinkel (1977), Card and Krueger (1992), Johnson and Stafford (1973), Link and Ratledge (1975), and Rizzuto and Watchel (1980). Card and Krueger (forthcoming) provide a review.)

What accounts for the discrepancies between these two literatures? A number of alternatives have been suggested (Betts (1996), Burtless (1996)). First, the two literatures examine different outcomes. Measured school quality may increase the impact of schools on earnings, even if it does not increase their impact on academic achievement (Card and Krueger (1992)).[1] Second, studies that use earnings as the dependent variable and find positive effects may have insufficient controls for family background. This potential problem is particularly relevant in light of findings that there is generally a positive raw correlation between spending and academic achievement but that this relationship usually disappears after family background characteristics are taken into consideration (Hanushek (1986)). Third, studies that focus on earnings have either ignored the potential effects of local labor markets (Johnson and Stafford (1973)) or, as in the case of the work by Card and Krueger (1992), have identified effects by comparing the earnings of individuals who grew up in different locations, but currently work in the same area. The first of these procedures is potentially problematic since the returns to skills learned in school may vary by location. The latter circumvents this problem, but may be biased since migrants are self-selected (Heckman, Layne-Farrar and Todd (1995)).

Alternatively, the conflict between the findings of positive school quality effects on earnings but not on academic achievement may be due to differences in the nature of the data used. The data in the two literatures have consistently differed in two ways. First, income analyses have tended to use historical information on school characteristics, while achievement analyses have tended to use contemporaneous data. This difference arises because more reliable data on achievement is available for recent cohorts, while income information is more stable for older cohorts. Second, earnings focused research almost exclusively has used aggregate measures of school inputs, matching workers with the average inputs in the school district or state in which they grew up.[2] Studies using achievement as the outcome measure, on the other hand, have tended to use micro level data, matching students to the precise school or classroom they attended.

In the study presented here, we use aggregate historical data on school inputs to predict the scholastic achievement of people in the same three cohorts used by Card and Krueger (1992) in their assessment of the effects of school quality on income returns to education—those born in the 1920s, the 1930s and the 1940s. By doing this we will be better able to distinguish among some of the possible causes of

Received for publication February 24, 1995. Revision accepted for publication April 8, 1996.

* University of Michigan.

We have benefited from comments by Charles Brown, Arline Geronimus, Caroline Minter Hoxby, and Robert Margo, as well as by seminar participants at the University of Michigan and at the National Science Foundation and this *Review*'s Conference on School Quality and Educational Outcomes (Dec. 1994, Boston). This paper was completed while the second author was a visiting scholar at the Russell Sage Foundation.

[1] This explanation seems possible given the common findings (e.g., Griliches and Mason (1972)) that there is only a weak relationship between academic achievement and earnings once educational attainment is controlled for.

[2] Two exceptions to this trend are recent studies by Betts (1995) and Grogger (forthcoming). Betts uses the National Longitudinal Survey of Youth and Grogger uses High School and Beyond and the National Longitudinal Survey of the High School Class of 1972 to look at earnings of young workers. Individuals in both studies are matched with the specific high school attended. Neither study finds significant effects of measurable school inputs on students' later earnings; however, the instability of wages of young workers is a potential problem for both of these studies.

the conflicting findings on school input effects. Our measure of scholastic achievement is derived from a simple, ten-item, vocabulary test administered as part of NORC's General Social Surveys. Unlike most studies of academic achievement, our estimates suggest powerful effects of measured school inputs on school effectiveness. As such, these results would seem to suggest that the difference between the two literatures arises, at least partially, from differences in the nature of the data used: either differences in the historical period covered or the extent of aggregation involved in measuring school inputs.

II. Model Specification and Data

The Data

The data source for information on individuals is the NORC General Social Surveys (GSS) (Davis and Smith (1992)). Each survey in the GSS consists of detailed interviews of an independently drawn sample of English-speaking persons 18 years of age or over, living in non-institutional arrangements within the United States. Since 1972 there have been a total of 29,388 interviews, of which 6,032 were of people born in the 1920s, 1930s and 1940s and included the relevant information for this study.[3] The most recent survey currently available was conducted in 1993. The GSS contains information on the geographic division where the respondent lived at age 16, but not the school district or the state. Thus, we run our analyses aggregated to the divisional level. Since most of the variance in school quality among states is due to variation among the nine divisions, the use of divisions instead of states should not substantially affect the results.[4]

We aggregate the state-level school quality data used by Card and Krueger (1992) (originally obtained from the Biennial Survey of Education (1918–1958) and the Digest of Education Statistics (1968–present)) to the divisional level. For the aggregation, we weight using the average number of teenagers in each state during the decade in question. For the three regions that had segregated school systems, we obtain separate aggregate school quality measures for black school districts and white school districts. Nine geographic divisions, three of which had segregated school systems, give us a total of 12 geographic/racial regions, each with three cohorts.

The school quality measures that we use are student-teacher ratio and term length. These measures should not necessarily be thought of as capturing the same characteristic of schools. Student–teacher ratio or class size[5] has been used commonly in studies of the effect of school quality on both achievement (Finn and Achilles (1990), Boozer and Rouse (1995)) and income (Card and Krueger (1992)). Current education research shows, at best, only moderate impact of these variables on achievement. Term length, on the other hand, appears to have played an important role historically in the increased achievement of black students in segregated school systems (Margo (1990), Orazem (1987)). Due to limited variation, term length has not been used in recent work on school achievement. However, "time on task" *has* consistently been shown to matter (Denham and Lieberman (1980)), and presumably term length is highly correlated with "time on task." Thus, while a finding using historical data that the student-teacher ratio matters does contradict current research, a similar finding for term length does not.

Figures 1a and 1b present average school inputs, student–teacher ratio and term length respectively, for each cohort in each division. From these figures we can see that school inputs tended to increase during the time period in question. For those born in the 1920s the average student-teacher ratio across divisions was 32.68. This measure decreased to 28.66 students per teacher for those born in the 1930s and to 26.54 for those born in the 1940s. Similarly, the average term length across divisions for those born in the 1920s was 164.31 days, while average term length was 172.94 and 177.08 days for those born in the 1930s and 1940s.

While the overall level of inputs increased over time, the variation across divisions decreased dramatically as school systems, to a considerable extent, converged in terms of measured inputs. The standard deviation of our student–teacher ratio measure declined from 7.4 to 2.5 between the cohorts born in the 1920s and those born in the 1940s. Similarly, the standard deviation of term length declined from 17.7 to 2.7. The most dramatic changes occurred in the segregated black school systems, but the pattern holds for the white and integrated school systems as well.

Correlation between the two school input measures is high (-0.91) but appears to have decreased over time. For the 1920s' cohort the correlation is -0.94, while for the 1930s' and 1940s' cohorts it is -0.79 and -0.54, respectively. These high correlations are, to an extent, being driven by

[3] The sample of 6,032 is limited to those respondents born in the 1920s, 1930s, and 1940s who have information on both educational attainment and test scores. The vocabulary test used in this study to measure achievement was included in the 1974, 1976, 1978, 1982, 1984, 1987, 1988, 1989, 1990, 1991, and 1993 surveys only.

[4] Univariate regressions of the state-level quality measures on the division dummy variables by cohort give R^2s of between 0.61 and 0.74.

[5] Student–teacher ratio and average class size are not the same measure. Average class size will almost always be larger than student–teacher ratio since more than one teacher, but never less than one, may be assigned to a classroom; and because many teachers have duties outside of the classroom (Odden (1990)). In Ferguson's (1991) study of Texas schools, for example, a student–teacher ratio of 18 corresponded to average class sizes in the low 20s. In addition, the relationship between class size and teacher–student ratio may be different for different groups of students. Boozer and Rouse (1995) find that while student–teacher ratio is the same for black and white students in their sample, black students tend to be in larger classes. Because of these differences between average class size and student–teacher ratio, one measure is not a perfect substitute for the other. However, especially for historical data in which there is much larger (and, likely, more correlated) variation in class size and student-teacher ratio than exists today, the two measures should be good proxies for each other.

FIGURE 1a.—AVERAGE NUMBER OF STUDENTS ENROLLED PER TEACHER BY COHORT AND DIVISION

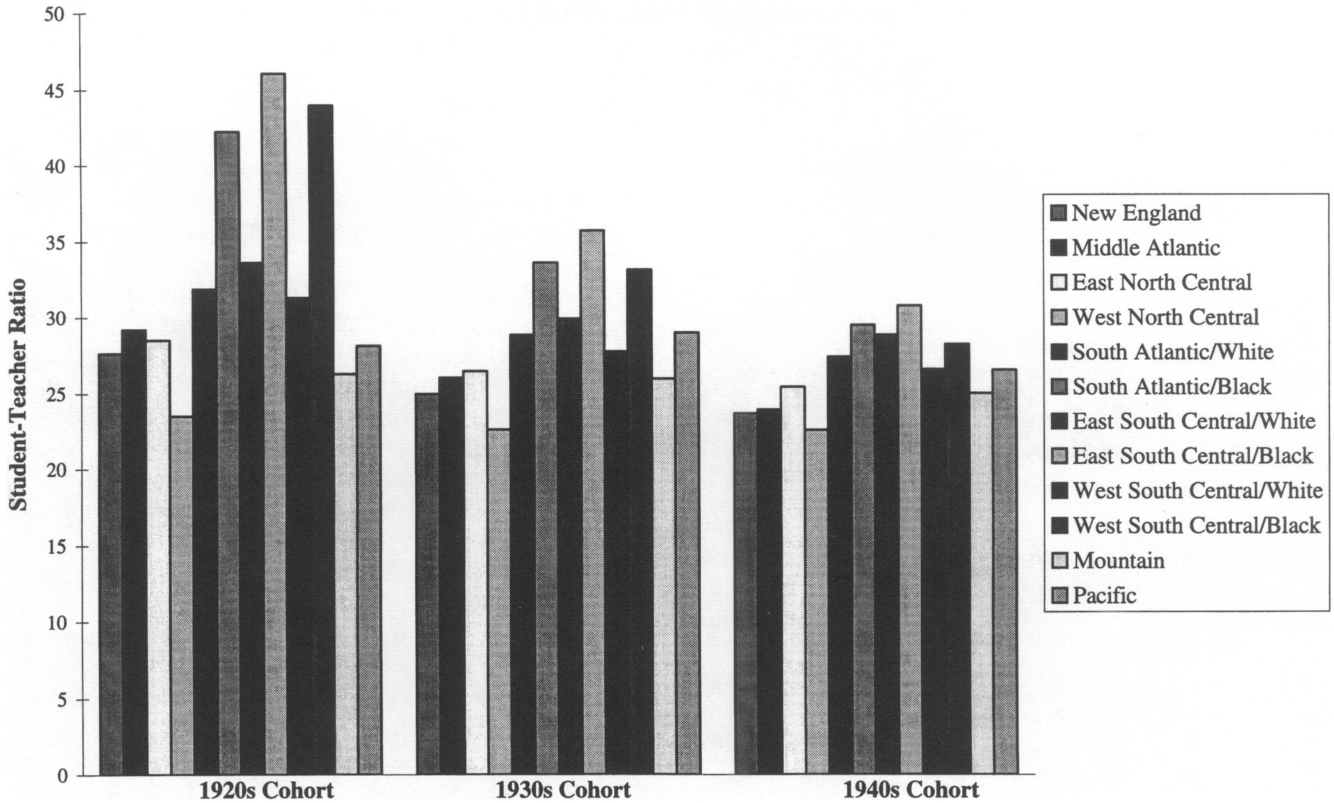


FIGURE 1b.—AVERAGE LENGTH OF THE SCHOOL TERM IN DAYS PER YEAR BY COHORT AND DIVISION

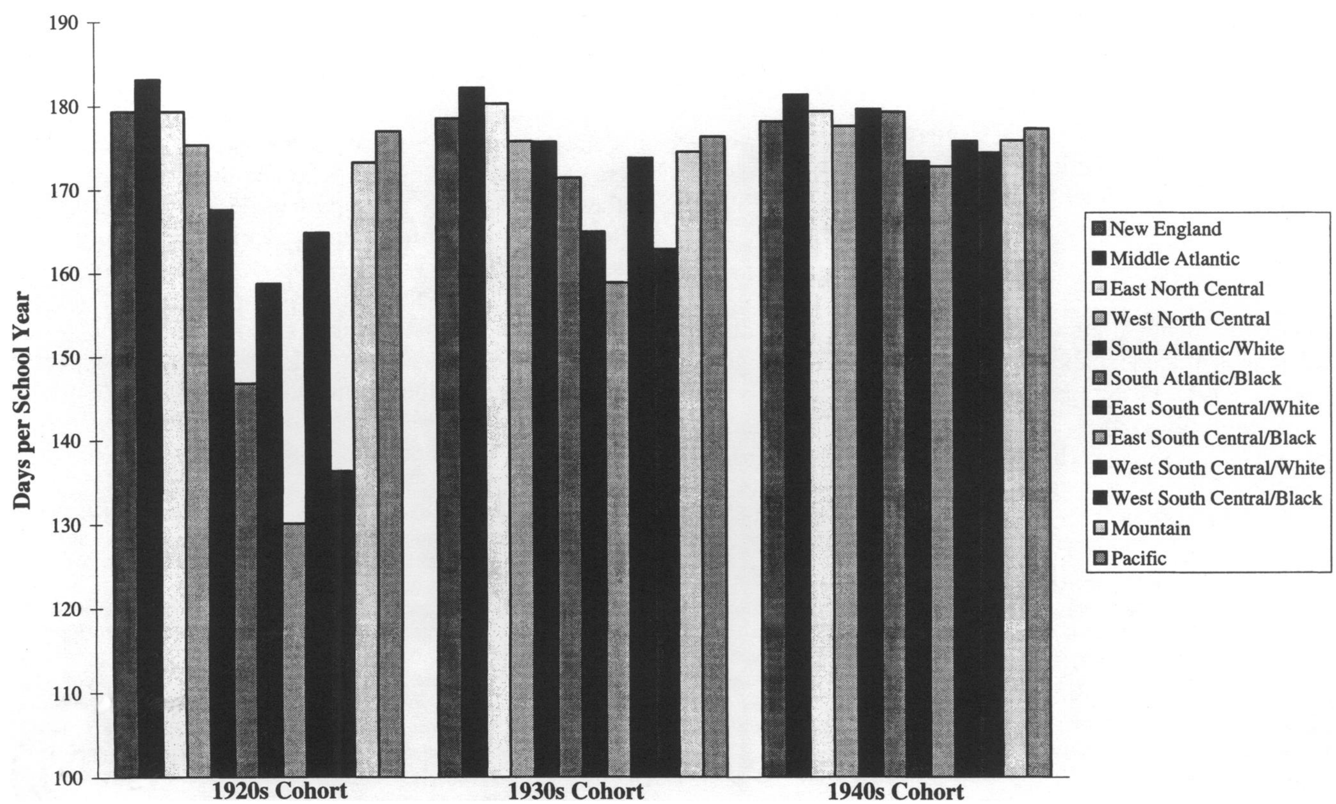
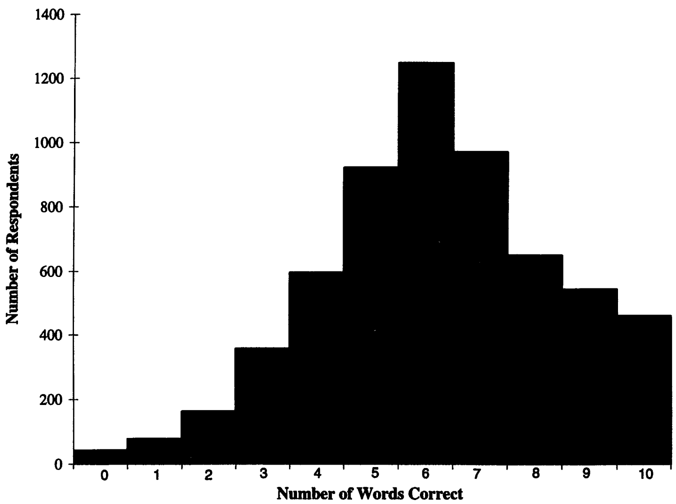


FIGURE 2.—THE DISTRIBUTION OF THE VOCABULARY TEST SCORE MEASURE



the inclusion of the segregated black districts. The correlations for the white and integrated districts are -0.63 for the full sample, and between -0.65 and -0.50 for the three cohorts separately.[6]

The achievement measure provided by GSS is a ten-item test of vocabulary knowledge. For each vocabulary word, the respondents are asked to choose which of five other words is closest in meaning. While the test words are not publicized, they come from a slightly longer vocabulary test developed by Thorndike and Gallup (Thorndike (1942), Thorndike and Gallup (1944)). The longer test was designed to measure learning (Thorndike and Gallup (1944)) and has been found to have a correlation with tests of general intelligence of approximately 0.8 or higher (Miner (1957)). The shorter version used by the GSS has an internal consistency reliability of 0.71 (Alwin (1991)). The mean score on the test for those used in this study is 6.21, with a standard deviation of 2.16. Figure 2 shows the overall distribution of this variable.[7] A standardized version of the number of correct answers to the ten questions is used for the analyses presented here.[8]

The Model

Similar to many previous studies we use a two-tiered model for estimating school input effects. The first stage of the analysis is an OLS regression in which test score for the 6,032 respondents is modeled as a linear function of years of education, demographic characteristics, family background measures, cohort-division group dummies, and interactions between the respondent's cohort-division group and education.[9] We include all 36 cohort-division dummy variables and interaction terms and then restrict the two groups of coefficients to sum to zero. As a result, the coefficients on the dummy variables can be interpreted as deviations from the unweighted average of mean cohort-division group test scores. Similarly, the coefficient on the educational attainment variable measures the unweighted average of mean test score returns to an extra year of schooling for each cohort-division group, while the interaction terms give the deviations for each group around this average. The interaction terms then capture the influence of each cohort-division group on the achievement returns to years of schooling. The main portion of the analysis includes three first-stage regressions: one without family background controls, one with linear family background controls only, and one with interactions between all family background measures and years of education completed as well as linear controls.[10]

The first-stage regressions control for three demographic characteristics: gender, race, and age. We included a dummy variable for gender to adjust for possible differences in education aside from years of schooling and for potential differences in experience that could affect vocabulary knowledge. A dummy variable for whether the respondent is black is used to adjust for a number of factors including potential differences in education quality that white and black students received in non-segregated districts and potential racial bias in the vocabulary test. We include age at the time of the interview in order to adjust for possible age differences in vocabulary memory that could bias cohort effects. Since the interviews were conducted over a 20-year period this control is not equivalent to the cohort controls.

Family background adjustments include measures of parents' education, family size and whether the respondent was raised in a single-parent household, as well as a rough assessment of family income when the respondent was 16. The parents' education variable is the average of the respondent's answers when asked the years of education received by his/her mother and father.[11] The family size variable used is the log of the total number of children in the family. The family income measure in the GSS is a five-level variable of responses to, "Thinking about the time when you were 16 years old, compared with American families in general then, would you say your family income was—far below average, below average, average, above average, or far above average?" We treat this income measure as categorical, including indicators for each level of income in our models.

[6] The correlations between student–teacher ratio and term length are lower at the state level than aggregated to the division level. For example, among states the correlation excluding the segregated black districts is -0.49 overall, -0.54 for the 1920s' birth cohort, -0.33 for the 1930s' birth cohort and -0.20 for the 1940s' birth cohort. A decrease in the correlation over time is evident.

[7] Appendix table A1 gives the means and standard deviations by division and cohort for this and all other variables used in the analyses.

[8] For more complete discussion of the vocabulary test see Miner (1957) and Alwin (1991).

[9] The GSS consists of a random sample of households in the United States, not a random sample of individuals. Within each household, one respondent was chosen at random from the eligible adults. In order to assess the possible difference between this sample and a random sample of adults, we do our analyses both unweighted and weighted by the number of eligible adults in the respondent's household. Our findings are similar in both cases. We report only the unweighted results.

[10] Due to the limited nature of the dependent variable, an ordered probit might be a more appropriate first-stage method than OLS. We ran the analyses using both methods and found virtually no difference in the results. The results from the linear regressions are somewhat easier to interpret and so they are presented.

[11] For those respondents who only answered the question for one parent, the average parents' education variable is the given parent's reported education.

Of the 6,032 respondents used in this analysis, 601 did not report information on either parent's educational attainment, 16 did not report number of siblings, 118 did not report on the family structure of the household they grew up in, and 44 did not respond to the income question. We handled this missing data by including four missing-data indicators in our models.

The second stage of the analysis addresses systematic variation across divisions and cohorts in the impact of years of education on test scores due to measurable school characteristics. We use the coefficients on the interactions between educational attainment and cohort-division group from the first-stage regressions as the outcome measure for assessing the effects of school quality. For each of the first-stage analyses we run regressions that include one of our two school quality measures and controls for cohort and/or division.

Due largely to differences in underlying sample sizes, there is considerable variation across divisions and cohorts in terms of the reliability of our first-stage estimates of the impact of education on test scores. For this reason, there are potential efficiency gains to be derived from weighting the data by an estimate of the covariance matrix. If our model is not perfectly specified, however, weighted and unweighted regressions will also represent, conceptually, somewhat different effects. The weighted estimate will downplay the coefficients of the smaller segregated southern school systems. For this reason, we present both weighted and unweighted second-stage estimates. The weighted results closely resemble feasible GLS one-step estimates.[12]

In addition to performing the above analyses, we estimate two other sets of school quality effects. First, we rerun the regressions on whites alone. Second, we add controls for current geographic division of residence in each of the first-stage analyses. Since local labor markets affect achievement to a much smaller extent, if at all, than they affect earnings, we do not need to control for current state of residence in order to accurately assess the effects of school inputs on achievement. By including controls for the local labor market we can estimate the impact of such controls on studies that have looked at the effect of school inputs on later earnings.

III. Results

Table 1 summarizes the results of the first-stage regressions for the full sample both with and without family-background controls. Although cohort-division dummies were included in the regression, for simplicity their coefficients as well as the coefficients on the interaction terms and on the no-response dummy variables are not reported in the table. Table 1 shows that, on average, an extra year of education corresponds to a 0.17 standard deviation increase in the test score without adjustments for family background, and a 0.15 standard deviation increase once family background is controlled for. Overall, women scored approximately 0.18 standard deviations higher than men both with and without the family background variables included. Whites scored 0.38 and 0.35 standard deviations higher than blacks with and without family background controls, respectively. The results indicate no significant relationship between age and test score in the multivariate framework.

Table 1.—First Stage Regression Results

Variables	Means	Model 1	Model 2
Education	12.60	0.17	0.15
	(3.03)	(0.00)	(0.00)
Age	46.83	−0.002	−0.002
	(10.86)	(0.021)	(0.002)
Black	0.14	−0.38	−0.35
	(0.35)	(0.05)	(0.05)
Female	0.57	0.19	0.18
	(0.50)	(0.02)	(0.02)
Average Parents' Education	9.59		0.025
	(3.42)		−(0.004)
Income			
Below Average	0.27		−0.027
	(0.45)		(0.039)
Average	0.49		−0.057
	(0.50)		(0.038)
Above Average	0.12		0.077
	(0.33)		(0.48)
Far Above Average	0.02		0.062
	(0.12)		(0.093)
Log (Number of Children)	1.39		−0.063
	(0.67)		(0.017)
Single Parent Family	0.17		−0.003
	(0.37)		(0.029)
R^2		0.39	0.40
F-Value		50.98	46.67

Note: 36 cohort-division dummy variables and 36 cohort-division by education level interaction terms were included in each of the regressions although their coefficients do not appear in the table. In addition, model 2 includes dummy variables for whether each of the family background measures is missing. The results for model 3, which includes interactions between educational attainment and each of the family background measures, is not reported here. The sample size is 6032.

Model 2 includes linear controls for family-background characteristics. Of the measures of family background included, only average parents' education and log of family size show significant effects. A one-year increase in parents' education corresponds to a 0.025 standard deviation increase in test scores. Those with larger families, on the other hand, tended to have lower test scores. A one-unit increase in log family size corresponds to a 0.063 standard deviation decrease in test scores. In addition, while none of the income variables was significant at the 0.05 level, the dummy variable for no response indicated that those who answered the question had on average 0.30 standard deviation higher test scores than those who did not. Model 3 includes family background–educational attainment interactions for all family-background measures. Since the inclusion of these interaction terms makes interpretation of the coefficients on the original variables difficult, the results of model 3 are not reported in table 1.

We use the coefficients on the interaction terms between education level and cohort-division group for each of these models as the dependent variable for the second stage. These coefficients measure the average additional gain in test scores per year of education associated with the division and cohort in which respondents were born. Figures 3a and 3b show the

[12] This result is hardly surprising. The models we estimated satisfy the conditions necessary for the optimal one- and two-step estimators to be identical (Amemiya (1978), Borjas (1982)).

coefficients from model 2, with linear family background controls, plotted against the average student-teacher ratio and the average term length of the division and cohort. The black districts are represented in these figures by square points, whereas white and integrated school districts have diamond-shaped points. These figures illustrate a positive, though imperfect, relationship between school inputs and the interaction coefficients, suggesting that cohort and division groups with higher average school inputs also tended to have greater returns in achievement for extra years of education. The figures make clear that the inclusion of the southern black districts adds important variation to the data.

The results of the unweighted second-stage regressions are summarized in table 2. This table reports the coefficients on the school quality measures and White's standard error estimates for each of the 12 models discussed above.[13] In all 12 models the regression coefficient for student–teacher ratio is approximately -0.4. This indicates that each extra student in a classroom decreases the test score returns to an extra year of schooling by 0.004 standard deviations. An increase in the student–teacher ratio of ten students, representing a drop in resources of between 20% and 30%, would be associated with a 0.04 standard deviation decrease in achievement returns to education. For the average respondent, this corresponds to a change from 0.15 to 0.11 standard deviations greater test score for each additional year of education, a decrease of just over 25%. While the coefficient size on student–teacher ratio is largely stable across the models, the standard errors increase as more controls are introduced into the model. This is hardly surprising. Cohort and division dummies explain 84% of the variation in the student-teacher ratio.

The coefficients on term length range from 0.19 to 0.10, depending on controls. These estimates imply that an increase in the length of the school year by ten days, representing a 6% to 8% change in the length of the school year, corresponds to a 7% to 13% increase in returns to education. Again, the limitations of the data are evident in these results. As more controls are included in the model, the standard errors increase, though not by as much as for student–teacher ratio. Unlike the coefficient on student–teacher ratio, the coefficient on term length tends to decrease as more controls are introduced.

The inclusion of the family-background measures has limited effect. When family background variables are interacted with educational attainment in model 3, the coefficients on the school quality measures drop substantially for those second-stage models that do not include division controls. However, once division is controlled for in the second-stage, the estimated effects of school inputs are essentially the same whether or not family-background measures are included in the first-stage regressions. Thus, once studies, such as Card and Krueger (1992), control for state and cohort of schooling, their lack of control for family background or their use of rough aggregate proxies for such control may not substantially bias their results.[14] At the aggregate level, controls for cohort and division appear to take account of most of the relationship between family background and measurable school quality. In contrast, micro-level research has tended to find that family-background measures often account for any univariate relationship between school inputs and student outcomes (Hanushek (1986)). This paradox points to the potential influence of data structure.

Results from the weighted second-stage estimates are presented in table 3. While conceptually the error from the second-stage equations includes both equation and measurement error, our computations imply that measurement error represents the bulk of the total. For this reason, we simply use the inverse of the estimated variances of the first-stage estimates to form the weights. The reliability of the first-stage interaction coefficients varies considerably; thus, it is not surprising that the point estimates and their estimated reliability are somewhat different for the weighted than for the unweighted second-stage analyses. In general, weighting reduces the absolute value of the coefficients as well as their standard errors. These reductions most likely result from down-weighting the smaller segregated black districts. While weighting does change the estimates, the substance of the findings remains.

Overall, though the reliability of some of the estimates we report is not high, the point estimates suggest very large effects of measured school quality. The implied elasticities of the effect of changes in measured school quality on the value of education hover around one! These effects are substantially stronger than those found in most other studies of student–teacher ratio or class size.[15] Glass and Smith (1979), in a meta-analysis of class-size effects, find only an approximately one five-hundredth of a standard deviation difference in achievement between 20 and 40 student classes. Hanushek (1986), in a summary of the research on school input effects, notes that of 112 studies of teacher–pupil ratio, only nine found positive significant results. Fourteen of these studies showed negative significant results and the rest had insignificant findings. Studies of term-length effects are not as prevalent, presumably because term length no longer varies significantly across schools.

The results also suggest that studies that use the same types of data and similar models to assess the impact of

[13] Due to the small sample sizes involved in our estimates, white standard errors are likely to underestimate the actual sampling variability of our estimates somewhat (Chesher and Jewitt (1987)), though by less than OLS estimates.

[14] Card and Krueger use average per capita income and median education for white persons over 25 years of age in each state in each cohort as proxies for individual family-background effects. Aggregated from the state to the division level, this measure has correlation coefficients of 0.80, 0.45 and -0.04 with the division-level aggregates of our parents' education, single-parent-family and family-size measures, respectively. When the residuals of a regression of Card and Krueger's family-background variable on cohort dummies and division dummies is compared with the residuals of a similar regression using our parents'-education measures, we find no relationship (a correlation coefficient of -0.13).

[15] As a note of caution, it is questionable whether the comparison of effects on different test score measures is useful. Tests vary in what they measure and in how they are affected by specific classroom learning.

FIGURE 3a.—A PLOT OF STUDENT-TEACHER RATIO BY ADDITIONAL ACHIEVEMENT RETURNS TO EDUCATION FOR EACH COHORT-DIVISION GROUP AS ESTIMATED BY MODEL 2 (SEE TABLE 1)

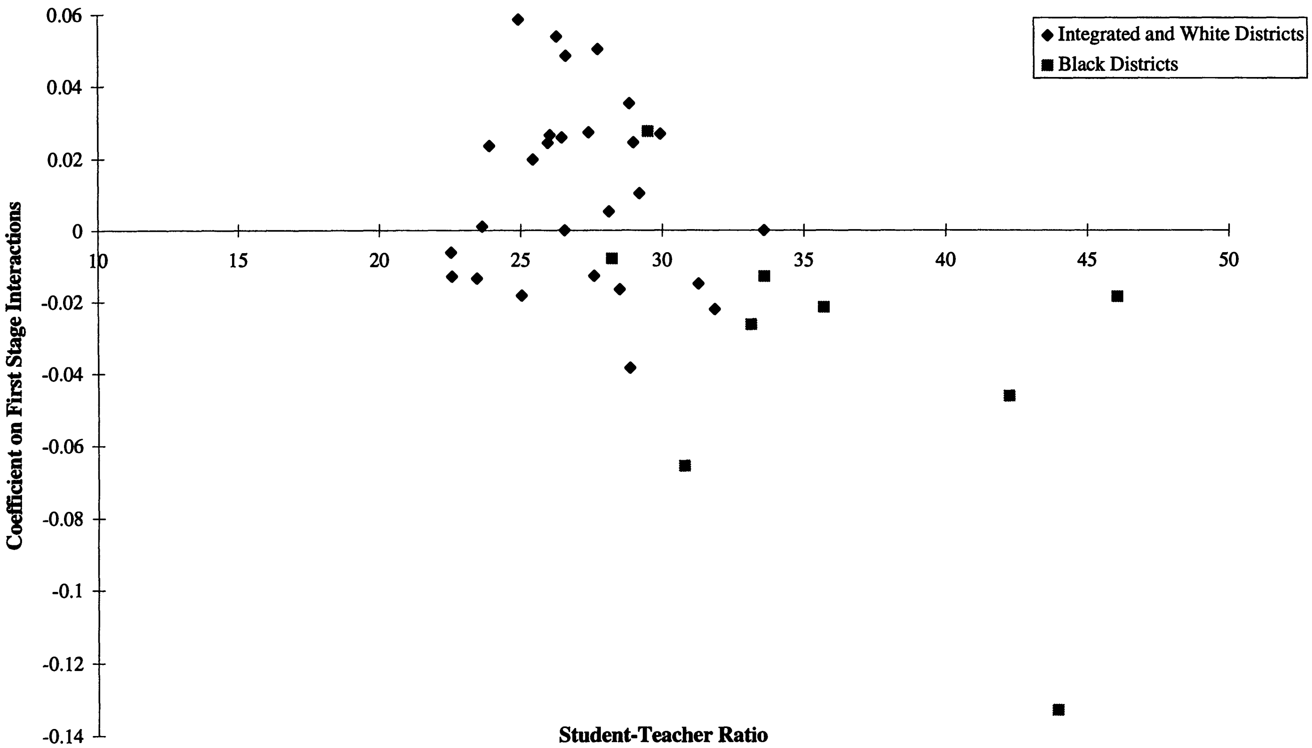


FIGURE 3b.—A PLOT OF LENGTH OF THE SCHOOL TERM BY ADDITIONAL ACHIEVEMENT RETURNS TO EDUCATION FOR EACH COHORT-DIVISION GROUP AS ESTIMATED BY MODEL 2 (SEE TABLE 1)

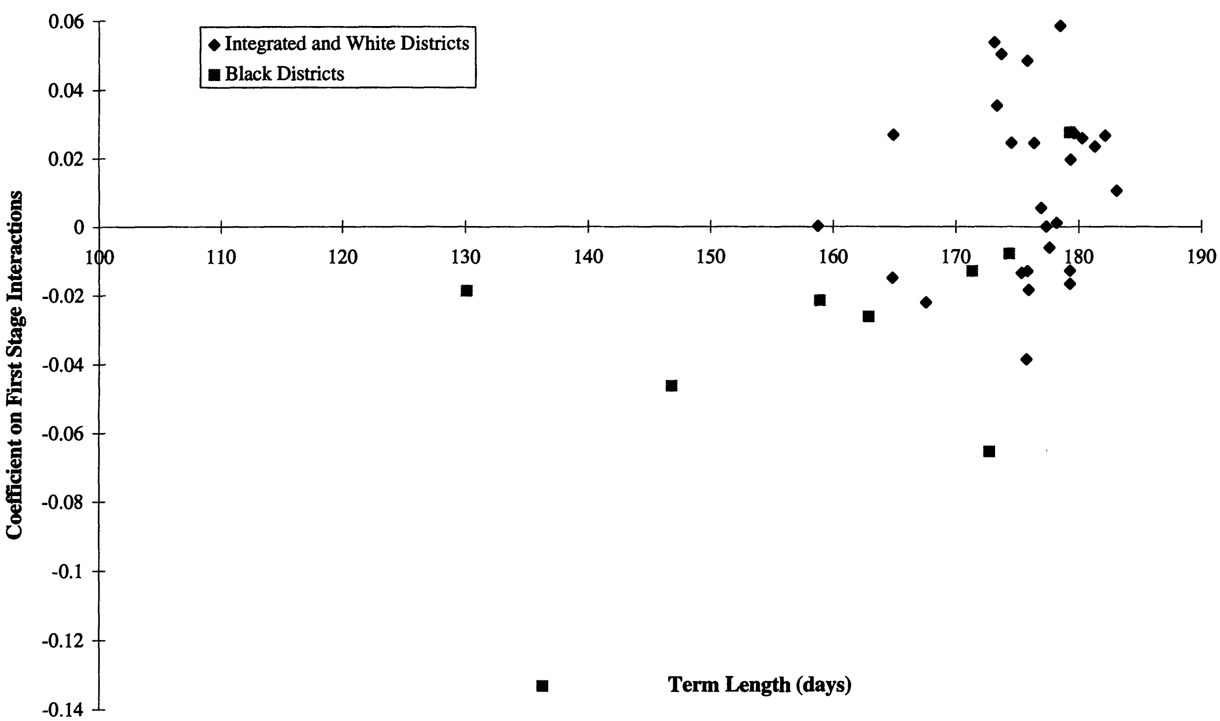


TABLE 2.—SECOND STAGE REGRESSION RESULTS: UNWEIGHTED

	Student–Teacher Ratio/100		Term Length/100	
	No Family Background Controls			
No Controls	−0.42 (0.13)	0.36	0.19 (0.07)	0.34
Cohort Controls	−0.41 (0.15)	0.37	0.18 (0.07)	0.36
Division Controls	−0.41 (0.17)	0.61	0.15 (0.07)	0.59
Cohort and Division	−0.35 (0.26)	0.63	0.10 (0.09)	0.61
	Linear Family Background Controls			
No Controls	−0.39 (0.14)	0.34	0.17 (0.07)	0.33
Cohort Controls	−0.37 (0.15)	0.36	0.17 (0.07)	0.35
Division Controls	−0.41 (0.18)	0.60	0.15 (0.08)	0.56
Cohort and Division	−0.38 (0.27)	0.62	0.10 (0.09)	0.60
	All Family Background Controls			
No Controls	−0.25 (0.14)	0.18	0.12 (0.07)	0.21
Cohort Controls	−0.20 (0.14)	0.24	0.10 (0.07)	0.26
Division Controls	−0.40 (0.19)	0.48	0.14 (0.08)	0.44
Cohort and Division	−0.38 (0.28)	0.51	0.10 (0.09)	0.49

Note: The top left hand number for each set of three numbers is the coefficient on the school quality measure in the second stage regression. The bottom left hand number is the White's estimate of the standard error. The right hand side number is the R^2 for the regression. The sample size for each regression is 36.

TABLE 3.—SECOND STAGE REGRESSION RESULTS: WEIGHTED

	Student–Teacher Ratio/100		Term Length/100	
	No Family Background Controls			
No Controls	−0.33 (0.09)	0.24	0.15 (0.05)	0.24
Cohort Controls	−0.25 (0.11)	0.30	0.12 (0.05)	0.32
Division Controls	−0.45 (0.15)	0.56	0.17 (0.07)	0.52
Cohort and Division	−0.20 (0.22)	0.59	0.08 (0.07)	0.60
	Linear Family Background Controls			
No Controls	−0.31 (0.09)	0.24	0.14 (0.04)	0.23
Cohort Controls	−0.23 (0.10)	0.30	0.11 (0.04)	0.32
Division Controls	−0.44 (0.14)	0.58	0.16 (0.07)	0.53
Cohort and Division	−0.23 (0.23)	0.61	0.07 (0.07)	0.61
	All Family Background Controls			
No Controls	−0.18 (0.09)	0.10	0.09 (0.04)	0.11
Cohort Controls	−0.08 (0.09)	0.22	0.05 (0.04)	0.24
Division Controls	−0.41 (0.14)	0.51	0.14 (0.07)	0.45
Cohort and Division	−0.24 (0.23)	0.54	0.07 (0.07)	0.54

Note: The top left hand number for each set of three numbers is the coefficient on the school quality measure in the second stage regression. The bottom left hand number is the White's estimate of the standard error. The right hand side number is the R^2 for the regression. The sample size for each regression is 36.

school inputs on student outcomes may find similar results, whether the outcome measure is achievement or later earnings. Our research, as did the Card and Krueger 1992 study, uses aggregate data on school characteristics to predict student level outcomes for cohorts born in the 1920s, 1930s and 1940s. In so doing, we find effects on achievement, if anything, somewhat larger in magnitude than the effects that Card and Krueger found using earnings as the outcome variable. Card and Krueger's results indicate that a decrease in student–teacher ratio of ten students would, on average, increase income returns to an extra year of education by between 10% and 20%, depending on the cohort. Results presented here estimate that a decrease in student–teacher ratio of 10 students increases the effectiveness of schools on raising achievement scores by over 20%.

We repeat the above analyses for whites only.[16] While the estimates are substantially less consistent than for the entire sample, the general trend for student–teacher ratio to have a negative effect and for term length to have a positive effect is evident. The lack of significant effects when only cohort controls are included, points to the importance of differences in achievement returns to years of education among divisions. Once division controls are included, the estimates are substantially larger than for the full sample and significant at standard levels (the coefficients on school characteristics range from −0.63 to −1.29 for student–teacher ratio and from 0.26 to 0.37 for term length depending on the model). These results indicate that the convergence of the school inputs in the segregated southern school districts is not the only force driving the positive findings on the impact of school inputs on achievement.

Lastly, we find little change in the effects of the measured school inputs when current division of residence is included as a control in the first stage, both linearly and interacted with educational attainment.[17] With this control included, the analysis, in effect, compares people who currently live in the same division but were born in other divisions. These "migrants" may be systematically different from people who remain in the division in which they were born. The results of studies, such as Card and Krueger's (1992), which control for current residence in order to adjust for local labor market effects may, then, be biased from these migration effects. The similarity between estimates that do and do not control for current division of residence in the first-stage would seem to indicate that migration effects are small and do not substantively change the findings. This result may apply especially to Card and Krueger's 1992 paper which uses the same year of birth cohorts and measures of school inputs as those in this study. Our analysis is not a direct test of migration bias for studies such as Card and Krueger's that use earnings as the dependent variable. Migrants may well select to move on the basis of wage differences that we are not be able to capture effectively using

[16] The results of the weighted and unweighted second-stage regressions for whites are reported in appendix tables A2 and A3, respectively.

[17] Appendix table A4 summarizes the second-stage results for models which include current division of residence as controls in the first-stage regressions.

a measure of verbal ability. Yet, our findings suggest that migration bias may not be the sole cause of the positive school input effects found.

IV. Discussion

Our results point to the importance of data characteristics in explaining the conflict in findings between income studies with positive school input effects and achievement studies without consistent significant effects. Data characteristics appear to be more important than differences in the outcome measure, biases from missing family background controls, or biases from labor market influences. Studies finding positive effects of school inputs typically use aggregate data on cohorts educated before 1960, while studies finding no effects tend to use micro-level data on more recent cohorts. Unfortunately, data limitations preclude us from directly assessing the extent to which cohort effects versus aggregation influence the results. While existing evidence supports the notion that both level of aggregation and cohort differences may contribute to the conflict in findings, only the aggregation effect appears potentially sufficient in magnitude to explain the extent of this conflict.

The findings on cohort effects are mixed. Economic historians have produced some direct evidence that school resources affected scholastic achievement during the early part of this century. The difference between these results and the inconclusive results of studies on more recent years suggests possible cohort differences in the effect of school inputs. Schmidt (1995), for example, using data from New York, finds evidence that increases in state funding for local schools during the early 1920s had a positive effect on outcomes. In addition, both Orazem (1987) and Margo (1990) report positive associations between term length and scholastic achievement; though, recall that findings of positive effects of term length should probably not be seen as conflicting with current evidence on achievement.

Non-linearities in the effects of school inputs on student outcomes are a possible explanation for stronger finding using older cohorts. Ferguson (1991) found that increasing student–teacher ratios above 18 had a negative effect on outcomes, but that decreases below this level had no effect. If such threshold effects exist they would provide a possible mechanism to explain systematic variation between results of studies based on older versus more recent cohorts. However, Glass and Smith's 1979 meta-analysis of 77 studies found substantially greater effects of class size among classrooms with less than 20 students than among those with 20–40 students, which would work in the opposite direction, enhancing the contrast between our results and those based on more recent data.[18]

[18] An alternative explanation for the source of possible cohort effects is that the greater variation in school resources that existed at the beginning of the century may make it easier to detect significant effects using historical data. Yet, diminished variation should not bias estimates. In addition, there continues to be considerable variation in inputs across schools and school districts. A more compelling argument for cohort effects, especially when considering comparisons across states, is that the greater variation in school quality may imply that measured inputs were more reliable indicators of school quality at the beginning of the century than they are today.

While the evidence suggesting the importance of cohort differences in explaining the conflicting findings is limited, there is mounting evidence that estimates of the effects of measured school inputs on school effectiveness are substantially larger when researchers use states rather than school districts, schools or classrooms as the unit of analysis (Betts (1996)). The design of several recent studies allows the direct assessment of the importance of aggregation. Each of these studies finds evidence suggesting that aggregation increases the estimated impact of measured school characteristics (Betts (1995), Grogger (1996), Hanushek, Rivkin and Taylor (1996)).

The finding that relationships are stronger at the aggregate level is not specific to this literature. When findings based on aggregate versus micro data conflict, the presumption is usually in favor of the more micro analysis. Yet, estimates based on micro data are not always preferred to ones based on aggregate data (Grunfeld and Griliches (1960)). In the current context, it has been argued that aggregation may decrease the endogeneity problem of parents selecting the schools that their children attend since parents are more likely to choose a district or a particular school than they are to choose the state or division in which they live because of the school resources available (Card and Krueger (forthcoming)). For this reason, state or division level studies may estimate more accurate effects than district or school level analyses. Moreover, aggregation can serve to mitigate the biases due to errors-in-variables. For example, micro-level studies assessing the impact of school inputs tend only to have measures of that input for the particular classroom in the particular year that the data cover. Yet, students attend numerous classrooms during their schooling experience. An average measure of the input for the district, or even for the state, may be a more accurate assessment of the average school inputs received by the student over his/her total schooling than is a single micro-level measure that matches students directly to classrooms at one point in time.

While state and division level studies may have some benefits over more micro-level studies that look at absolute measures of achievement, most recent micro-level studies use value-added measurements, in which prior achievement is controlled for either by introducing measures of prior achievement as a left-hand-side variable or by using gain in achievement from being in a particular classroom with specific inputs as the dependent variable. Studies with this value-added framework appear to be preferable to aggregate studies by most measures. The controls for initial achievement should diminish the endogeneity problem evident with other methods, while data on the specific inputs that the student receives substantially reduce errors-in-variables bias.

While criticisms of the micro studies in the literature are largely addressed by the value-added approach, the problems that plague aggregate studies remain. For example, arguments in favor of using micro-level data are based on the fact that if there are aggregate effects that are correlated with the explanatory variables of interest, aggregation will exacerbate the bias due to the omission of these factors

(Hammond (1973)). In the current context, it is natural to imagine that measured school inputs are endogenous and may serve as proxies for other characteristics of the environment in which students grew up.[19] Increases in the investment in education that occurred in the first half of the twentieth century did not occur randomly, but reflected increased awareness across various groups and organizations of the importance of education (Goldin (1994))!

In this regard, it is important to distinguish two potential aspects of the environment that average school input measures may be capturing. First, school input variables may be picking up features of the communities that are unrelated to the schools themselves. As the southern economy came to more closely resemble the northern economy (Wright (1986)), higher expectations about the future and increased orientation toward the value of academic achievement may have caused the increased test scores, quite apart from direct effects of the schools. On the other hand, differences in measured school inputs may reflect actual school quality difference. Findings of positive school input effects, in this case, would reflect actual school quality differences, though not necessarily the impact of the particular input measure in question. In fact, the crudeness of the school quality measures used in studies such as this and the likely inaccuracies in the Survey of Education data,[20] together with the strong correlation among school inputs, make it hard to imagine that measured differences in such things as the average student–teacher ratio reflect only such differences and not also other school quality differences.[21]

In either case, it may still be appropriate to think of the kinds of school inputs used by Card and Krueger as reflecting the level of human capital investments. In this sense, increases in the length of the school year or decreases in the pupil-teacher ratio that occurred over the first half of the twentieth century are part of the dramatic increases in human capital investment that were occurring over this period of time. There would seem to be little doubt that, in general, such investment contributed importantly to economic growth (Griliches (1970), Denison (1985)). However, while aggregate studies can point to the importance of human capital investments, such studies *cannot* provide the detail on the benefits of specific school inputs that is needed for informed policy decisions.

[19] For example, some quality of the state that one is raised in may both affect student outcomes and be collinear with school inputs. If a state or region has a culture in which the importance placed on education and academic achievement is growing over time, then citizens of that state are likely to vote to increase teacher salaries and the length of the school term. However, if students educated in this region after the growth have higher achievement or greater later earnings than earlier cohorts, it is impossible to separate out the effect of increased emphasis on education from the effect of increased school inputs.

[20] Robert Margo (in personal communications, February 1995) cited such data problems as double counting of students and teachers who move across districts and unweighted aggregation of district level data.

[21] For example, given the high correlation between our two school quality measures, it seems plausible that both the student-teacher ratio and term length are proxying for the time spent on learning, "time on task."

REFERENCES

Akin, John S., and Irwin Garfinkel, "School Expenditures and the Economic Returns to Schooling," *The Journal of Human Resources* 12 (1977), 460–481.

Alwin, Duane F., "Family of Origin and Cohort Differences in Verbal Ability," *American Sociological Review* 56 (Oct. 1991), 625–638.

Amemiya, Takeshi, "A Note on a Random Coefficients Model," *International Economic Review* 19 (1978), 793–796.

Betts, Julian, "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," this REVIEW 77 (May 1995), 231–247.

———, "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in Gary Burtless (ed.), *Does Money Matter? The Effects of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution, 1996).

Boozer, Michael, and Cecilia Rouse, "Intraschool Variation in Class Size: Patterns and Implications," Princeton Industrial Relations Section Working Paper No. 344 (1995).

Borjas, George J., "On Regressing Regression Coefficients," *Journal of Statistical Planning* 7 (1982), 131–137.

Burtless, Gary "Does Money Matter? The Effects of School Resources on Student Achievement and Adult Earnings," in Gary Burtless (ed.), *Does Money Matter? The Effects of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution, 1996).

Card, David, and Alan Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100 (1992), 1–40.

———, "Chapter 6, The Economic Return to School Quality," in William E. Becker and William J. Baumol (eds.), *Assessing Educational Practices: The Contribution of Economics* (Cambridge, MA: MIT Press, forthcoming).

Chesher, Andrew, and Ian Jewitt, "The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator," *Econometrica* 55 (1987), 1217–1222.

Childs, T. Stephen, and Charol Shakeshaft, "A Meta-Analysis of Research on the Relationship between Educational Expenditures and Student Achievement," *Journal of Education Finance* 12 (Fall 1986), 249–263.

Coleman, James S., et al., *Equality of Educational Opportunity* (Washington, D.C.: Government Printing Office, 1966).

Davis, James A., and Tom W. Smith, *The NORC General Social Survey: A User's Guide* (Newbury Park: Sage, 1992).

Denham, Carolyn, and Ann Lieberman (eds.), *Time To Learn* (California Commission for Teacher Preparation and Licensing, 1980).

Denison, Edward F., *Trends in American Economic Growth, 1929–1982* (Washington, D.C.: The Brookings Institution, 1985).

Ehrenberg, Ronald G., and Dominic J. Brewer, "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review* 13 (1994), 1–17.

Ferguson, Ronald F., "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal on Legislation* 28 (1991), 465–498.

Finn, Jeremy D., and Charles M. Achilles, "Answers and Questions about Class Size: A Statewide Experiment," *American Educational Research Journal* 27 (Fall 1990), 557–577.

Fuchs, Victor A., and Diane M. Reklis, "Mathematical Achievement in Eighth Grade: Interstate and Racial Differences," National Bureau of Economic Research Working Paper No. 4784 (1994).

Glass, Gene V., and M. L. Smith, "Meta-Analysis of Research on Class Size and Achievement," *Educational Evaluation and Policy Analysis* 1 (1979), 2–16.

Goldin, Claudia, "How America Graduated From High School: 1910 to 1960," National Bureau of Economic Research Working Paper No. 4762 (1994).

Griliches, Zvi, "Notes on the Role of Education in Production Functions and Growth Accounting," in W. Lee Hansen (ed.), *Education, Income and Human Capital: Studies in Income and Wealth* 35 (New York: Columbia University Press for the National Bureau of Economic Research, 1970), 71–114.

Griliches, Zvi, and William M. Mason, "Education, Income and Ability," *Journal of Political Economy* 80 (Spring 1972), 74–103.

Grogger, Jeff, "School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond," this REVIEW 78 (Nov. 1996).

———, "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* (forthcoming).

Grunfeld, Yehuda, and Zvi Griliches, "Is Aggregation Necessarily Bad?" this REVIEW 42 (Feb. 1960), 1–13.

Hammond, John L., "Two Sources of Error in Ecological Correlation," *American Sociological Review* 38 (Dec. 1973), 764–777.

Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24 (Sept. 1986), 1141–1147.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor, "Aggregation and the Estimated Effects of School Resources," this REVIEW 78 (Nov. 1996).

Heckman, James J., Anne Layne-Farrar, and Petra Todd, "The Schooling Quality–Earnings Relationship: Using Economic Theory to Interpret Functional Forms Consistent with the Evidence," unpublished manuscript, University of Chicago (June 1995).

Hedges, Larry V., Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher* 23 (1994), 5–14.

Johnson, George E., and Frank F. Stafford, "Social Returns to Quantity and Quality of Schooling," *Journal of Human Resources* 8 (Spring 1973), 139–154.

Link, Charles R., and Edward C. Ratledge, "Social Returns to Quantity and Quality of Education: A Further Statement," *Journal of Human Resources* 10 (1975), 78–89.

Margo, Robert A., *Race and Schooling in the South 1880–1950: An Economic History* (Chicago: The University of Chicago Press, 1990).

Miner, John B., *Intelligence in the United States* (New York: Springer Publishing Company, 1957).

Odden, Allan, "Class Size and Student Achievement: Research-Based Policy Alternatives," *Educational Evaluation and Policy Analysis* 12 (Summer 1990), 213–227.

Orazem, Peter, "Black-White Differences in Schooling Investment and Human Capital Production in Segregated Schools," *American Economic Review* 77 (Sept. 1987), 714–723.

Rizzuto, Ronald, and Paul Watchel, "Further Evidence on the Returns to School Quality," *Journal of Human Resources* 15 (1980), 240–272.

Schmidt, Stefanie, "Do School Inputs Matter? Historical Evidence from New York State," unpublished manuscript, Massachusetts Institute of Technology (Feb. 1995).

Thorndike, Robert L., "Two Screening Tests of Verbal Intelligence," *Journal of Applied Psychology* 26 (1942), 128–135.

Thorndike, Robert L., and George H. Gallup, "Verbal Intelligence of the American Adult," *Journal of General Psychology* 30 (1944), 75–85.

Wright, Gavin, *Old South, New South: Revolutions in the Southern Economy Since the Civil War* (New York: Basic Books, 1986).

APPENDIX

TABLE A1.—MEANS AND STANDARD DEVIATIONS OF ANALYSIS VARIABLES BY COHORT AND DIVISION

Division/Cohort	Cohort	Test Score	Education	Age	Percent Black	Percent Female	Parents' Education	5 Levels Income	No. of Children in Family	Single Parent
New England	1	6.68 (2.09)	12.06 (2.95)	58.94 (6.55)	0.80	65.81	8.60 (3.73)	2.69 (0.77)	5.24 (3.08)	15.04
	2	7.08 (2.28)	13.53 (2.66)	48.53 (6.52)	1.20	57.83	10.34 (3.33)	2.96 (0.74)	3.89 (2.88)	16.05
	3	6.91 (2.08)	13.69 (2.94)	38.08 (6.41)	3.64	55.45	10.70 (3.06)	2.83 (0.80)	3.85 (2.14)	12.84
Middle Atlantic	1	6.56 (2.14)	12.15 (3.10)	58.87 (6.84)	5.28	53.47	8.57 (3.50)	2.66 (0.93)	4.75 (3.07)	16.16
	2	6.85 (1.97)	12.91 (2.90)	47.26 (6.82)	7.19	55.23	9.65 (3.01)	2.82 (0.79)	4.30 (3.15)	16.61
	3	6.97 (1.94)	13.67 (2.57)	37.60 (6.54)	8.50	57.29	10.97 (2.62)	2.86 (0.79)	3.93 (2.56)	13.51
East North Central	1	6.39 (2.02)	12.19 (2.91)	57.26 (6.76)	8.04	56.57	8.72 (3.11)	2.62 (0.86)	5.04 (3.13)	16.21
	2	6.20 (2.00)	12.57 (2.48)	47.98 (6.82)	8.00	56.86	10.02 (2.68)	2.71 (0.83)	4.87 (3.09)	14.93
	3	6.56 (1.89)	13.56 (2.46)	38.38 (6.65)	14.03	52.46	10.68 (2.84)	2.81 (0.76)	4.67 (2.75)	12.50
West North Central	1	6.30 (2.02)	12.45 (2.81)	59.15 (6.44)	4.79	56.91	9.18 (3.15)	2.65 (0.82)	5.20 (3.45)	12.09
	2	6.53 (2.09)	13.12 (2.79)	48.71 (6.42)	9.32	53.42	9.85 (3.13)	2.63 (0.86)	4.81 (3.06)	19.62
	3	6.76 (1.86)	13.56 (2.58)	37.84 (6.99)	5.15	58.80	10.99 (2.85)	2.82 (0.88)	4.47 (2.78)	13.79
South Atlantic White	1	5.67 (2.15)	11.18 (3.55)	58.81 (6.83)	0.00	62.38	8.50 (4.13)	2.63 (0.83)	5.68 (3.25)	11.73
	2	5.67 (1.94)	11.76 (3.14)	47.19 (6.86)	0.00	56.41	8.77 (3.68)	2.70 (0.77)	5.20 (3.44)	14.51
	3	6.37 (2.03)	12.96 (2.89)	37.91 (6.57)	0.00	54.04	10.39 (3.60)	2.80 (0.80)	4.50 (2.94)	13.31
South Atlantic Black	1	4.03 (2.04)	9.51 (3.81)	58.32 (6.17)	100.00	57.97	7.44 (3.85)	2.08 (0.89)	6.59 (3.90)	37.88
	2	4.79 (2.04)	11.01 (3.23)	48.85 (6.25)	100.00	54.88	7.18 (3.37)	2.10 (0.86)	7.45 (5.13)	34.57
	3	4.66 (2.06)	12.42 (2.72)	38.57 (6.54)	100.00	60.18	8.40 (3.30)	2.45 (0.87)	6.79 (4.21)	36.04
East South Central White	1	5.02 (2.08)	10.41 (3.19)	60.00 (6.79)	0.00	63.93	7.56 (3.48)	2.61 (0.89)	6.18 (3.39)	17.50
	2	5.62 (2.28)	12.14 (3.10)	48.75 (6.77)	0.00	58.82	8.87 (3.10)	2.68 (0.85)	5.34 (3.48)	14.00
	3	5.59 (2.08)	12.36 (2.90)	38.34 (7.17)	0.00	59.33	9.38 (3.22)	2.73 (0.73)	4.97 (3.50)	8.78
East South Central Black	1	3.92 (1.96)	9.72 (3.70)	59.49 (5.74)	100.00	59.02	7.02 (3.64)	2.13 (1.01)	6.62 (4.36)	39.66
	2	4.28 (1.93)	10.78 (2.96)	49.23 (5.94)	100.00	65.00	7.21 (3.09)	1.98 (0.85)	7.23 (4.82)	31.66
	3	4.77 (1.96)	12.60 (2.81)	39.17 (6.20)	100.00	63.33	8.17 (3.92)	2.07 (0.92)	7.30 (3.87)	30.51
West South Central White	1	5.92 (2.12)	11.89 (3.16)	58.39 (7.06)	0.00	57.26	8.41 (3.85)	2.59 (0.88)	5.21 (3.11)	23.08
	2	5.89 (2.40)	12.10 (3.54)	48.24 (7.14)	0.00	57.02	7.90 (3.93)	2.58 (0.92)	5.41 (2.99)	14.53
	3	5.96 (2.31)	12.75 (3.07)	38.14 (6.71)	0.00	56.12	9.49 (4.17)	2.78 (0.94)	4.77 (3.01)	13.92
West South Central Black	1	3.67 (2.20)	9.37 (3.50)	57.70 (7.36)	100.00	60.00	6.90 (3.42)	2.33 (0.80)	7.07 (3.72)	37.93
	2	4.67 (1.75)	11.46 (2.87)	48.31 (6.83)	100.00	64.10	7.55 (4.09)	2.23 (0.99)	7.50 (3.65)	28.21
	3	5.10 (1.91)	12.31 (2.60)	37.43 (5.95)	100.00	69.05	7.68 (3.23)	2.19 (0.86)	8.67 (4.32)	31.71
Mountain	1	6.18 (2.54)	12.02 (3.28)	59.93 (6.16)	0.00	70.18	9.34 (3.43)	2.79 (0.88)	5.79 (3.69)	7.14
	2	6.12 (2.36)	12.56 (3.87)	48.05 (7.51)	1.82	52.73	8.87 (2.97)	2.51 (0.80)	6.07 (3.91)	15.69
	3	6.51 (1.83)	13.43 (2.39)	39.49 (6.87)	1.22	50.00	10.79 (2.92)	2.74 (0.86)	5.46 (3.61)	8.97
Pacific	1	6.72 (2.03)	12.59 (2.62)	57.51 (6.00)	1.69	55.08	9.60 (3.24)	2.76 (0.87)	4.53 (2.99)	25.86
	2	6.34 (2.21)	13.00 (2.69)	47.84 (7.57)	6.72	46.27	9.68 (3.28)	2.76 (0.76)	4.35 (2.81)	21.05
	3	6.73 (2.02)	13.70 (2.75)	38.22 (6.96)	6.83	55.82	10.95 (3.07)	2.88 (0.83)	4.33 (2.64)	16.26
Overall		6.21 (2.16)	12.60 (3.03)	46.83 (10.46)	13.98	56.66	9.59 (3.42)	2.69 (0.85)	4.99 (3.28)	16.66

TABLE A2.—SECOND STAGE REGRESSION RESULTS, WHITES ONLY: UNWEIGHTED

	Student–Teacher Ratio/100		Term Length/100	
	No Family Background Controls			
No Controls	−0.10	0.01	0.09	0.03
	(0.17)		(0.06)	
Cohort Controls	0.05	0.12	0.05	0.13
	(0.21)		(0.05)	
Division Controls	−0.70	0.38	0.37	0.39
	(0.23)		(0.09)	
Cohort and Division	−1.29	0.46	0.36	0.45
	(0.50)		(0.11)	
	Linear Family Background Controls			
No Controls	−0.11	0.01	0.08	0.03
	(0.17)		(0.06)	
Cohort Controls	0.01	0.13	0.05	0.14
	(0.20)		(0.04)	
Division Controls	−0.66	0.44	0.31	0.41
	(0.20)		(0.09)	
Cohort and Division	−1.27	0.54	0.29	0.48
	(0.42)		(0.11)	
	All Family Background Controls			
No Controls	−0.11	0.01	0.07	0.02
	(0.16)		(0.06)	
Cohort Controls	0.00	0.14	0.04	0.14
	(0.18)		(0.04)	
Division Controls	−0.63	0.46	0.28	0.43
	(0.18)		(0.09)	
Cohort and Division	−1.23	0.58	0.26	0.51
	(0.35)		(0.11)	

Note: The top left hand number for each set of three numbers is the coefficient on the school quality measure in the second stage regression. The bottom left hand number is the White's estimate of the standard error. The right hand side number is the R^2 for the regression. The sample size for each regression is 27.

TABLE A3.—SECOND STAGE REGRESSION RESULTS, WHITES ONLY: WEIGHTED

	Student–Teacher Ratio/100		Term Length/100	
	No Family Background Controls			
No Controls	−0.18	0.03	0.12	0.06
	(0.16)		(0.06)	
Cohort Controls	0.09	0.22	0.06	0.22
	(0.18)		(0.05)	
Division Controls	−0.70	0.47	0.40	0.50
	(0.23)		(0.09)	
Cohort and Division	−0.73	0.49	0.30	0.55
	(0.46)		(0.11)	
	Linear Family Background Controls			
No Controls	−0.21	0.05	0.12	0.07
	(0.17)		(0.06)	
Cohort Controls	0.01	0.15	0.07	0.23
	(0.18)		(0.05)	
Division Controls	−0.67	0.55	0.34	0.53
	(0.18)		(0.09)	
Cohort and Division	−0.81	0.58	0.23	0.59
	(0.36)		(0.10)	
	All Family Background Controls			
No Controls	−0.21	0.05	0.10	0.06
	(0.16)		(0.06)	
Cohort Controls	0.00	0.21	0.05	0.22
	(0.18)		(0.05)	
Division Controls	−0.64	0.57	0.31	0.53
	(0.16)		(0.09)	
Cohort and Division	−0.87	0.61	0.20	0.60
	(0.31)		(0.10)	

Note: The top left hand number for each set of three numbers is the coefficient on the school quality measure in the second stage regression. The bottom left hand number is the White's estimate of the standard error. The right hand side number is the R^2 for the regression. The sample size for each regression is 27.

TABLE A4.—SECOND STAGE RESULTS INCLUDING CURRENT DIVISION OF RESIDENCE IN THE FIRST STAGE

	Student–Teacher Ratio/100		Term Length/100	
	No Family Background Controls			
No Controls	−0.39	0.27	0.18	0.28
	(0.15)		(0.07)	
Cohort Controls	−0.37	0.28	0.18	0.30
	(0.16)		(0.07)	
Division Controls	−0.41	0.62	0.15	0.60
	(0.18)		(0.08)	
Cohort and Division	−0.38	0.64	0.11	0.63
	(0.28)		(0.09)	
	Linear Family Background Controls			
No Controls	−0.35	0.24	0.17	0.27
	(0.15)		(0.07)	
Cohort Controls	−0.33	0.26	0.16	0.29
	(0.17)		(0.07)	
Division Controls	−0.42	0.61	0.15	0.59
	(0.20)		(0.09)	
Cohort and Division	−0.41	0.64	0.11	0.62
	(0.29)		(0.10)	
	All Family Background Controls			
No Controls	−0.21	0.11	0.12	0.16
	(0.15)		(0.07)	
Cohort Controls	−0.16	0.17	0.10	0.20
	(0.16)		(0.07)	
Division Controls	−0.40	0.53	0.14	0.50
	(0.20)		(0.09)	
Cohort and Division	−0.41	0.56	0.11	0.54
	(0.30)		(0.10)	

Note: The top left hand number for each set of three numbers is the coefficient on the school quality measure in the second stage regression. The bottom left hand number is the White's estimate of the standard error. The right hand side number is the R^2 for the regression. The sample size for each regression is 36.