MODELS OF THE RELATIONSHIP BETWEEN EDUCATION SPENDING AND THE SOCIAL DISTRIBUTION OF ACHIEVEMENT

Harold Wenglinsky

September 1998
Models of the Relationship between Education Spending
and the Social Distribution of Achievement

Harold Wenglinsky

Educational Testing Service
Abstract

This study relates various types of educational spending to mean mathematics achievement and its social distribution among students. A nationally representative database of 7,217 twelfth graders was developed by linking the 1992 National Assessment of Educational Progress in mathematics to the U.S. Common Core of Data and the School District Data Book. Through applying Hierarchical Linear Modeling to the database, the study found that spending on instruction and capital expenditures, while not related to mean achievement levels, were related to differences in achievement between SES groups; lower spending levels are associated with greater achievement gaps within schools.

Keywords: Production functions; school finance; multilevel modeling; National Assessment of Educational Progress
One of the most common proposals to promote equity in education is finance equalization. "Equity," for purposes of this paper, refers to a situation in which students' educational outcomes are affected as little as possible by their socio-economic status (SES). The finance equalization movement arose in the late 1960s in response to the financial inequalities inherent in school funding systems. The bulk of education was financed through local property taxes, which depend upon the level of wealth in the school district. This wealth level, in turn, depended to some extent on the relative affluence or poverty of the residents of the district, as well as on the value of commercial property. This meant that the quality and level of services school districts could afford could be related to the SES of students' families; students from poor families might thus have fewer educational opportunities than those from affluent families.\(^1\) The finance equalization movement, through litigation and other forms of public pressure, called upon states to increase the flow of state dollars to low-spending school districts in order to reduce spending disparities. A series of court cases ensued, leading state legislatures to increase these funds, known as equalization funds. This movement has not been without effect. Currently, states pay 45.6% of school spending as opposed to 38.3% in 1970 (General Accounting Office, 1995). This change in the financing of education may have reduced disparities in resources between school districts, by making those resources less dependent upon local wealth (General Accounting Office, 1997).

Ironically, although finance equalization is well known for promoting equity, it concerns itself only with inequities between school districts, ignoring inequities within school districts, let alone schools. Reducing differences in spending between school
districts is intended to reduce differences in educational quality and thus in student achievement between school districts; student achievement would be less dependent upon the affluence or poverty of students’ parents. Yet finance equalization ignores disparities within schools, that is, disparities in educational quality among classrooms in the same schools and academic achievement among students in the same school. The level of spending a district allocates to a given school may have ramifications for the distribution of resources and achievement within it. It may be that schools that have additional resources due to finance equalization spend these funds on honors programs or Advanced Placement courses, items that might increase average student achievement in the school but also increase disparities between high- and low-SES students. On the other hand, it may be that schools with additional resources are better able to serve all of their students, thus reducing achievement differences between low and high SES students.

The current study assesses the consequences of finance equalization on both between-district and within-school equity. The novel nature of the study involves the fact that it applies multilevel statistical techniques to a nationally representative sample of twelfth graders to measure the relationship between various types of spending (instruction, administration, etc.) and both mean school achievement and the social distribution of achievement within schools. Before discussing the study, however, it is worthwhile to review the results of previous work.
Review of the Literature

While there is no prior research on the relationship between school district spending and the social distribution of achievement within schools, there is a great deal on the relationship between school district spending and mean school achievement. This research, known as production function research, uses regression analysis to relate spending and resources to student achievement while controlling for student background characteristics. The studies indicate that spending matters in some cases but not others.

The earliest major study of the spending-achievement relationship was the Equality of Educational Opportunity Study, otherwise known as the Coleman Report (Coleman et al., 1966). That study collected and analyzed data on a nationally representative sample of students and schools. It found that, by and large, when student SES and other background characteristics are taken into account, aggregate per-pupil spending is not significantly related to achievement. The study did, however, analyze the spending-achievement relationship for various regional and ethnic subgroups of students, and found that some resources did evince strong associations with achievement for some subgroups. Reanalyses of the EEOS data found that resources were particularly important for African-American students in the South (Armor, 1972). Nonetheless, the Coleman Report itself concluded that, for the most part, student background rather than school inputs was the key influence on academic achievement.

Over the course of the next thirty years, researchers sought to measure the relationship between spending and achievement by conducting smaller-scale studies of a
particular state or school system. The results of these studies have been meta-analyzed by Hanushek (1997, 1996, 1994, 1989) and Hedges and his colleagues (Hedges, Laine & Greenwald, 1994; Greenwald, Hedges, & Laine, 1996; Hedges & Greenwald, 1996) with the two arriving at opposite conclusions. Hanushek identified the different spending and resource measures included in the production functions, and counted the number of each that were positively and significantly related to achievement, positively and non-significantly related to achievement, negatively and non-significantly related to achievement, negatively and significantly related to achievement, and non-significantly related to achievement but of unknown sign. In his most recent tally, Hanushek (1997) identified 163 studies that included a measure of expenditures per pupil. Of these, 27% were positive and statistically significant, 7% were negative and statistically significant, and 66% were non-significant. Hanushek concluded that "[t]here is no strong or consistent relationship between school resources and student performance (1997:148)."

Hedges and his colleagues analyzed a somewhat different sample of production function studies. They also tallied different types of expenditures and resources. They found, in their most recent tally of spending-achievement relationships (Greenwald, Hedges & Laine, 1996), that 44% were positive and significant, 3% were negative and significant, and 53% were non-significant. They also performed a combined significance test on the 27 studies of per-pupil expenditures for which sufficient information was available, and found the hypothesis of a positive relationship supported and that of a negative relationship unsupported. Finally, they calculated the median regression coefficients for the production function studies of spending and found a coefficient of .0003, meaning that
a $1,000 increase in spending translates into a .3 standard deviation increase in achievement. They concluded that spending is related to achievement.

The Coleman Report and the small-scale studies that were meta-analyzed, however, suffer from methodological problems that make it difficult to determine from them whether or not spending is related to achievement. First, these studies used only aggregate per-pupil expenditures as their measure of school district spending; they did not attempt to distinguish between different types of spending. To the extent that some types of spending are significantly related to achievement and others are not, the effects of those that are could appear non-significant when combined with non-significant types of spending in a single aggregate measure. Second, these studies did not adjust for regional variations in the cost of education. It requires more dollars to buy the same services in some regions than it does in others; a dollar may go further in Montgomery, AL than in New York City (Barton, Goertz & Coley, 1991). Consequently, variations in achievement that are attributed to spending may be due to variations in the cost of education. Third, the studies after the Coleman Report tended to be small in scope, studying only a state or a school system. To the extent that the effect of spending on achievement differs from state to state, these small-scale studies might come to different and apparently contradictory conclusions (see Fortune & O’Neil, 1994; Monk, 1992 for problems with production function studies).

Wenglinsky (1997a, 1997b) attempted to address these problems in studies of fourth and eighth graders. In both studies, Wenglinsky applied structural equation modeling techniques to relate spending to achievement in mathematics. Rather than using
an aggregate measure of spending, he distinguished between spending on instruction, central office administration, principal's office administration and capital outlays. He also adjusted these expenditure measures by the cost of education, using the Teacher Cost Index (see below for a discussion of the index). The database he analyzed was national in scope, consisting of nationally representative samples of fourth and eighth graders who had taken the National Assessment of Educational Progress (NAEP) in mathematics. Wenglinsky found that some expenditures mattered and some did not. Instructional and central office spending were positively related to student achievement; principal’s office and capital spending were not.

From this series of studies, it can be seen that existence of a spending-achievement relationship depends on many factors. The meta-analyses do not show either that money matters always or that it never matters; rather, they show that money matters sometimes. This finding leaves unanswered the question of under what circumstances the relationship exists. The Wenglinsky studies suggest that one such circumstance is the type of spending being measured; some types have an effect and some do not. Wenglinsky also suggests that the spending-achievement relationship may depend upon the developmental level of the students. Spending was more strongly related to achievement for fourth graders than for eighth graders. The fit between spending and educational practices may also influence the spending-achievement relationship. Murnane and Levy (1996), for instance, found that when 15 schools were given additional resources to hire teachers, these resources only led to achievement increases in the two districts where they were accompanied by other educational reforms designed to take advantage of the smaller classes.
These production function studies measure the association between spending variations between school districts and achievement variations between schools. They do not, however, examine associations between spending variations between school districts and variations in the distribution of achievement within schools. Yet research has shown that there are inequities within schools that are just as significant as inequities between schools. Not only do students of low-SES tend to go to lower quality schools, but within those schools they can receive a lower quality education than that of high-SES students in the same schools. Tracking research indicates that schools are often organized into academic, general and vocational course sequences, and that low-SES students tend to be steered away from the academic sequence. Yet it is the academic sequence that demands rigorous coursework and encourages high achievement. Curricular organization, then, can strengthen the tendency of low-SES students to perform less well than their counterparts rather than weaken it (Spade, Columba & Vanfossen, 1997; Oakes & Guiton, 1995; Jones, Vanfossen & Ensminger, 1995; Gamoran & Berends, 1987; Oakes, 1985). Research on Catholic schools suggests that precisely because these schools place almost all of their students in the academic track, they evince less of an association between SES and student achievement than do public schools (Bryk, Lee & Holland, 1993; Gamoran, 1992).

Tracking is not the only traditional school practice that increases the SES gap in achievement. Other policies, such as departmental organization of subjects and competitive (as opposed to cooperative) classrooms also reinforce the tendency of low-SES students to lag behind their high-SES counterparts (Lee, Smith & Croninger, 1997; Lee & Smith, 1995). Schools seem to be able to affect the social distribution of
achievement within them; it is therefore important to examine the role that school funding plays in this process.

Hypotheses, Data, Method

Hypotheses

The current study analyzes a nationally representative database of twelfth graders to relate four types of school district spending to mean school achievement levels and the social distribution of achievement within schools. Two hypotheses are tested. First it is hypothesized that none of the types of spending are related to mean achievement. The ability of schools to change student achievement declines as students age; the earlier the intervention, the greater the results. This notion was confirmed for the relationship between spending and mean achievement levels in studies of fourth and eighth graders (Wenglinsky, 1997a, 1997b). The association between spending types and achievement was lower for eighth graders than it was for fourth graders. It is therefore likely that for twelfth graders, the association between mean district spending and mean achievement will be still lower.

Second, it is hypothesized that some types of spending will have an effect on the social distribution of achievement. School districts, like all organizations, engage in triaging behavior. When schools are functioning well, they are able to provide a rigorous education to all students. A lack of funds, however, creates a situation in which schools must reduce support for some of their programs; the result is a situation in which students
are left to their own devices to pursue their education. The less affluent students may "fall through the cracks," increasing disparities between their performance and that of the more affluent students. Possible explanations of how this process unfolds in schools are presented at the end of this article.

Data

To test these hypotheses it is necessary to have data on various types of school district spending, student achievement, student SES, and the appropriate statistical controls. Unfortunately, no single nationally representative database includes all of these variables. A database was therefore synthesized from four sources: the National Assessment of Educational Progress (NAEP), the U.S. Common Core of Data (CCD), the Teacher Cost Index (TCI) and the School District Data Book (SDDB).

NAEP provides information on student achievement and SES. It is a nationally representative database of students and schools, collected by the Educational Testing Service under a contract from the National Center for Education Statistics (NCES). It is administered every two years to nationally representative samples of fourth, eighth and twelfth graders and their teachers and principals. The subject areas tested vary, but have included at one time or another mathematics, reading, writing, history, geography, civics and science. The information collected is used to assess what students around the country know, to compare levels of knowledge of various ethnic, socio-economic and gender subgroups; and to measure the progress of students in the nation, both over time and between grades (see Johnson, 1994 for an overview of NAEP and Mullis et. al., 1993 for
a report on the 1992 mathematics assessment. Previous studies of school spending utilized samples of fourth and eighth graders who took the 1992 mathematics assessment. The current study analyzes twelfth graders who took the 1992 mathematics assessment. The examination provides information on student achievement and a background questionnaire administered to students provides information on student SES.

CCD includes data on school district spending. It is a survey of financial and demographic information administered to the universe of public school districts every year by the U.S. Department of Education. Although the information provided can be used to measure district by district per-pupil expenditures in broad spending categories, such as instruction or central office administration, CCD cannot be relied upon for more detailed information because differences in the charts of accounts of school districts result in their categorizing specific expenses in different ways. Therefore, CCD was used to provide measures of spending on instruction, central office administration, principal’s office administration and capital outlays. It was used even though the school district was the lowest level of aggregation because such information is not available at the school level.

The TCI provides an adjustment to the per-pupil expenditure measures for regional variations in the cost of education. It is the result of a study, conducted by NCES (1995b), to develop an index of the cost for a particular state of hiring teachers. This cost can be expected to vary geographically, even for teachers of similar levels of education and experience, because the cost of living, quality of life and other factors that potentially affect the cost of hiring teachers vary in this way. The TCI was developed by applying regression analysis to the Schools and Staffing survey, a NCES survey conducted in 1990-
91. The regression analysis estimates the influence of various factors on teacher salaries, including factors under the control of schools and school districts, such as teachers' experience and education, and factors not under their control, including the cost of living and quality of life. The resulting estimates of the impact of these non-discretionary characteristics on teacher salaries can then be used as estimates of the cost of teachers in a particular region, holding the discretionary factors constant. TCI scores for each state were used in this analysis to adjust the per-pupil expenditure measures for geographical variations in the cost of education.5

The school district database provides demographic information on school districts that can be used as statistical controls in the analysis. SDDB was developed over a six year period through the efforts of NCES, the MESA Group, the Census Bureau and the Council of Chief State School Officers (CCSSO). It began in 1988 when the CCSSO and NCES developed maps for 15,274 school districts in the United States. The Census Bureau conducted a school district special tabulation and transferred the data to the appropriate school districts based upon the maps. The MESA group then assembled this information into databases and added information from the 1989-90 Common Core of Data and the 1989-90 Survey of School District Finances. The resulting database includes up to 200,000 data items for each school district or county (National Center for Education Statistics, 1994). From these items, indicators of the degree to which a school district was in an urban area and its median personal income were chosen to act as controls for the school district financial information from CCD.
To analyze these databases, the relevant variables had to be linked to one another for the same set of cases. Since all databases were universes except for NAEP, the resulting database had the sampling characteristics of NAEP. First, NAEP was linked to CCD. For the 9,499 NAEP twelfth graders, each was matched to a CCD school district with the FIPS ID, which is an NCES assigned identification number generated for CCD and included with NAEP. For students in the remaining 56 NAEP school districts that could not be matched in this way, Westat, Inc. identified those who were attending private or Catholic schools (students in 48 districts) and for those in the remaining eight provided the addresses of their school districts. These eight districts were then matched to a CCD school district using common address information. The state of each district was identified, and the TCI estimate for that state entered manually into the database. The resulting database was linked to SDDB using the FIPS ID and address information, since SDDB had these pieces of information as well. Six districts were eliminated due to missing SDDB data, leaving 182 school districts with 7,217 students. The final database included the following variables (see Appendix A for descriptions):

1. Mathematics achievement
2. Student SES
3. Capital per-pupil expenditures
4. Central office administration per-pupil expenditures
5. Instructional per-pupil expenditures
6. Principal’s office administration per-pupil expenditures
7. District median personal income
8. Urbanness

Method

Hierarchical Linear Modeling (HLM) was employed in the current study for two reasons. First, the study seeks to measure "school district effects" on student achievement. Multilevel techniques such as HLM provide better information on school district effects than conventional techniques such as regression analysis (Raudenbush & Willms 1995). Studies of school effects measure relationships between variables at two levels of analysis: the school and the student. Students are nested within schools. Conventional regression analysis can measure such relationships either by aggregating student-level variables into school-level means, or by treating each school-level variable as a student characteristic. In the former case, important information about variation among students is lost and the possibility of aggregation bias is raised; in the latter case, observations are treated as independent when in fact observations of students in the same school are highly dependent upon one another. This can result in inappropriately small standard error estimates, which treat school residuals as independent for each student when, in fact, they are highly dependent. HLM is preferable to conventional regression analysis for the study of school district effects because it explicitly distinguishes between student- and school-level variables in its models (Bryk & Raudenbush 1992). Separate equations are estimated for the effect of student-level variables on students and of school-level variables on the average of student-level variables. Error terms are partitioned into within- and between-school variance.
The relationship between a single student-level outcome and a single school-level input can be modeled and estimated in two equations as follows:

(1) \[ Y_y = \beta_{0j} + r_y \]

(2) \[ \beta_{0j} = \gamma_{00} + \gamma_{01} W + u_{0j} \]

where \( Y_y \) is the outcome to be modeled for the ith student at the jth school;

\( \beta_{0j} \) is the mean level of the outcome in school j;

\( r_y \) is the error unique to student i in school j;

\( \gamma_{00} \) is the mean outcome absent school effect \( W \);

\( \gamma_{01} \) is the effect of an increment change in \( W \) on an increment change in \( \beta_{0j} \);

and \( u_{0j} \) is the error unique to the school j conditional on school effect \( W \) for mean achievement.

It can be seen that the two equations allow data to be modeled explicitly as student and school-level effects, and with distinct error terms for each level of analysis (\( u_{0j} \) and \( r_y \)).

Second, the study is not only interested in the effect of school district characteristics on an outcome (achievement) but also their effect on the relationship between a student-level characteristic and that outcome (the slope of the SES-achievement relationship). Conventional techniques do not permit the simultaneous measurement of intercepts and slope as outcomes. HLM makes this possible by allowing multiple independent variables in the student-level equation (1) and including an additional equation in which the independent variable parameter is an outcome (Bryk & Raudenbush, 1992). That is, equation (1) is modified to include an additional term,
(1a) \[ Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \]

where \( \beta_{1j} \) is the effect of an increment change in \( X_{ij} \) on an increment change in \( Y_{ij} \).

and \( X_{ij} \) is a student-level characteristic,

and a new equation is added,

(3) \[ \beta_{1j} = \gamma_{10} + \gamma_{11} W + u_{ij} \]

where \( \gamma_{10} \) is the mean SES-achievement relationship absent school effect \( W \)

\( \gamma_{11} \) is the effect of an increment change in \( W \) on an increment change in \( \beta_{1j} \).

and \( u_{ij} \) is the error unique to the school \( j \) conditional on school effect \( W \) for the SES-achievement relationship.

It can be seen that the three equations permit the student-level relationship to be treated as an outcome; the error terms for the three equations can be estimated separately as well \((r_{ij}, u_{0j}, u_{ij})\).

The current analysis applies HLM in three steps. First, the database was split into student- and school-level variables.\(^7\) The student-level database consisted of the FIPS identifiers, the SES measure produced from the six SES indicators in NAEP, and five "plausible values" of student achievement.\(^8\) The school-level variables consisted of the FIPS identifier, the four measures of per-pupil expenditures, median personal income and urbanness. There were no missing values for the school-level data. For student-level data, only the SES measure had missing values, and for these the mean was substituted.\(^9\)

In addition, the student-level database included a weighting variable. Since NAEP is a stratified clustered sample, students in the population have unequal probabilities of
selection. Consequently, a student weight was calculated as the inverse of the probability of selection of a given student. Second, data were combined into a system file for analysis. Cases were weighted by the student base weight, thus producing estimates analogous to those that would be produced in a sample with equal probabilities of selection (The weight also takes into account student nonresponse and poststratification; Johnson, 1989; Johnson, Rust & Wallace, 1994). Descriptive statistics were also calculated for the variables. Third, a series of HLMs was run. For each, sub-models were developed for each of the five plausible values and then combined. Standard errors are upwardly adjusted to take into account variability between the plausible value estimates (Bryk, Raudenbush & Congdon, 1996; Johnson et. al., 1994). The first model, an ANOVA, estimates two equations. One is equation (1) above, with the outcome variable \(Y_{ij}\) representing student achievement as measured by each of the five plausible values. The other equation is as equation (2) above, but without any term accounting for a school effect, or

\[
(2a) \quad \beta_{0j} = \gamma_{00} + u_{0j}.
\]

This model will reveal the grand mean achievement score \(\gamma_{00}\), the reliability of the model as a sample of schools, between-school variation in achievement \(\text{Var} (u_{0j})\), or \(\tau_{00}\) and within-school variation in achievement \(\text{Var} (r_{ij})\), or \(\sigma^2\). From these variance components, the proportion of variance attributable to between-school variation can also be calculated \((\tau_{00}/(\tau_{00} + \sigma^2))\).
The second model, a random coefficients model, consists of three equations. One is as equation (1a) above, with \( Y_{ij} \) representing achievement and \( X_{ij} \) representing student SES. The other is equation (2a) above, with no school district effects parameters. This model produces an estimate of the relationship between SES and achievement within schools (\( \beta_{1j} \)), the reliability of the SES-achievement relationship, the between-school variance associated with the SES-achievement relationship (\( \tau_{1} \)) as well as re-estimates of the ANOVA terms conditional upon the SES data. The amount of variance in the random coefficient model above and beyond that of the ANOVA model is also calculated.

The third model presents the full HLM of three equations. The student-level equation is (1a), as in the random coefficient model. The school-level equations are expanded forms of (2) and (3) as follows:

\[
(2b) \quad \beta_0 = \gamma_{00} + \gamma_{01}W_1 + \gamma_{02}W_2 + \gamma_{03}W_3 + \gamma_{04}W_4 + \gamma_{05}W_5 + \gamma_{06}W_6 + u_0
\]

\[
(3a) \quad \beta_1 = \gamma_{10} + \gamma_{11}W_1 + \gamma_{12}W_2 + \gamma_{13}W_3 + \gamma_{14}W_4 + \gamma_{15}W_5 + \gamma_{16}W_6 + u_1
\]

where \( W_1 = \) Capital per-pupil expenditures

\( W_2 = \) Central office per-pupil expenditures

\( W_3 = \) Instructional per-pupil expenditures

\( W_4 = \) Principal’s office per-pupil expenditures

\( W_5 = \) Median personal income

and \( W_6 = \) Urbanness

These models make it possible to estimate the effect of each school district characteristic on mean achievement (\( \beta_0 \)) and the social distribution of achievement (\( \beta_1 \)) conditional on
the other characteristics. It also produces estimates of reliability and the amount of
variance in $\beta_0$ and $\beta_1$ explained by the model above and beyond that in the random
coefficients models (which had no school-level predictors).

The hypothesis can be restated in terms of the HLMs analyzed here. It is expected
that there will be significant between-school differences in achievement and the social
distribution of achievement, reflecting a lack of equity between and within schools; this
will be tested by the ANOVA and random coefficient models. It is further expected that
achievement and SES will be significantly related to one another within schools; this will
be indicated in the random coefficients model. Third it is expected that only demographic
characteristics of school districts will affect variations in achievement; equation (2b) of the
full model will test this proposition. Finally, it is expected that some school spending
measures will affect variations in the social distribution of achievement; equation (3a) of
the full model will test this proposition.

Results

The descriptive information confirms what is generally known (e.g. Odden et al.,
1995) about spending and achievement distributions in U.S. high schools (Table 1). The
mean achievement score in mathematics is 290 points, with 96% of students scoring
between 255 and 325 points. For the per-pupil expenditure measures, the largest share
goes to instructional spending, at $3,104 per pupil.11 The next largest amount was spent
on capital outlays, $497 for the twelfth grade sample. Next was school level
administration where $295 were spent per pupil. Finally, $116 were spent per pupil on central office administration. It should be noted that this amount may seem small given the perception that administrative expenditures are large and support a "bloated bureaucracy." The central office administration measure, however, includes only the superintendent and his or her staff, and does not include support services, such as transportation and student lunches, that are often categorized as administrative costs and do constitute a large share of school budgets.

[Table 1 about here]

The ANOVA model reveals that schools are important contributors to variation in achievement (Table 2). The grand mean for achievement ($\gamma_{00}$) is 292.65, and is obviously significantly different from zero ($p<.01$). The sampled scores are highly reliable (.86).¹² School-level variation ($\tau_{00}$) is significant, as measured by a chi-square test ($p<.01$) and represents 17.25% of the total variance in achievement. This percentage indicates that nearly one-fifth of the variation in achievement can be attributed to differences between schools, whereas four-fifths can be attributed to differences among students, irrespective of their schools.¹³

[Table 2 about here]

The random coefficient model indicates that there is a significant relationship between achievement and student SES within schools, and that this relationship also varies significantly. The achievement intercept ($\gamma_{00}$) and its statistics remain largely the same as in the ANOVA model. The coefficient measuring a relationship between student SES and student achievement ($\gamma_{10}$) is 4.48, suggesting that this coefficient can add almost five
points to the score of above-average-SES students, and subtract almost five points from the scores of below-average-SES students. The coefficient is significant (p<.01) and moderately reliable (.30). The variation in this relationship (τ₁₁) is significant according to the chi-square test (p<.01), but suggests that there is a relatively small amount of variance to be explained above and beyond that of the ANOVA model (3%).

The full model finds no relationship between the mean school achievement and district expenditure measures when district-level demographics are included as controls (Table 3). Median personal income proves significant, with higher income being associated with higher mean achievement (γ₅₅ = .0018, se = .0002, p < .05). Urbanness is also significant, with more heavily urban areas evincing lower mean achievement scores (γ₆₆ = -.0551, se = .0320, p < .10). The expenditure measures are, however, not significantly related to mean achievement. The model explains 35% more variation in mean achievement than the random coefficient model does, although there is a significant amount of achievement remaining to be explained (τ₆₆ = 143, p < .01).

[Table 3 about here]

On the other hand, the full model does find relationships between two expenditure measures and the social distribution of achievement. As with the mean achievement model, district demographics play a role. The higher the level of median income, the more pronounced a relationship exists between SES and achievement within the district (γ₁₅ = .0005, se = .0001, p < .05). Expenditures can compensate for the effect of median income. More spending on instruction is associated with a weakening of the SES-achievement relationship (γ₁₃ = -.1810, se = .0885, p < .05). Higher capital outlays are
associated with a lower SES-achievement relationship as well ($\gamma_{11} = -.1615$, $se = .0680$, $p < .05$).\textsuperscript{14} The two types of administrative expenditures are unrelated to the social distribution of achievement. The model explains 34% more of the variation in the social distribution of achievement than the random coefficient model although here too a significant amount of variation in the relationship remains to be explained ($\tau_{11} = 10$, $p < .01$).

The hypotheses, then, appear to be confirmed. School district spending at the twelfth grade level can affect the social distribution of achievement within schools, but not mean levels between schools. Further, it is only certain types of spending that affect the social distribution of achievement.

Methodological Caveats and the Need for Further Research

Before drawing conclusions from these findings some methodological issues should be noted. One is that the study used district-level expenditure estimates to measure the relationship between school and student characteristics. The study was interested in discovering how the resources available to each NAEP high school affected the mean level and social distribution of scores on the assessments. Yet no expenditure data were available for these high schools, since most data on expenditures is collected at the district rather than the school level. Thus, it needed to be assumed that variations in expenditures at the district level could at least serve as a proxy for variations in expenditures at the school level. One possible line of future inquiry would be to collect school-level finance
data and relate it to student achievement, to see if this assumption is supported. Texas
and Ohio have recently begun collecting such data (National Center for Education

A second issue is that the analysis does not take into account selection effects
caused by student attrition. Unlike with elementary and middle schools, high schools
experience significant rates of dropping out. The students remaining in twelfth grade,
when NAEP is administered, are thus not broadly representative of students entering high
school, or of the student-age population. Yet, in the case of the link between district
expenditures and SES differences in achievement, these selection effects are probably
attenuating rather than exaggerating it. Presumably, students drop out of resource-poor
schools at greater rates than they do from resource-rich ones. Thus, schools with high
levels of resources would be expected to retain more low SES low achieving students, and
therefore the SES-achievement relationship should be stronger in such schools. Thus, if
the study had not found a relationship between expenditures and the SES-achievement
relationship, this might be because it failed to account for student attrition; however, given
that the study did find a relationship, selection bias probably did not change the results.

To confirm this notion, it would be necessary to have data on dropouts, which is not
available from NAEP. Such an analysis would be another promising area for further
research.

A third caveat lies in the measurement of student SES. The measure is based upon
student self-reports of possessions in the home and parents’ education levels. Yet self-
reporting of SES can be highly unreliable, and some dimensions of SES, such as family
income and occupation, are not included in the measure. The self-reporting problem is mitigated to some extent by the fact that the sample consists of twelfth graders, who provide more reliable self-reports than do younger students. Yet the lack of income and occupational measures of SES does mean that some of the relationship between district expenditures and the SES-achievement relationship may be attributable to low-SES students in high-spending schools actually being of higher SES based upon unmeasured dimensions, and low-SES students in low-spending schools actually being of lower SES. Correcting this problem would require richer SES data.\textsuperscript{15}

A fourth problem is that the relationship between SES and achievement within schools may be nonlinear. If high-spending districts tend to have more high-SES students, then the relationship between expenditures and the social distribution of achievement could be spurious. This possibility was tested in this study through a sensitivity analysis that added the square of SES to the within-school model. This addition did modestly diminish the relationships, but did not eliminate them. An additional model was developed in which the square of SES was related to the same district-level independent variables to which the other outcome variables had been related, and it was found that even the square SES-achievement relationship evinced the same associations with capital and instructional per-pupil expenditures. These findings suggest that the nonlinearity of the SES-achievement relationship does not seem to undermine the study’s results.

A fifth issue is that the data analyzed here are cross-sectional. All information was collected during the 1991-92 school year. Thus the variables being related to one another, district spending and student achievement, occur at the same time. This fact is
problematic for two reasons. First, it may be that the relationship between district spending and the social distribution of achievement is spurious; both variables may be influenced by the prior social distribution of achievement. Second, there may be a time-lag in the relationship; for instance, capital investment might not be expected to be related to achievement in the year in which the money was spent, but, rather in the year in which the capital projects are completed. These problems can only be addressed through the use of longitudinal data. Although there is no single nationally representative longitudinal database that includes the requisite expenditure and achievement data, it can be synthesized from multiple sources.

A sixth issue is that the TCI does not completely adjust per-pupil expenditure measures for the cost of education. The TCI measures regional variations in one important cost of education, the cost of hiring teachers. It does not, however, adjust for regional variations in other costs of education, such as the cost of hiring administrators. The TCI, however, is the best measure available at this time, and does account for more than half of the expenditures of the typical school district.

A seventh issue is that the finding of no relationship between educational expenditures and mean achievement levels is a null finding, and therefore should be treated with caution. Significance tests are not designed to confirm the lack of a relationship but, rather, to confirm the existence of one. In the case of a null finding, it may be that, under slightly different circumstances (such as a larger sample size or different statistical controls) a relationship could be discerned. Because the relationship to mean achievement levels is just such a finding, it should be regarded as tentative.
A final issue is that the study does not provide information on how various school policies and practices might mediate between district spending and student achievement. While, as the next section indicates, prior research does make it possible to speculate on differences in the policies of schools at various resource levels and the relationship of these policies to the social distribution of achievement, confirming these speculations requires models that actually measure them. These models can be produced using various structural equation modeling packages, but usually only for a single level of analysis. Fortunately, recent developments in structural equation modeling make it more feasible to develop multilevel structural equation models (Muthen, 1994; Gustafsson & Stahl, 1997).

Conclusions

Even with these caveats, certain inferences can be made from the current study about the interrelationships between district expenditures, mean school achievement levels and the social distribution of achievement. First, the analysis suggests that there is no relationship between any of the spending categories and mean school achievement for twelfth graders. This finding can be explained in the context of the findings of earlier studies regarding fourth and eighth graders. The fourth grade study found that more spending on instruction and central office administration were associated with smaller class size which, in turn, was associated with higher mean district achievement. For eighth graders the effect was somewhat less direct. More spending on instruction and central office administration were associated with smaller class size, but smaller class size was
only associated with higher mean achievement scores by virtue of its effect on the social environment of the school. Smaller class sizes were conducive to more positive social environments, which, in turn, improved mean achievement. The size of the spending-achievement relationship was also weaker for eighth graders than it was for fourth graders. Spending that is linked to a doubling of teacher-student ratios would be associated with seven points (or a little over half a grade level) in achievement gains for eighth graders, as opposed to 15 points (or a little over a full grade level) in achievement for fourth graders.\textsuperscript{16} It should therefore not be surprising that for twelfth graders effect sizes could dwindle further, perhaps to the point of statistical non-significance.

Second, the analysis may be construed to mean that some types of spending are related to the social distribution of achievement while others are not. The key difference, it appears, between those types that evince such a relationship (instruction and capital) and those that do not (the two types of administration), is that the first two can more easily be targeted to certain students.\textsuperscript{17} The offices of principal or superintendent can be large or small, and this size, presumably, affects all students equally. Instructional spending, on the other hand, can significantly affect the direction of a school because the school can target instructional resources to different students. For instance, instructional dollars could go to hiring Advanced Placement teachers, which would help more advanced students, or to remedial programs, which would help less advanced students. Capital spending can also be targeted. Money could be spent on a state-of-the-art biology lab for honor students, or for a computer lab designed to support multi-media literacy programs. It is only those
dollars that can be differentially spent that can be associated with the distribution of achievement.

The type of relationship they do evince suggests that when schools lack sufficient funds, their capacity to educate all students toward a common yardstick may be reduced. Students enter high school with different levels of preparation, depending upon their SES and various other factors. To reduce these inequalities in preparation to the point that both low- and high-SES students become proficient in the requisite subject matter requires the active intervention of the school; when the school lacks adequate funds, its ability to intervene is compromised and, as a result, students will be more likely to advance based upon their past preparation, that is, a situation of within-school inequity.

These findings support certain speculations regarding the allocative principles by which policymakers should be guided. First, the pattern of findings between this study and those of fourth and eighth graders may support the notion that dollars should be targeted towards instructional spending. Higher levels of such spending were associated with improvement in some achievement outcome at all three grade levels; it is therefore plausible that if school districts allocate more dollars to instruction, they will experience improvements in academic achievement. Second, the pattern of findings between this study and those of fourth and eighth graders may support the notion that dollars should be targeted to earlier grades. The strength of the relationship between spending and mean achievement was largest for fourth grade, smaller for eighth grade, and non-significant for twelfth grade. It is therefore plausible that if school districts allocate additional dollars to fourth grade, achievement will increase more than would be the case if those dollars were
allocated to the eighth or twelfth grades. Finally, the findings for twelfth graders suggest that if spending is increased, policymakers should not expect gains in mean achievement, but instead perhaps a reduction in achievement disparities among students. While this study found no significant relationship between the different types of spending and the average achievement level in the school, it did find a link between certain types of spending and the social distribution of achievement; it is with such a finding in mind that policymakers should shape their expectations regarding the potential consequences of allocating additional dollars to twelfth graders.
References


TABLE 1
*Descriptive Statistics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics Achievement</td>
<td>7,217</td>
<td>289.89</td>
<td>35.79</td>
</tr>
<tr>
<td>Socio-economic Status</td>
<td>7,217</td>
<td>10.23</td>
<td>.93</td>
</tr>
<tr>
<td><strong>District Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital PPE (Hundreds of Dollars)</td>
<td>182</td>
<td>4.97</td>
<td>7.70</td>
</tr>
<tr>
<td>Central Office Administration PPE (Hundreds of Dollars)</td>
<td>182</td>
<td>1.16</td>
<td>.86</td>
</tr>
<tr>
<td>Median Personal Income (Hundreds of Dollars)</td>
<td>182</td>
<td>145.44</td>
<td>54.24</td>
</tr>
<tr>
<td>Instructional PPE (Hundreds of Dollars)</td>
<td>182</td>
<td>31.04</td>
<td>8.38</td>
</tr>
<tr>
<td>Principal Administration PPE (Hundreds of Dollars)</td>
<td>182</td>
<td>2.95</td>
<td>.99</td>
</tr>
<tr>
<td>Urban Percentage</td>
<td>182</td>
<td>69.00</td>
<td>36.02</td>
</tr>
</tbody>
</table>
# TABLE 2

**ANOVA and Random Coefficient Models**

<table>
<thead>
<tr>
<th>Variable</th>
<th>ANOVA Model</th>
<th>Random Coefficient Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>$\gamma_{n0}$</td>
<td>292.65</td>
<td>292.64</td>
</tr>
<tr>
<td>SE ($\gamma_{n0}$)</td>
<td>1.21</td>
<td>1.20</td>
</tr>
<tr>
<td>T ratio ($\gamma_{n0}$)</td>
<td>242.67</td>
<td>242.87</td>
</tr>
<tr>
<td>P value ($\gamma_{n0}$)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Reliability</td>
<td>.86</td>
<td>.87</td>
</tr>
<tr>
<td>$\tau_{nn}$</td>
<td>219.62</td>
<td>220.25</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>1570.72</td>
<td>1606.98</td>
</tr>
<tr>
<td>P value ($\chi^2$)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1053.40</td>
<td>1021.21</td>
</tr>
<tr>
<td>Explained Variance</td>
<td>17.25</td>
<td>3.06</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td>Mean Achievement ($\beta_0$)</td>
<td>Social Distribution ($\beta_1$)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>Coefficient</td>
<td>(SE)</td>
</tr>
<tr>
<td>$\gamma_{n0}$ (Intercept)</td>
<td>271.0152**</td>
<td>(4.8167)</td>
</tr>
<tr>
<td>$\gamma_{n1}$ (Capital PPE)</td>
<td>.0658</td>
<td>(.1341)</td>
</tr>
<tr>
<td>$\gamma_{n2}$ (Central PPE)</td>
<td>2.1390</td>
<td>(1.3884)</td>
</tr>
<tr>
<td>$\gamma_{n3}$ (Instruction PPE)</td>
<td>-.2049</td>
<td>(.1583)</td>
</tr>
<tr>
<td>$\gamma_{n4}$ (Principal PPE)</td>
<td>.8197</td>
<td>(1.2208)</td>
</tr>
<tr>
<td>$\gamma_{n5}$ (Median Income)</td>
<td>.0018**</td>
<td>(.0002)</td>
</tr>
<tr>
<td>$\gamma_{n6}$ (Urban)</td>
<td>-.0551*</td>
<td>(.0320)</td>
</tr>
<tr>
<td>Reliability</td>
<td>.811</td>
<td>.224</td>
</tr>
<tr>
<td>$\tau_{mn}$</td>
<td>142.7231</td>
<td>10.3513</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>1061.2666</td>
<td>221.4259</td>
</tr>
<tr>
<td>P value ($\chi^2$)</td>
<td>.000</td>
<td>.008</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1021.1812</td>
<td></td>
</tr>
<tr>
<td>Explained Variance</td>
<td>.3520</td>
<td>.3358</td>
</tr>
</tbody>
</table>

*p<.10; **p<.05
Appendix A: Variable Definitions

Mathematics Achievement: Taken from NAEP data for mathematics for 1992. Consists of five plausible values for students responding to NAEP. For all HLM estimates, plausible values were analyzed in accordance with plausible values methodology.

Student SES: Derived from NAEP data for mathematics for 1992. Calculated as the summated scale of the following six items: The highest level of education attained by the mother; the highest level of education attained by the father; and whether or not (1) the family receives a newspaper; (2) there is an encyclopedia in the home; (3) there are more than 25 books in the home; and (4) the family subscribes to magazines. Measured as the total of that scale.

Capital PPE: Derived from data in CCD for fiscal year 1992. Calculated by dividing total capital outlays for each school district by the number of students in the school district and the TCI. Measured in hundreds of dollars.

Instructional PPE: Derived from data in CCD for fiscal year 1992. Calculated by dividing total expenditures on instruction for each school district by the number of students in the school district and the TCI. Measured in hundreds of dollars.

Central Administration PPE: Derived from data in CCD for fiscal year 1992. Calculated by dividing total expenditures on central administration for each school district by the number of students in the school district and the TCI. Measured in hundreds of dollars.
Principal Administration PPE: Derived from data in CCD for fiscal year 1992. Calculated by dividing total expenditures on school-level administration for each school district by the number of students in the school district and the TCI. Measured in hundreds of dollars.

District Median Personal Income: Taken from SDDB, census data for 1990. Consists of the median personal income for the school district in which the student is attending school. Measured in dollars.

Urbanness: Taken from SDDB, census data for 1990. Consists of the proportion of the school district in which the student is attending school that can be classified as urban. Measured as a percentage.
Harold Wenglinsky is Associate Research Scientist, Policy Information Center, Educational Testing Service, Princeton, NJ 08541; e-mail hwenglinsky@ets.org. His areas of specialization are sociology of education, educational policy and quantitative methods.

The research reported here was supported, in part, by grant REC-9628157 from the National Science Foundation. The author thanks James Carlson, William Fowler, Frank Jenkins, Eugene Johnson, Alfred Rogers, Keith Rust and Larry Suter for their assistance at various stages of this project. Any errors of fact or interpretation included in this article are the responsibility of the author.
Notes

1 Studies have found that SES is strongly related to school resource levels (National Center for Education Statistics, 1995a). In some cases, however, this relationship may not hold. A student from a poor family living in an area with a large amount of commercial property might find schools with high levels of property tax revenue.

2 Within-district but between-school equity is not addressed here. There may be significant disparities in resources and achievement between schools within the same school district; there is, however, a lack of data, both in NAEP and generally, that measure school-level resources, making it difficult to study this issue.

3 “Social distribution of achievement,” for purposes of this article, refers to the extent to which student achievement depends upon student SES.

4 Exceptions to this are some of the analyses and re-analyses of the Coleman Report, discussed above, which found strong input-output relationships for minority students than for white students (e.g. Armor, 1972).

5 For a discussion of the theory of cost-of-education indices, see Barro (1994).

6 “School effects” refer to associations between school characteristics and student outcomes. The association is viewed as an “effect” only by hypothesis; the data can confirm an association, but not an effect. Because the school characteristics in the study are actually district characteristics, the term “school district effects” is generally used.

7 For purposes of this study, the terms “school-level” and “district-level” are used interchangeably. Because NAEP includes only one or two schools in a school district, the
models cannot distinguish between school- and district-level effects. Hence the district-level effects studied here are assumed to be divisible into their component school-level effects.

8 NAEP does not provide a single score for each student but, rather, a series of five estimates. Of all the questions in the survey designed to measure achievement in a subject area, only a sub-sample is administered to any given student. Inferences of what a particular student would have scored on questions not administered to him or her, and the resulting total score that the student would have received are, therefore, complex. NAEP uses Item Response Theory (IRT) to summarize student performance (see Hambleton, Swaminathan & Rogers, 1991 for discussion of IRT). Because individual student performance is not well determined, due to small numbers of items administered to any student, traditional IRT point estimates of student performance ("scale scores") are inappropriate for estimating population distributions. Special procedures are used to account for the fact that individual proficiencies are not well determined. These procedures, known as plausible values methodology, use the IRT model for the relationship of the item responses to proficiency, a statistical model relating proficiency to background variables, and the observed data (item responses and background variables for each individual) to simulate five sets of plausible values for proficiencies (Johnson, Mislevy & Thomas, 1994).

9 Sensitivity analyses, in which all cases with missing values for SES were deleted, were also conducted, and produced the same results.
10 Plausible values are combined by calculating the mean of the point estimates and calculating adjusted variances through the formula

\[ V = U^* + (1 + M)^{-1} B \]

Where \( V \) = the adjusted variance

\[ U^* \] = the average sampling variance of the plausible values

\( M \) = the number of plausible values

and \( B \) = the variance among the plausible values

(Johnson, Mislevy & Thomas, 1994).

11 Because the spending measures were collected at the district level, they represent average spending at all grade levels in a given district. Thus, the measures should be read to mean per-pupil expenditures in the districts where twelfth graders were sampled, not per-pupil expenditures for twelfth graders.

12 Reliability here refers to the between-group variance as a percentage of the total variance, rather than more conventional definitions.

13 It should be noted that district variations include not only institutional characteristics of schools, but also demographic characteristics of school districts, such as the degree of urbanness, which Table 1 found to be substantial (SD=36.02% with a mean of 69.00%).

14 A negative sign indicates a reduction in the coefficient \( \beta_1 \), the within-school relationship between SES and achievement.
15 For a discussion of the weaknesses of the NAEP measures of SES, see Berends & Koretz (1995).

16 These estimated point increases in achievement assume that the relationships are linear for those particular changes in class size, which may not necessarily be the case; thus the size of the gains mentioned here should be seen as purely illustrative.

17 Although the study suggests that expenditures should be “targeted” to certain areas, it does not suggest whether targeting should occur through reallocation of existing funds or the provision of new funds. It should be noted, however, that the funds that could be reallocated from the two types of administration are quite small (see Table 1); therefore, dollars would have to come primarily from new funds, support service expenditures, or a combination of the two.