

# *School District Expenditures, School Resources and Student Achievement: Modeling the Production Function*

*Harold Wenglinsky*

*Educational Testing Service*

## About the Author

Harold Wenglinsky is an Associated Research Scientist in the Policy Information Center at Educational Testing Service. The Policy Information Center conducts original research as well as syntheses of existing research on important questions of educational policy for dissemination to policymaking audiences. Recent studies Dr. Wenglinsky has authored at the Center include *When Money Matters: How Educational Expenditures Improve Student Performance and How They Don't* and *Students at Historically Black Colleges and Universities: Their Aspirations and Accomplishments*. He is currently studying

the application of multilevel structural equation modeling to data on school effects. Before joining the Policy Information Center, Dr. Wenglinsky was the ETS National Assessment of Educational Progress Visiting Scholar. He has been the recipient of numerous grants and awards from various organizations including the National Science Foundation and the American Educational Research Association. (The paper presented here was funded in part by a grant from NSF, REC-9628157.) Dr. Wenglinsky received his Ph.D. in Sociology from New York University in 1996.



# *School District Expenditures, School Resources and Student Achievement: Modeling the Production Function*

*Harold Wenglinsky*  
*Educational Testing Service*

## **Introduction**

After more than 30 years of research, social scientists have made little progress in identifying the educational production function. “Production function” studies are those that use some form of multivariate analysis, such as regression analysis, to measure associations between various educational inputs, such as per-pupil expenditures, and outputs, such as academic achievement as measured by standardized tests.<sup>1</sup> One of the earliest studies of this type was the Equality of Educational Opportunity Report, commonly referred to as the Coleman Report (Coleman et al. 1966). This study found little association between inputs and outputs for a nationally representative sample of students and schools. Since the publication of the Coleman Report, nearly 400 additional studies of this sort have been conducted. Their results have been mixed, fueling, rather than resolving, the debate as to whether money matters to educational achievement (see Hanushek 1997 for list of studies).

Because of the mixed results of this very large number of studies, some researchers have concluded that the production function approach is flawed and should be abandoned. In their view, production function studies suffer from a multitude of problems, including their failure to analyze different types of educational expenditure (such as spending on instruction and administration) and their failure to adjust for regional variations in the cost of edu-

cation (Fortune and O’Neil 1994). Some researchers suggest alternate approaches to estimating the relationship between expenditures and achievement. Monk (1992) suggests conducting small-scale studies at low levels of aggregation, such as the classroom level; Fortune and O’Neil (1994) suggest comparing the achievement levels of specific subgroups, such as high-spending and low-spending urban school districts.

This paper contends that the production function approach is salvageable; the problems researchers have identified can be addressed, producing meaningful results. The present study provides an example of how this may be done. It applies structural equation modeling and multilevel modeling to recently developed databases of fourth-graders. The study is national in scope, distinguishes between different types of spending, and adjusts for regional variations in the cost of education, thus addressing many of the issues raised by critics of the production function approach. The study finds that, at least for fourth-graders, some inputs are strongly associated with academic achievement, while others are not: Instructional expenditures, central office administration expenditures, and teacher-student ratios are all associated with achievement; principal’s office expenditures, capital outlays, and teacher education levels are not. Before discussing these results and their derivation,

---

<sup>1</sup> For the purposes of this paper, “expenditures” refers to actual dollars spent by school districts, “resources” to quantifiable goods made available to schools, and “inputs” to both expenditures and resources.

however, it is necessary to touch upon the methodological issues involved in production functions.

## Background

Production function studies of education have been undertaken for more than 30 years. By one estimate nearly 400 production function studies have been conducted and published since the Coleman Report of 1966 (Hanushek 1997, 1996). These studies have tended to use samples that are smaller in their geographical scope than the national Coleman Report, and have studied the same sorts of inputs that report did (aggregate per-pupil expenditures). These studies have come to different conclusions regarding the production function, some finding relationships between a given input and academic achievement, and others finding no such relationship.

More recently, studies known as “meta-analyses” have applied statistical techniques to synthesize the findings from production function studies; these too arrived at contradictory conclusions. Hanushek (1989) conducted a meta-analysis covering both expenditure and resource measures, and found no relationship between these inputs and academic achievement. Hanushek synthesized the findings of 187 production function studies using the technique of vote counting. He first divided each study into its component inputs. A study that related class size and teacher education to achievement, for example, was divided into those two inputs. Each input was then placed in one of seven categories: per-pupil expenditures, teacher experience, teacher education, teacher salary, teacher-student ratio, administrative inputs, and facilities. Within each category, the relationship of the input to the studied output was classified as positive and statistically significant, positive and statistically non-significant, negative and statistically significant, negative and statistically non-significant, and non-significant but of unknown direction. Hanushek found most relationships to be non-significant. Of 65 aggregate per-pupil expenditure relationships, for example, he

found 13 to be positive and significant, 3 negative and significant, and 49 to be non-significant. He concluded that “there is no strong or systematic relationship between school expenditures and student performance (1989, 47).”

Hedges, Laine, and Greenwald (1994) reanalyzed most of the same studies, and drew the opposite conclusion. They first excluded from their analysis the relationships Hanushek had classified as non-significant but of unknown direction. For the remaining relationships, they reinterpreted Hanushek’s vote counting in the context of rules regarding statistical significance. They argued that, if the relationships are treated as a sample, in order to draw the conclusion that there is no relationship between an input and achievement, no more than 5 percent of the relationships could be significant, and these relationships would have to be equally divided

between the positive and negative directions. Yet, in fact, if relationships of unknown direction are excluded, many more than 5 percent of the relationships are significant (up to 30 percent for per-pupil expenditures); most of the significant relationships are in the positive direction. The bulk of insignificant relationships are also in the positive direction.

After reinterpreting the vote count, Hedges, Laine, and Greenwald (1994) applied a significance test, the inverse chi-square, to combine the relationships for each input into a single significance measure. They tested two hypotheses: that each input is positively related to achievement, and that each is negatively related to achievement. They found, for the full sample of relationships (as well as for various subsamples), that almost all relationships were significant in the positive direction, with a few others being significant in the negative direction. Finally, for each input, Hedges and his colleagues combined the coefficients from those studies that reported them by calculating their median. They found positive coefficients for per-pupil expenditures, teacher experience, teacher salary, administrative inputs,

*More recently, studies known as “meta-analyses” have applied statistical techniques to synthesize the findings from production function studies; these too arrived at contradictory conclusions.*

and facilities, and mixed results for class size, and concluded that resources affect achievement.

Hanushek (1996) continued the debate, countering the meta-analysis of Hedges, Laine, and Greenwald. He updated his sample of studies to include those published after his 1989 meta-analysis, making a total of 377 studies. Hanushek again found, when he classified relationships into the seven categories, that the bulk of studies indicated no significant relationship between resources and achievement. In a counter-study of their own, Greenwald, Hedges, and Laine (1996) created their own sample of studies, and placed the relationships they identified from the studies into seven somewhat different categories: per-pupil expenditures, teacher ability, teacher education, teacher experience, teacher salary, teacher-pupil ratio and school size. They again found, for both this new sample and for various subsamples, that the combined significance test and median effect sizes supported the hypothesis that resources affect achievement. Most recently, Hanushek (1997) has compared his sample of 377 studies to the sample of Greenwald, Hedges, and Laine (1996), and found that the latter sample systematically over-represented positive relationships.<sup>2</sup>

The fact that different meta-analyses can reach different conclusions from similar sets of studies indicates that the underlying studies are quite volatile in their results when subjected to different assumptions. This volatility was even revealed within the meta-analyses. For instance, Hedges, Laine, and Greenwald (1994) were able to find support both for the hypothesis of a positive relationship and that of a negative relationship between a given resource and achievement when using combined significance tests. Both

Greenwald, Hedges, and Laine (1996) and Hanushek (1997) found that the results from a subsample of longitudinal studies differed markedly from those of the full sample. What the meta-analyses reveal most clearly, then, is that the original studies do not provide a clear answer to the question of whether or not money matters.

This lack of consensus among the meta-analyses reflects to some degree shortcomings in the methods of the original studies. Six shortcomings have been commonly noted.

First, unlike the Coleman Report, most subsequent studies were not nationally representative, but instead studied a particular state or school district. This hampers development of a consensus, because different regions of the country may have different spending patterns and different relationships between these spending patterns and student achievement.

Second, the studies did not distinguish among different types of spending. While they measured multiple inputs, such as teacher experience and teacher-student ratios, the only expenditure measure used was aggregate per-pupil expenditures. Using such a gross measure risks missing certain dynamics in the relationship between school spending and academic achievement, as increases in some types of spending may have an effect

while increases in others may not. For instance, increased spending on administration may not significantly raise achievement, while increased spending on instruction may. If these types of spending are not measured separately, the apparent effects of spending on instruction will be reduced or eliminated when combined with the lack of effects from administration.

*The fact that different meta-analyses can reach different conclusions from similar sets of studies indicates that the underlying studies are quite volatile in their results when subjected to different assumptions.*

<sup>2</sup> Other meta-analyses have also arrived at contradictory conclusions. With regard to class size, Glass and Smith (1979) found a clear and consistent relationship while Odden (1990) did not. The effect of class size on student achievement has also been the subject of a controlled experiment in which students in kindergarten and first grade were randomly assigned to small and large classes. The study found significant achievement differences that persisted even after the students in small classes were returned to large ones (Finn and Achilles 1990; Mosteller 1995; Mosteller, Light, and Sachs 1996). This finding, like those of production function research, has been the subject of controversy (Hanushek 1997).

Third, the studies did not take into account the ways in which other influences on the process of schooling may mediate between spending and achievement. Effective schools research suggests that certain aspects of the school environment, particularly supportive relations between teachers and principals, positively influence achievement.<sup>3</sup> Yet none of the prior research has sought to measure the influence of school spending patterns on school environment.

Fourth, not all of the studies provided rich measures of student background.<sup>4</sup> While the research on measures of the socio-economic characteristics of students indicates that a single measure, socio-economic status (SES), can be generated by adding together responses to a relatively small number of questions, many studies did not include such questions. If SES is poorly measured, it is difficult to determine if relationships between spending and achievement are attributable to some degree to SES differences between students in high- and low-spending districts.

Fifth, most studies did not control for variations in cost between regions. The cost of living in New York City is higher than the cost of living in Montgomery, Alabama, and presumably this difference means that teachers paid the same actual dollars in the two cities are not able to maintain the same standard of living; a dollar will buy less in New York City. As a result, New York City would have to offer higher salaries to recruit successfully the same teachers as Montgomery.<sup>5</sup> Other factors may also influence the cost of hiring comparable teachers, including union pres-

sure to increase wages and the overall quality of life in the region. Most studies did not take these factors into account, and they may be as important as SES, in that differences in achievement between two districts may be due to some degree to differences in how much it costs to hire teachers.<sup>6</sup>

Sixth, many of the measures of achievement used by earlier studies were unsophisticated. Some did not use achievement measures at all but merely relied on proxies, such as graduation rates. Some used measures as simple as whether or not a student passed a minimum competency test. Few took into account modern developments in test theory, such as Item Response Theory (IRT).<sup>7</sup>

Finally, the prior research has not taken into account the multilevel nature of school effects. Measuring the relationship between school characteristics and student achievement entails relating variables whose level of analysis is the school or school district to an outcome whose level of analysis is the student. Various estimation techniques have been developed that take the multilevel nature of school effects into account, and it has been found that these techniques sometimes produce results that differ from more conventional techniques. In particular, conventional techniques often underestimate standard errors and, in some cases, fail to identify important components of school effects (Raudenbush and Willms 1995; Bryk and Raudenbush 1992). Production function models have generally not made use of estimation techniques that are sensitive to multilevel data, and consequently may produce inaccurate results.

*Effective schools research suggests that certain aspects of the school environment, particularly supportive relations between teachers and principals, positively influence achievement.*

---

<sup>3</sup> Despite some early criticism of effective schools research (e.g., Cuban 1984; Purkey and Smith 1983), later large scale multivariate studies have persuaded most researchers that there is a social dimension to school life that plays some independent role in student achievement. The extent of this role is, however, still being debated (Lee, Bryk, and Smith 1993).

<sup>4</sup> This was pointed out by Hedges, Laine, and Greenwald (1994, 12).

<sup>5</sup> When cost of living is taken into account, differentials in per-pupil expenditures between high-spending and low-spending states decrease markedly, indicating that states with fewer resources often tend to be states with lower costs of living (Barton et al. 1991).

<sup>6</sup> This was pointed out by Fortune and O'Neil (1994, 24).

<sup>7</sup> For a discussion of this shortcoming in production function research, see Fortune and O'Neil (1994, 24). For a discussion of IRT, see Hambleton et al. (1991).

Some researchers doubt that these problems can be addressed, and have argued that the production function approach should be abandoned, and alternate approaches explored. Monk (1992) proposes to shift the unit of analysis for school resource studies to the classroom level. He notes that prior research has found a great deal of variation in the efficacy of teachers within the same school, as well as variation in the efficacy of a particular teacher during different classes. He classifies teachers as being of two types, those who are engaged with their classes, trying actively to address any problems in them, and those who are accommodating, seeking only to avoid dealing with problems. He views whether the teacher chooses the engagement or accommodation route as dependent upon a number of factors, including resource decisions made at the school level. Monk calls for a research program in which teachers are interviewed to provide retrospective information on the problems they face, their responses to these problems, and the degree to which resources operate as a constraint.

Another alternative to production function research is the threshold approach, proposed by Fortune and O'Neil (1994). They argue that the key problem with production function research is its use of linear models. They hypothesize that input-output relationships occur in a punctuated manner, with small increments of inputs having no effect on achievement, but large increments having a large effect. To estimate this effect, they propose comparing the mean achievement levels of school districts that are in the top 30 percent in terms of spending to school districts that are in the bottom 30 percent. To address the problem that demographic variables might be at the root of achievement differences, they propose using demographically similar school districts for the comparison. They also propose eliminating outlying cases. In applying this approach to samples of school districts in Missouri and Ohio, they find

*Such alternatives, however, raise their own methodological issues. The most significant is that in both cases it is difficult to separate the factors contributing to student achievement.*

that while correlation coefficients rarely uncover an input-output relationship, the threshold approach often finds one.

Such alternatives, however, raise their own methodological issues. The most significant is that in both cases it is difficult to separate the factors contributing to student achievement. In the classroom-based approach, efficacious and non-efficacious teachers are identified and the resource constraints traced. Yet, the availability of one type of resource tends to be highly correlated with availability of another. Since the efficacious teacher may have many resources available at once, it will be difficult to determine which is the basis of high teacher efficacy. In addition, it will be difficult to determine whether high student achievement is primarily attributable to teacher efficacy or student characteristics.<sup>8</sup> In the threshold approach, only school districts with extremely high and extremely low levels of expenditures are compared. In many areas the high expenditure districts will have high levels of resources and the low expenditure districts low levels of resources. It will thus be difficult to determine which resource is responsible for achievement levels. In addition, it will be difficult to determine whether or not the resource levels in the schools or the resource levels in the communities account for achievement differences.<sup>9</sup>

Given that these alternate approaches raise their own difficulties, it may also be worthwhile to salvage the production function approach through addressing its problems. The present study is an attempt to do just that.

## The Design of the Study

### Hypotheses

This study hypothesizes that there are various potential paths through which school district expen-

<sup>8</sup> It is also worth noting that to draw conclusions about expenditures and resources, the classroom approach will still have to collect district- and school-level data, because expenditure and resource decisions are made at those levels, not the classroom level. Thus many of the methodological problems of production function studies will also hold for the classroom approach.

<sup>9</sup> NCES (1995a) found that school district expenditures and the average SES of the districts were strongly related.

ditures and school resources can influence student achievement (figure 1). These paths occur in three basic steps. First, the allocation of money at the school district level influences the availability of resources at the school level. Most decisions about how to spend money are made by school superintendents and their staffs. These spending decisions determine how much of each school resource is purchased, and therefore what is available in the school. For instance, more spending on instruction will lead to some combination of more teachers per student, higher teacher salaries and more instructional materials.

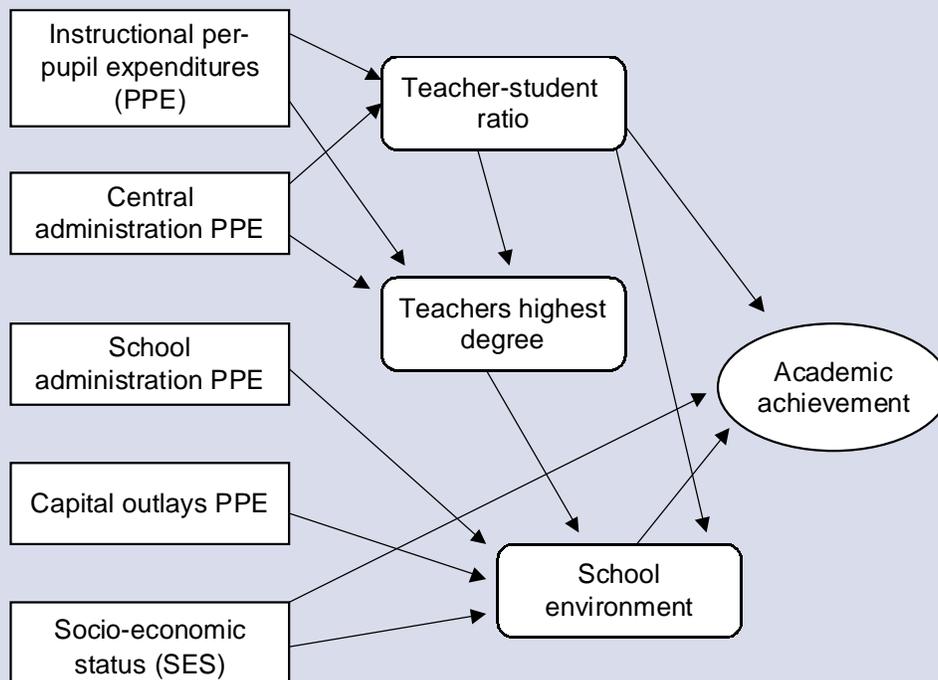
Second, the availability of resources has consequences for the school climate. Schools vary widely in environment, some possessing low levels of student and teacher absenteeism, collegial relationships between teachers and principals, and a lack of disruptive and delinquent behaviors, and others possessing the opposite. In part, environment is influenced by the availability of resources; teachers who

are paid lower salaries, for example, might be expected to be more frequently absent.

The third step involves the influence of school climate on student achievement. Effective schools research suggests that school climate strongly influences student performance (Lee, Bryk, and Smith 1993; Austin and Garber 1985; Brookover et al. 1979; Edmonds 1979). Disruptive students, high levels of student and teacher absenteeism, and frayed principal-teacher relations can be expected to interfere with the ability of teachers to instruct and students to learn.

It is hypothesized here that some educational expenditures influence achievement via these steps. Four types of expenditures are considered: instructional expenditures, central office administration expenditures, principal's office administration expenditures, and capital outlays. The first two, it is hypothesized, will directly affect school resources.

Figure 1.—Hypothesized paths to achievement



SOURCE: Wenglinsky, unpublished.

Research has shown that expenditures are typically invested in one of two resources, increasing the number of teachers per student or improving teacher quality (Odden and Clune 1995). It is therefore expected that instructional and central office administration expenditures will influence one or both of these resources. These resources will, in turn, affect the school climate, which will itself affect student achievement. It is also expected that capital outlays and principal's office administration will play a role in the learning process. While it is unlikely that spending in either area would directly affect the number of teachers in the classroom or the types of teachers hired, it is expected that they will influence the school climate which will itself influence student achievement.

The model hypothesized here also must take into account the role of two factors outside the school in the spending-achievement relationship. First, student SES can be expected to affect the school climate and student achievement; students from more affluent backgrounds will be more likely to meet the social demands of the school, develop a rapport with teachers, and be better prepared to achieve at high levels (Hauser, Sewell, and Alwin 1976; Jencks et al. 1972; Coleman et al. 1966). Second, the cost of education can be expected to affect the ability of expenditures to purchase school resources and influence the school climate. A given level of expenditures will not go as far in a high-cost region.

## Data

The data employed to test this model are drawn from three sources: the National Assessment of Educational Progress (NAEP), the Common Core of Data

(CCD), and a Teacher's Cost Index (TCI). NAEP is a nationally representative database of students and schools collected by the Educational Testing Service (ETS) under a contract from the National Center for Education Statistics (NCES); CCD is a database consisting of the universe of school districts in the United States, collected by NCES; and the TCI was developed by NCES to measure regional variations in the price of teachers. Three data sources had to be used because none contains all of the necessary measures.

NAEP is administered by ETS every 2 years to nationally representative samples of fourth-, eighth- and twelfth-graders, and to their teachers and principals. The subject areas tested vary, but have included at various times mathematics, reading, history, geography, and science. The information collected by NAEP is used to assess the knowledge of

students throughout the country; to make comparisons in the levels of knowledge of various regional, 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 overview of NAEP; Mullis, Dossey, Owens, and Phillips 1993 for report card for 1992 mathematics assessment.) The 1992 mathematics assessment of students attending fourth grade was used in this study.<sup>10</sup> It contains measures of mathematics achievement, school environment, teacher education levels, teacher-student ratios,

and student- and school-level SES.<sup>11</sup>

CCD is a database of financial information provided by the universe of U.S. school districts. All school districts send this information to the U.S. Department of Education on a yearly basis. While

*. . . unlikely that spending in either area . . . directly affect the number of teachers in the classroom or the types of teachers hired, . . . will influence the school climate which . . . influence student achievement.*

<sup>10</sup> Eighth-graders are analyzed in another study (Wenglinsky 1997).

<sup>11</sup> The NAEP SES variables have been criticized for relying on student self-reports and not including a family income measure. In their comparison of various large scale databases that used both a student and parent self-report, however, Berends and Koretz (1995) found little difference between the two types of reports, suggesting that the use of student self-reports is not problematic. In terms of the lack of a family income measure, while this may be true on the student level, there is an indicator of family income at the school level—the percentage of students receiving free or reduced-price lunches, which is used as part of the SES measure in this study.

the information provided can be used to measure district-by-district per-pupil expenditures in broad spending categories, such as instruction or capital outlays, it cannot be relied upon for more detailed information because differences in the charts of accounts of school districts result in their categorizing specific expenses differently. Therefore, CCD was used to provide measures of expenditures on instruction, central office administration, school-level administration, and capital outlays only. CCD was used here, even though the district level is its lowest level of aggregation, because no nationally representative database exists that measures different types of expenditures at a lower level of aggregation.<sup>12</sup>

The TCI is the result of a study by NCES. NCES has conducted analyses to develop an index of the cost of hiring teachers for particular regions of the country (NCES 1995b). This cost can be expected to vary by region, even for teachers of similar levels of experience and education, because the cost of living, quality of life, and other factors all differ by region. The TCI was developed by applying regression analysis to the Schools and Staffing Survey (SASS), an NCES survey, conducted in 1990–91. The regression analysis estimates the influence of various factors on teacher salaries; these include factors that are under the control of schools and school districts, such as teacher experience and education, as well as those that are not, such as 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 teacher costs in a particular region, holding constant the discretionary factors. TCI scores have been estimated for each state, and

*For this study the NAEP data were for fourth-graders taking the 1992 mathematics assessment. This sample consisted of 9,414 students in 270 school districts around the United States.*

these are used in this analysis to adjust the per-pupil expenditure measures (NCES 1995b, 51).<sup>13</sup>

To analyze data from these sources, all three needed to be linked together. For this study the NAEP data were for fourth-graders taking the 1992 mathematics assessment. This sample consisted of 9,414 students in 270 school districts around the United States. Of the school districts, 48 were private schools and therefore no corresponding information was available in CCD. Of the remainder, 195 school districts were linked to CCD through common identification numbers, 8 were linked through common address information, and 19 (7 percent of the sample) could not be matched. State-level TCI scores were linked to CCD and NAEP by locating the state in which each school district was located and entering the appropriate TCI score. These linking procedures were used to produce two databases, one at the district level and one at the student level. The district-level database was produced by aggregating NAEP data to the district level and linking it to the already district-level CCD. The student-level database was produced by disaggregating CCD to the student level and linking it to the already student-level NAEP. The district-level database was used for all analyses except the multilevel one, for which the student-level database was used. Because NAEP is a sample while CCD and TCI are universes, the two databases took on the sampling characteristics of NAEP; this means that the databases are nationally representative samples of public schools and their students and that the weighting techniques and standard error adjustments required for NAEP apply.<sup>14</sup>

---

<sup>12</sup> Few school systems collect budget information at the school level. For a recent study addressing this issue in Texas and Ohio, see NCES (1996).

<sup>13</sup> The TCI is a cost-of-education index. It differs from a cost of living index in that a cost of living index only measures the cost of living, while a cost-of-education index measures other factors that affect the cost of education as well. See Barro (1994) for a discussion of the differences between cost-of-living indices and cost-of-education indices.

<sup>14</sup> Because NAEP is not nationally representative at the district level, all results refer to students and schools in districts, and not the districts themselves (Johnson, Rust, and Wallace 1994).

The district-level database was then used to produce measures of the variables needed to test the hypotheses (see table 1 for means and standard deviations and appendix A for full definitions). The database included the four expenditure measures, the number of pupils in the school district and the TCI score for that state. Cost-adjusted per-pupil expenditures in the four areas were calculated by dividing each by the number of pupils and the TCI. The database also included seven SES measures summed to create an SES variable; seven school environment measures summed to create a school environment variable; a measure of teacher’s highest degree attained; the number of full-time teachers and students in the school, used to calculate the school teacher-student ratio; and five measures of mathematics achievement known as “plausible values,” the use of which will be discussed below.

## Method

The bulk of analyses were conducted on the district-level database using a structural equation modeling program, LISREL 8. LISREL requires as input rules regarding which variables are allowed to be related to one another and which are not, and a covariance matrix calculated from data. The program then estimates parameters relating the variables allowed to be related while maximizing the goodness of fit between the covariance matrix these parameters imply and the input covariance matrix. LISREL produces three principal outputs: the estimates of the direct effects between variables; estimates of the total effects between variables; and the goodness of

fit as measured by adjusted goodness-of-fit and normed goodness-of-fit indices (Joreskog and Sorbom 1993). Models are considered to have a satisfactory fit when the chi-square is statistically insignificant (indicating that there is no significant difference between the input covariance matrix and the implied covariance matrix) and the adjusted and normed goodness-of-fit indices are more than 0.9 (Bentler and Bonnett 1980). LISREL also allows for the comparison of goodness of fit between the hypothesized model (referred to here as the full model) and a model in which the relationships found to be significant in the full model are fixed as being unrelated to one another (referred to here as the nested model). By running such a nested model and comparing its chi-square to that of the full model, it is possible to reject the nested model in favor of the full one (Hayduk 1987).

First, full and nested models were designed to test the hypothesized relationships. For the full model, the four cost-adjusted per-pupil expenditure measures and the SES index were treated as exogenous variables; their values were not allowed to depend on those of the other variables. Per-pupil expenditures on instruction and central office administration were allowed to affect school environment; SES was allowed to affect school environment and academic achievement; teacher-student ratio was allowed to affect teacher education, school environment and academic achievement; teacher education was allowed to affect school environment and academic achievement; and school environment was

	Mean	Standard deviation
Instructional per-pupil expenditures (PPE)* (dollars)	2999.89	754.47
Central administration PPE* (dollars)	113.38	99.29
School administration PPE* (dollars)	287.72	89.65
Capital outlays PPE* (dollars)	499.54	548.66
Socio-economic status (summated scale)	13.58	2.36
Teacher-student ratio (number of teachers/students)	0.05	0.01
Teacher’s highest degree (1=<BA, 2=BA, 3=MA, 4=>MA)	2.53	0.40
School environment (summated scale)	22.39	2.14
Mathematics achievement (mean for five plausible values)	210.65	17.41

\*Adjusted for regional variations in the cost of education.  
 SOURCE: Wenglinsky, unpublished tabulations.

allowed to affect academic achievement.<sup>15</sup> For the nested model, the relationships that were found to be significant in the full model and were either directly or indirectly associated with achievement were fixed at zero (making them unrelated to one another).

A design effect was then calculated through running a series of preliminary LISREL models. LISREL parameter and standard error estimates assume a simple random sample, and since NAEP is a clustered, stratified sample, these estimates are inaccurate (Johnson 1989). To adjust parameters for the NAEP sample design, covariance matrices used in all analyses were weighted by a student base weight, provided by the NAEP database. Covariance matrices were also weighted by the number of students in each school district. To adjust standard errors for the NAEP sample design, a design effect that estimated the amount by which the standard error estimate was downwardly biased in assuming a simple random sample had to be calculated. This was accomplished by first running a LISREL analysis for the full model on a covariance matrix weighted by only the student base weight and the number of students per school district, thus producing baseline estimates. LISREL analyses were then conducted for the full model on 56 covariance matrices, each weighted by the jackknife replicate weight provided by the NAEP database. For three representative relationships, the variance of the 56 estimates was calculated and the variance for the baseline model was divided by this jackknife variance, producing three estimated design effects, the most conservative of which was used for subsequent analyses (1.75).

Five full models were then run on five covariance matrices. Five models needed to be run to take into account “plausible values” methodology in the

measurement of academic achievement. Students who take the NAEP examination each receive only a subset of the items. In order to impute total scores, it is necessary to use models that take into account other information about the students, including their demographic characteristics. Five achievement scores are produced for each student, each based upon slightly different models. The variability of the scores needs to be taken into account in the estimation of standard errors of all coefficients in which achievement scores are involved (Johnson, Mislevy, and Thomas 1994).<sup>16</sup> This analysis employed a standard methodology, conducting five LISREL analyses for the full model on five covariance matrices, each using one of the plausible values as its achievement measure; calculating parameters as the mean of those for the five analyses; and then adjusting the mean of the standard errors for the five analyses by multiplying by the square root of the design effect and, for the parameters involving achievement, adding the product of 1.2 and the variance of the five parameter estimates (O’Reilly et al. 1996, 78–79). In order to assess goodness of fit, five nested models were run on the same covariance matrices as were used for the full models, and the mean of the goodness-of-fit statistics for the five full models were compared to the mean of the goodness-of-fit statistics for the five nested models.

*Finally, a multilevel estimation program, Hierarchical Linear Modeling (HLM) was applied to the student-level database to test the sensitivity of the LISREL model to multilevel data structure.*

Finally, a multilevel estimation program, Hierarchical Linear Modeling (HLM) was applied to the student-level database to test the sensitivity of the LISREL model to multilevel data structure. Much of the LISREL model involved a single-level data structure and was therefore not re-estimated as a multilevel model; the relationships among the first three steps of the model, expenditures, resources and social environment, all involve district- or school-level variables. The relationships between resources and student achievement, however, involve school-

---

<sup>15</sup> Teacher education was not allowed to reciprocally affect teacher-student ratio, in order to keep the model recursive. The choice of having teacher-student ratio precede teacher education was arbitrary, but, as indicated by modification indices, did not significantly affect the goodness-of-fit of the model.

<sup>16</sup> Researchers have recently proposed an alternate approach to plausible values, known as direct estimation (Cohen 1998).

level independent variables and a student-level dependent variable, the situation under which multi-level techniques are appropriate. The HLM thus consisted of student achievement as the dependent variable and the two resources (teacher's highest degree and teacher-student ratios) as independent variables. As in the LISREL model, SES was incorporated as a statistical control. School-level SES, the school-level aggregate of student-level SES, was included as an independent variable, and the student-level relationship between SES and achievement was included as an additional dependent variable. Plausible values methodology is handled automatically by HLM, which ran separate models for each plausible value and combined them into a single model (Bryk, Raudenbush, and Congdon 1996). The resulting model thus takes the underestimation of standard errors due to measurement variability into account, although it does not take the underestimation of standard errors due to sampling variability into account.

To confirm that a particular expenditure or resource is part of the production function, four results must occur. First, the direct effects measured that trace a path to student achievement must be statistically significant; if they are not, it brings into doubt the reliability of the model. Second, the goodness-of-fit measures for the full models must all confirm the models, while those for the nested models must be unsatisfactory; if not, the null hypothesis may hold. Third, the total effects should be substantial enough for a feasible level of investment to produce marked improvements in student performance; if not, the inputs are not of interest from a policy standpoint. Fourth, the HLM results should be consistent with the LISREL results; otherwise, the latter may be rejected for failing to take into account the multilevel nature of the data.

## Results

The expenditure and resource variables measured in the structural equation model are consistent with what is generally known (table 1).<sup>17</sup> Instructional per-pupil expenditures are, on average, \$3,000 per student, and 68 percent of the school districts in the sample spend between \$2,620 and \$3,380. This spending level constitutes 60 percent of current per-pupil expenditures. Central administration per-pupil expenditures are \$113 per student, and school administration per-pupil expenditures are \$288, constituting 3 percent and 6 percent of current per-pupil expenditures, respectively. These amounts for administrative expenditures might appear low, but are in fact consistent with estimates from other studies. Administrative expenditures refer only to superintendents, principals and

their staffs, and so do not include support services, from student transportation to janitorial services, that are often perceived as being part of the administrative category. Five hundred dollars are spent per-pupil on capital outlays, and here there is wider variation than with the other expenditure variables; the standard deviation is nearly \$550. The average teacher-student ratio is 0.05 teachers per student, which means 1 teacher for every 20 students. This seems to be a low number, except that it also includes special education classes, which may have

teacher-student ratios as low as 1:1. The average teacher's highest degree is somewhere between a bachelor's and a master's.

The estimates from the full structural equation model reveal that some expenditures and resources are part of the production function while others are not (table 2). Instructional and central office admin-

*The expenditure and resource variables measured in the structural equation model are consistent with what is generally known.*

<sup>17</sup> For examples of distributions of expenditures found in other studies that conform to those found here, see Adams (1994) and Miles (1995).

**Table 2.—LISREL estimates of direct effects**

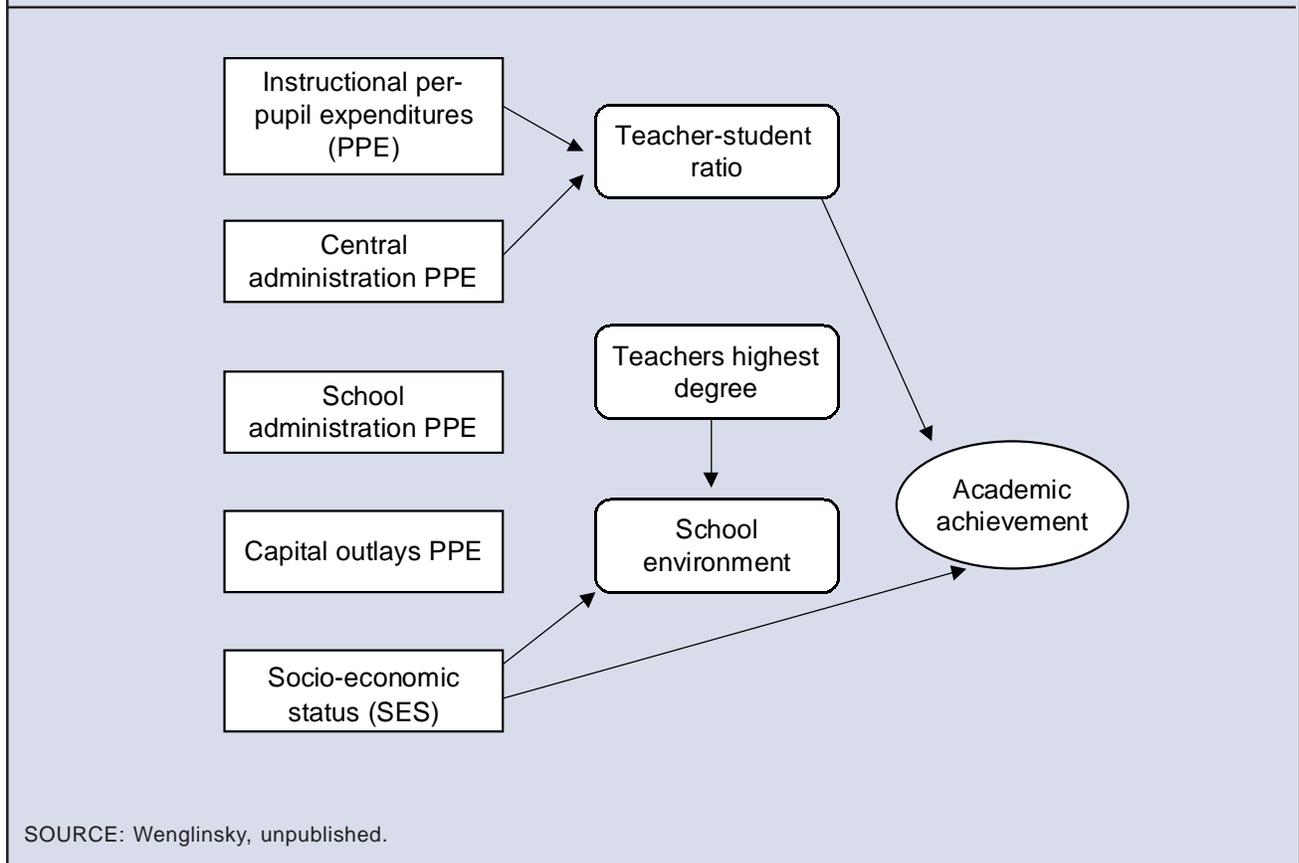
	Teacher-student ratio	High degree	School environment	Mathematics achievement
Instructional per-pupil expenditures (PPE) (thousands of dollars)	0.0000051** 0.0000015 0.3000870	0.0000787 0.0000524 0.1484077	—	—
Central administration PPE (thousands of dollars)	0.0000368** 0.0000108 0.2850938	0.0000938 0.0003962 0.0232586	0.0023330 0.0019294 0.1080068	—
School administration PPE (thousands of dollars)	—	—	0.0019241 0.0020389 0.0804268	—
Capital outlays PPE (thousands of dollars)	—	—	-0.0001847 0.0003314 -0.0472520	—
Socio-economic status (summated scale)	—	—	0.3679910** 0.0778429 0.4416251	5.6298038** 0.4050784 0.6395601
Teacher-student ratio (number of teachers/students)	—	-2.1375013 3.2277413 -0.0683488	-17.6074397 15.0059229 -0.1050825	152.1944001** 74.5571103 0.1119985
Teacher's highest degree	—	—	-0.8180802* 0.4473701 -0.1526883	1.5975948 2.4070308 0.0367765
School environment (summated scale)	—	—	—	0.5822595 0.4903571 0.0718114
Mathematics achievement (plausible values)	—	—	—	—
— Relationship fixed at zero.				
* p<.10				
** p<.05				
NOTE: Cells contain unstandardized parameters, standard errors, and standardized parameters.				
SOURCE: Wenglinsky, unpublished tabulations.				

istration expenditures do result in improved achievement. They positively affect teacher-student ratios, with standardized coefficients of 0.30 for instruction and 0.29 for central office administration. Teacher-student ratios, while not being associated with school environment as was expected, are directly associated with mathematics achievement, with a standardized coefficient of 0.11. On the other hand, school-level administration and capital outlays proved not to be related to school climate or

mathematics achievement. Teacher's highest degree is weakly related to school environment (albeit in the counterintuitive direction), but school environment appears not to be related to mathematics achievement. Thus instructional expenditures, central office administration expenditures and teacher-student ratios appear to be part of the production function, while school-level administration, capital outlays and teacher's highest degree are not (see figure 2 for a schematic representation of results).<sup>18</sup>

<sup>18</sup> The analysis of eighth graders found the same three input variables to be components of the production function. It differed from the fourth grade analysis in that school climate mediated between the inputs and achievement. Instructional and central office administration expenditures were positively related to teacher-student ratios, which, rather than being directly related to achievement, were directly related to school climate. School climate, in turn, was related to achievement (Wenglinsky 1997).

Figure 2.—Paths to fourth grade achievements



To confirm this set of findings, goodness of fit was measured and compared to the goodness of fit of a model in which instructional and central office administration expenditures and teacher-student ratios were eliminated from the production function. In the full model, the chi-squares proved statistically insignificant, indicating good fit, with a mean chi-square of 25.67 across the five plausible values and a significance level of 0.06. The goodness-of-fit indices were also of sufficient size, with a mean adjusted goodness-of-fit index of 0.925 across the five plausible values and a mean normed goodness-of-fit index of 0.936. In the nested model, the chi-squares proved statistically significant, with a mean chi-square of 78.73 and a significance level better than 0.0001. The goodness-of-fit indices were of insufficient size, with a mean adjusted goodness-

of-fit index of 0.817 and a mean normed goodness-of-fit index of 0.804. The goodness-of-fit measures, then, confirm that the model with the three production function components has an adequate fit and that an alternate model that excludes the components does not.

Estimates of the total effects of the production function components indicate that their effect on achievement can be substantial (table 3). The total effect of instructional per-pupil expenditures on mathematics achievement is statistically significant and amounts to 3.2 points of achievement for every \$4,000 dollars. The total effect of central office administration on mathematics achievement is 3.3 points for every \$500. Given that 12 points repre-

**Table 3.—LISREL estimates of total effects**

	Teacher-student ratio	High degree	School environment	Mathematics achievement
Instructional per-pupil expenditures (PPE) (thousands of dollars)	0.0000051**	0.0000679	-0.0001452	0.0007987*
	0.0000015	0.0000499	0.0000951	0.004659
	0.3000870	0.1278971	-0.0510623	0.0346460
Central administration PPE (thousands of dollars)	0.0000368**	0.0000152	0.0016734	0.0065922*
	0.0000108	0.0003787	0.0018442	0.0034844
	0.2850938	0.0037728	0.0776724	0.0376322
School administration PPE (thousands of dollars)	—	—	0.0019241	0.0011203
			0.0020389	0.0015197
			0.0804268	0.0057755
Capital outlays PPE (thousands of dollars)	—	—	-0.0001847	-0.0001076
			0.0003314	0.0002136
			-0.0472520	-0.0033932
Socio-economic status (summated scale)	—	—	0.3679910**	5.8440701**
			0.0778429	0.4059927
			0.4416251	0.6395601
Teacher-student ratio (number of teachers/students)	—	-2.1375013	-15.8587921	139.5456070*
		3.2277413	0.0000951	75.2240084
		0.1278971	-0.094645	0.1026882
Teacher's highest degree	—	—	-0.8180802*	1.1212598
			0.4473701	2.3910553
			-0.1526883	0.0258117
School environment (summated scale)	—	—	—	0.5822595
				0.4903571
				0.0718114
Mathematics achievement (plausible values)	—	—	—	—

— Relationship fixed at zero.  
 \* p<.10  
 \*\* p<.05  
 NOTE: Cells contain unstandardized parameters, standard errors, and standardized parameters.  
 SOURCE: Wenglinsky, unpublished tabulations.

sents a grade level, these effects are fairly substantial.<sup>19</sup> The effect of teacher-student ratios is still stronger. The total effect of teacher-student ratios on student achievement is 140 points for an increase

of 1 teacher per student. Translated into class sizes, this means that a reduction in class size from 25 students to 15 students would result in an achievement gain of 14 points, well over a grade level.<sup>20</sup>

<sup>19</sup> It could be argued that a \$4,000 increase in instructional expenditures is infeasible. Yet, a district would not need to raise all \$4,000; some money could be obtained through the reallocation of existing funds. Thus, if a school district is currently spending \$7,000, of which it allocates \$3,000 for instruction, it could potentially increase spending on instruction by \$4,000 by increasing aggregate expenditures by \$2,000, to \$9,000 and reallocating \$2,000 of existing funds. It should also be noted that translating dollars into achievement assumes linearity, which may not be the case. It may be that only spending changes of a certain threshold translate into achievement changes. It may also be that only spending changes for school districts that begin at a certain level of expenditure result in achievement changes.

<sup>20</sup> It should not be surprising that the effect of teacher-student ratios is stronger than the effects of the two expenditure measures. To the extent that instructional and administrative dollars are spent on teacher-student ratios, they are conducive to academic achievement. Yet, not all instructional and administrative dollars are invested in ways that raise teacher-student ratios. Thus, while the most effective investment strategy to increase achievement would be to raise directly teacher-student ratios, where this is not feasible it is still possible to produce gains through allocating expenditures to the two areas known to raise these ratios.

Finally, the HLM analyses of the student-level database are consistent with the LISREL findings (table 4). As in the LISREL model, teacher-student ratios are significant related to achievement levels. The unstandardized coefficient is 153.8, as opposed to 152.1 for the LISREL model. Also, as in the LISREL model, socio-economic status is significantly related to achievement levels. The unstandardized coefficient is 6.01, as opposed to 5.63 in the LISREL model. Further, teacher’s highest degree is not significantly related to achievement levels as in the LISREL model. It is also interesting to note that the only independent variable found to be significantly related to the SES-achievement relationship is district-level SES, suggesting that while resources can be associated with the level of achievement, they cannot affect its social distribution, at least for the population of fourth graders.

In sum, a series of structural equation models made it possible to identify some expenditures and resources that affect student achievement. Expenditures on instruction and central office administration affect teacher-student ratios, which, in turn affect student achievement. On the other hand, capital outlays, school-level administration and teacher education levels were found not to be associated with student achievement. These relationships persisted when subjected to multilevel analysis using HLM. It remains to discuss the implications of these results and the techniques employed to obtain them for the viability of the production function approach.

## Conclusions

The models described here show that the key shortcomings of production functions can be addressed. First, the study was able to produce results that are national in scope. Since no single national database contains all of the variables needed for a production function, data were drawn from two universes and a sample and linked to one another. Second, the study distinguished between different types of expenditure. CCD made it possible to measure four types of expenditure, and the structural equation model made it possible to relate these to different parts of the learning process, such as school climate. This proved an important innovation because not all expenditures had an effect on achievement; those for the central office and instruction did, but those for capital and the principal’s office did not. Third, the study took into account the role of school climate. NAEP provided a set of indicators of school climate that could be used to create a scale, and the structural equation model made it possible to measure both the influence of expenditures and resources on school climate and the influence of school climate on student achievement. In this study, however, the innovation proved of limited utility, since school climate was found not to play a mediating role in the production function. Fourth, the study measured student SES in a reasonably robust fashion, using a scale calculated from the measures provided by NAEP. Structural equation modeling made it possible to measure its influ-

**Table 4.—Hierarchical linear modeling (HLM) unstandardized estimates of direct effects on intercept and slope of mathematics achievement**

	Relationship to achievement intercept	Relationship to achievement slope
Teacher-student ratio (number of teachers/students)	153.764687*	-5.686118
	61.900057	28.600768
Teacher’s highest degree	3.028100	0.826238
	2.533728	0.976926
Socio-economic status (summated scale)	6.008911*	0.444126*
	0.333620	0.139585

\*p<.05

NOTE: Cells contain unstandardized parameters and standard errors.

SOURCE: Wenglinsky, unpublished tabulations.

ence on two variables, school climate and student achievement. This proved important because both relationships were significant. Fifth, the study adjusted the expenditure measures by the cost of education, using the TCI. This proved important as well, since the relationships would have been markedly different without these adjustments. Sixth, the study used a sophisticated achievement measure, drawn from NAEP, and applied it appropriately through adapting plausible values methodology to structural equation modeling. This innovation also proved important, as illustrated by the fact that many of the relationships which were found to be statistically insignificant would have appeared significant using the unadjusted mean of the plausible values. Even slight changes in the measurement of achievement can have significant effects on production function results. Finally, the study applied HLM to student-level data. This innovation actually proved unimportant; the HLM results did not differ substantially from the LISREL results.

Much more remains to be done, however. First, there were important differences in the findings from this study of fourth graders and a similar study of eighth graders. It therefore cannot be presumed that the production function for one grade level is the same for all; other grade levels should be studied. Second, many resource variables that might affect achievement were omitted from this analysis. The study used teacher education as a measure of teacher quality and found no relationship. Other measures need to be tested, however, before researchers arrive at the counterintuitive finding that teacher quality does not matter; for instance, teacher experience, teacher proficiency on standardized tests, and teachers having majored in the subject matter they are teaching, all may potentially influence student achievement. Finally, the current study uses cross-sectional data; meta-analyses (Hanushek 1997; Greenwald, Hedges, and Laine 1996) suggest that longitudinal data produces somewhat different findings. It is therefore important that a database be developed that tracks both inputs and outputs for a sample of students and schools over time.

## References

- Adams, Jacob E. 1994. "Spending School Reform Dollars in Kentucky: Familiar Patterns and New Programs, But Is This Reform?" *Educational Evaluation and Policy Analysis*. 16 (4): 375–390.
- Austin, Gilbert R. and Herbert Garber (eds.). 1985. *Research on Exemplary Schools*. New York: Academic Press, Inc.
- Barro, Stephen M. 1994. *Cost-of-Education Differentials Across the States*. Washington, D.C.: U.S. Department of Education.
- Barton, Paul, Margaret Goertz, and Richard Coley. 1991. *The State of Inequality*. Princeton, NJ: Educational Testing Service.
- Bentler, Peter M. and Douglas G. Bonnett. 1980. "Significance Tests and goodness of fit in the analysis of covariance structures." *Psychological Bulletin*. 88: 588–606.
- Berends, Mark and Daniel M. Koretz. 1995. "Reporting Minority Students' Test Scores: How Well Can the National Assessment of Educational Progress Account for Differences in Social Context?" *Educational Assessment*. 3(3): 249–285.
- Brookover, Wilbur, Charles Beady, Patricia Flood, John Schweitzer, and Joe Wisenbaker. 1979. *School Social Systems and Student Achievement: Schools Can Make a Difference*. Brooklyn, NY: J.F. Bergin Publishers.
- Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Bryk, Anthony S., Stephen W. Raudenbush, and Richard T. Congdon. 1996. *Hierarchical Linear and Nonlinear Modeling with HLM/2L and HLM/3L Programs*. Chicago, IL: Scientific Software International.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Cohen, Jon. 1998. "Redesigning NAEP to Increase Its Usefulness." San Diego, CA: American Educational Research Association Annual Meeting.
- Cuban, Larry. 1984. "Transforming the Frog into a Prince: Effective Schools Research, Policy and Practice at the District Level." *Harvard Educational Review*. 54: 129–151.
- Edmonds, Ronald. 1979. "Effective Schools for the Urban Poor." *Educational Leadership*. 37 (1): 15–24.
- Finn, Jeremy D. and Charles M. Achilles. 1990. "Answers and Questions about Class Size: A Statewide Experiment." *American Educational Research Journal*. 27 (3): 557–577.
- Fortune, Jim C. and John S. O'Neil. 1994. "Production Function Analyses and the Study of Educational Funding Equity: A Methodological Critique." *Journal of Education Finance*. 20 (Summer): 21–46.

Glass, Gene V. and Mary Lee Smith. 1979. "Meta-analysis of Research on Class Size and Achievement." *Educational Evaluation and Policy Analysis*. 1 (1): 2-16.

Greenwald, Robert, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research*. 66(3): 361-396.

Hambleton, Ronald K., H. Swaminathan, and H. Jane Rogers. 1991. *Fundamentals of Item Response Theory*. London: Sage Publications.

Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher*. 18 (4): 45-65.

Hanushek, Eric A. 1996. "School Resources and Student Performance." In Gary Burtless (ed.). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Pp. 43-73.

Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis*. 19 (2): 141-164.

Hauser, Robert M., William H. Sewell, and Duane F. Alwin. 1976. "High School Effects on Achievement." In William H. Sewell, Robert M. Hauser and David C. Featherman (eds.). *Schooling and Achievement in American Society*. Pp. 309-342. London: Academic Press.

Hayduk, Leslie A. 1987. *Structural Equation Modeling with LISREL: Essentials and Advances*. Baltimore, MD: Johns Hopkins University Press.

Hedges, Larry V., Richard D. Laine, and Robert Greenwald. 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher*. 23 (3): 5-14.

Jencks, Christopher, Marshall Smith, Henry Ackland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.

Johnson, Eugene. 1989. "Considerations and Techniques for the Analysis of NAEP Data." *Journal of Educational Statistics*. 14 (4): 303-334.

Johnson, Eugene. 1994. "Overview of Part I: The Design and Implementation of the 1992 NAEP." In Eugene G. Johnson and James E. Carlson (eds.). *The NAEP 1992 Technical Report*. Pp. 9-32. Princeton, NJ: Educational Testing Service.

Johnson, Eugene, Robert J. Mislevy, and Neal Thomas. 1994. "Scaling Procedures." In Eugene G. Johnson and James E. Carlson (eds.). *The NAEP 1992 Technical Report*. Pp. 241-256. Princeton, NJ: Educational Testing Service.

Johnson, Eugene G., Keith Rust, and Carol Wallace. 1994. "Weighting Procedures and the Estimation of Sampling Variance." In Eugene G. Johnson and James E. Carlson (eds.). *The NAEP 1992 Technical Report*. Pp. 193-239. Princeton, NJ: Educational Testing Service.

- Joreskog, Karl G. and Dag Sorbom. 1993. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago: Scientific Software International.
- Lee, Valerie E., Anthony S. Bryk, and Julia B. Smith. 1993. "The Organization of Effective Secondary Schools." *Review of Research in Education*. 19: 171–267.
- Miles, Karen H. 1995. "Freeing Resources for Improving Schools: A Case Study of Teacher Allocation in Boston Public Schools." *Educational Evaluation and Policy Analysis*. 17 (4): 476–493.
- Monk, David H. 1992. "Educational Productivity Research: An Update and Assessment of Its Role in Education Finance Reform." *Educational Evaluation and Policy Analysis*. 14: 307–332.
- Mosteller, Frederick. 1995. "The Tennessee Study of Class Size in the Early School Grades." *The Future of Children: Critical Issues for Children and Youths*. 5(2): 113–127.
- Mosteller, Frederick, Richard J. Light, and Jason A. Sachs. 1996. "Sustained Inquiry in Education: Lessons from Skill Grouping and Class Size." *Harvard Educational Review*. 66(4): 797–842.
- Mullis, Ina V.S., John A. Dossey, Eugene H. Owen, and Gary W. Phillips. 1993. *NAEP 1992 Mathematics Report Card for the Nation and the States*. Princeton, NJ: Educational Testing Service.
- National Center for Education Statistics. 1995a. *Disparities in Public School District Spending: 1989–90*. Washington, D.C.: Government Printing Office.
- National Center for Education Statistics. 1995b. *Public School Teacher Cost Differences Across the United States*. Washington, D.C.: Government Printing Office.
- National Center for Education Statistics. 1996. *Assessment and Analysis of School-Level Expenditures: Working Paper 96–19*. Government Printing Office: Washington D.C.
- O'Reilly, Patricia E., Christine A. Zelenak, Alfred M. Rogers, and Debra L. Kline. 1996. *1994 Trial State Assessment Program in Reading Secondary-Use Data Files User Guide*. Washington, D.C.: U.S. Department of Education.
- Odden, Allan, and William H. Clune. 1995. "Improving Educational Productivity and School Finance." *Educational Researcher*. 24 (9): 6–10,22.
- Odden, Allan. 1990. "Class Size and Student Achievement: Research-Based Policy Alternatives." *Educational Evaluation and Policy Analysis*. 12 (2): 213–227.
- Purkey, Stewart C. and Marshall S. Smith. 1983. "Effective Schools: A Review." *Elementary School Journal*. 83: 427–452.
- Raudenbush, Stephen W. and J. Douglas Willms. 1995. "The Estimation of School Effects." *Journal of Educational and Behavioral Statistics*. 20 (4): 307–355.
- Wenglinsky, Harold H. 1997. "How Money Matters: Models of the Effect of School District Spending on Academic Achievement." *Sociology of Education*. 70 (3).

## **Appendix A: Variable Definitions**

**Capital Outlays Per-pupil Expenditures:** Derived from data in CCD for Fiscal Year 1992. Calculated by dividing total capital outlays, as defined in CCD, for each school district by the number of students in the school district and the Teacher Cost Index. Measured in thousands of dollars.

**Central Administration Per-pupil Expenditures:** Derived from data in CCD for Fiscal Year 1992. Calculated by dividing total expenditures on central administration, as defined in CCD, for each school district by the number of students in the school district and the Teacher Cost Index (TCI). Measured in thousands of dollars.

**Highest Degree:** Taken from NAEP data for mathematics for 1992. Consists of the highest level of education attained by teacher responding to NAEP on behalf of a student. Responses were coded "1" for less than a Bachelor's degree, "2" for a Bachelor's degree, "3" for a Master's degree and "4" for more than a Master's degree.

**Instructional Per-pupil Expenditures:** Derived from data in CCD for Fiscal Year 1992. Calculated by dividing total expenditures on instruction, as defined in CCD, for each school district by the number of students in the school district and the Teacher Cost Index. Measured in thousands of dollars.

**Mathematics Achievement:** Taken from NAEP data for mathematics for 1992. Consists of the five plausible values for students responding to NAEP. Means and standard deviations presented in this paper are means of these statistics for the five plausible values. For all maximum likelihood estimates, plausible values were analyzed in accordance with plausible values methodology. Measured on common proficiency scale for all grades (fourth, eighth and twelfth).

**School Administration Per-pupil Expenditures:** Derived from data in CCD for Fiscal Year 1992. Calculated by dividing total expenditures on school-level administration, as defined in CCD, for each school district by the number of students in the school district and the Teacher Cost Index (TCI). Measured in thousands of dollars.

**School Environment:** Derived from NAEP data for mathematics for 1992. Calculated as summated scale of the following items: for each school in NAEP the degree to which teacher absenteeism is not a problem; the degree to which student tardiness is not a problem; the degree to which student absenteeism is not a problem; the degree to which class cutting is not a problem; and the degree to which there is a regard for school property; for each teacher in NAEP the degree to which teachers have control over instruction; and the degree to which teachers have control over course content. Measured as total of that scale.

**Socio-economic Status (SES):** Derived from NAEP data for Mathematics for 1992. Calculated as summated scale of the following items: for each student whether or not family receives newspaper; whether or not there is an encyclopedia in the home; whether or not there are more than 25 books in the home; whether or not the family subscribes to magazines; the highest level of education attained by the mother; the highest level of education attained by the father; and for each school in NAEP the percentage of students who receive reduced price or free lunches. Measured as total of that scale.

**Teacher-Student Ratio:** Derived from NAEP data for mathematics for 1992. Calculated by dividing total number of teachers in school by total number of students in school.