School Consolidation and Educational Performance: An Economic Analysis of West Virginia High Schools

A Monograph Prepared for

The West Virginia School Building Authority

May 2004

Michael J. Hicks & Viktoriya Rusalkina

Center for Business and Economic Research
Marshall University
One John Marshall Drive
Huntington, West Virginia 25755

The results and interpretation of data contained in this study are entirely those of the authors and do not reflect the opinions of Marshall University, its governing body, or sponsors of this research.

CONTENTS

1. INTRODUCTION

The relative effectiveness and efficiency of public services guides a number of important policy decisions regarding the deployment of public resources. Policymakers naturally desire to allocate scarce resources efficiently (at low cost) and effectively (impact per dollar spent). With the high proportion of tax revenues dedicated to public education, and the importance of quality schooling it is reasonable that efficiency and effectiveness in public education be considered.

This analysis focuses a larger research effort to that end. (see Rusalkina and Hicks, 2003a; Hicks and Rusalkina, 2003b). In this report we isolate school size and place in context the debate over school size and student performance through empirical analysis of all West Virginia high schools.

First it is important to reveal what this study will not do. We will not evaluate the efficiency of schools within this analysis. More clearly, we do not estimate economies of scale. A common error in the education literature is to measure scale economies through evaluations of production outcomes. While this is appropriate in firms that operate in competitive markets it is a major technical mistake when applied to public schools (which enjoy few of the efficiency characteristics of competitive firms). [1]

[1] The technique which permits the linkage of efficiency (cost functions) and effectiveness (production functions) is known as duality theory.

This study will confine itself to measurement of effectiveness of education as it relates to school size when controlled for other factors. Efficiency studies of West Virginia schools have been proposed to the legislature as joint studies by West Virginia University and Marshall University. These are both important policy issues for the state, but will not be concluded for several months.

We proceed with a brief, largely non-technical analysis of the debate over school size in West Virginia. This is followed by a review of the scholarly literature on school size and performance, econometric and measurement considerations, our analysis and findings, a discussion and our findings as they relate to linear and more complex relationships (including those that we do not yet know). We conclude with a summary of our findings.

2. SCHOOL SIZE RESEARCH

There is a large body of scholarly literature on school performance. The majority of this work evaluates school size within the context of other issues. The quality of the analysis is improving as personal computers and the availability of data remedy some of the drawbacks of earlier research. Chief among the weaknesses of early studies are omitted variable bias. Omitted variable bias, as we describe in more detail later, occurs when an important modifier (or its correlate) to an empirical study is not included.

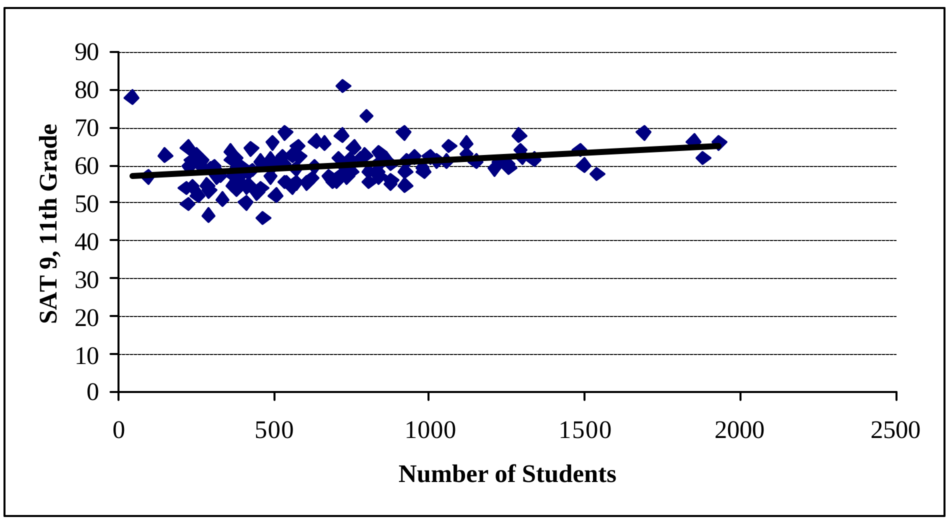
As an example of omitted variable bias, we illustrate the relationship between school size (number of students) and SAT9 test scores for 11th grade students in all West Virginia High Schools (see Figure 1). In this graph it seems clear that bigger schools are, on average, better schools. While this may be the ultimate conclusion of this study, without better analysis that includes student and regional demographics, other school and teacher characteristics, we should conclude that this graph tells us little about the true relationship of size and performance. Importantly, this type of error is ultimately self correcting (through better analysis and peer review) and may be unrelated to researcher bias.

Alas, these types of studies do not inform much of the debate regarding school size and performance in West Virginia. Indeed, we have been unable to uncover a convincing empirical study of West Virginia schools that informs this debate. Most arguments regarding school size and performance is rhetorical, not empirical.

Rhetorical analysis may be useful in two related contexts. First, it often serves as the precursor to formal hypothesis tests, which can then be subjected to scientific analysis. Second, rhetorical analysis is useful where data is not yet available. Rhetorical analysis that is not expeditiously subjected to the rigor of empirics should be viewed as suspect. This is especially true in the field of education where data is unusually abundant.

These observations suggest we will limit our analysis to a narrow set of questions regarding performance and school size. This does not mean the analysis will be trivial or incomplete. On the contrary, we present a robust model of school performance estimates performance in schools, with a large set of control variables.

Figure 1, Performance and School Size in West Virginia High Schools



3. A Review of the Literature

Much research has been attempted to explain the effects of school size on student academic performance. The studies are typically of two types: those that concentrate on educational outputs such as test scores, attendance rates, graduation and dropout rates among others, and those that examine costs and address such issues as economies of scale. This literature review will focus on studies examining the effects of school size on schooling outputs, specifically student academic performance (Stiefel et al, 2000).

While the majority of studies support the idea that students perform better in smaller elementary and middle schools, the results for high schools provide mixed results. Some studies found that students in smaller high schools demonstrate higher levels of academic achievement, while others either do not find any effect of school size on achievement levels, or their results indicate better academic performance for students attending larger high schools. Yet other studies consider medium-sized high school as most favorable for academic achievement (Office of Instructional and Accountability Services, 2000). Clearly, it is essential to define what number of students comprises "a large", "a small", or a "medium-size school" and whether these sizes are comparable across different states and different regions of the country.

An important analysis of school size and student performance was conducted by Driscoll, Halcoussis, and Shirley Svorny [2003]. In their study, "School District Size and Student Performance", the authors examine the impact of school district as well as school sizes on student academic performance. Driscoll et al. use 1999 school data from California schools because this state has numerous schools varying in size, quality, and student demographics.

The authors evaluate 5,525 schools in 755 districts in California. The advantage of this study is that it examines size effects at three levels: district, school, and class. The authors also include population density as a regressor because district size and density are correlated (Driscoll et al., 2003). They separately estimate regressions for elementary, middle and high schools. Among major variables included in the analyses are: district size, school size, class size, median household income, and population density. Driscoll et al. use production function approach with the school level standardized test scores as the dependent variable in the regressions (Driscoll et al., 2003:196).

This study found that "district size has a negative effect on student performance, as measured by standardized scores"[2] (Driscoll et al., 2003: 199). The school size also has a significant negative effect on student performance at elementary school level, but

[2] The coefficient for district size was negative and statistically significant at 1% error level for both elementary and middle school, but it was statistically insignificant for high school regression.

no significant effect on the middle school and high school levels. Similarly, class size is negatively correlated with academic attainment only on elementary school level, and not on the secondary level (Driscoll et al., 2003:199).

In "Revisiting Economies of Size in American Education: Are We Any Closer to a Consensus", the authors take a close look at school consolidation and attempt to come to a consensus on how school and district size affect costs and student performance (Andrews et al., 1999). The extensive literature review of the existing studies on economies of scale in education is also included in this paper.

Andrews et al. examine 15 cost function studies and 12 production function studies to answer the following questions: do school size and school district size matter and is consolidation generally an effective policy? They conclude,

> "…moderation in district and school size may provide the most efficient combination. Under some conditions, consolidation of very small rural districts may save money, as long as schools are kept moderate size and transportation times remain reasonable" (Andrews et al, 2002:256).

Cost functions used in the research, for the most part, lead to a conclusion that there is an opportunity to save significant administrative and instructional costs when moving from a small district with 500 or less students to a larger district with 2,000-4,000 students (Andrews et al, 2002). Andrews et al. note that per student costs may also continue to decline until the enrollment reaches approximately 6,000 students. That is the point where economies of scale are exhausted (Andrews et al, 2002).

Since the studies using cost function do not consider the opportunity costs of increased travel time for students and parents in the case of consolidation, the optimal enrollment, according to the authors, is in fact lower than the studies suggest. This leads to the recommendation that any school district considering consolidation should determine total travel times. If those times are too high, then it could be concluded that any potential cost savings due to consolidation will result not from savings in teacher salaries or maintenance and capital costs from consolidation of school buildings, but from cutting the administrative expenditures, and support staff and services costs. (Andrews et al, 2002:255).

Production function analysis shows that large schools in many cases negatively affect student performance, especially if those schools have a sizable number of disadvantaged students. However, the authors warn that many of the studies do not consider a nonlinear relationship between enrollment and student performance. Andrews et al. also find that “decreasing returns to size may begin to emerge for high schools above 1000 students and elementary schools above 600 students” (Andrews et al., 2002:255). This is a cost, not performance finding.

In their study “The Effect of School Size on Exam Performance in Secondary Schools”, Bradley and Taylor examine the effect of school size on the exam performance in all secondary schools in England for the period of 1992-1996 (Bradley and Taylor, 1998:318). The authors utilize regression analysis using the exam performance of all secondary schools in England as a dependent variable. Independent variables include the following: type of school, selection policy, school gender[3], school age range, teaching staff, pupils, family background, and school size (Bradley and Taylor, 1998:324). Specifically, the model utilizes the following equation: $y^{*} = \mathbf{b'x} + e$, where y^{*} is the unobserved dependent variable[4], and x is the vector of explanatory variables listed above (Bradley and Taylor, 1998:303).

The authors’ findings support arguments by advocates of larger schools. The results indicate, “exam performance increases with school size but at a decreasing rate” (Bradley and Taylor, 1998:318). Specifically, “exam performance is maximized at a school size of around 1,200 for schools with 11-16 years age range, and 1,500 for schools with 11-18 years school range (Bradley and Taylor, 1998: 318). Authors conclude that although some educational benefits unrelated to academic performance may be higher in smaller schools, the results of academic performance are higher in larger schools (Bradley and Taylor, 1998:318).

Several well-known studies on school size and its effects on student achievement were conducted by Valerie Lee and Julia Smith. In “Effects of High School Restructuring and Size on Early Gains in Achievement and Engagement”, the authors specifically address the effects of schools size on student academic performance using organizational structure framework. They utilize existing research on two types of organizational forms-

[3] Some schools were all male/female.
[4] Exam performance of the school (Bradley and Taylor, 1998:303).

bureaucratic and communal[5] to classify schools into unresutructured schools, schools with traditional practices, and schools with restructuring practices (Lee and Smith, 1995:248). In addition, Lee and Smith classify schools according to their size considering "smaller size as a feature of school structure that moved schools toward a communal form" (Lee and Smith, 1995:258).

The study includes a sample of 11,794 sophomores in 830 high schools from the National Educational Longitudinal Study of 1988 (Lee and Smith, 1995:241). The majority of schools in the sample are public (717), 54 are Catholic schools, and 49 are independent secondary schools. To make the sampling procedure more accurate, the authors designed their own school weights (Lee and Smith, 1995:247). Descriptive and multivariate analyses were used in this study. Variables utilized in this paper include: average socioeconomic status (SES), percent minority enrollment, school sector, average achievement, average number of mathematics and science courses, variability in mathematics and science course taking, and school size among others (Lee and Smith, 1995:253). The average school size is comparable across groups (1,000-1,500 students), although traditional–practice schools are typically smaller than restructuring or non-reform schools (lee and Smith, 1995:252).

Lee and Smith conclude, "Students learn more in smaller schools and in schools with several practices that are consistent with restructuring" (Lee and Smith, 1995:259). They also point out that "students in small schools are more engaged in their courses" (Lee and Smith, 1995: 259). However, the authors warn that their findings do not immediately lead to a conclusion that reducing the size of the high school will increase academic performance of the students. On the contrary, Lee and Smith emphasize, that the size of the school has only an indirect effect on student's learning (Lee and Smith, 1995: 262). Specifically, Lee and Smith point out,

> "…the size of enrollments acts as a facilitating or debilitating factor for other desirable practices…. Reducing the size of schools, although a potential

[5] Bureaucratic organizations are characterized by specialized and differentiated work roles, a top-down hierarchy of decision-making, and a formalization of goals and expectations into affectively neutral roles and codes of behavior. Communal organizations are characterized by higher task uncertainty and unpredictability, emphasizing shared responsibility for work, shared commitment to a common set of goals, lateral communication and power in decision making and greater personalization and individual discretion. (Lee and Smith, 1995:243).

structural reform in its own right, would not increase students' learning per se" (Lee and Smith, 1995:262).

Later Lee and Smith used their research described above to conduct another study "High School Size: Which Works Best, and for Whom?" Once again, the authors evaluate the relationship between the size of the school students attend and their academic achievement in reading and mathematics. They analyze data from the National Educational Longitudinal Study of 1988, and utilize it in developing hierarchical linear modeling methods. In this paper, besides examining the relationship between school size and academic performance, the authors also attempt to find out what is the most equitable and the most effective school size in terms of student learning (Lee and Smith, 1997: 205).

The results demonstrate that the "ideal" high school enrolls 600-900 students (Lee and Smith, 1997: 217). Lee and Smith point out that although students in schools with less than 600 students demonstrate lower academic achievement, students attending very large schools with more than 2,100 students demonstrate substantially lower academic achievement. In addition, the findings of this study indicate, "the ideal size of high school holds across schools regardless of their students' social class and ethnic background" (Lee and Smith, 1997:217). Finally, the authors determine that learning is more equitable in smaller schools: "Students learn more in relatively smaller high schools; learning is more equitable in small places" (Lee and Smith, 1997:217).

Barnett et al. (2002), in their study "Size, Performance and Effectiveness: Cost-Constrained Measures of Best-Practice Performance and Secondary-School Size", utilize data envelopment analysis (DEA) to examine the relationship between school size and academic performance in the set of Northern Ireland secondary schools. The authors chose to use DEA method because " it can be used to investigate relative school performance with regard to both higher-ability and lower-ability student outcomes" (Barnett et al, 2002:295).

The data used in this study includes output-input-cost information for the 152 Northern Ireland secondary schools for two academic years: 1994-1995 and 1995-1996. The authors classify these 152 schools according to their size, and according to student performance. There are seven different groups of schools, distinguished according to their size, with the largest being over 1000 students and the smallest being 0-299 students (Barnett et al, 2002:298).

The authors confirm the existence of a very strong relationship between school size and academic performance. The results of the study indicate that the largest two size groups demonstrate considerably higher results in academic performance than do the smaller schools. Specifically, they conclude:

> "Smaller size implies lower specialization effects, lower performance, and hence less opportunities for the students. These findings suggest that, where possible, policy should be directed towards securing larger school size and thus better performance" (Barnett et al, 2002:308).

Lamdin, in his article "testing for the effect of school size on student achievement within a school district", examines the existence and significance of school size effect on student performance. Although his study concentrates on elementary schools, the findings may be very useful in analysis of other school levels and especially in the future research. Lamdin utilizes 1990 data provided by the Baltimore Citizens Planning and Housing Association (CPHA), which include 97 public elementary schools in Baltimore, Maryland (Lamdin, 1995:34).

Lamdin uses the production function approach. Specifically, the independent variables include various school inputs such as socio-economic measure, teacher-pupil ratio, operating expenditure per pupil, among others. The dependent variable is the results of student performance on the California Achievement Test (CAT) (Lamdin,1995:35).

The results of regression analysis demonstrate that school size is negative, but not statistically significant, indicating that school size does not have an unambiguous effect on student performance. However, the author warns, "there could be a non-linear effect masked by the linear functional form" (Lamdin, 1995:43).

In conclusion, Lamdin confirms the significance of socio-economic variables. He also points out that the positive relationship between student performance and school inputs either does not exist, or, as shown in some cases, is negative, and statistically significant. In addition, there is "no significant influence of school size on student achievement" (Lamdin, 1995:43).

"School Reform, School Size, and Student Achievement," ca study performed by Eberts and Schwartz, is in many ways similar to that of Lamdin. However, the authors' conclusions provide different results. Eberts and Schwartz investigate the effect of school size and other educational inputs on mathematics test scores. They use data from the report "Sustaining Effects Study", conducted by the Systems Development Corporation for the former Office of Education (Eberts and Schwartz, 1990: 4).

Similarly to Lamdin's study, the authors use the production function method in their analysis. The sample includes 14,000 fourth-grade students attending the 287 schools. The variables utilized in this study include: student background characteristics, teacher characteristics, principal characteristics, number of administrators, teachers, and office personnel per students, time teachers spend on instruction, preparation, and administration, and others. Major dependent variable is student performance in mathematics (Eberts and Schwarz, 1990: 5). The results suggest that students in large and medium size schools have higher academic performance results than students in smaller schools.

A study conducted by Public Schools of North Carolina State Board of Education "School Size and Its Relationship to Achievement and Behavior", examines the effect of school size in North Carolina. The authors use 1997-98 and 1998-99 "End-of-Grade" (EOG) and "End-of Course" (EOC) test data, dropout data, and school violence data for public schools in North Carolina. The results are then compared to the analogous data nationwide (Office of Instructional and Accountability Services, 2000). Although students in smaller elementary and middle schools demonstrate higher academic performance, this finding did not hold for high schools.

In this study, High schools are initially classified into four categories, according to their size: schools with less than 700 students, schools with 700-1000 students, schools with 1001-1500 students, and schools with more than 1500 students (Office of Instructional and Accountability Services, 2000). Next, student performance in each of these groups is compared in five major subject areas: Algebra I, English I, U.S. History, Biology I, and Economic, Legal, and Political Systems (Office of Instructional and Accountability Services, 2000). The results demonstrate that there are "no differences in test scores in any of the five courses among the four groups of high schools" (Office of Instructional and Accountability Services, 2000).

Stiefel et al., in their paper " High School Size: Effects on Budgets and Performance in New York City", utilize budget and academic performance data to compare large and small high schools in New York City in terms of their cost-effectiveness and academic achievement. Data was provided by Board of Education of the City of New York, and includes information on 121 New York City high schools for the 1995-1996 school year (Stiefel et al., 2000). The authors use ordinary least squares regression (OLS) to build model of budget per pupil and model of budget per graduate. Variables used in these models include the following: budget per student, budget per graduate, size of school[6], poverty[7], limited English proficiency, part-time special education[8], and RCT math test[9] (Stiefel, 2000:32).

The results of the study indicate that both budget and performance measures for small and large high schools in New York City are similar. In fact, the authors point out that both, small and large schools are equally cost-effective due to the high school choice policy implemented in New York City which allows students to choose their school on the basis of its size (Stiefel, 2000:37). The authors' findings do not demonstrate significant differences between the different types of schools (Stiefel, 2000:36). Clearly, an examination of West Virginia High Schools is warranted.

5. AN APPLICATION TO WEST VIRGINIA

Employing the information gleaned from these studies we attempt to isolate a particular factor –school size – in our analysis of school performance in West Virginia. This factor is the direct measurement of consolidation in West Virginia schools. As the literature review presented above should make clear, isolating a single factor contributing to educational outcome is possible, but requires complex methods of estimation. Simple correlations (and sometimes more complex models) risk omitted variable bias. Frankly,

[6] Variable "school size" is divided into three categories: small schools (schools with 0-600 students), medium-sized school (schools with greater than 600-2,000 students; subdivided into greater than 600 to 1,200 and greater than 1,200 to 2,000 students), and large schools (schools with greater than 2,000 students) (Stiefel, 2000:32).
[7] Percentage of students eligible for free lunch in 1994-1995 (Stiefel, 2000:32).
[8] Percentage of students who receive resource room or related services in 1995-1996 (Stiefel,2000:32)
[9] Percentage of students who passed a Regents competency math test in 1995-1996 (Stiefel, 2000:32).

any study that offers conclusions on such complex issues as these through simple correlation analysis is, at best, not useful.

Thus, while our goal here is to evaluate school size and performance, to do so we must also include estimates of other critical variables in order to minimize omitted variable bias. While we can never fully exclude this possibility we can be comfortable that we have made every possible effort to do so. This began with an extensive data collection effort.

One area we will not explore in this analysis is technical efficiency (i.e. economies of scale). That is, we will not estimate whether or not schools are combining inputs in the most efficient manner. This is an important question that goes to the heart of public investment in education. We save this analysis for later research.

The data we collected for this project included all publicly available data from each West Virginia middle and high school. The data collected were from 1997 through 2001 and were available from the West Virginia Department of Education. These data included, but are not limited to all available test scores, attendance and enrollment data for each of the years. Data on number of teachers, administrators and other staff members and their average salaries were collected. The numbers of teachers who met certain categories of educational achievement (proportion with MA's, Ed.D.'s), information regarding advanced placement and enrollment in languages, mathematics, science, and social studies were also available and were collected. The Department of Education also made available information regarding the number of students taking the SAT 9 tests under standard conditions and those that missed the examination. School construction dates (from which we calculated school ages) were also provided by the Department of Education. All of these data permitted us to make additional variables through averaging and three year changes to the levels.

We made several assumptions about each of these data elements that are central to our analysis. We use teacher education as a proxy for teacher quality, other useful measures being largely unavailable. In addition to the number or percentage of teachers in each reported instructional category (e.g. BA+15 or MA degree) we were able to compute a mean number of post secondary years of education for each school.

We use test scores of differing types to measure school quality. We also did this with attendance, though it is clear that the direction of causation may occur in either direction. Similarly, as we estimated the duration a teacher had served we recognized that good schools might be magnets for better teachers, so the direction of causation is reversed in our analysis.[10] In cross sectional analysis it is not possible to establish this direction of causation or endogeneity.[11] It remains a theoretical, not empirical issue for which clear determination is not forthcoming at present.

We matched these data on individual schools with demographic information for the surrounding region. Here we used the local zip codes in which the school was located to proxy school district demographics. Matching zip codes to school districts in a consistent fashion proved too costly an approach. Demographic variables included all Census data for 2000. Here again, it is not always clear in which direction causality flows. For example, median house value is often employed as a measure of wealth, but a number of studies have found (to no one's surprise) that school quality affects home prices.

In some instances we also used county or binary variables for such things as rural/urban dichotomy and county population density. We were also able to combine or scale variables to create such variables as proportion of college graduates in a region. As with any statistical study, the application of proxy variables and assumptions suggests that careful interpretation of the results is warranted. We will endeavor to make these interpretations clear in later sections. There are well over 175 variables available for analysis. All received some level of review (and happily, many were rejected early in this process). We will not present an exhaustive discussion of each variable not used in the final results presented later. We will discuss those not used due to an absence of statistical significance which itself may have important policy implications. The individual variables and summary statistics are available in Hicks and Rusalkina, 2003.

---

[10] This may be an important policy consideration for recruitment and retention of teachers.
[11] This issue has spawned considerable research into both technical and theoretical issues arising from this problem (see Ericsson and Irons, 1994 and Pearl, 2000).

5. MODELING EDUCATION INPUTS IN WEST VIRGINIA

A number of options are available to the economist modeling the connection between various institutional and individual factors that influence educational outcome. The vast economic literature on this issue was reviewed in Rusalkina and Hicks [2002] and in the preceding section. Here we briefly discuss the options available along with our decision to choose the method we describe below. While there are theory driven reasons to select each approach, any selection is also heavily predicated on data availability.

The human capital approach views educational decisions as a function of rational consumers. In this approach the decision on quality, quantity and type of education achieved is the result of a number of factors influencing individuals. This method obviously benefits from its ability to analyze individual decisions. However, data limitations make this technique limited to small samples. This approach is perhaps most appropriate when the available data is at the individual level, it is less appropriate when examining regions.

A reduced form model is a flexible approach that imposes no restrictions on the available theory. This loose approach is often employed when data is aggregated to the school district, county or state level. The reduced for model is especially useful when trying to answer broader questions involving economic growth or migration with regional education as an aggregate input to these decisions. Our data is sufficiently disaggregated that a more sophisticated method is available.

As mentioned earlier the production function approach permits testing various 'inputs' of education on 'outputs' and is thus very appropriate when attempting to measure the influence of a particular 'input' on 'outputs.' The production function approach in education is not typically characterized by the use of specific functional form as often occurs in other industries. This is often a drawback to modeling when an issue, such as scale or scope economies, is estimated. As noted earlier, we will not be estimating scale or scope economies directly since these are issues of technical efficiency (though scale benefits in production will be estimated).[12] This largely removes concern

[12] To be clear, whether or not economies of scale or economies of scope exist is a different question than whether or not scale influences outcomes. The former set of questions must be framed within a cost analysis (since duality theory is not applicable in this type of public good setting). A number of authors in the education literature have employed these terms incorrectly.

for development of specific non-linear functional form in this model. We chose not to use the Data Envelope Analysis since the application of a non-stochastic model would not answer the most important questions in this research.

The modeling of these data that are currently being performed is a production function approach at the individual school and grade level. The production function approach is one of three main modeling methods employed by economists to measure the relationship between school inputs and outputs. It is the most appropriate method of modeling data on aggregate school performance data. The model functionally relates school outputs (or performance) based upon inputs and control variables (such as teacher quality, regional demographics, etc.). We express this as:

Equation 1

$$Y = f(X, M, C)$$

where **Y** is a matrix of school quality measures (such as test scores), **X** a vector of teacher quality measures (education levels), **M** a matrix of other school variables (such as school size and age), and **C** a matrix of control variables (such as per capita income within the school district).

This approach is flexible, comprehensive and suggested by an understanding of the extensive research on educational performance. However, selecting the appropriate specification among many alternatives is an econometric issue of some magnitude.

6. Issues in Econometric Analysis

We test this model using several multivariate statistical techniques including ordinary, weighted and non-linear least squares estimates, instrumental variable, principal components and simple correlation analysis. This analysis is extensive, since we have at least 21 school performance indicators and over 170 explanatory variables from which to estimate the impact of teacher quality on school performance. This potentially results in $21*(170^{169})$ total possible specifications making selection of the most appropriate model challenging. Even calculating the number of possible combinations is not computationally feasible in most settings, and so calls for some selection criterion to generate useful results.

In order to best represent the correlation between various explanatory variables and school performance we employ several test statistics that point to the best fitting model. This approach involves testing and comparing each equation against all possible variations. The chief method for selecting the most appropriate model is the use of the Akaike Information Criterion (AIC). The AIC balances the variance of the estimated equation with a penalty for over use of variables (absence of parsimony). This test statistic is widely used in advanced time series and large data set estimations. The AIC takes the form:

Equation 2

$$AIC = -2\left(\frac{l}{t}\right) + 2\left(\frac{k}{t}\right)$$

where there are *k* parameters and *t* observations estimating the log likelihood function *l*.[13]

The process also includes other test statistics that measure particular elements of goodness of fit or correct for common problems in multivariate estimation. Chief among these are the significance tests (F-statistic) and the Durbin-Watson (D-W) statistic. The Durbin-Watson statistic is a serial correlation test that, in this instance, is a generally understood measure of model selection where serial autocorrelation is not an issue (as with this data).[14]

Also, all variances are treated by White's [1980] heteroscedasticity invariant variance-covariance matrix. This is recommended by the observation of severe heteroscedasticity among some studies mentioned in the literature review. This matrix takes the form:

Equation 3

$$M = \frac{t}{t-k}(X'X)^{-1}\left(\sum_{i=1}^{T} u_i^2 x_i x_i'\right)(X'X)^{-1}$$

[13] $l(b) = \sum_{i=0}^{n} y_i \log\left(1 - F\left(-x_i'b\right)\right) + \left(1 - y_i\right)\log F\left(-x_i'b\right)$ establishes the log likelihood function which is estimated using a second derivative iterative algorithm (Bernt, Hall, Hall and Hausman algorithm).

[14] The Durbin Watson statistic is calculated as $u_t = ru_{t-1} + e_i$ which should, under ideal conditions roughly be equal to two.

where t are the number of observations on k regressors and u is the ordinary least squares residual term.

The chief concern facing this analysis is the clear likelihood of multicollinearity among the regressors. The existence of this condition will certainly prove fatal to a number of specification options. For example, it is clearly impossible to regress four regional income variables without generating this problem to such a degree that the results are meaningless.[15]

Multicollinearity can be handled by a variety of means. First we can eliminate variables that are collinear through simply choosing that variable which minimizes the AIC. This is how we intend to pick the best model, which by definition would mitigate (if not formally minimize) collinearity.[16] This is a preferred approach for demographic variables, but is not a solution for several other variables such as our measures of teacher quality. Clearly, levels and hours of schooling in a particular school are likely to be collinear, so that selecting some alternative is necessary. One method is the principal components method. A principal components regression is a statistical device that allows variables that are linear combinations to present a subset of relationships that describe the underlying variability in the data. This method was employed in the model selection with less success than the actual average number of hours of teacher education created from the underlying data. These criterion combined with correction for typical concerns of ordinary least squares provides a basis for analyzing the data collected on schools.

7. ESTIMATION RESULTS

For our modeling efforts we faced considerable specification choices that reduced to a single model applied across all outcome measures. All of these variables (or some linear combination of them) have been reviewed in the preceding literature. This model appears as:

---

[15] Multicollinearity exists when two or more regressors are linear functions of each other. In this condition, the variance between estimators is low (asymptotically approaching zero in some cases). When computing least squares estimators we use these between variable variances, which when very low may make the estimate unsolvable.

[16] Minimizing multicollinearity involves deriving first and second order conditions of eigenvectors, which are here not computationally feasible. Mitigating the ever present problem of collinearity is sufficient for our purposes.

Equation 4

$$\begin{aligned}Outcome = f(&Teacher\ \text{Education, Percent of Families in Poverty, Percent of Adults with College Degree,}\\ &\text{Age of School, Enrollment, Average Class Size, Average Attendance, Drop Out Rate,}\\ &\text{Population Density, stochastic error, intercept)}\end{aligned}$$

The outcome measures we employ include SAT 9 test scores for grades 7 through 11, PSAT tests for 10th and 11th Grade Students, ACT test scores, enrollment in English, Foreign Languages, Math, Science and Social Studies and advanced placement examinations for grades 10 through 12.

As previously mentioned, this model specification is the result of eliminating available variables in each of our three categories through the use of simple correlation measures (competing variables in the same category were selected by the highest simple correlation). So, for example, when faced with median household income and per capita income in the same region as a proxy for regional economic conditions we chose the variable or variables that exhibited statistically significant correlation with outcome measures. The second step was in choosing the combination of these variables that minimized the AIC. The results appear in the following tables. Analysis of results follows.

TABLE 1, REGRESSION RESULTS FOR SAT 9 TESTING

	9th Grade	10th Grade	11th Grade
C	-90.21067**	-16.1401	17.97182
TEACHERED	0.000788	0.003181***	0.002452**
PERFAMINPOVERTY	-7.5867	-12.40336**	-9.54453**
PERCOLLGRAD	18.93412*	23.94337***	23.67902***
AGEOFSCHOOL	-0.01502	0.002089	0.002338
ENROLL	0.002418	0.000747	0.001878
AVGCLASSSIZE	-0.589524**	-0.24358	-0.394027**
AVGATTEND	1.690819***	0.790491**	0.466266
DROPOUTPERCENT	-0.738763***	-0.533839*	-0.535813*
POP/SQUAREMILES	0.012672**	0.005919[a]	0.006195
R-squared	0.372292	0.511019	0.507367
Adjusted R-squared	0.316906	0.471726	0.467781
S.E. of regression	6.310056	4.226905	3.848594
Sum squared resid	4061.314	2001.073	1658.908
Log likelihood	-360.004	-343.753	-332.314
Durbin-Watson stat	1.799698	2.356341	2.018412
Mean dependent var	57.95387	58.83333	59.79508
S.D. dependent var	7.634716	5.815577	5.275417
Akaike info criterion	6.607211	5.799229	5.611705
Schwarz criterion	6.849935	6.029067	5.841543
F-statistic	6.721766	13.00531	12.81665
Prob(F-statistic)	0	0	0

*denotes statistical significance to the .10 level, ** statistical significance to the .05 level and *** statistically significance to the .01, [a]statistically significant to the .15 level employing asymptotic t-statistics. All standard errors are treated with White's [1980] matrix.

In these sets of results, between 21 and 51 percent of the variation in SAT 9 test scores are explained by the variables we present above. The better performing models offer considerable explanatory power for cross sectional analysis.

Importantly, the teacher education enjoyed a positive and statistically significant impact on SAT 9 scores for 10^{th} and 11^{th} grade students. While the magnitude of the impact is not large, it is important to realize that this is a rough proxy for teacher quality. This is consistent with the findings of both Hanushek and Kreuger. For this variable we used both aggregate years of education and mean years of education. Both results were provided almost identical results. From this evidence alone it is clear that a link between teacher quality and educational outcomes is important, though it calls for considerable additional study.

Not surprisingly, and consistent with all other studies we have observed, measures of income and education play the dominant role in overall explanations of educational outcome. Here, the percent of families in poverty explains a considerable proportion of educational outcome. Also, the percentage of adults in the zip code in which the school is located possessing a college degree explains much of the variation in test scores. This finding is consistent with virtually all earlier research. The direction of these impacts is as expected, and are illustrative of the persistence effect felt by school performance in regions that suffer poor educational achievement.

School size plays a small, positive role in higher test scores among high school students. The effect is modest (and linear in follow up tests) and without significant interaction with other variables. Also, the common levels of statistical significance are not present. The clearest interpretation of school size impact on SAT 9 test scores is that there are none at the high school level.

Importantly, this study does not estimate scale economies so the efficiency of larger schools cannot be determined without costs data. It should be again noted that several other studies have attempted to measure scale economies, with mixed success. Without a better understanding of efficiency gains associated with scale in schools, no policy recommendations are supportable. However, it is equally clear that among middle and high school students, larger schools are not adversely impacting SAT 9 test scores. This is consistent with other economic studies that find only very large school districts suffering ill effects of size on educational outcome.

Class size was statistically significant and negative in this analysis. For high school students, a one-pupil reduction in the average class size resulted in a half a point increase in the average SAT 9 test score. These findings strongly support the work of Krueger. There was no apparent interaction between class size and teacher education, thus no evidence exists that teacher education and class size are either substitutes or complements.

School age was found not be statistically significant or economically significant in any model. This is likely due to the inability of our study to measure classroom characteristics more specifically. A newly renovated school may improve the educational climate, and hence not be captured by these data. There are also other potential interpretations such as the possibility that an older school enjoys more community support, or that the failure of a school levy to pass may suggest a less educationally friendly environment within a community. In any case, these data are silent on this issue.

Average attendance rates are positively correlated with SAT 9 scores, though the relationship is not large, and not subject to clear policy recommendations beyond the obvious, keeping children in school improves performance, even when measured at the aggregate levels. Also, we cannot establish the direction of causation among these variables.

Lower drop-out rates were also strongly correlated with SAT 9 test scores, but as with attendance, the magnitude of the impact is not large. The primary culprit in constraining the usefulness of adjusting these variables is that there is little variation in these variables across schools, and the drop out and rates of absence are not large. This does not mean they cannot be improved, only that the impact of a percentage point change in either variable will have a limited impact on SAT 9 scores. Notably, a one-percentage point improvement in either represents a fairly substantial change in the total.

Population density also affects SAT 9 scores. Population density is a continuous (not dichotomous) representation of rurality. This variable has also been employed as a proxy for district size (Driscoll, et. al., 2003). So, schools in more rural areas are associated with lower SAT 9 scores in 9^{th} and 10^{th} grades (though the latter does not enjoy the commonly reported level of statistical significance). Also, this affect is not large, nor is its cause clear. A number of factors that are correlated with rural areas may lead to this result, though we have attempted to correct for these through our demographic data. It may be also that some unmeasured variable such as duration of school bus rides (to pick a popular topic) generates this result.

These results suggest that there are several factors at issue in determining educational outcomes. Indeed, the appropriate measure we use for outcomes should be evaluated for robustness. At issue is whether several proxies for educational outcome are similarly correlated with the inputs we employ. If they are, we can feel more certain in interpreting the results of our estimation. To this end we estimate the impacts of ACT component testing. See Table 4 below.

These results offer much of the same evidence as does the earlier results. Here the ACT composite results are less well explained than the proportion of total students taking the exam. For the proportion of students taking the test, this model explains more than a third of variation between schools. As with the earlier estimates, teacher education, the percent of adults with college degrees in the zip code in which the school is located, average class size, and percentage of students dropping out the previous year all enjoy strong statistical significance. Poverty, rurality and enrollment are not clearly significant across the board.

Composite ACT scores are correlated with poverty and school size. In this case the impact of enrollment enjoys statistical significance, but as with the impacts of school size on SAT 9 testing, the impact is so small as to be insignificant.

Table 4, Regression Results for ACT Testing

	% ACT Takers	ACT Composite
C	17.45026	23.56969*
TEACHERED	0.007626***	-0.000493
PERFAMINPOVERTY	-11.32774	-2.041008*
PERCOLLGRAD	29.37365**	1.56048
AGEOFSCHOOL	-0.028572	-0.008635
ENROLL	0.002551	0.000752*
AVGCLASSSIZE	-0.948661***	0.223576
AVGATTEND	0.510277	-0.076096
DROPOUTPERCENT	-2.630301***	0.077861
POP/SQUAREMILES	-0.002202	0.001535
R-squared	0.40032	0.274299
Adjusted R-squared	0.351255	0.214378
S.E. of regression	8.967562	1.822494
Sum squared resid	8845.888	362.0419
Log likelihood	-428.2856	-235.0556
Durbin-Watson stat	1.937067	1.861153
Mean dependent var	55.90583	19.71485
S.D. dependent var	11.13365	2.056172
Akaike info criterion	7.30476	4.118581
Schwarz criterion	7.537051	4.35212
F-statistic	8.15901	4.577724
Prob(F-statistic)	0	0.000039

*denotes statistical significance to the .10 level, ** statistical significance to the .05 level and *** statistically significance to the .01, [a]statistically significant to the .15 level employing asymptotic t-statistics. All standard errors are treated with White's [1980] matrix.

Overall, these second set of results supports the findings in the first set of results. There are still additional data to explore, the following table illustrates the results from the combined SAT scores.

Table 5, SAT Results

	Coefficient
C	102.2105
TEACHERED	-0.00204
PERFAMINPOVERTY	-28.53001***
PERCOLLGRAD	63.21019***
AGEOFSCHOOL	0.05767
ENROLL	0.007201***
AVGCLASSSIZE	0.032002
AVGATTEND	-0.99239
DROPOUTPERCENT	-0.65109
POP/SQUAREMILES	0.052419***
R-squared	0.602967
Adjusted R-squared	0.570483
S.E. of regression	8.333429
Sum squared resid	7639.063
Log likelihood	-419.485
Durbin-Watson stat	2.192583
Mean dependent var	12.54889
S.D. dependent var	12.7155
Akaike info criterion	7.158082
Schwarz criterion	7.390373
F-statistic	18.56171
Prob(F-statistic)	0

*denotes statistical significance to the .10 level, ** statistical significance to the .05 level and *** statistically significance to the .01, [a] statistically significant to the .15 level employing asymptotic t-statistics. All standard errors are treated with White's [1980] matrix.

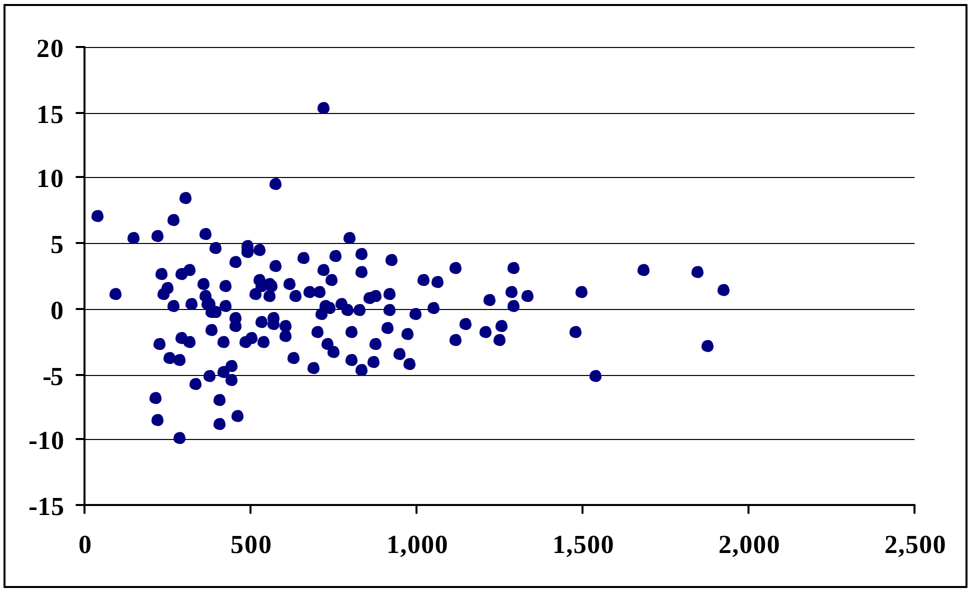
These results also support the earlier findings, albeit with some notable exceptions. The impact of school size on performance is positive and statistically significant, but the magnitude of the impact remains small. Regional poverty, percentage of college graduation and rurality appears to have strong impacts on SAT scores.

The findings with respect to enrollment in foreign languages, math, science, English and social studies all provide similar findings. So, too does the PSAT testing. In each of these cases there is some variation in the size and significance of the impacts. Generally the same variables matter: educational achievement of adults in the region, poverty rates, class size, size of school, rurality and, most importantly for our purposes teacher education. To place the impacts in relative size we should point out that for our best estimates, we can account for only a little more than half the variation in test scores. This is better than most of the other studies reviewed above, but it is clear that there's much more research needed in support of policy.

Importantly, these impacts are not the direct correlations, but are controlled for other variables. These are impacts at the margin, not the average. Notably, the size of these variables should offer some pause for policymakers. For example, increasing the proportion of parents with a college degree by one percentage point is, in some districts, a 20 percent change. Also, it is not clear that this variable, and not some underlying characteristic of this variable influences educational outcome. We discuss this issue more in a later section.

One other feature of these estimation results is that they may be employed to evaluate the unmeasured components of educational performance. The goodness of fit statistics (such as the R^2) tell us what proportion of the variability in the observed data the model explains. For the SAT-9 tests scores we are explaining between 37 percent and 51 percent of the variation in test scores. This is an unusually strong model in a cross sectional analysis. However, this leaves a number of other factors unexplained. These factors may be data errors, misreported or manipulated data or factors we have not measured such as the performance of the school principal in the setting of standards. One common interpretation of the unexplained variation is that it provides an Adjusted Performance Measure (APM). This APM may be interpreted as how well the school is performing after controlling for the variables in the model (over which an individual principal or staff may have little or no control). Thus, the APM explains aggregates the unexplained variables into a single metric which then can be compared across schools. Notably, it is only in comparison to the population of schools that this APM is useful within this context. A high APM is interpreted as suggesting that the school is doing better than would be expected given its measured characteristics. An APM value of zero suggests that the school is performing as expected given its measured characteristics while a negative valued APM suggests the school is performing less well than expected given its measured characteristics. Figure 2 illustrates the relationship between school size and the APM in West Virginia's high schools (using 11th grade SAT-9 scores as the baseline comparison).

Figure 2, Adjusted Performance Measures and School Size



Clearly, there is no correlation between school size and the APM. This is to be expected since school size was employed as a measured variable. Among the high schools in West Virginia, the unmeasured performance of the school influences test scores dramatically. The range in impacts are from just over 15 extra points to roughly ten points lower than predicted by the measured variables on the SAT-9 for 11^{th} grade students. The appendix lists each West Virginia high school, their APM's and rank.

This section has outlined findings from these models (and some we do not illustrate for brevity). But what we have found is not significant in the modeling process is also potentially of some importance. These other findings are reviewed below.

8. Other Findings

There are considerable issues not thus far addressed in this analysis. These may be characterized in three areas: answers we did not find because we did not ask; answers to questions we do not know; and variables we found do not impact education and didn't make it into our analysis. We will address them in reverse order.

We found that none of the three-year changes in inputs explained any significant issue in educational output. The main reason is that three-year changes in these variables are largely non-existent. That is to say, that while there have been changes in the actual values, they are not statistically significantly different over the three-year period at the school level. This means that any measured improvement in the inputs that affected West Virginia school quality during the study period is largely a chimera. The same can be said for aggregate test scores, which do not show statistically significant improvement over the study period. The absence of findings in these areas is perhaps one of the leading results of this study.

Like Hanushek (1995) we found no correlation between teacher experience and educational outcomes. Similarly, teacher salaries were not correlated with improved educational outcomes. This differed from other studies and may partly be explained by a highly centralized compensation system in West Virginia. Simply, pay and outcomes are not designed to be connected in the State funding formula, and they are not.

We also could find no strong relationship between measures of staff experience and anything measuring quantity of administrators to school performance. In our quest to employ every possible variable we even tested the impact of teacher gender on school performance. We posited the possibility that gender may be correlated with some other unobserved variable that influences the school environment (such as prior military service in male principals). While that possibility may exist, there is no statistically significant relationship in these data between principal gender and educational outcomes.

As with other researchers we sought to estimate the potential for interaction effects. Interaction effects are often employed to evaluate whether or not there are combinations of variables that better explain the dependent variable. The key variable we were interested in (which has appeared in several other studies) is whether or not school performance is influenced jointly by size and demographic conditions. What we sought to specifically examine is whether there is a joint relationship between school size and poverty directly on school performance. This method tests for what is sometimes referred to as non-linear relationships between size, poverty and performance. We found no such relationships in any of the specifications we tested.

We sought also to examine whether or not school size and performance were related in a higher order functional form. This is an additional type of test, which evaluates whether or not there is a constant relationship between size and performance. To perform this test we estimated, in the original model, the relationship between size and performance in subsets of the data, separated by school size. In each instance, without regard to our labeling of size we were able to reject a difference in the parameter estimates. This suggests that, as Figure 1 illustrates, the relationship between size and performance remains constantly irrelevant in overall test scores across the size range. This differs from some studies that find that there is a non-linear relationship between size and performance. Or, more clearly, some studies find small schools under performing, with mid-range schools performing better, and with performance declines among larger schools. The most probable explanation for this is that, in West Virginia, even the largest schools are small by national standards.

We do not yet know which schools deliver educational services most efficiently. We know which schools combine inputs most effectively (that is get the greatest impact from the inputs and situations they have been given) but we have not yet assigned a cost to this finding. That is a wholly additional study.

Importantly, though we have not performed a cost study we can conclude some cost related issues from this study. Since the funding formula is designed to equalize funding per student there is little room for savings from most activities, including school consolidation. This does not mean that there are not efficiencies generated from undertaking policies that improve school performance. Reducing educational costs while keeping quality unchanged or improving school performance while keeping costs constant are both efficiency enhancing outcomes.

Also, a good many variables that may be good proxies for regional income or demographics are not significantly correlated in this study. Much to our surprise the final specification of the model appears very similar to the more extensive studies reviewed in the early part of this monograph. This is helpful for two reasons. Firstly, these results provide considerable support for our estimation. Secondly, these specifications provide strong support for a Bayesian approach to modeling education. In the Bayesian approach we would be modeling teacher quality with strong expectations regarding the specification of the model and the parameter estimates. Had we estimated these models in Bayesian setting the prior selection of a positive relationship between teacher quality and outcomes would have been supported. This is strong support for continued research into this area and is also robust support for the findings of this study.

We found modest correlations between the density of private schools and public school performance. This is potentially important because it supports the main argument for private schools (they stimulate regional competition in school performance) and refutes the main argument against private schools (they are cherry picking students). However, these results are very tentative.

Also importantly, we found some puzzling relationships between the way SAT-9 tests are administered and the scores of these tests. We do not understand the results. The proportion of students taking the SAT-9 test under non-standard conditions ranges, at the school level from between 4 and 25 percent. This degree of variation is, as its most charitable characterization, puzzling. There is little to explain this degree of variation beyond program failure. At the county level, high rates of non-standard test taking result in higher individual school test scores. The link between non-standard testing and scores at the school level is more complex. For example both high and very low ls of non-standard test taking are correlated with better educational outcomes (though this result is blurred by small sample sizes).[17] Overall, these statistical relationships should not occur if the program is effectively administered.

Keeping in mind that there should be, in actuality, little difference between schools in the proportion of students eligible for non-standard testing there are a number of explanations for a relationship between test scores and the proportion of students taking the tests in a non-standard setting. In the extreme, out-migration may have generated real differences in the demographic characteristics of a school that would generate high levels of students appropriately provided non-standard testing environments. Schools with fewer resources may not be able to identify students effectively to permit them to participate in the program. Parents may feel compelled to direct students into, or away from the program in different regions, thus providing higher variance in the data. However, the demographic homogeneity in the State combined with a school funding formula designed to mitigate differences in school funding suggests that these reasons are not likely causes of these correlations.

[17] Notably, the levels of non standard test taking could vary dramatically by school due to placement of special needs students in particular schools.

More likely explanations for the high variability in non-standard testing rates and the relationships between rates and test scores are that the data are manipulated to provide either increased financial resources to individual schools (due to programmatic issues) or simply to manipulate scores. This is an important additional avenue of research.

Finally, there are two important issues we did not explore in this study. The first, is the aforementioned estimate of technical efficiency in schools. This study permits us to rank schools based upon their performance after correcting for different variables beyond their control. This is the Adjusted Performance Measure mentioned earlier (but without the problems mentioned by Rubenstein). By this we mean that through the process outlined in this study we can rank schools on how well they are doing given the demographics, poverty, education, rurality and other factors beyond their control. We have not included these results at this juncture since we propose to use these findings in a later double-blind analysis of the top and bottom performing schools.

However, simply ranking schools or establishing the effectiveness of the different variables does not establish whether or not schools are efficient. More complex modeling is needed to perform this analysis. For example, we find that teacher education and larger schools are both correlated with higher test scores among certain groups. This is important, but it does not tell us if this would be an efficient use of resources or what schools are combining resources most efficiently.

9. SUMMARY

The results presented in this work closely mirror those presented in Hicks and Rusalkina [2003]. Within this study, we have focused on the potential effect of consolidation on school performance. We have sought to measure the effects of consolidation through the school and district effect. In so doing, we have performed the first true empirical test of this issue specifically tailored to West Virginia. Our results mirror those of most studies.

Insofar as West Virginia's policymakers are considering policy adjustments in education it is important to understand that, at the State level, there are currently few direct policy tools available to influence performance. Also, the effect of these policies are dwarfed by the impact of variables that cannot be directly controlled – primarily the contribution of families to educational performance. Thus, any policy effort that results in even modest impacts on some attribute of school performance is likely to comprise the bulk of available policy options at the State administrative level. Class size, teacher education, and size of the school impact educational achievement minimally but represent virtually all the tools the State directly enjoys.

First, we find that there are few schools in West Virginia that are large. Indeed, the largest school in West Virginia is only 40 percent as large as biggest high schools in the nation. Second, we find, unambiguously, that school size does not reduce any performance measure, and may actually enhance some areas, though the impact is too small to be the prime guide of policy, though clearly better performance per dollar expended is a useful policy goal. Third, we find a modest negative impact of rurality on school performance. This variable serves as a proxy for district size and bus trip times. Here too we find that though there is a statistically significant impact in some schools, the impact is too small to matter. The sum of these findings suggests that school consolidation has not had a large positive or negative impact on school performance in West Virginia.

10. REFERENCES

Altonji, Joseph G. 1998. "Effects of Personal and School Characteristics on Estimates of the Return to Education," *Economic Perspectives* 22, pp.65-80.

Andrews, Matthew; William Duncombe and John Yinger.2002. "Revisiting Economies of Size in American Education: Are We Any Closer to a Consensus?" *Economics of Education Review* 21, pp. 245-263.

Angrist, Joshua D. and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, pp. 979-1014.

Arabsheibani, Reza G. and Lamine Manfor. 2001. "Non-Linearities in Returns to Education in Libya," *Education Economics* 9, pp.139-144.

Arias, Omar; Kevin F. Hallock and Walter Sosa-Escudero. 2001. "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data," *Empirical Economics* 26, pp.7-40.

Ashenfelter, Orley and Alan Krueger. 1994. " Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review* 84, pp.1157-1174.

Ashenfelter, Orley; Colm Harmon and Hessel Oosterbeek. 1999. "A Review of Estimated of the Schooling/Earnings Relationship, with Tests for Publication Bias," Industrial Relations Section working paper, Princeton: Princeton University Press, pp.425-449.

Black, Sandra E. 1998. "Measuring the Value of Better Schools," *Economic Policy Review* 4, pp. 87-94.

Boissiere, M; J.B. Knight and R.H. Sabot. 1985. "Earnings, Schooling, and Cognitive Skills," *The American Economic Review* 75, pp.1016-1030.

Bradbury, Katharine L. 2002. "Education and Wages in the 1980s and 1990s: Are All Groups Moving Up Together?" *New England Economic Review*, pp.19-46.

Card, David. 1998. "The Casual Effect of Education on Earnings," The Center for Labor Economics working paper 2.

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

Casse, Daniel and Bruno V. Manno. 1998. "The Cost and Price of College and the Value of Higher Education," *Academic Questions* 11, pp. 38-55.

Chakraborty, Kalyan; Basudeb Biswas and W. Cris Lewis. 2001. " Measurement of Technical Efficiency in Public Education: A Stochastic and Nonstochastic Production Function Approach," *Southern Economic Journal* 67, pp. 889-905.

Census: More Degrees Equal Bigger Bucks. CNN.com. 07/18/02. fyi.cnn.com/2002/fyi/teachers.ednews/07/18/degree.dollars.ap/index.html Access Date: 07/18/02.

Chu, Hong-Yih. 2000. "The Impacts of Educational Expansion and Schooling Inequality on Income Distribution," *Quarterly Journal of Business & Economics* 39, pp. 39-50.

Consumer Price Index. Federal Reserve Bank of Minneapolis. http://minneapolisfed.org/research/data/us/calc/hist1913.cfm Access Date: 09/14/02.

Cosca, Theresa. 1998. "Earnings of College Graduates in 1996," *Occupational Outlook Quarterly* 42, pp. 20-29.

Datcher, Linda. 1982. "Effects of Community and Family Background on Achievement," The *Review of Economics and Statistics* 64, pp.32-41.

Dewey, James; Thomas A. Husted and Lawrence W. Kenny. 2000. "The Ineffectiveness of School Inputs: A Product of Misspecification?" *Economics of Education Review* 19, pp.27-45.

Dinopoulos, Elias and Paul Segerstrom. 1999. "A Schumpeterian Model of Protection and Relative Wages," *The American Economic Review* 89, pp. 450-472.

Dothan, Uri and Joseph Williams. 1981. "Education as an Option," *The Journal of Business* 54, pp.117-139.

Driscoll, Donna; Dennis Halcoussis and Shirly Svorny. 2003. "School District Size and Student Performance," *Economics of Education Review* 22, pp. 193-202.

Eide, Eric R; Mark H. Showalter and David P. Sims. 2002. "The Effects of Secondary School Quality on the Distribution of Earnings," *Contemporary Economic Policy* 20, pp. 160-170.

Ericsson, Neil R and John S. Irons (ed). 1994. Testing Exogeneity. Oxford University Press, Oxford. 1994.

Glewwe, Paul. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes," *Journal of Economic Literature* XL, pp. 436-482.

Grosskopf, Shawna; Kathy Hayes, Lori L. Taylor and William L. Weber. 1998. "Allocative Inefficiency and School Competition," *Proceedings of the 91 Annual Conference on Taxation, National Tax Association*, pp.282-290.

Haider, Steven J. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991," *Journal of Labor Economics* 19, pp.799-837.

Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries," *The World Bank Research Observer* 10, pp. 227.

Hanushek, Eric A.; Alan B. Krueger and Jennifer King Rice. 2002. *The Class Size Debate*, Washington, D.C: Economic Policy Institute.

Harmon, Colm and Ian Walker. 1995. "Estimates of the Economic Return to Schooling for the United Kingdom," *American Economic Review* 85, pp. 1278-1287.

Haveman, Robert; Barbara Wolfe and James Spaulding. 1991. "Childhood Events and Circumstances Influencing High School Completion," *Demography* 28, pp. 133-157.

Haveman, Robert and Barbara Wolfe. 1995. "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature* 33, pp. 1829-1878.

Hecker, Daniel E. 1992. "College Graduates: Do We Have Too Many or Too Few? College Earnings and Why They Vary," *Occupational Outlook Quarterly* 36, pp. 13.

Heckman, James. 1998. "What Should Be Our Human Capital Investment Policy?" *Fiscal Studies* 19, pp.103-119.

Hedges, Larry V. and Rob Greenwald. 1996. Have Times Changed? The Relation between School Resources and Student Performance. Brookings Institution Press.

Hicks, Michael J. and Viktoriya Rusalkina and (2003b) Regional Correlates of Educational Performance IN WEST VIRGINIA MIDDLE AND HIGH SCHOOLS: A Study of Demographics, Parental Factors, Teacher Education and School Characteristics. The Senate Education Committee.

Hoxby, Caroline M. and Bridget Terry Long. 1999. "Explaining Rising Income and Wage Inequality Among the College-Educated," Harvard University, NBER working paper.

Imazeki, Jennifer and Andrew Reschovsky. 1998. "Measuring the Costs of Providing an Adequate Public Education in Texas," *Proceedings of the 91Annual Conference on Taxation, National Tax Association*, pp.275-281.

Jaeger, David A. and Marianne E. Page. 1996. "Degree Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics* 78, pp.733-740.

Jacques, Charles and Wade B. Brorsen. 2002. "Relationship Between Types of School District Expenditures and Student Performance," *Applied Economics Letters* 9, pp. 997-1003.

Jolliffe, Dean. 2002. "Whose Education Matters in the Determination of *Household Income? Evidence from a Developing Country," Economic* Research Service, U.S. Department of Agriculture, working paper.

Juhn, Chinhui; Kevin M. Murphy and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill," *The Journal of Political Economy* 101, pp.410-442.

Kane, Thomas J. and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-Year College," *American Economic Review* 85, pp.600-615.

Kremer, Michael R. 1995. "Research on Schooling: What We Know and What We Don't – A Comment on Hanushek," *The World Bank Research Observer* 10, pp. 247.

Lassibille, Gerard and Lucia Navarro Gomez. 1998. "The Evolution of Returns to Education in Spain 1980-1991," *Education Economics* 6, pp. 3-10.

Levin, Jesse. 2001. "For Whom the Reductions Count: A Quintile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement," *Empirical Economic* 26, pp. 221-246.

Mallier, Tony and Timothy Rogers. 1995. "Measuring Value Added in Higher Education: A Proposal," *Education Economics* 3, pp. 119-133.

Marlow, Michael L. 2000. "Spending, School Structure, and Public Education Quality. Evidence from California," *Economics of Education Review* 19, pp. 89-106.

Mariani, Matthew. 1999. "High-Earning Workers Who Don't Have a Bachelor' Degree," *Occupational outlook Quarterly* 43, pp.8-15.

McHaffie, Patrick H. 1998. "Contingency in the Local Provision of Public Education," *Growth and Change* 29, pp. 196-215.

Miller, Paul and Charles Mulvey. 1995. "What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings," *American Economic Review* 85, pp. 586-600.

Moll, Peter G. 1998. "Primary Schooling, Cognitive Skills and Wages in South Africa," *Economica* 65, pp.263-284.

Newell, Andrew and Barry Reilly. 1999. "Rates of Return to Educational Qualifications in the Transitional Economies," *Education Economics* 7, pp. 67-85.

Odden, Allan. 1990. " School Funding Changes in the 1980s," *Educational Policy* 4, pp. 33-48.

Pearl, Judea. 2000. Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge. 2000

Pritchett, Lant. 2001. "Where Has All the Education Gone?" *World Bank Economic Review* 15, pp.367-391.

Pryor, Frederic L. and David Schaffer. 1997. "Wages and the University Educated: A Paradox Resolved," *Monthly Labor Review* 120, pp.3-15.

Psacharopoulos, George. 1994. "Returns to Investment in Education: A Global Update," *World Development* 22, pp. 1325-1344.

Psacharopoulos, George and Eduardo Velez. 1996." Returns to Education During Economic Boom and Recession: Mexico 1984, 1989, and 1992," *Education Economics* 4, pp. 219-231.

Psacharopoulos, George. 1995. "The Profitability of Investment in Education." Working Paper.

Rosenzweig, Mark R. 1995. "Why Are There Returns to Schooling?" *The American Economic Review* 85, pp.153-158.

Rubenstien, Ross, Amy Ellen Schwartz and Leanna Stiefel.2000. "Conceptual and Empirical Issues in the Measurement of School Efficiency" Proceedings of the National Tax Association, v 91. pp. 267-274.

Rupert, Peter and Mark E. Schweitzer. 1996. "Earnings. Education, and Experience," *Economic Review* 32, pp. 2-13.

Rusalkina, Viktoriya and Michael J. Hicks (2003a) *Individual Returns to Educational Acheivement*. The Education Alliance

Selected Interest Rates. Federal Reserve Statistical Release http://www.federalreserve.gov/releases/h15/data.htm. Access Date: 09/12/02.

The Data Page. University of New York. http://pages.stern.nyu.edu/~adamodar/New_Home_Page/data.html Access Date: 09/10/02.

Toh, Mun Heng and Chai Shing Wong. 1999. "Rates of Return to Education in Singapore," *Education Economics* 7, pp.235-253.

Vila, Luis E. 2000. "The Non-Monetary Benefits of Education," *European Journal of Education* 35, pp. 21-33.

Weinberger, Catherine J. 1998. "Race and Gender Wage Gaps in the Market for Recent College Graduates," *Industrial Relations* 37, pp. 67-84.

White, Halbert (1980) "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Wilson, Kathryn. 2001. "The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions," *Southern Economic Journal* 67, pp. 518-551.

Wolter, Stefan C. 2000. "Wage Expectations: A Comparison of Swiss and US Students," *Kyklos* 53, p

Appendix:
Adjusted Performance Measure, School Size and 11th Grade SAT-9 Scores

APM Rank	School Name	APM Value	Enrollment	11th Grade SAT-9
1	GEORGE WASHINGTON HIGH SCHOOL	15.25685025	723	65
2	RITCHIE COUNTY HIGH SCHOOL	9.384288326	577	57
3	BIG CREEK HIGH SCHOOL	8.31028781	308	58
4	PICKENS ELEMENTARY/HIGH SCHOOL	7.062718433	45	66
5	WILLIAMSON HIGH SCHOOL	6.734666474	270	57
6	FAYETTEVILLE HIGH	5.561618275	365	51
7	EAST HARDY HIGH SCHOOL	5.516398138	225	62
8	HUNDRED HIGH SCHOOL	5.321744038	151	58
9	BRIDGEPORT HIGH SCHOOL	5.314228886	799	58
10	IAEGER HIGH SCHOOL	4.72298484	491	60
11	RICHWOOD HIGH SCHOOL	4.537424121	395	62
12	INDEPENDENCE HIGH SCHOOL	4.468301884	532	55
13	TYLER CONSOLIDATED HIGH SCHOOL	4.346456369	496	62
14	TUG VALLEY HIGH SCHOOL	4.32610119	496	59
15	NITRO HIGH SCHOOL	4.110179372	840	66
16	BLUEFIELD HIGH SCHOOL	3.982974895	757	56
17	CLAY COUNTY HIGH SCHOOL	3.75962897	664	54
18	RIPLEY HIGH SCHOOL	3.660895064	929	60
19	CALHOUN COUNTY HIGH SCHOOL	3.502913552	458	47
20	JAMES MONROE HIGH SCHOOL	3.255462838	581	63
21	MARTINSBURG HIGH SCHOOL	3.025164237	1293	63
22	PRINCETON SENIOR HIGH SCHOOL	2.964529109	1120	62
23	MORGANTOWN HIGH SCHOOL	2.870610174	1689	54
24	WINFIELD HIGH SCHOOL	2.811261092	720	59
25	MONTCALM HIGH SCHOOL	2.80679916	318	61
26	HUNTINGTON HIGH SCHOOL	2.78542292	1850	68
27	SOUTH CHARLESTON HIGH SCHOOL	2.764299475	840	61
28	VALLEY HIGH SCHOOL	2.575250419	298	61
29	VAN JUNIOR/SENIOR HIGH SCHOOL	2.57021368	237	60
30	FRANKFORT HIGH SCHOOL	2.170879026	527	61
31	NORTH MARION HIGH SCHOOL	2.152061084	1025	60
32	BRAXTON COUNTY HIGH SCHOOL	2.149066462	744	56
33	HURRICANE HIGH SCHOOL	1.951906644	1064	81
34	MOOREFIELD HIGH SCHOOL	1.867115935	361	64
35	WEBSTER COUNTY HIGH SCHOOL	1.807276913	558	56
36	WAYNE HIGH SCHOOL	1.759192187	623	62
37	PENDLETON COUNTY HIGH SCHOOL	1.716191551	564	61
38	DODDRIDGE COUNTY HIGH SCHOOL	1.660213122	431	57
39	MAGNOLIA HIGH SCHOOL	1.62696448	536	60
40	PADEN CITY HIGH SCHOOL	1.570066948	250	52
41	WHEELING PARK HIGH SCHOOL	1.357865359	1930	55
42	OAK GLEN HIGH SCHOOL	1.265703602	709	53
43	UNIVERSITY HIGH SCHOOL	1.236642765	1290	46
44	JEFFERSON HIGH SCHOOL	1.188939079	1497	52

APM Rank	School Name	APM Value	Enrollment	11th Grade SAT-9
45	MOUNT VIEW HIGH SCHOOL	1.178088559	679	56
46	SHADY SPRING HIGH	1.09471783	519	65
47	LOGAN SENIOR HIGH SCHOOL	1.077786754	923	62
48	GAULEY BRIDGE HIGH	1.042793466	239	62
49	PAW PAW HIGH SCHOOL	1.028911904	97	57
50	BURCH HIGH SCHOOL	0.89135515	367	58
51	WILLIAMSTOWN HIGH SCHOOL	0.886455532	636	53
52	ROANE COUNTY HIGH	0.853520887	878	50
53	POCA HIGH SCHOOL	0.837168322	559	60
54	HEDGESVILLE HIGH SCHOOL	0.819344536	1337	61
55	POINT PLEASANT HIGH SCHOOL	0.752470481	860	55
56	BUCKHANNON UPSHUR HIGH SCHOOL	0.522087199	1224	55
57	GREENBRIER WEST HIGH SCHOOL	0.341477543	372	66
58	MEADOW BRIDGE HIGH	0.327567844	327	61
59	WIRT COUNTY HIGH SCHOOL	0.254428229	379	59
60	SAINT ALBANS HIGH SCHOOL	0.222941473	779	60
61	BROOKE HIGH SCHOOL	0.164930395	1298	62
62	BUFFALO HIGH SCHOOL	0.127403859	272	60
63	SCOTT HIGH SCHOOL	0.102092158	726	56
64	ST MARYS HIGH SCHOOL	0.076395495	428	50
65	LIBERTY HIGH SCHOOL	0.026575393	742	58
66	MUSSELMAN HIGH SCHOOL	-0.06705241	1056	61
67	WEIR HIGH SCHOOL	-0.10750611	793	56
68	OAK HILL HIGH	-0.12456454	832	57
69	HAMPSHIRE SENIOR HIGH SCHOOL	-0.24716563	921	57
70	OCEANA HIGH SCHOOL	-0.27857775	400	66
71	MIDLAND TRAIL HIGH	-0.36505576	386	54
72	GREENBRIER EAST HIGH SCHOOL	-0.41056806	1002	63
73	WYOMING COUNTY EAST HIGH SCHOOL	-0.49241845	719	62
74	GILMER COUNTY HIGH SCHOOL	-0.75472085	572	63
75	SHERMAN HIGH SCHOOL	-0.85546696	460	62
76	WEST HAMLIN HIGH	-1.12891375	538	59
77	SPRING VALLEY HIGH SCHOOL	-1.27386405	1151	64
78	RAVENSWOOD HIGH SCHOOL	-1.31792027	570	62
79	PARKERSBURG SOUTH HIGH SCHOOL	-1.35159894	1257	61
80	GUYAN VALLEY HIGH SCHOOL	-1.35796009	606	69
81	WAHAMA HIGH SCHOOL	-1.38223109	458	63
82	FAIRMONT SENIOR HIGH SCHOOL	-1.59329547	917	59
83	CLAY-BATTELLE HIGH SCHOOL	-1.75668379	383	64
84	PHILIP BARBOUR HIGH SCHOOL COMPLEX	-1.83886473	808	66
85	WOODROW WILSON HIGH SCHOOL	-1.88549315	1210	66
86	SISSONVILLE HIGH SCHOOL	-1.8965328	703	61
87	PARKERSBURG HIGH SCHOOL	-1.8978738	1483	78
88	ROBERT C. BYRD HIGH SCHOOL	-2.01156891	978	58
89	BERKELEY SPRINGS HIGH SCHOOL	-2.2171931	609	60
90	GILBERT HIGH SCHOOL	-2.35277176	293	54
91	DUVAL HIGH SCHOOL	-2.36619268	507	58
92	JOHN MARSHALL HIGH SCHOOL	-2.45160208	1119	60

APM Rank	School Name	APM Value	Enrollment	11th Grade SAT-9
93	CAPITAL HIGH SCHOOL	-2.50527603	1253	57
94	TUCKER COUNTY HIGH SCHOOL	-2.55614635	419	53
95	TYGARTS VALLEY HIGH SCHOOL	-2.57296663	490	59
96	SUMMERS COUNTY HIGH SCHOOL	-2.57590005	540	73
97	HARTS HIGH SCHOOL	-2.58255177	322.0	62
98	PETERSBURG HIGH SCHOOL	-2.78620713	735	65
99	HARMAN ELEMENTARY/HIGH SCHOOL	-2.81785579	228	55
100	HERBERT HOOVER HIGH SCHOOL	-2.83945552	877	58
101	CABELL MIDLAND HIGH SCHOOL	-2.87404752	1880	54
102	PIKEVIEW HIGH SCHOOL	-3.31043822	750	56
103	EAST FAIRMONT HIGH SCHOOL	-3.49081548	954	54
104	HANNAN HIGH SCHOOL	-3.87426384	257	56
105	NICHOLAS COUNTY HIGH SCHOOL	-3.90372232	630	55
106	MATEWAN HIGH SCHOOL	-3.93897441	286	59
107	GRAFTON HIGH SCHOOL	-4.02400369	805	69
108	LEWIS COUNTY HIGH SCHOOL	-4.1606995	876	68
109	ELKINS HIGH SCHOOL	-4.28365938	982	59
110	LIBERTY HIGH SCHOOL	-4.41236767	449	58
111	LINCOLN HIGH SCHOOL	-4.60407577	692	62
112	KEYSER HIGH SCHOOL	-4.73929638	839	69
113	POCAHONTAS COUNTY HIGH SCHOOL	-4.89125479	422	61
114	CAMERON HIGH SCHOOL	-5.17646661	381	63
115	PRESTON HIGH SCHOOL	-5.25599361	1539	62
116	CHAPMANVILLE SR HIGH SCHOOL	-5.58315055	447	60
117	BAILEYSVILLE HIGH SCHOOL	-5.7802745	336	57
118	VALLEY HIGH SCHOOL	-6.96553179	218	62
119	SOUTH HARRISON HIGH SCHOOL	-7.03625919	412	55
120	TOLSIA HIGH SCHOOL	-8.27082894	466	65
121	MOUNT HOPE HIGH SCHOOL	-8.64928276	225	63
122	MAN SENIOR HIGH SCHOOL	-8.84149567	410	64
123	MARSH FORK HIGH	-9.9139051	290	57