

Vogt, William B.; Strauss, Robert P.

Working Paper

It's What You Know, Not How You Learned to Teach It: Evidence from a Study of the Effects of Knowledge and Pedagogy on Student Achievement

Working Paper, No. 168

Provided in Cooperation with:

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

Suggested Citation: Vogt, William B.; Strauss, Robert P. (2001) : It's What You Know, Not How You Learned to Teach It: Evidence from a Study of the Effects of Knowledge and Pedagogy on Student Achievement, Working Paper, No. 168, The University of Chicago, George J. Stigler Center for the Study of the Economy and the State, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/262570>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

It's What You Know, Not How You Learned to Teach It:
Evidence from a Study of the Effects of
Knowledge and Pedagogy on Student Achievement

Robert P. Strauss and William B. Vogt

July 3, 2001

Abstract

While there is growing national concern about inadequate student achievement in the US compared to that of our trading partners, and a general awareness that what a classroom teacher knows and how they teach it must make a difference in student achievement, there is little systematic evidence on the relationship between such teacher characteristics and student achievement. This paper reports the results of estimating a multi-equation model of hiring policies, teacher characteristics, and student achievement at the district level in Pennsylvania.

We find that, before correcting for endogeneity of teacher quality *viz a viz* the hiring decision, the *elasticity* of median ETS's National Teacher Exam General Knowledge test scores on multidimensional measures of student achievement is about .8. However, *after* correcting for endogeneity of this teacher quality measure *viz a viz* the hiring decision, the elasticity is very large in absolute value (from about 8.0 to 12.0) and statistically significant. We also find that the median Professional Knowledge test score is typically negatively related to student achievement. After correcting for endogeneity, the elasticity remains negative, but becomes much larger in absolute value, although it is typically not statistically significant.

Although problems with instrument strength preclude very strong conclusions, the very large positive effect of General Knowledge on student achievement, and very large negative effect of Professional Knowledge on student achievement warrant further investigation by the researcher community. These findings also suggest that those institutions responsible for preparing classroom teachers might be able to do a better job *viz a viz* student achievement if they emphasized General Knowledge rather than Professional Knowledge as they prepare classroom teachers.

Contents

1	Introduction	3
2	Earlier and Related Studies on the Effect of Teacher Quality on Student Achievement	4
3	Model, Empirical Measures, and Estimation Considerations	5
3.1	A Model of Employment and Teacher and Student Achievement	5
3.2	Empirical Measures	7
3.3	Statistical Estimation Considerations	9
4	Estimation Results	11
4.1	Basic Results	11
4.2	Robustness	17
4.3	Selection	18
4.4	Effect Sizes and Policy Experiments	21
5	Conclusions	21
6	Bibliography	23

List of Tables

1	Variable Descriptions	8
2	Descriptive Statistics	8
3	Reduced form equations and (standard errors)	12
4	Structural Equations and (standard errors), by OLS	13
5	Structural Equations: Insularity, General Knowledge and Achievement	14
6	Structural Equations, Professional Knowledge	15
7	Structural Equations, Both Knowledge Measures	16
8	Robustness to Additional Variables	18
9	Selection Effects: Sample Means	19
10	Selection-Free 2 Equation Model	20

1 Introduction

Successful public school reform in the U.S. is increasingly described by educators and commentators on education in terms of increased student learning, both in terms of greater proportions of students testing at grade level, and in terms of generally higher levels of student achievement. Mechanisms to accomplish this include reformed curriculum that emphasizes higher order thinking and problem solving, associated improvements in teacher preparation, and decentralization of management and the further empowerment of building level administrators and teachers to address individual student learning issues. In addition to strengthening the traditional three legs of the stool (curriculum, teachers, and administrators) at the building level, school reform increasingly embraces repeated assessment of students and teachers as diagnostic and management devices, adoption of the view that all students can achieve at high levels¹, the increased use of technology to complement traditional classroom pedagogy, increased parental involvement in the educational process, and the reduction in class size at least in earlier grades to cement basic learning skills.² Some also suggest improvements in school governance³, and greater attention to the teacher selection *process*⁴.

Many educators conclude that such innovations require different and higher levels of public financing to finance these initiatives, and that teacher salaries should be materially be increased. Our court systems continue to opine on whether or not the structure and level of public support for public education is adequate to meet state constitutional obligations for universal access to public education that is "...thorough and efficient."

Finally, for those inclined to implement school reform through market solutions or at least systematic financial incentives to reward those who accomplish increased student learning, notions of competition are embraced to encourage traditional school institutions to improve at a more rapid pace.

While improving teacher quality is now generally embraced as central to successful educational reform, there remains significant disagreement among those who study teacher preparation about how important content knowledge is viz. a viz. pedagogical knowledge in the preparation of teachers. Historically, a prospective teacher would take general coursework in the first two years of an undergraduate curriculum, and then take specialized coursework in a department or school of education the last two years of college. At issue is the balance in coursework during these last two years between content or substantive courses (e.g. more advanced coursework in English literature for a prospective English teacher) in the area the prospective teacher intends to teach, and courses about the educational process (e.g. more advanced courses on the correct method to teach English) or pedagogy. Given the realities of limited class time during the undergraduate experience, and the fact that advanced content courses

¹Levin()

²Although see Hanushek(1999) for cautionary results.

³Strauss(2000)

⁴See Ballou and Podgursky(1997a), and Strauss, Bowes, Marks and Plesko(2000)

are typically offered in departments and schools outside of education, *per se*, this is more than a passing matter because it ultimately affects the net number of enrollment hours and therefore monies, which different parts of a college or university realize. In turn, this can impact on the ultimate number of faculty needed in different parts of the same college or university, and have implications for faculty tenure and growth in the size of departments.

Our purpose in this paper is several fold:

1. to examine the importance of teacher quality as measured by various teacher test scores on student learning outcomes and thus investigate whether or not the very large effects of such test scores found by Strauss and Sawyer(1986) in North Carolina schools are sustained,
2. to distinguish among types of teacher proficiencies (general vs. pedagogical knowledge) while holding constant other factors such as the teacher selection decision, and
3. suggest a methodological innovation in characterizing student learning outcomes by constructing an *index* of learning through the use of factor analysis.

The paper is organized as follows: Section 2 reviews earlier findings on the relationship between teacher quality and student achievement; Section 3 specifies the structural model of the teacher selection and student achievement process and reviews the measures of student and teacher achievement and other empirical data, and discusses statistical estimation issues resulting from incomplete data on some variables in the model; Section 4 presents and interprets the statistical estimation results, and Section 5 concludes.

2 Earlier and Related Studies on the Effect of Teacher Quality on Student Achievement

There is a small academic literature on the effect of teacher quality, as measured by standardized teacher test scores, and substantive preparation on student performance in the US. In an early study, Lins(1946) found with a sample of teachers in 27 classes that the correlation between average National Teacher Exam scores and residual pupil gains scores was .45.

In an examination of the statistical relationship between NTE scores and student competency and student achievement in North Carolina, Strauss and Sawyer(1986) found very strong evidence of a sizable link between core battery NTE test scores and 11th grade reading and math competency and achievement scores.⁵ In that study, a 1% relative increase in the average of core battery scores at the district level was associated with a 3 to 5% relative decline in the fraction of students who fell below grade level in reading and math; this result

⁵See Strauss and Sawyer(1986).

was after controlling for ethnicity, student teacher ratio, college going plans, and per capita income of the school district.

Webster (1988) found a significant relationship between teachers' scores on the Wesman Personnel Classification test, a test of verbal and quantitative ability, and middle school students' scores on the Iowa Tests of Basic Skills.

Loadman and Deville (1990) demonstrated a stronger relationship between ACT scores and NTE, then between GPA and NTE. One interpretation of this empirical relationship is that teacher preparation institutions may not be adding particular value through approved courses of studies.

Ferguson(1991) found a similar relationship, although not as large, between measures of teacher quality and student achievement in Texas, and Ferguson and Ladd(1996) found similar relationships in Alabama.

Monk and King (1995) investigated the effects of subject-specific teacher preparation on student performance in secondary math and science. They find that students whose sophomore-year teacher possessed relatively high levels of subject-matter preparation in mathematics (more than 9 mathematics courses) scored significantly higher than corresponding juniors whose sophomore-year teacher possessed relatively low levels of subject-matter preparation. One more semester of a mathematics course translated to a 1.5 percent improvement in performance, independent of the student's initial pretest score.⁶

Sanders(1994) found that students who scored at the same level on mathematics tests in third grade were separated subsequently by differences of as much as 50 percentage points on sixth grade tests depending on the quality of the teachers to whom they were assigned.

3 Model, Empirical Measures, and Estimation Considerations

3.1 A Model of Employment and Teacher and Student Achievement

Our model of the relationship between teacher characteristics and student achievement follows the typical cross-sectional production function model at the district level. Strauss and Sawyer(1986) specify and estimate a relationship in which student achievement depends on various measures of factor quality and intensity: average National Teacher Exam score at the district level, student teacher ratio, insured capital per student, percent of the student body non-white, and fraction of the student body with post-secondary educational plans. Spending per capita is not included in such a specification because it is viewed as a monotonic transformation of input quantities already specified. If input prices are uniform across districts, then multi-collinearity will result in the estimation process.

⁶Monk and King(1995, p.46)

$$\begin{aligned}\overline{\text{Achievement}} &= \gamma_1 + \gamma_2 \overline{NTE} + \gamma_3 \frac{\text{Capital}}{\text{Students}} + \gamma_4 \frac{\text{Students}}{\text{Teachers}} \\ &+ \gamma_5 \% \text{Nonwhite} + \nu_1\end{aligned}\tag{1}$$

Additionally, we account for teacher selection process and resulting insularity (or nepotism) in the hiring reported in Strauss, Bowes, Marks and Plesko(2000). That 2 equation model specified that the fraction of teachers employed in a district who graduated high school in that district, taken to be a measure of hiring insularity, is a function of the unemployment rate, the level of educational attainment in the district. Student achievement in turn depends on the level of insularity, and measures of student poverty and general educational attainment of population:

$$\text{Insularity} = \gamma_6 + \gamma_7 \text{Urate} + \gamma_8 \% \text{BA} + \nu_2\tag{2}$$

$$\overline{\text{Achievement}} = \gamma_9 + \gamma_{10} \text{Insularity} + \gamma_{11} \% \text{BA} + \gamma_{12} \% \text{AFDC} + \nu_3\tag{3}$$

In this study, we combine the considerations from the modeling of the hiring decision and educational production function literature in a three equation model of the teacher hiring, teacher quality, and student achievement process.

In our first equation, districts choose how insular a hiring process to pursue. We measure insularity by the percentage of a district's teachers who received their high school credential from that district. The idea is that district boards and administrators are, in part at least, pursuing personal and political goals via the hiring process and that their ability to do this is constrained by voter monitoring and by the costs of their actions. So, teaching jobs will be a more valuable commodity in poor districts with high unemployment, so that we should see boards more motivated to distribute these jobs on other-than-merit bases in such districts. Districts in which parents are better able or more willing to monitor board actions should have less scope to be insular. Districts which have a broader pool of applicants will find it more costly in terms of "opportunity quality" to pursue insular policies, so that these districts should be less insular. Finally, rural districts may be more insular simply because they find it more difficult to attract candidates (and the candidates they do attract are more likely to be attached to the district already).

The quality of the employed teachers are reflected by the median General Knowledge and Professional Knowledge test scores in the district of those who took the test, the number of schools of education in the county as a proxy for the general level of supply, the unemployment rate, and the general educational background of the population. We expect that as the level of insularity and unemployment goes up in the district, the general level of teacher quality will decline as non-academic considerations dominate the hiring decision. As a result, and the quality of those hired is expected to decline. We would expect that as the general level of educational attainment rises in the district, that the

general quality of the teachers hired would rise, and that as the number of education schools in the area rises, there will be more and higher quality teachers to choose among by hiring districts.

Finally, we expect that student achievement will rise with the general educational attainment of the population, fall as poverty increases, and rise with various measures of teacher quality e.g. be positively related to higher median General and Professional Knowledge test scores.

The structural model thus is:

$$\text{Inslr} = X_1\beta_1 + \epsilon_1 \quad (4)$$

$$*\text{Kn} = \rho_1 \text{Inslr} + X_2\beta_2 + \epsilon_2 \quad (5)$$

$$\text{Achvm} = \rho_2 (*\text{Kn}) + X_3\beta_3 + \epsilon_3 \quad (6)$$

Below we estimate two sets of (4)-(6), one with GenKn endogenous, and the other with PrfKn endogenous. Then, we estimate a four equation variant of (4)-(6) which includes joint estimation of GenKn *and* PrfKn and includes both in equation 6.

Note that since the quality of teacher hires is endogenous, we must estimate the model using systems estimation techniques. The estimation problem is complicated by the fact that we do not have data on all 501 school districts in Pennsylvania for equations (4)-(6).

3.2 Empirical Measures

Data to estimate the above structural models by school district come from a variety of sources in Pennsylvania.

School District Insularity in hiring process: Inslr. This is measured by percentage of employed teachers in 1996-7 who obtained their high school diploma from that high school as reported to a special survey conducted by the Pennsylvania State Board of Education in July, 1997;

Teacher Quality by school district of employment: GenKn, PrfKn The two teacher quality measures are median teacher test scores on Educational Testing Service's National Teacher Exam core battery tests for General Knowledge and Professional Knowledge. The scores ranged from 250 to 990. Since 1987, the Pennsylvania Department of Education has required anyone aspiring to be certified as a school teacher to pass ETS core and specialized tests.

The NTE exams have been used since 1940 to assess the knowledge of prospective teachers in many states. They were first administered by the American Council on Education, and in 1950 became the responsibility of ETS. The NTE contained common or core examinations in professional education and general education.⁷ The General Knowledge examination measures general background knowledge while the Professional Knowledge measures knowledge about

⁷See Quirk, Witten and Weinberg(1973).

Table 1: Variable Descriptions

Variable	Description
Inslr	Insularity: the log of the % of teachers who graduated from the same district
GenKn	General Knowledge: the log of the mean score of district teachers on general knowledge portion of National Teacher Examination
PrfKn	Professional Knowledge: the log of the mean score of district teachers on professional knowledge portion of National Teacher Examination
Achvm	Achievement: the log of the first principal factor of achievement test scores in the district
Urate	Unemployment: the log of the unemployment rate
BApct	Parents' Schooling: the log of the percent of school district residents who have at least a bachelor's degree
Edsch	Education Schools: the number of schools of education in the same county as the district
Prprl	Rural: the proportion of the district's population who live in rural areas.
Afdc	Parents' Poverty: the log of the percentage of the district's population which received AFDC.

Table 2: Descriptive Statistics

Variable	Mean	Std Dev	25%	75%	N
Inslr	3.34	0.94	2.94	3.91	209
GenKn	6.50	0.01	6.49	6.50	487
PrfKn	6.50	0.01	6.50	6.50	487
Achvm	7.76	0.06	7.73	7.80	487
Urate	1.68	0.47	1.31	2.03	487
BApct	2.51	0.53	2.11	2.82	487
Edsch	2.71	2.91	0.00	4.00	487
Prprl	0.52	0.39	0.09	1.00	487
Afdc	1.49	1.08	0.74	2.25	487

pedagogy and general psychology. The median score was calculated from individual test results for any teacher who took the examinations between 1987 and 1997, and who was ever employed by a public school district in Pennsylvania. The median was calculated across 1987-99 for each district;

Unemployment Rate: Urate The unemployment rate by school district as tabulated and reported by the 1995 edition of the Pennsylvania Educational Policy data base, maintained at the University of Pittsburgh.

Percentage of the Adult Population in 1990 with a bachelor's degree or more as reported in the 1990 Census of Population and tabulated at the school district level: BApct.

Number of schools of education in same county as school district: Edsch. Authors' tabulation by county of 1997 list of approved programs in teacher preparation.

Percent of 1990 Population living in areas of 50,000 or more population in school district: Prpl;

Percentage of school age children in families receiving cash assistance in each school district: Afdc Collected by the Pennsylvania Department of Education in conjunction with administering 1995 student achievement tests;

Composite measure of academic achievement and competency by school district: Achvm This measure is the result of applying factor analysis to a variety of student achievement (1995) and competency test (1990) results by school district in Pennsylvania since 1989. Both the competency and achievement tests are for reading and mathematics for grade school, middle school and high school. Twelve scores per district are summarized in this composite measure. In particular, we utilized achievement test scores in reading and math taken in 19XX at the 5th, 8th, and 11th grade levels (6 measures total). In addition, we utilized competency tests taken in 19XX at the 3rd, 5th, and 8th grade levels also in reading and math (6 measures total). The data we have for the competency tests is the percentage of students who failed. For our student achievement measure, we use the first principal factor of these twelve scores. The factor loadings were quite intuitive. The achievement test scores all received positive loadings and the competency test scores all received negative loadings.⁸

3.3 Statistical Estimation Considerations

Typically, one would estimate (4)-(6) via three-stage least squares. Denoting Y as the stacked left-hand-side variables, X as a block diagonal matrix with X_1, X_2, X_3 on the block diagonals, β as the β_i and ρ_i stacked, Σ is the variance-covariance matrix of the stacked error terms, and W as the matrix of instruments, the 3SLS estimate is:

⁸The factor loadings were: 5th grade reading achievement (0.304), 8th grade reading achievement (0.320), 11th grade reading achievement (0.307), 5th grade math achievement (0.315), 8th grade math achievement (0.310), 11th grade math achievement (0.265), 3rd grade reading (in)competency (-0.274), 5th grade reading (in)competency (-0.291), 8th grade reading (in)competency (-0.277), 3rd grade math (in)competency (-0.259), 5th grade math (in)competency (-0.255), and 8th grade math (in)competency (-0.278).

$$\hat{\beta}_{3SLS} = \left(X' \hat{\Sigma}^{-1/2'} P_W \hat{\Sigma}^{-1/2} X \right)^{-1} X' \hat{\Sigma}^{-1/2} P_W \hat{\Sigma}^{-1/2} Y \quad (7)$$

where $P_W = \hat{\Sigma}^{-1/2'} W \left(W' \hat{\Sigma}^{-1} W \right)^{-1} W' \hat{\Sigma}^{-1/2}$.

In the usual 3SLS setup, there are an equal number of observations per equation. This, along with the assumption that the error terms are homoscedastic and uncorrelated across observations but freely heteroscedastic and correlated within observations leads to the error structure $\Sigma = \Omega \otimes I$. An estimate of Σ is then formed by estimating Ω from residuals in the obvious way, after performing 2SLS equation-by-equation.

We cannot simply use equation 7 in the usual way to perform our estimation because our system of equations is “unbalanced.” There are only 209 observations with usable data on all analysis variables. Most of the problems arise due to missing values for *Inslr*. Many school districts did not report their insularity. Were there not to be missing values for *Inslr*, there would be 487 useful observations. On tack we could take in our estimation would be to drop the observations with missing *Inslr* entirely and use 3SLS in a straightforward manner. We did pursue this strategy for some of our robustness checks in section 4.2, and our baseline estimation of equations 4 through 6 using only the 209 observations does appear there. However, this procedure amounts to throwing away a lot of potentially useful data. We have all the necessary data to estimate equation 6 using 487 observations, and we have 487 observations on every variable save *Inslr* in equation 5.

We pursue a strategy designed to use this information. First, notice that the utility of the estimator in equation 7 depends in no way upon having an equal number of observations per equation. It is only the elegant formula $\Sigma = \Omega \otimes I$ which depends on this assumption. With different numbers of observations per equation, the formula for getting Σ from Ω is merely uglier and the calculation more tedious. So, it would be relatively straightforward to use 209 observations on equations 4 and 5 and 487 observations on equation 6. No “canned” statistical package that we know of can handle the problem, but with a matrix programming language the task is not too difficult. Even this strategy throws away some useful observations, however. We would be using only 209 observations on equation 5 when we almost have enough information to use all 487 — only *Inslr* is missing.

To be able to use all 487 observations on equation 5 we pursue the following strategy. First, observe that we can substitute equation 4 into equation 5 to get:

$$*Kn = \rho_1 (X_1 \beta_1) + X_2 \beta_2 + \epsilon_2 + \rho_1 \epsilon_1 \quad (8)$$

If we have consistent estimates of β_1 in hand, we can construct $X_1 \hat{\beta}_1$ for the observations which have missing *Inslr* and use equation 8 in place of equation 5

in the estimation. This introduces two distinct additional complexities. First, the Σ matrix is rendered even more complex. Now, the covariance between the error in the achievement equation and the error in the *Kn equation is Ω_{23} for observations in which *Inslr* is not missing but $\Omega_{23} + \rho_1\Omega_{13}$ for observations for which *Inslr* is missing (since the error term for equation 8 is different from the error term for equation 5. Of course, the variance term for the *Kn equation is similarly affected. The second issue which arises is that $X_1\hat{\beta}_1$ is an *estimated* quantity: it is not equal to the correct right-hand-side variable $X_1\beta_1$. This means that the standard errors of the estimates must be corrected, as described in Pagan(1984?). In our main results reported later, we use this strategy along with the corrections to Σ and the standard errors of the $\hat{\beta}_1$. To begin the process, we need consistent estimates of β_1 , and these can be had via OLS on equation 4. To summarize our estimation procedure:

1. We estimate equation 4 by OLS, yielding consistent estimates of $\hat{\beta}_1$.
2. We construct $X_1\hat{\beta}_1$.
3. We run 2SLS on equations 5 (using only 209 obs) and 6 to get consistent estimates of β_2 and β_3 .
4. We use the residuals from the three above regressions to estimate Ω .
5. We construct the complicated $\hat{\Sigma}$ taking account of the facts we have 209 observations on the *Inslr* equation, 487 observations on the *Kn equation, 487 observations on the *achvm* equation, and that for 276 observations on the *Kn equation we are using equation 8 and for the other 209 we are using equation 5.
6. We apply equation 7 to get the 3SLS estimator.
7. We calculate the variance of our estimator using Pagan's(1984?) methods.

The above procedure produces consistent estimates of parameters and variance matrix as long as the selection of which observations are present for the variable *Inslr* is random (i.e. independent of the errors in the equations). If the selection is non-random, our estimates will be biased. We explore the possibility of selection below. We find that there is not compelling evidence that there is economically significant selection in our data, so we do not make an effort to address this problem further.

4 Estimation Results

4.1 Basic Results

We estimate the relationship between student achievement and teacher quality in several steps. First, we estimate reduced form equations for the three

Table 3: Reduced form equations and (standard errors)

eq:	(3.1)	(3.2)	(3.3)	(3.4)
	Inslr	GenKn	PrfKn	Achvm
const	3.441 (0.657)	6.49773 (0.00365)	6.49679 (0.00289)	7.640 (0.017)
Urate	0.237 (0.233)	-0.00183 (0.00121)	-0.00143 (0.00096)	-0.003 (0.006)
BApct	-0.168 (0.178)	0.00156 (0.00989)	0.00216 (0.00079)	0.0060 (0.005)
Edsch	-0.738 (0.030)	-0.00019 (0.00015)	0.00004 (0.00012)	-0.001 (0.001)
Prprl	-0.169 (0.229)	0.00144 (0.00115)	0.00022 (0.00091)	0.025 (0.006)
Afdc	0.110 (0.115)	-0.00939 (0.00055)	-0.00085 (0.00043)	-0.022 (0.003)
N	209	487	487	487
R^2	0.154	0.089	0.154	0.652

endogenous variables (Inslr, GenKn and PrfKn), and our index measure of student achievement (Achvm). Second, we estimate our structural equations without taking into account endogeneity through the use of ordinary least squares (OLS). Third, we estimate our system model with general knowledge as the key explanatory variable, and then, separately, professional knowledge as the key explanatory variable. Finally, we estimate our system model with general knowledge and professional knowledge both specified as endogenous and simultaneously estimated. Table 3 - 7 contain this sequence of estimation results.

In Table 3 we find that with regard to explaining Inslr, only the number of schools of education has a statistically significant result, and the elasticity of effect is -.738. Thus as the number of schools of education increases near to a school district, the extent to which districts simply hire their own former graduates diminishes. With regard to the two measures of teacher quality, the median General and Professional Knowledge test scores, we find that only poverty inversely affects both median test scores of employed teachers, although the elasticities are extremely small. The median Professional Knowledge test score is positively affected by the extent to which the district's population holds a bachelor's degree or more; however, the effect is quite small. School districts with higher unemployment rates tend to hire teachers with lower median Professional Knowledge, but, again, the effect is extremely small. Finally, with regard to the effect of exogenous factors on our index measure of student achievement, only the percent rural positively and poverty measure negatively affect student achievement. Again, the effects are quite small, although 65% of the variation in

Table 4: Structural Equations and (standard errors), by OLS

eq:	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)
	GenKn	PrfKn	Achvm	Achvm	Achvm
const	6.498 (0.005)	6.495 (0.004)	3.090 (1.432)	7.286 (1.830)	4.928 (1.929)
Inslr	-2.63×10^{-4} (6.28×10^{-4})	1.98×10^{-4} (4.82×10^{-4})			
GenKn			0.71 (0.22)		0.871 (0.248)
PrfKn				0.063 (0.282)	-0.446 (0.314)
Urate	-2.37×10^{-3} (1.39×10^{-3})	-2.53×10^{-3} (1.07×10^{-3})			
BApct	2.19×10^{-3} (1.41×10^{-3})	2.94×10^{-3} (1.08×10^{-3})	0.043 (0.004)	0.043 (0.004)	0.044 (0.004)
Edsch	-4.83×10^{-4} (2.52×10^{-4})	-1.19×10^{-4} (1.93×10^{-4})			
Afdc			-0.026 (0.002)	-0.027 (0.002)	-0.026 (0.002)
N	209	209	487	487	487
R^2	0.0573	0.1285	0.6374	0.6296	0.6389

Table 5: Structural Equations: Insularity, General Knowledge and Achievement

eq:	5.1	5.2	5.3
	Inslr	GenKn	Achvm
const	3.882 (0.412)	11.417 (0.004)	-57.755 (0.023)
Inslr		-0.012 (0.005)	
GenKn			10.070 (4.579)
Urate	0.081 (0.137)	2.00×10^{-3} (2.22×10^{-3})	
BApct	-0.238 (0.117)	-3.06×10^{-4} (1.53×10^{-4})	0.039 (0.007)
Edsch	-0.071 (0.015)	-9.72×10^{-4} (3.69×10^{-4})	
Prprl	-0.242 (0.188)		
Afdc	0.147 (0.081)		-0.008 (0.009)

achievement is explained, compared to under 16% of the three other endogenous measures.

Estimating our structural model (See Table 4), but without taking into account endogeneity provides a more satisfactory pattern of explanation, although the effects of unemployment and the level of educational attainment are quite small. However, we begin to see that General Knowledge has a positive, statistically significant effect on student achievement; however, Professional Knowledge does not. A one percent increase in the median General Knowledge test score is associated with a .7 percent increase in student achievement. Greater educational attainment in the community has a 59% greater positive effect on student achievement than the depressing effect of poverty (compare coefficients respectively of .043 and -.027) on student achievement.

When General Knowledge and Professional Knowledge are used simultaneously to explain student achievement (see eq 4.5 in Table 4), we find that the effect of General Knowledge *increases* compared to when it is used by itself to explain student achievement; compare General Knowledge's coefficient in eq 4.5 of .871 to its coefficient of .71 in eq 4.3. Moreover, when both measures of teacher quality are entered together to explain student achievement, the measure of pedagogical knowledge, Professional Knowledge, has a *negative* effect on student achievement, although the calculated t-ratio is only -1.42 and is thus not statistically different from zero at a typical confidence level.

Table 6: Structural Equations, Professional Knowledge

eq:	6.1	6.2	6.3
	Inslr	PrfKn	Achvm
const	3.425 (0.475)	6.510 (0.003)	-0.074 (0.020)
Inslr		-3.83×10^{-3} (2.81×10^{-3})	
PrfKn			-1.184 (3.805)
Urate	0.027 (0.166)	-1.25×10^{-3} (1.45×10^{-3})	
BApct	-0.131 (0.129)	1.50×10^{-3} (0.87×10^{-3})	0.046 (0.010)
Edsch	-0.060 (0.020)	-1.56×10^{-4} (2.03×10^{-4})	
Prprl	-0.005 (0.174)		
Afdc	0.231 (0.084)		-0.028 (0.005)

Table 7: Structural Equations, Both Knowledge Measures

eq:	7.1	7.2	7.3	7.4
	Inslr	GenKn	PrfKn	Achvm
const	3.789 (0.392)	6.536 (0.005)	6.511 (0.003)	-39.323 (0.043)
Inslr		-0.011 (0.005)	-3.95×10^{-3} (3.03×10^{-4})	
GenKn				12.661 (4.638)
PrfKn				-5.432 (6.232)
Urate	0.036 (0.132)	1.31×10^{-3} (2.50×10^{-3})	-1.34×10^{-3} (2.50×10^{-3})	
BApct	-0.211 (0.109)	-6.09×10^{-4} (15.02×10^{-4})	1.65×10^{-3} (0.91×10^{-3})	0.055 (0.020)
Edsch	-0.067 (0.016)	-8.65×10^{-4} (3.78×10^{-4})	-2.44×10^{-4} (2.19×10^{-4})	
Prprl	-0.166 (0.164)			
Afdc	0.184 (0.073)			-0.011 (0.011)

Properly estimating the effects of the endogenous variables in our system, radically increases the estimated effect of General Knowledge (See Table 5) and Professional Knowledge (see Table 6) on student achievement. Both estimated elasticities increase by a factor of 10 in the case of General Knowledge and a factor of 2.6 in the case of Professional Knowledge; however, the reliability of the effect of Professional Knowledge on student achievement remains low.

Table 7 contains our preferred results and indicate that the elasticity of student achievement with respect to General Knowledge is 12.66, and highly significant. The elasticity of student achievement with respect to Professional Knowledge is -5.43, although statistically insignificant. Finally, the effect of community educational attainment on student achievement is positive, and statistically significant, although small in absolute effect; the elasticity is only .055.

A word of warning is appropriate in interpreting Table 7, however. Notice that the standard error of the estimates of the key parameters also rise dramatically between Table 4 and Table 7. Our instruments for GenKn and PrfKn evidently are not strong. Indeed, looking at Table 3, we can get a feel for instrument strength. The exogenous variables which influence *Kn either directly or through Inslr but which are not present in equation 6 are Urate, Edsch, and Prprl. None of these three are statistically significant in moving either GenKn or PrfKn. Jointly, the three are significant at the 10% level in the case of general knowledge, but the p-value in the case of professional knowledge is only 47%. Absent a compelling formal theory of the behavior of school boards and especially of the hiring process, it is difficult to know what other variables could profitably be exploited as instruments, however.

4.2 Robustness

Our specification of the achievement equation, equation 6, is fairly narrow. Other authors have used a number of additional variables to explain achievement. If we have omitted relevant variables and these relevant variables are correlated with our instruments, then our results could be biased as a result. In this section, we check for such biases by running our model with a number of additional variables included. In all cases, we use only the observations for which we have complete data.

The results are described in Table 8. Each column of the table contains results from re-running our specification on the 209 observations with complete data and adding the variables named in the column heads. The rows identify the coefficients being reported. The base analysis is 3SLS on equations 4-6, except using only the 209 observations for which we have complete data. This specification yields the same qualitative results as our analysis above, but with larger standard errors: the elasticity of achievement with respect to general knowledge is estimated to be 8.7 and is significant at 5%, the elasticity of achievement with respect to professional knowledge is estimated to be about -26 and is also significant. The next three specifications are estimated by including each of the named variables (spending per student, per capita income in the district, average daily enrollment in the district) in each of the four equations of the model

Table 8: Robustness to Additional Variables

Variable	base	\$/student	income	students	all
Inslr in GenKn	−0.014 (0.010)	0.011 (0.008)	0.009 (0.012)	−0.024 (0.030)	−0.004 (0.012)
Inslr in PrfKn	−0.010 (0.006)	0.009 (0.006)	−0.000 (0.006)	−0.019 (0.021)	0.008 (0.012)
GenKn in Achvm	8.733 (4.385)	8.833 (3.982)	7.196 (4.389)	9.508 (4.822)	5.399 (5.943)
PrfKn in Achvm	−26.151 (9.026)	−21.237 (8.927)	−16.371 (16.758)	−27.951 (9.566)	−11.904 (18.582)

and in the instrument set. The final specification has all of these three measures included in each of the four equations of the model and in the instrument set. We also tried the same systems except including the named variables only in the achievement equation and in the instrument set. The results from these additional specifications were essentially the same.

We read these results as being consistent with our results above. At point estimates: insularity reduces teacher test scores, general knowledge increases achievement, and professional knowledge reduces achievement. Significance is often lost here both because there are fewer observations and because we are including more variables. Furthermore, it is interesting to note that in not a single one of the eight specifications were the additional variables in the achievement equation significant at conventional levels. The lowest p value for testing the null hypothesis that the additional variable could be dropped was 0.35 and more typically, the p value was in the neighborhood of 0.8.

4.3 Selection

Nonrandom selection has long been understood to bias the coefficient estimates in linear models such as ours. Since we are using 209 out of an otherwise usable sample of 487 in equation 4 due to selective reporting of insularity, it is reasonable to be quite concerned about selectivity here. We address the problem in two ways. First, we look for non-random selection directly, by examining the differences in observable characteristics between reporters and non-reporters of insularity.

Table 9 presents a table of means of observables broken out by reporters of insularity and non-reporters. The differences between reporters and non-reporters are, for the most part, small. Non-reporting districts have about a 10% greater proportion of college graduate residents than do non-reporters. Also, non-reporters are considerably more likely to be rural. Interestingly, none of the more performance oriented measures show a significant difference, either in the economic or statistical sense. Although there is some evidence here of non-random selection, it does not look very large.

Table 9: Selection Effects: Sample Means

Variable	reporters	non-reporters	difference	t statistic	P value
Urate	1.720 (0.487)	1.657 (0.461)	0.063 (0.472)	1.45	0.147
BApct	2.449 (0.524)	2.549 (0.534)	-0.101 (0.529)	-2.09	0.037
Prprl	0.590 (0.372)	0.470 (0.397)	0.120 (0.386)	3.41	0.001
Afdc	1.508 (1.050)	1.478 (1.102)	0.030 (1.080)	0.30	0.761
GenKn	6.497 (0.008)	6.498 (0.007)	0.001 (0.008)	-0.77	0.440
PrfKn	6.498 (0.006)	6.499 (0.006)	0.001 (0.006)	-1.32	0.189
Achvm	7.759 (0.058)	7.764 (0.058)	0.006 (0.058)	-0.83	0.405
N	209	278			

We also estimated a logit model in which the left-hand-side variable was a dummy equal to one if the district was a non-reporter. In this regression, the only statistically significant variable (of Urate, BApct, Prprl, and Afdc) was Prprl. It had a probability derivative (at sample means) of 0.002. So, a one percentage point increase in the percent of a district's population which is located in a rural area raises the probability of reporting by 0.2 percentage points. The logistic regression overall was significant at the 2% level. However, the logistic regression did not do well at predicting whether a district reported. The regression predicted correctly 58.9% of the time. Since 57.1% of the districts did not report, one could achieve 57.1% correct predictions by simply predicting non-reportage always.

For our second check of the importance of selection bias, we are able to estimate a smaller model on the full sample even in the absence of observations on insularity. Observe that we can substitute equation 4 into equation 5 to get the two-equation (three-equation when we are including both types of knowledge in equation 6) system:

$$*Kn = X_4\beta_4 + \epsilon_2 + \rho_1\epsilon_1 \quad (9)$$

$$Achvm = \rho_2(*Kn) + X_3\beta_3 + \epsilon_3 \quad (10)$$

The matrix X_4 contains the “union” of elements from X_1 and X_2 . The elements of β_4 are equal to the corresponding elements of $\rho_1\beta_1$ for variables appearing only in equation 4. They are equal to the corresponding elements of

Table 10: Selection-Free 2 Equation Model

Variable	GenKn	PrfKn	Achvm
const	6.4962 (0.0035)	6.4972 (0.0029)	-1.912 (50.482)
GenKn			9.932 (4.120)
PrfKn			-8.460 (8.158)
Urate	-0.0016 (0.0012)	-0.0015 (0.0010)	
BApct	0.0018 (0.0010)	0.0021 (0.0008)	0.059 (0.021)
Edsch	-0.0001 (0.0001)	0.0000 (0.0001)	
Prprl	0.0022 (0.0010)	0.0000 (0.0009)	
Afdc	-0.0009 (0.0005)	-0.0009 (0.0004)	-0.020 (0.011)

β_2 for variables appearing only in equation 5. They are equal to $\rho_1\beta_1 + \beta_2$ for variables appearing in both equations. Obviously, we will no longer be able separately to identify many of the β , and we will not be able separately to identify ρ_1 . However, equations 9 and 10 can be estimated using all 487 observations, obviating selection issues. Furthermore, the most interesting parameters, ρ_2 are still identified and can be reported.

Results from this estimation appear in Table 10. The results in that table agree qualitatively with the results reported above. The elasticity of median score on the general knowledge portion of the NTE has an achievement elasticity of about 10 and is significant. The elasticity of median score on the professional knowledge portion of the NTE has a negative but insignificant coefficient.

We have examined the effect of selection in another way as well. Recall that the selection occurs because *Inslr* is missing. In our reduced form regressions reported earlier in Table 3, we used all 487 observations in the equations for GenKn, PrfKn, and Achvm. These equations were re-run using only the 209 observations for which *Inslr* is available. The results were qualitatively similar; although we have not yet run a formal statistical test to determine whether they are significantly different.

It is our conclusion from the results presented in this section that non-random selection probably does not play a large role in biasing our results.

4.4 Effect Sizes and Policy Experiments

Simply reading off the results in Table 7 or Table 5, we find that the elasticity of achievement with respect to general knowledge is in the neighborhood of ten. On the surface, this would appear to be a very large, even implausibly large, result. Of course, given the standard errors, this elasticity could be in the neighborhood of 3 and still be in the 95% confidence interval. However, the results are no so large when put into the context of the data used to estimate them. The general knowledge measure has a very small percentage variation in the data. Looking at Table 2, the standard deviation of GenKn (logged general knowledge) is only 0.01. Furthermore, the interquartile range is only 0.01. Not reported in the table is the range between the 95th percentile of (not-logged) general knowledge (671) and the 5th percentile (655), at 16. The range between the 95th and 5th percentiles is less than 3%! So, in these data, a change of 1% in the median general knowledge score is a very large change indeed. By contrast the interquartile range in the logged achievement index is 0.07 and the spread between the 95th and 5th percentiles amounts to slightly more than 19%.

So, one policy question we might ask ourselves is: what would happen, *ceteris paribus*, if a school district at the 25th percentile, in terms of general knowledge NTE, were to adjust its hiring practices in order to rise to the 75th percentile. This would amount to changing hiring practices to raise its median general knowledge NTE score upwards by about 1%.⁹ This change, by our estimates, would cause our achievement index to rise by about 10% — by slightly more than its interquartile range. Even this effect is quite large, but if teacher quality is important and if it is best measured by general knowledge NTE, it is not as implausibly large as the shocking elasticity might lead one to believe.

If we were to look at an elasticity value of 3, still in the 95% confidence interval of the estimator, we still get a pretty large effect. Increasing GenKn across its interquartile range (increasing it by 1%) increases achievement by 3% which is substantially less than its interquartile range.

5 Conclusions

Our objective in this paper has been to examine whether or not various measures of teacher quality affect student achievement once the nature of the hiring process has been accounted for. We find that the more highly educated the residents of a school district are, the less likely the district will hire its former graduates as teachers in the classroom. Also, the greater the number of schools of education nearby, the less likely it is that a district will hire its former graduates. On the other hand, the greater the fraction of children in poor families, the more likely it is that a district will hire its former graduates.

With regard to the determinants of the level of professional and general knowledge an employed teacher has, we find that districts with more schools of

⁹It is worth noticing that hiring policy affects flows while the policy goal reflects a stock, so that it would doubtless take some time for a changed hiring policy to affect the overall level of the NTE scores.

education nearby paradoxically hire teachers with *lower* general knowledge.

Perhaps our most interesting finding is that hiring teachers with greater general knowledge has a very large effect on our composite measure of student achievement; the elasticity is 12.66. On the other hand hiring teachers with greater professional knowledge has a negative, but statistically insignificant effect on student achievement.

These findings for Pennsylvania in the mid-1990's are consistent with the earlier findings of Strauss and Sawyer(1986) for North Carolina in 1979, although they are more focused in differentiating between general knowledge vs. professional knowledge. The empirical results do not answer the question of whether or not emphasizing content knowledge, *per se*, as contrasted with educational school coursework in pedagogy, will directly improve student learning outcomes. If the NTE General Knowledge scores are highly correlated with such college entrance screening devices as SAT or ACT scores, then there may be an immediate, policy-operational interpretation of our results.

As has been discussed elsewhere¹⁰, various commercial SAT preparation services only hire SAT high scoring teachers to prepare students for such tests. This market validated practice in conjunction with the empirical results above may suggest that states and districts trying to sort through mounds of teacher applications make a primary distinction at a fairly high cutoff level based on such test scores before moving on to other intangible factors such as personality.¹¹.

¹⁰See Strauss(1999a).

¹¹For example, Stanley Kaplan Educational Services will not hire anyone below the 90'th percentile in the SAT distribution – about 1400 combined math and verbal – to teacher in their SAT preparation courses

6 Bibliography

- Dale Ballou and Michael Podgursky, "The Case Against Teacher Certification," *The Public Interest*, No. 132, Summer 1998, pp. 17-29.
- Ballou, Dale and Michael Podgursky (1997a). *Teacher Pay and Teacher Quality*. s(Kalamazoo, Michigan: Upjohn Institute for Employment Research, 1997)
- Ballou, Dale and Michael Podgursky (1997b), "Reforming Teacher Training and Recruitment: A Critical Appraisal of the Recommendations of the National Commission on Teaching and America's future," mimeo, Department of Economics, University of Missouri, September, 24, 1997.
- Bishop, John H. "Signaling, Incentives, and School Organization in France, the Netherlands, Britain, and the United States," in Eric A. Hanushek and Dale W. Jorgenson, (editors) *Improving Americas Schools:The Role of Incentives* (Washington, D.C.: National Academy Press, 1996).
- Brooks. *The Incompetent Teacher*
- Darling-Hammond, L. and E. M. Sclan (1996), "Who Teaches and Why," in J. Siula (editor), *Handbook of Research on Teacher Education*. New York: Macmillan.
- Ehrenberg, Ronald G. and Dominic J. Brewer(1994), "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review* 13 (1) 1-17.
- Elsbree, Willard S. (1939). *The American teacher; evolution of a profession in a democracy* (New York, Cincinnati: American Book Company)
- Ferguson, Ronald (1991). "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal of Legislation*, 28, (2), Summer, 1991, 465-98.
- Ferguson, Ronald and Helen F. Ladd (1996). "How and Why Money Matters: An Analysis of Alabama Schools," in Helen F. Ladd (editor) *Holding Schools Accountable*. (Washington, D.C.: The Brookings Institution), pp. 265-299.
- Hanushek, Eric A., and Richard Pace (1995), "Understanding Entry into the Teaching Profession, " in Ronald Ehrenberg (editor) *Choices and Consequences: Contemporary Policy Issues in Education*, (Cornell University: ILR Press).
- Hanushek, Eric A., "The Evidence on Class Size," in Susan E. Mayer and Paul Peterson (ed.), *Earning and Learning: How Schools Matter*. (Washington, D.C.: Brookings Institution, 1999), pp. 131-168.

- The Holmes Group(1986). *Tomorrow's Teachers: A Report of the Holmes Group*. (Lansing, Michigan: November, 1986).
- Kolb, Ruth and Robert P. Strauss (1999), "A Survey of State Laws Governing School Board Ethics," Paper Presented at the American Education Finance Association Annual Research Conference, Seattle, Washington, March 19, 1999.
- Lins, E. (1946), "The Prediction of Teaching Efficiency," *Journal of Experimental Education*," 16, (15), 2-60.
- Loadman, William E. and Craig W. Deville(1990), "The NTE as a Measure of General Academic Ability: Implications for Teacher Education," presented at Mid-Western Education Research Association, (Chicago, Illinois, October, 1990).
- Manski, Charles E. (1987), "Academic Ability, Earnings, and the Decision to become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972," in David A. Wise (editor), *Public Sector Payrolls* (Chicago: University of Chicago Press), 291-237.
- Monk, David H. and Jennifer A. King (1995), "Multilevel Teacher Resource Effects on Pupil Performance in Secondary Mathematics and Science: The Case of Teacher Subject-Matter Preparation," in Ronald Ehrenberg (editor) *Choices and Consequences: Contemporary Policy Issues in Education*. (Cornell University: ILR Press).
- Monk, David H.(1992), "Education Productivity Research: An Update and Assessment of its Role in Education Finance Reform," *Educational Evaluation and Policy Analysis* 14, (4): 307-332.
- Monk, David H. and Jennifer King Rice (1997), "Modern Education Productivity Research: Emerging Implications for the Financing of Education," in *Selected Papers in School Finance* pp. 5-28.
- Murnane, R.J. and R. Olsen (1987), "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching," *Journal of Human Resources*, 25, 1 (Winter).
- Murnane, R.J. (1991). *Who Will Teach? Policies that Matter*. (Cambridge, Mass.: Harvard University Press).
- Quirk, Thomas J., Barbara J. Witten, and Susan F. Weinberg (1973), "Review of Studies of the Concurrent and Predictive Validity of the National Teacher Examinations," *Review of Educational Research*, 43, (1), 89-113.
- Sanders, William L. (1994), "The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment," *Journal of Personnel Evaluation in Education* 8: 299-311.

- Strauss, Robert P. and Elizabeth Sawyer(1986), "Some New Evidence on Teacher and Student Competencies," *Economics of Education Review* 5, 1, 41-48.
- Strauss, Robert P. (1993). *Who Should Teach in Pennsylvania's Public Schools?*. Carnegie-Mellon University, Center for Public Financial Management, pp. 154.
- Strauss, Robert P., Lori Bowes, Mindy Marks, and Mark Plesko (1998). *Teacher Preparation and Selection in Pennsylvania*. (Harrisburg, Pennsylvania: Pennsylvania State Board of Education, June, 1998), pp. 233. (ERIC: Clearinghouse Number TM029186).
- Strauss, Robert P. (1999a) "A Reverse Engineering Approach to Improving Teacher Quality: The Hiring Decision and State Laws Governing School Board Conduct and Ethics." *Testimony* before the Committee on Education and the Workforce Subcommittee on Postsecondary Education, Training, and Life-Long Learning U.S. House of Representatives May 13, 1999
- Strauss, Robert P. (1999b) "Who Gets Hired to Teach? The Case of Pennsylvania." in Marci Kanstoroom and Chester E. Finn, Jr. (editors) *Better Teachers, Better Schools*. (Washington, D.C.: Fordham Foundation Press, July, 1999), pp. 175-201.
- Strauss, Robert P., Lori Bowes, Mindy Marks, and Mark Plesko (2000), "Improving Teacher Preparation and Selection: Lessons from the Pennsylvania Experience," with Lori Bowes, Mindy Marks, and Mark Plesko, *Economics of Education Review*, 19,4 (July, 2000), 387-415.
- Weaver, W. T. (1983). *America's Teacher Quality Problem: Alternatives for Reform*. (New York: Praeger Publishing Co.).
- Webster, W.J (1988), "Selecting Effective Teacher," *Journal of Educational Research*.