The Potential for System-Friendly K–12 Reform

John Merrifield

Numerous empirical models connect individual student test scores or average test scores to theoretically plausible policy and socioeconomic variables. Although the models were created to test for the effect of a specific factor like funding levels or teacher training on student performance, the fully specified models also have important implications for the ability of the current K–12 school system to significantly improve its performance. This article examines various student performance models and the potential for system-friendly K–12 reform.

Substituting potential, best-case values of independent variables into fully specified models of average student performance allows one to discover how much improvement might be achieved within the constraints imposed by factors that cannot be represented by explanatory variables. The fixed factors are the governance and funding processes of the school systems that generated the numbers that are the basis for the parameter estimates. The specific aim of this article is to reveal what existing student performance models say about the potential for improvement, and thus what they say about the significance of the fixed factors.

This is a suggestive assessment of the qualifying studies. Since the major school reform policy issues are aggregate performance measures like average test scores, I largely excluded models with individual student performance as the dependent variable. I mention the

Cato Journal, Vol. 29, No. 2 (Spring/Summer 2009).

John Merrifield is Professor of Economics at the University of Texas, San Antonio.

two that assess the significance of determinants of the National Assessment of Education Progress (NAEP) scores that are the most trusted consistent measures of student skill. Since my concern is the absolute level of test scores, I also excluded "value-added" models—that is, models that explain changes in performance. Incomplete documentation of results or unavailability of one or more variables means the forced exclusion of some otherwise qualifying studies.[1]

The next section uses hypothetical highly favorable (definition depends on available data) values for the independent variables to solve the regression equations for the predicted average level of each qualifying study's student performance measure. The result is a best-case outcome estimate of the performance variable. I also make a second performance variable estimate with control variables at their median or mean values, and only the policy variables at their highly favorable levels. The third section assesses the predicted best-case values of the student performance variables and discusses interpretation issues. For example, suppose the predicted maximum NAEP score is 300. The total possible score on a NAEP exam is 500, so a score of 300 is only 60 percent. This score is in the "advanced" range for fourth grade math and reading scores, but only in the "proficient" range for eighth grade scores.[2] Indeed, the hypothesis that motivated this article is that large gaps between the predicted "best case" and the total possible score will be the norm. The fourth section discusses some of the costs of changing real-world average values to the highly favorable values inserted into the equation. The final section provides some concluding remarks.

Predicted Maximum Student Performance

Blair and Staley (1995) used data for 266 Ohio urban school districts. Their dependent variable (SCORE) is a three-year (fourth, sixth, and eighth grades), three-test (reading, math, and language arts) district average composite test score ranging from zero to 100. Multicollinearity problems prompted them to omit some theoret-

[1]The following studies were excluded: Hanushek (1996) because of the lack of student data; Dee (2005) because of the lack of data on academic achievement and the omission of α; and Betts, Danenberg, and Rueben (2000) because of the lack of data on schools and the fact that the dependent variable is relative to a threshold.
[2]For details, see www.nagb.org/pubs/readingbook.pdf.

ically plausible variables. The model that included the remaining explanatory variables is

$$(1)\ \text{SCORE} = 20.98 + 0.34\text{SAL} - 0.13\text{PTRATIO} + 0.15\text{AGI} - 3.73\text{MINOR} - 17.28\text{ADC} + 0.43\text{NEIGHBOR} - 0.80\text{COUNT}.$$

Where (highly favorable value = average +/– 2 standard deviations, in parenthesis):

SAL = Average teacher salaries in thousands (38.53)
PTRATIO = Pupil-teacher ratio (14.4)
AGI = Family adjusted gross income, thousands (57.35)
MINOR = % Minority (0 > av. – 2 s.d.)
ADC = % getting "Aid to Families with Dependent Children" (0 > av. – 2 s.d.)
NEIGHBOR = Average SCORE of contiguous school districts (66.7)
COUNT = Number of adjacent school districts (9.66 = av. + 2 s.d.).

To avoid the bias that might result from excluding insignificant variables, the predicted academic performance calculations exclude only variables with an implausibly signed coefficient. So, in the Blair and Staley (1995) model, only COUNT did not factor into the estimate of the predicted best-case value of SCORE. Substituting the highly favorable values into all of the independent variables with plausibly signed coefficients yields a predicted, best-case SCORE = 69.5 percent. With the control variables AGI, MINOR, and ADC set at their average values, and only the policy variables set at highly favorable values, the predicted best-case SCORE = 58.7 percent.

Gamrat (2002) examined 1999 Pennsylvania System School Assessment (PSSA) scores for the fifth, eighth, and eleventh grades. His model is

$$(2)\ \text{PSSA} = 1{,}259.86007 + 0.00735\text{OEPPS} - 2.37297\text{LI} - 0.00012\text{ENR} + 0.00086\text{ATS} + 1.45744\text{SPCT} - 20.87379\text{DLMA} + 1.24332\text{JU}.$$

Where (highly favorable = actual max or min, in parenthesis):

OEPPS = Operating expenditures per pupil spending ($13,170)
LI = Percentage of low-income students (3.9)

ENR = District enrollment (1,715)
ATS = Average teacher salary ($64,338)
SPCT = Students per classroom teacher (12)
DLMA = Whether or not the district is located in a metro area (0)
JU = District gives its teachers the option of joining a union (1).

Solving for PSSA with the designated highly favorable values, the predicted average PSSA score is 1,403.8 out of a total possible score of 2,200 to 2,500;[3] roughly in the middle of the range of scores Pennsylvania now defines as proficient. With only the policy variables set at their highly favorable values, the predicted PSSA is 1,328.6, which is in the lower end of the range Pennsylvania currently defines as proficient.

Mensah, Schoderbek, and Werner (2005) used 2000–2001 New Jersey school district data to specify models for per pupil expenditure and test scores at various levels. Since the most important test score is the one for the oldest children, I report only the TEST3_AV model—that is, district average scores for combined language and mathematics for the eleventh grade. The model is

$$\begin{aligned}(3)\ \text{TEST3\_AV} = {} & 5.223 + 0.01\text{YR01} - 0.022\text{POOR} + 0.002\text{STU\_POP} \\ & + 0.071\text{EXPPP} + 0.195\text{CS\_INST} \\ & - 0.171\text{CS\_ADMIN} - 0.302\text{CS\_OPMAIN} \\ & + 1.124\text{CS\_EXTCUR} - 0.054\text{ABBOT\_DIST} \\ & - 0.022\text{LOWINC} - 0.150\text{COSTIDX} \\ & + 1.325\text{ATTD\_RATE}.\end{aligned}$$

Where (highly favorable values = observed max or min, in parenthesis):

YR01 = Dummy variable that equals one for year 2001
POOR = Dummy variable that equals one if the proportion of STU_POP in the school district receiving federal meal aid is over 25 percent (0)
STU_POP = Number of students enrolled in the school district expressed in natural log terms (41,378 in 2000)

[3]The total possible score varies. The cut scores that define the ranges were not available for 1999. The scores 1,403 and 1,329 are in the range currently defined as proficient.

EXPPP = Natural log of expenditures per-pupil, defined as total operating expenditures divided by average daily enrollment ($13,981 in 2000)
CS_INST = Cost share for instruction, computed as total instructional divided by total operating expenditures (0.73)
ABBOT_DIST = Dummy variable = 1 for Abbott school districts
CS_ADMIN = Cost share for administration computed as total administrative expenditures divided by total operating expenditures (0.06)
CS_OPMAIN = Cost share for operating and maintenance, equals total
operating and maintenance spending divided by total
operating expenditures (0.07)
COSTIDX = Geographic cost of living index for the school district, expressed in natural log terms (0.90)
CS_EXTCUR = Cost share for extracurricular activities, computed as total extracurricular expenditures divided by total operating expenditures (0.06)
LOWINC = Proxy for the income level of the district, measured by the pct of STU_POP in the school district receiving meal aid under the federal school lunch program (0)
ATTD_RATE = Average class attendance rate of STU_POP in the school district, in natural log terms (0.97).

With highly favorable values for the independent variables with correctly signed coefficients, the predicted TEST3_AV is 437 out of 450. Independent variable outliers can produce unrealistic results, so where the median plus or minus two standard deviations is less favorable, I also use them to calculate TEST3_AV. That adjustment lowers the predicted TEST3_AV to 403. With the maximums and minimums for highly favorable values for policy variables and median values for the control variables POOR, STU_POP, LOWINC, COSTIDX, and ATTD_RATE, the predicted TEST3_AV score for New Jersey is 426 out of 450. With the control variables set at their median values, and only the policy variables set at their highly favorable values (median + or – 2 standard deviations), the predicted TEST3_AV score for New Jersey is 368 out of 450 (81.8 percent).

Nelson, Rosen, and Powell (1996) specified equations for statewide average 1995 SAT (Math + Verbal) scores and fourth grade reading scores from the 1994 National Assessment of Educational Progress (NAEPRead4). Depending upon the sign of the coefficient, the highly favorable value is twice the mean or half the mean. The SAT equation is

$$\text{(4) SAT} = 1038 - 5.9\text{TT} + 0.0424\text{TT}^2 + 0.0023\text{PCI} + 0.588\text{PRIV} - 0.0005\text{PUP} - 0.23\text{MINOR} - 0.278\text{U} + 0.357\text{BARG} + 0.528\text{CONFER} - 9.61\text{SOUTH}.$$

Where (highly favorable value in parenthesis):

TT = % that take the test (17.6)
PCI = per capita income ($41,916)
PRIV = private HS graduate percentage (19)
PUP = $ per pupil expenditure ($10,618)
MINOR = % minority (10.3)
U = % Urban (34)
BARG = Collective bargaining percentage (100)
CONFER = Teacher – District meet and confer agreement (100 – BARG)
SOUTH = 1 for southern region, 0 otherwise (0).

The NAEPRead4 equation is:

$$\text{(5) NAEPRead4} = 195.4 - 0.04\text{U} - 0.466\text{PRIV} - 0.001\text{PUP} + 0.026\text{LEP} - 0.271\text{ABS} + 0.001\text{PCI} - 0.136\text{MINOR} + 0.0933\text{BARG} + 0.125\text{CONFER} + 0.202\text{CS}{<}25.$$

Where (highly favorable value in parenthesis):

LEP = % of pupils with limited English proficiency (0)
PRIV = Private HS graduate percentage (5.1)
ABS = Absence rate (0)
CS<25 = % of class sizes under 25 (100).

In equations (4) and (5), PUP's coefficient has the wrong sign. In equation (5), LEP has the wrong sign. They are not part of the computation of the predicted test scores. Because PRIV's opposite coefficients in (4) and (5) are plausible, PRIV was part of the best-case computation. Highly favorable values for the independent variables with correctly signed coefficients yield predicted SAT and NAEPRead4 statewide average scores of 1,079 and 260 (proficient), respectively. With the control variables set at their mean values, and only the policy variables set at their highly favorable values, the predicted SAT and NAEPRead4 statewide average scores are 949 and 232 (basic), respectively.

Sander and Krautman (1991) used 1986–87 American College Test scores for Illinois counties to assess the importance of spending and some schooling variables. Their model is

$$(6)\ \text{ACT Score} = 19.7 - 0.0002\text{SE} + 0.00004\text{SS} + 0.00002\text{TSA} + 0.004\text{TSC} - 0.07\text{TE} + 0.02\text{PTR} - 0.024\text{PT} + 0.001\text{URB} - 0.00034\text{DEN} + 0.000084\text{FI} + 0.07\text{CP} - 0.15\text{FHF}.$$

Where (highly favorable = average +/– 2 standard deviations, in parenthesis):

SE = Schooling expenditures/pupil ($4,261)
TSA = Average annual teacher's salary ($30,721)
TSC = Teacher's schooling – % with master's degree (59.3%)
TE = Teacher's experience (17.4)
PTR = Pupil-teacher ratio (11.4)
SS = School size (1,581)
FI = Median family income ($26,010)
FHF = Female-headed family (4.0%)
PT = Percent taking (40.6%)
DEN = Density (0 > av. – 2 s.d.)
URB = Urban (94.2%)
CP = % of the population 25 years or older with a college degree (20.4%).

In equation (6), SE's coefficient has the wrong sign. So, after excluding SE, and solving for ACT with the highly favorable values, the predicted, best-case average ACT score is 21.8 out of a total possible of 36. With the control variables set at their average values, and only highly favorable values for the policy variables, the predicted best-case average ACT is 19.2.

Walberg and Fowler (1987) used 1983–84 New Jersey data for district average scores on a variety of reading, writing, and math tests to assess the significance of socioeconomic status (SES), per-child expenditure (SPEND), and school district size (SIZE = enrollment). Among their empirical results, I report only the ninth grade reading model; the test scores for the most important subject for the oldest children in their data set. Their model is

$$(7)\ \text{READ9} = 82.9 + 2.47\text{SES} - 0.00064\text{SPEND} - 0.00031\text{SIZE}.$$

Walberg and Fowler (1987) tested three measures of per student expenditure. But since none had the predicted sign, SPEND's coefficient did not affect the predicted value of READ9. Using different measures of SPEND did not markedly impact the coefficients of SES or SIZE. Defining highly favorable as the mean +/– two standard deviations, the highly favorable values of SES and SIZE are 3.92 and 0 (> av. – 2 s.d.), respectively. Zero is not a plausible value for district enrollment, so half the mean (1,002.27) was the highly favorable value. The predicted READ9 with the highly favorable values of SES and SIZE is 92.3 percent. With SES at its average value, and only SIZE at its highly favorable value of 1,002.27, READ9 = 82.6 percent.

Assessment

Table 1 summarizes the findings of the previous section. Its overriding message is that even highly favorable circumstances produce mostly disappointing NAEP, ACT, and SAT results. Even with highly advantaged students, generous school funding, and highly qualified teachers, the average level of academic performance is nothing to get excited about. The Mensah, Schoderbek, and Werner (2005) and Walberg and Fowler (1987) results (both from New Jersey data) are high predicted outcomes. But some high state test scores are to be expected. Noted education historian Diane Ravitch (2006: 58) notes, "School officials and editorial writers know by now not to trust local or state claims of progress until the NAEP results for the states are released every other year, thus verifying or rejecting state claims." Harvard's Paul Peterson (2006: 12) notes, "The best available yardstick, NAEP, is widely considered to be the nation's report card." The 1983–84 and 2000–01 New Jersey exams may be too easy.

The lower predicted state test scores of Blair and Staley (1995) and Gamrat (2002) are more consistent with the NAEP, SAT, and ACT results reported by Nelson, Rosen, and Powell (1996) and

TABLE 1
PREDICTED TEST SCORES
(PERCENTAGE OF TOTAL POSSIBLE)

Study (Lead Author)	All Variables			Only Policy Variables		
	Most Favorable	Highly Favorable	Official Characterization[a]	Most Favorable	Highly Favorable	Official Characterization
Blair	69.5%	N.A.	N.A.	58.7%	N.A.	N.A.
Gamrat[b]	≤63.8%	N.A.	Proficient	≤60.4%	N.A.	Proficient
Mensah	97.1%	89.6%	N.A	94.7%	81.8%	N.A.
NelsonSAT	67.4%	N.A.	N.A.	59.3%	N.A.	N.A.
NAEP	52.0%	N.A.	Proficient	46.4%	N.A.	Basic
Sander	60.8%	N.A.	N.A.	53.3%	N.A.	N.A.

[a]If any. The official possibilities are below basic, basic, proficient, and advanced.
[b]"≤" because the calculation is based on the 2,200 low end of the range of the total possible.

Sander and Krautmann (1991). According to the Blair and Staley (1995) model, D/C– average outcomes would result from highly favorable values for all explanatory variables, and "flunk" is the average outcome for average socioeconomic conditions with highly favorable policy values. Nelson's NAEPRead4 model indicates that average socioeconomic conditions combined with highly favorable policy values only yields an average outcome of "basic" competency (a score of 46.4 percent).

NAEP results are seen as more reliable, but the NAEP exams are still criticized for being too easy. For example, Loveless (2004: 12) states:

> The problem solving items on NAEP are not very challenging—at least not in the arithmetic required to answer them. Content taught in first and second grades is at least two years below grade level for fourth graders. But that is the level of difficulty of more than four out of ten (43.6 percent) problem solving items on NAEP. The second surprising finding is that even though the NAEP items are so easy, fourth graders do not do very well on them. . . . The eighth grade items are only slightly more difficult than those on the fourth grade test (3.4 mean grade level). Almost four out of ten items (39.6 percent) address arithmetic skills taught in first and second grade—six years below the grade level of eighth graders taking the test. Indeed, more than three-fourths of the items (33/43) are at least four years below grade level, taught in the fourth grade or lower. Yet the percentage of eighth graders answering problem-solving items correctly is an unimpressive 41.4 percent. Problem solving items on the eighth grade NAEP only require knowledge of very simple arithmetic. Despite this, eighth graders have trouble getting them right.

Still, the predicted NAEP scores under highly favorable conditions are not that high, even for the lower grades where U.S. students are closer to international norms. Even SAT scores have become suspect for recent changes in what they aim to measure, and for becoming easier in 1994, and harder again in 2005 (Kahn 2006). Note that Nelson, Rosen, and Powell's SAT model yielded a disappointing prediction for the highly favorable scenario even though Nelson used data generated by the easier post-1993 SAT.

Two additional NAEP models bolster the conclusion that with present fixed factors held constant there is a low upper limit on potential academic gains. Darling-Hammond (1999) specified the relationship between various state-level average NAEP scores, teacher characteristics, and control variables. Of the six equations reported by Darling-Hammond, the equations for the most recent math and reading performance are

$$(8)\ \text{NAEPMath8} = \alpha_m + 0.79\text{QUAL} + 0.157\text{MA} - 0.034\text{NEWUNCERT} - 0.032\text{CLASSZ} - 0.353\text{POV} + 0.391\text{LEP}$$

$$(9)\ \text{NAEPRead4} = \alpha_r + 0.636\text{QUAL} + 0.103\text{MA} - 0.199\text{NEWUNCERT} - 0.091\text{CLASSZ} - 0.166\text{POV} - 0.058\text{LEP}.$$

Where (highly favorable value in parenthesis):

NAEPMath8 = 1996 eighth grade math scores
NAEPRead4 = 1994 fourth grade reading scores
QUAL = % teachers certified, and with a major in their teaching field (100)
MA = % teachers with a masters degree (100)
NEWUNCERT = % new teachers that are uncertified (0)
CLASSZ = Average class size (10)
POV = % of pupils from families with incomes below poverty (0)
LEP = % of pupils with limited English proficiency (0).

In equation (8), LEP's coefficient has the wrong sign. Substituting the highly favorable present values into all of the independent variables with correctly signed coefficients yields predicted NAEPMath8 and NAEPRead4 statewide average scores of [α_m + 94.4] and [α_r + 73.0], respectively. With the control variables POV and LEP set at their average values, and only the remaining policy variables set at their highly favorable values, the predicted NAEPMath8 and NAEPRead4 statewide average scores are [α_m + 89.5] and [α_r + 70.4], respectively. Using Nelson, Rosen, and Powell's (1996) NAEPRead4 α_r = 195.4 as a proxy for Darling-Hammond's α_r, the additional 70.4 points achieved under highly favorable conditions attains an average total score of only 265.4 (53.1 percent), which is barely in the advanced range according to the NAEP cut scores. The small size of the coefficients indicates that it will take very large changes in the independent variables to yield a noteworthy change in the test scores.

QUAL = 100 accounts for the lion's share of those additional 70.4 points. Though mandated by the Federal No Child Left Behind (NCLB) law, QUAL = 100 will be a huge achievement. In 2002, "fewer than 36 percent [of teachers] feel very well–prepared" (U.S. Department of Education 2008: 1). The story is similar for math. Using Johnson's (2000) α_m of 253.2 as a proxy for Darling-Hammond's, $\alpha_m + 90 = 340.2$ (68.1 percent), which is barely in the advanced range of the official cut scores.

Johnson (2000) studied NAEP data for fourth and eighth grade math (1996 data) and reading (1998 data) to assess the importance of teacher training. Johnson's results were not included earlier because his observations were of individual students rather than district or state average performance. I computed the predicted NAEP values for highly favorable conditions only for the eighth grade results. The reading and math equations are

$$
\begin{aligned}
(10)\ \text{NAEPREAD} = {} & 243.4966 - 21.3058\text{BC} - 12.7026\text{HP} \\
& - 1.1072\text{ONWC} + 14.0562\text{PC} \\
& + 6.7694\text{ARMH} - 6.0616\text{PPLP} - 12.7874\text{GM} \\
& - 1.0460\text{BDE} + 2.1112\text{BDEL} + 6.8530\text{ADEL} \\
& + 2.8666\text{BDOS} + 4.5590\text{ADOS}.
\end{aligned}
$$

$$
\begin{aligned}
(11)\ \text{NAEPMATH} = {} & 253.2074 - 33.1248\text{BC} - 17.7678\text{HP} \\
& - 2.84800\text{NWC} + 14.3722\text{PC} + 4.9266\text{ARMH} \\
& - 9.6302\text{PPLP} - 0.2478\text{GM} + 1.6916\text{BDE} \\
& + 5.9142\text{BDEL} + 9.3040\text{ADEL} + 4.8690\text{BDOS} \\
& + 3.9020\text{ADOS}.
\end{aligned}
$$

Where (highly favorable value in parenthesis):

BC = Black communities (0)
HP = Hispanic communities (0)
ONWC = Other non-white communities (0)
GM = Gender = Male (0)
PC = Parents attended at least some college (1)
ARMH = Students with additional reading material in home (1)
PPLP = Students participate in the free/reduced-price lunch program (0)

BDE = Teacher bachelor's degree in education (1)
BDEL= Teacher bachelor degree in education or language arts (1)
ADEL = Teacher advanced degree in English or language arts (1)
BDOS = Teacher bachelor degree in other subjects (1)
ADOS = Teacher advanced degree in other subjects (1).

Solving for the NAEP reading score with the highly favorable values, the predicted best-case average NAEP score is 281 (56.2 percent); the minimum score of NAEP's proficient level for eighth grade reading. Solving for the NAEP math score with the highly favorable values, the predicted average NAEP score is 298, just short of the minimum score of NAEP's proficient level for eighth grade math. Again, highly favorable conditions yield only mediocre performance. Because all of Johnson's explanatory variables were binary, it didn't make sense to produce the second set of predicted values with only favorable policy-sensitive variables, and with average values for the socioeconomic variables.

Attaining Highly Favorable Values

I already noted that the NCLB-mandated QUAL = 100 will be difficult to achieve. We are far below 100 percent now, and obstacles like teacher tenure and single salary schedules will slow progress toward that goal. There are other problems as well.

The test score scenario derived from the Blair and Staley (1995) model assumed a 25.1 percent rise in the average teacher salary and a 20.8 percent drop in class size. By themselves, they would be quite costly, but combined (more teachers at a much higher salary) they would be prohibitively costly. Likewise, the predicted best-case average test score derived from Gamrat's (2002) model assumed a 81.5 percent rise in per pupil expenditures and a 28.1 percent drop in class size, which also would be prohibitively costly strategies. The Mensah, Schoderbek, and Werner (2005) model's predicted best-case average test score assumed a 65.1 percent increase in average per pupil expenditures, and the Nelson, Rosen, and Powell (1996) model assumed a doubling in per pupil expenditures. The Sander and Krautman (1991) model's predicted best-case average test score assumed a 30 percent increase in average per pupil expenditure. The Walberg and Fowler (1987) model says that spending levels don't matter.

Conclusion

Highly favorable, probably mostly unattainable, independent variable values mostly yield test scores indicating mediocre average academic performance. Though even the most trusted indicator tests, including NAEP, are widely seen as easy, many of the predicted best-case average scores would not enjoy designations like proficient or advanced. And except for the New Jersey models, the absolute level of the scores is quite low. In the models and in the real world, attainable values of policy variables produced inadequate, continued "Nation at Risk" performance levels. This suggestive mining of information from studies produced for other reasons indicates that studies need to be commenced that specifically aim at determining the upside potential of schooling, holding constant the current fixed factors—namely, the governance and funding processes substantially shared by all U.S. states and school districts.

Many people will see these results as a confirmation of the longstanding "schools don't matter" finding.[4] But the right conclusion to draw from econometric findings in which policy variables have small or insignificant coefficients is that those variables don't matter much given the fixed background factors that all of the observations have in common. Indeed, the coefficient of a policy variable indicates how much it affects performance now, not necessarily how much it could affect performance with different underlying fixed factors.

Econometric analysis explains variability in the dependent variable around its mean, but it does not explain the mean level of the dependent variable. Without variability in key factors like common governance and funding practices, empirical measurement of those factors' importance to academic performance is not possible. The small regression coefficients of the policy variables say that those school characteristics don't matter much now, but they could matter in an environment lacking in some of the current system's widely shared key characteristics. Repeated urgent calls for reform indicate that the United States needs to find a way to make school characteristics matter.[5] In other words, fixed factors are needed that will sub-

[4]There is a1966 federal report that many scholars consider the most famous study of education, *Equality of Educational Opportunity*, otherwise known as the "Coleman Report" for the sociologist James Coleman. It was widely interpreted as finding that schools have no significant effect on student learning.
[5]See U.S. Commission on National Security (2001) and U.S. Department of Education (1983).

stantially elevate the average level of test scores around which non-school factors may still explain most of the variability, and perhaps fewer fixed factors so that variable factors become more significant. Research is needed to identify the appropriate fixed factors, and how to vary other factors within and among schools to best serve a highly diverse student population.

References

Betts, J. R.; Danenberg, A.; and Rueben, K. S. (2000) *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California.* San Francisco: Public Policy Institute of California. Available at www.ppic.org/content/pubs/report/R_200JBR.pdf.

Blair, J. P., and Staley, S. (1995) "Quality Competition and Public Schools: Further Evidence." *Economics of Education Review* 14 (2): 193–98.

Darling-Hammond, L. (1999) *Teacher Quality and Student Achievement: A Review of State Policy Evidence.* Seattle: University of Washington Center for the Study of Teaching and Policy.

Dee, T. S. (2005) "Expense Preference and Student Achievement in School Districts," *Eastern Economic Journal* 31 (1): 23–44.

Gamrat, F. (2002) "Explaining Variability in School Performance: The Case of Pennsylvania." Allegheny Institute for Pubic Policy, Report No. 02–04.

Hanushek, E. (1996) "School Resources and Student Performance." In G. Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,* 43–73. Washington: Brookings Institution.

Johnson, K. (2000) *The Effects of Advanced Teacher Training in Education on Student Achievement.* Washington: Heritage Foundation. Available at www.heritage.org/Research/Education/CDA00-09.cfm.

Kahn, D. S. (2006) "How Low Can We Go?" *Wall Street Journal* (26 May): W11.

Loveless, T. (2004) *How Well Are American Students Learning?* Washington: Brookings Institute.

Mensah, Y. M.; Schoderbek, M. P.; and Werner, R. H. (2005) "Public School Spending, Functional Cost Classifications, and Student

Performance: A Simultaneous Equation Approach." New Brunswick, N. J.: Rutgers University Department of Accounting and Information Systems, Rutgers Business School.

Nelson, F. H.; Rosen, M.; and Powell, B. (1996) *Are Teachers' Unions Hurting American Education?* Milwaukee: Institute for Wisconsin's Future.

Peterson, P. E. (2006) "The Use of Market Incentives in Education." In P. E. Peterson (ed.) *Choice and Competition in American Education,* 3–12. Lanham, Md.: Rowman and Littlefield.

Ravitch, D. (2006) "National Standards." *Education Week* (5 January): 58.

Sander, W., and Krautmann, A. C. (1991) "Local Taxes, Schooling, and Jobs in Illinois." *Economics of Education Review* 10 (2): 111–21.

U.S. Commission on National Security for the 21st Century (2001) *Road Map for National Security: Imperative for Change—Phase III Report.* Washington: Government Printing Office.

U.S. Department of Education, National Commission on Excellence in Education (1983) *A Nation at Risk: The Imperative for Educational Reform.* Washington: U.S. Department of Education.

U.S. Department of Education (2008) *Student Achievement and School Accountability Conference, October 2002.* Available at www.ed.gov/admins/tchrqual/learn/hqt/edlite-slide006.html.

Walberg, H. J., and Fowler, W. J. (1987) "Expenditure and Size Efficiencies of Public School Districts." *Educational Researcher* 16 (7): 5–13.