DOCUMENT REVIEWED:	**Report Card on American Education**
AUTHOR:	**Andrew T. LeFevre**
PUBLISHER/THINK TANK:	**American Legislative Exchange Council**
DOCUMENT RELEASE DATE:	**November, 2006**
REVIEW DATE:	**January 8, 2007**
REVIEWER:	**Gene V Glass**
E-MAIL ADDRESS:	**glass@asu.edu**
PHONE NUMBER:	**(480) 965-2692**

Summary of Review

The "Report Card on American Education," published by the American Legislative Exchange Council,[1] uses poor and misleading methods to draw some very controversial findings. The report presents readily available statistics to generate hundreds of tables and figures concerning each state's education "inputs," "outputs," and demographics. Interspersed among these tables are a mere dozen pages of analysis intended to support the conclusion, in the words of ALEC Executive Director Lori Roman, that per-pupil spending increases, pupil-to-teacher ratio reductions and raises for teachers "… are *not* going to make the difference in raising American student achievement to international standards. Empowering parents will" (p. 1). But ineptness and naiveté in measurement and data analysis have thwarted any attempt to legitimately derive such conclusions.

Review

I. INTRODUCTION

The *Report Card on American Education* attempts to touch all the bases in contemporary education policy: education finance, teacher preparation and compensation, tuition tax credits, charter schools, and vouchers. Little of importance escapes author Andrew T. LeFevre in this wide-ranging assessment of the nation's K-12 public education system. If the quality of the recommendations matched the report's ambitions, then policy makers might be wise to embark on the complete revolution in public education that would result. Budgets would be slashed; public monies for educating children would go directly from the government to children's parents; private profit-making companies would provide the bulk of the nation's teaching; and the training, licensing, and pay schedules for teachers would be revamped from top to bottom.

II. REPORT'S FINDINGS AND CONCLUSIONS

The findings are reported in the scant 11-and-a-half pages of text that are contained in this 143-page document; the other 132 pages list literally tens of thousands of bits of undigested data, mostly organized by state, all of which could be downloaded from the internet.

The *Report Card on American Education* makes the following five assertions:

- In spite of increases in per-pupil expenditures greatly exceeding (by 77%) the rate of inflation since 1983, 71% of U.S. eighth graders "are still performing below proficiency" (p. 3) in mathematics according to the National Assessment of Education Progress (NAEP); the report's author sees a "growing consensus that simply increasing spending on education is not enough to improve student performance" (p. 3);
- There is no correlation—and presumably then no causal link—between pupil-to-teacher ratios (commonly discussed in terms of class size) and educational achievement (p. 3);
- There is no correlation between teachers' salaries and educational achievement (p. 3);
- "Strong accountability measures" (p. 3) will help focus resources where they are most needed; and
- Parental choice—as evidenced in the charter school system—will benefit a child's educational future (p. 3).

III. REPORT'S RATIONALES FOR ITS FINDINGS AND CONCLUSIONS

LeFevre presents a great deal of data, but the vast majority of these data are not analyzed. He bases his findings and conclusions loosely on the more than 50 tables and figures (or many times 50 depending on how one counts tables within tables) containing tens of thousands of pieces of raw data. More than 100 measures of educational "inputs" and "outputs" are arrayed in dozens and dozens of tables. Fifty pages are devoted to profiles of individual states—one page per state—where each state is described in terms of "outputs" (SAT, ACT, and NAEP averages), "inputs" (per-pupil spending, pupil-teacher ratio, average teacher salary), and student demographics (white, black, Hispanic, etc.).

The report's analysis leans heavily on an examination of the relationship of inputs and outputs on a composite measure of the author's own devising. To create a measure of educational achievement comparable across the 50 states and the District of Columbia, LeFevre formed an arithmetic composite based on NAEP (8th Grade Math), SAT, and ACT test score averages for each state. A state's ranking on NAEP (1 highest, 51 lowest) was divided by 51 to produce a scaled score ranging from .02 to 1.00. For 26 states reporting SAT scores, the average scores were similarly scaled (the state that ranks #10, for instance, would receive a scaled score of 10/26 = .38). A similar calculation was made for the 25 states reporting ACT averages. The three constituents were summed and ranked to determine a final achievement ranking for each state. Massachusetts ranked highest; the District of Columbia ranked lowest.

Having arrayed these data points across all 51 states, the author proceeds to examine the "vital question" (p. 102) of the relationship between inputs and outputs by placing them

> side-by-side on four different tables. Looking at these tables gives an idea of possible correlations between educational inputs and outputs. For example, if a state spends a relatively large amount of money per pupil and has a relatively high average SAT score, then it may be the case that spending large amounts of money leads to higher SAT scores. (p. 102)

These data displays cry out for formal, precise statistical analyses of the co-relationships between the scores and the expenditures, rather than just an eye-balling of 51 separate pairs of numbers. Such statis-

tical analyses would also allow researchers who are familiar with these relationships to compare these findings with existing research findings. The report's author, recipient of a B.A. in political science from Temple University, with apparently no formal training in statistics, is using methodology that is a century out of date. It is as if Karl Pearson (1857-1936) had never lived to invent the correlation coefficient.

As discussed below, the author also reports "two standard regression tests" in an appendix to "account for the possibility that several educational inputs are important to student achievement" (pp. 102-3). These models, too, have serious flaws.

IV. REVIEW OF THE REPORT'S USE OF RESEARCH LITERATURE

The *Report Card on American Education* fails to take advantage of the voluminous research literature on precisely the topics it regards as most important. In fact, it ignores, intentionally or unintentionally, the many studies that flatly contradict its findings and conclusions. Its bibliography lists only the sources of the myriad tabulations of raw data; no research studies are cited. Particularly for a report with such sweeping, far-reaching recommendations, this oversight is indefensible.

Relationship between Spending and Student Performance. Research on the relationship between education expenditures and achievement is decades old. Although truly experimental research is lacking, sophisticated statistical analytic methods have superseded the type of simple correlation studies presented in this report. Moreover, aggregation of study findings by meta-analysis has moved the debate off of simplistic questions such as "Are expenditures related to student achievement?" Those researchers

without an immoveable agenda have formed a consensus around the work of Greenwald, Hedges, and Laine,[2] who concluded that "…a broad range of resources were positively related to student outcomes, with effect sizes large enough to suggest that moderate increases in spending may be associated with significant increases in achievement" (p. 361).

As discussed below, Greenwald and his colleagues also stressed the importance of limiting analyses of these relationships to the school-district level, and that aggregating data at greater levels can lead to inaccurate conclusions. The *Report Card on American Education*, by using state-level analyses, runs afoul of this advice.

Class Size. LeFevre examines the class size question by reporting pupil-to-teacher ratios in apparent innocence of both nearly a century of experimental and quasi-experimental research on class size and achievement[3] and the exemplary and widely heralded Tennessee STAR experiment that conclusively demonstrated the benefit of reducing class size.[4]

Teacher Quality and Salary. The report ventures into the domain of teacher quality when it claims that teacher salaries are unrelated to educational "outputs," and *ipso facto* that such markers of higher salaries like certification and experience have no benefits in terms of achievement. Were this the case, it would be some comfort to charter school operators who typically hire uncertified and inexperienced personnel and pay them at lower rates than traditional public school teachers. But the report's claim lacks support and is inconsistent with other research.[5]

Any policy analyst who writes for a lay audience appreciates the need to hold in check the scholarly enthusiasm for citations to the research literature. To ignore widely accepted findings from peer-reviewed literature, however, marks a work as political polemic rather than a policy analysis.

V.   REVIEW OF THE REPORT'S METHODS

Measurement Methods. LeFevre's devising of a measure of educational "output" represents the only derived measure in the report. Essentially, the author calculated an arithmetic average of each state's percentile rank on average NAEP 8th Grade Math, SAT, and ACT scores. The resulting measure of achievement bears only a very weak relationship to the results of school teaching and learning. It essentially gives equal weight to the NAEP, which is a legitimate achievement measure, and the SAT and ACT, which are aptitude measures specifically designed so as not to be greatly influenced by schooling experience.[6] Varying participation rates make state-level SAT and ACT averages virtually useless even as measures of scholastic aptitude—and certainly as measures of achievement levels.[7] Test validity aside, the transformation of state averages into percentile ranks induces curvilinearity into any possible relationships among variables, rendering them inappropriate for correlation and regression analysis.

The other, non-derived measures are merely data downloaded from various government websites. Per-pupil expenditure data are taken from the U.S. Department of Education, National Center for Education Statistics, and Common Core of Data Surveys. As such, they reflect all expenditures in a state, including administrative and support personnel, and are poor proxies for the resources spent on classroom instruction. More careful research shows that only a small portion of increased spending has

gone to regular education—to the sorts of programs that are likely to show up in test scores. For example, Rothstein and Miles (1995) studied expenditures in nine typical U.S. school districts and found that "the share of expenditures going to regular education dropped from 80% to 59% between 1967 and 1991, while the share going to special education climbed from 4% to 17% …. Per pupil expenditures for regular education grew by only 28% during this quarter century—an average annual rate of about 1%" (p. 1).[8] In addition to special education, the new money has been focused on such items dropout prevention, transportation, health insurance, school lunch programs, and security.

Since one of the *Report Card*'s major contentions is that expenditures have risen historically while achievement "outputs" have not—a contention like others in the document that is unsupported by the document's own data and proven false by other sources of information—it would have been advisable for the author to at least attempt to determine the portion of expenditures spent on teaching.

Analysis Methods. Granted, a lay audience of legislators might have some difficulty with even middle-level statistical analyses, but the *Report Card on American Education* eschews even the simplest displays and calculations that would support or fail to support its points. Indeed, the predominant method of analysis might be called "juxtaposition," where numbers coming from variables purportedly related are listed side-by-side. Are expenditures and achievement correlated? Well, look at the numbers side-by-side, the report invites the reader. Of course, correlations often can not be seen even by experienced researchers scanning columns of side-by-side numbers. So Le-Fevre extracts a couple of examples: "Of the

ten states that increased their per pupil expenditures the most over the past two decades, … only New Hampshire (3rd) and Vermont (5th) ranked in the top ten in academic achievement" (p. 4). Such examples are offered to demonstrate a missing correlation, but in truth these facts are not inconsistent with a positive relationship between per-pupil expenditures and achievement for all 51 data points.

VI.   REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS

Relationship between Spending and Student Performance. The *Report Card* states that in spite of increases in per-pupil expenditures greatly exceeding (by 77%) the rate of inflation since 1983, 71% of U.S. eighth graders "are still performing below proficiency" (p. 3) in mathematics according to the National Assessment of Education Progress (NAEP). The report thus concludes that increasing costs of education are somehow associated with poor performance, or that increases over the past two decades should have produced a greater percentage of "proficient" eighth graders (and students at all grades). This correlation rests on a single data point that is itself an impossibility, since one has no knowledge of the math proficiency rate in 1983 and thus no knowledge of the level of improvement. Moreover, the statement relies heavily on the validity of the NAEP performance levels, pursuant to which various percentages of students are labeled "proficient," "advanced," or "basic." Unfortunately for this report's conclusions, the validity of the NAEP performance levels has been authoritatively condemned, both by scholars and by the federal General Accounting Office.[9] (NAEP has resisted to this day any fundamental changes in its flawed methods of establish-

ing performance levels.) It is therefore impossible to attach any significance at all to the juxtaposition of the two facts.

The Report Card's author assumes that increasing per-pupil expenditures in inflation-corrected dollars should produce greater academic achievement. As noted earlier, no attempt was made by this author to track whether those increasing dollars actually are spent on regular instruction of students. In fact, past studies that have gone to the trouble of tracing these dollars have reached very different conclusions. With the federal and state governments imposing increasingly onerous unfunded burdens over the last two decades—including recent requirements concerning student tracking and reporting— most increased expenditures appear to never reach the classroom, certainly not in ways that one should expect to directly increase a school's average test scores. This does not mean that expenses for dropout prevention, special education, or health insurance are unnecessary or not useful—only that a simple comparison of average spending to average test scores is not well-designed to detect such usefulness.

The Report Card's most important conclusion concerns dollar "inputs" and achievement "outputs": "The first conclusion of these [regression analyses] is that differences in educational inputs … (students per school, schools per district, student to teacher ratios, per-pupil expenditures, teacher salaries, and funds received from the federal government) taken together do not explain differences in student achievement" (p. 103). This conclusion is based on two regression analyses. The first is a regression analysis in which LeFevre's measure of educational achievement (a conglomerate of NAEP and aptitude scores) is predicted from per-pupil expenditures, among other things. The second analysis regressed changes in

SAT state averages between two dates—1983/84 and 2003/04—onto changes in per-pupil expenditures between those same dates, plus other variables. (As a side note, it appears that a log transformation was applied to some of these variables before analysis, but no rationale is given. Readers are expected to trust, but not verify, the author's modeling and conclusions.) Of all the possible analyses that could have been performed, only these two have been reported. No rationale is given for selecting only these to report. Yet one can hardly accuse the author of "cherry-picking" favorable results since his results bear no apparent relationship to any conclusion.

The second regression analysis is particularly egregious. Not only are SAT averages scarcely reflective at all of educational attainment, they are seriously confounded by self-selection. Most of the variability in state SAT averages is due to the percentages of students electing to take the SAT exam instead of the ACT test or no test at all. Further, as a larger portion of the U.S. population considers attending college, the tests are taken by an increasingly non-elite slice of the high school population. What might be considered to be a success by public school educators (pushing more students to consider college) is transformed by this analysis into something that looks like a failure (lower average test scores). This analysis is largely meaningless, and even its meaningless results seemed to bear no direct link to any of LeFevre's conclusions.

Not reported among the results—which led LeFevre to conclude that these inputs "do not explain differences in student achievement"—is the fact that his "per pupil expenditures" do in fact correlate at +.41 with NAEP 8th Grade Math state averages.[10] Correlations of this magnitude generally

constitute substantial evidence of a relationship between inputs and outputs.

Teacher Quality and Salary. The *Report Card* concludes that there is no correlation between teachers' salaries and educational achievement (p. 3). Yet the data presented in the report itself could have been used to show a +.20 correlation between average teacher salary (by state) with NAEP 8th Grade Math average score.[11]

State-Level NAEP Data. But set aside for a moment the fact that the conclusions of the report are inconsistent with its own data. Even the use of such data is ill-advised. Ironically, the *Report Card* appeared within days of the call for a moratorium on the use of state-level aggregate NAEP data by the editor of a leading, peer-reviewed scholarly journal.[12] Editor Sherman Dorn noted that NAEP data have been publicly available for some time at the level of individuals. Continuing to analyze NAEP data aggregated to the level of states is to continue to commit what is known in research methodology as the "ecological correlation" fallacy.[13]

For instance, imagine that Wyoming experiences an increase of \$X per pupil while achievement averages a decline of Y points. From this small amount of information, it can not be concluded that those students who scores declined were the recipients of the increased funding. It is entirely possible that those particular schools receiving the increased funding showed gains in achievement while their influence on the state average was offset by decreases in the other schools for different reasons. Negative correlations of aggregate data points are not inconsistent with positive causal relationships at the level of the constituent data. Greenwald, Hedges and Laine made this point in their 1996 review of research on school resources and achievement.[14]

Parental Choice. LeFevre concludes that parental choice—such as that seen in the charter school system—will benefit a child's educational future (p. 3). In Chapter Four of the *Report Card*, data on charter school enrollments are tabulated, documenting a rapid increase in numbers of schools and students. No attempt is made to relate charter school attendance to achievement or even to cite collateral research that might support a claim of superiority for charter schools. In language that would be appropriate if uttered from a politician's soapbox but not in legitimate reports of research, author LeFevre concludes:

> As more and more parents see that they can—and should—have a choice in their child's education, it causes more and more leaks in the dam that has been holding back real educational reform. And soon, the educational establishment will run out of fingers to plug those leaks and then the flood of educational reform and school choice will finally be free to flow all across this great nation—bringing liberation to many that have struggled far too long to escape from an educational system that has failed them all too often. (p. 131)

Choice and a glorious new day for American education become linked again by the mere fact of being juxtaposed in the same paragraph.

VII. REPORT'S USEFULNESS FOR GUIDANCE OF POLICY AND PRACTICE

In spite of being clad with myriad numbers and statistics, the *Report Card on American Education* is rhetoric, not research. Legisla-

tors may find value in looking up education statistics for their own state and comparing them with other states. But they will find neither credible findings nor any firmly established facts on which to base policy decisions.

NOTES & REFERENCES

[1] LeFevre, A. T. (2006, November). *Report Card on American Education: A State-by-state Analysis, 1982-1983 to 2004-2005*. Washington, D.C.: American Legislative Exchange Council

[2] Greenwald, R., Hedges, L.V. & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-96.

[3] Among others see Glass, G.V., Cahen, L.S., Smith, M.L. & Filby, N.N. (1982). *School Class Size: Research and Policy*. Beverly Hills, CA: SAGE Publications; and Finn, J. D. (2002). Class size reduction in grades K-3. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 27-48). Greenwich, CT: Information Age Publishing, Inc.

[4] Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children 5(2)*, (1995): 113-127.

[5] Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Education Policy Analysis Archives, 8(1). Retrieved January 2, 2007 from http://epaa.asu.edu/epaa/v8n1/. Darling-Hammond writes: "Quantitative analyses indicate that measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status."

[6] The SAT acronym originally stood for Scholastic Aptitude Test, but aptitude having fallen out of favor in the last several decades, the College Board gives no explanation of the acronym at all. Since 1995, the College Board refers to the old aptitude section as a "reasoning test."

[7] See Wainer, H. (1986). Five pitfalls encountered while trying to compare states on their SAT scores. *Journal of Educational Measurement, 23*(1), 69-81. Fetler, M. E. (1991). Pitfalls of using SAT results to compare schools. *American Educational Research Journal*, *28*(2), 481-491.

[8] Richard Rothstein & Karen Hawley Miles. (1995). *Where's the Money Gone?* Washington, DC: Economic Policy Institute.

[9] A General Accounting Office review of the NAEP proficiency levels was prompted by a report of a small group of scholars who labeled the National Assessment Governing Board (NAGB)—the body that sets the proficiency levels -- as "incompetent" and the levels themselves as "ridiculous." The GAO review concluded: "NAGB's… approach was inherently flawed, both conceptually and procedurally, …the approach [should] not be used further until a thorough review could be completed… These weaknesses are not trivial; reliance on NAGB's results could have serious consequences." (p. 38) (GAO (U.S. General Accounting Office). 1993. *Educational Achievement Standards. NAGB's Approach Yields Misleading Interpretations.* GAO/PEMD 93-12. Washington, D.C.: General Accounting Office.

[10] From this reviewer's own calculations extracting data from the report's Table 1.7, p. 72 and Table 2.1A, p. 88; District of Columbia was eliminated from the calculations because it is a 3.7 standard deviation outlier on the NAEP variable.

[11] From this reviewer's own calculations extracting data from the report's Table 1.10A, p. 76 and Table 2.1A, p. 88. Again District of Columbia data are eliminated because it is a 3.7 standard deviation outlier on the NAEP variable.

[12] Dorn, S. (2006). No more aggregate NAEP studies? *Education Policy Analysis Archives, 14*(31). Retrieved January 2, 2007 from http://epaa.asu.edu/epaa/v14n31/.

[13] Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-57.

[14] Greenwald, R., Hedges, L.V. & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-9