The Twisted Relationship Between School Spending and Academic Outputs: In Search of a New Metaphor

Stephen J. Ceci, Paul B. Papierno, and Katrin U. Mueller-Johnson
Cornell University, Ithaca, NY 14853, USA

In this commentary we explore both putative and potential relationships between school funding and educational outcomes, for which numerous indices and variables have been examined during the past two decades. We conclude that despite large increases in real dollars, there is no readily discernable link between inputs and outputs, although when statistical control is exerted over prior ability, school mix, and teacher turnover, certain causal inferences appear warranted. However, it can be argued that the dominant analogy should be reconsidered, and schools serving large numbers of needy children may require higher inputs to keep from further declines, much like certain departments of a hospital require disproportionately more resources than others in order to be successful because they deal with problems that are more severe.

Keywords: Relationship, School spending, Academic output.

For readers who have come to expect a certain depth and acumen from Rutter and his associates, Rutter and Maughan’s target article will not disappoint them. These authors present a fair and responsible portrayal of the school effectiveness literature, and readers will find no better or more concise analysis than in these pages. Their review is chock full of insights into the often confusing and outright contradictory assertions made by proponents and opponents of various school effectiveness interventions. In our commentary we will discuss the implications of several aspects of Rutter and Maughan’s analysis of the relationship between several inputs and cognitive outputs. The three types of inputs we focus on are class size, community resources, and financial expenditures.

No one doubts that school matters: Keeping a child out of school results in plummeting scores as a linear function of each year the child...
stays out of school or is late starting school (Ceci, 1991). And yet, among children who attend school, the search for input–output relationships has not yielded strong conclusions. Rutter and Maughan are well aware of the controversy surrounding the relationship between financial inputs and cognitive outcomes. They correctly summarize this large literature by noting that associations between school spending and pupil achievement are, with certain noteworthy exceptions, quite weak. As we show later, large differences in outcomes can be found at the same level of input, and similar outcomes can be found despite large differences in inputs. In their words,

Studies of school effectiveness have generally . . . shown that, at any given level of resource, schools differ in their capacity to make use of what is available, and also that schools with similar levels of resources vary in their degree of effectiveness. That has been an important finding . . .

Yet, Rutter and Maughan suggest that we are lacking the appropriate data to analyze whether or not the availability of exceptional levels of expenditures would lead to very significant changes in student outcomes:

. . . it is crucial to note that, in almost all cases, the schools that were studied differed little in their level of resources. It cannot be assumed that a major increase or decrease in resources would make no difference . . . The empirical evidence is distinctly thin and it is obvious that the effects of variations in resources requires further study. We suggest that it is likely that, as with class size, minor variations in the middle of the range will prove to be of little consequence but equally that extra resources may be critical in some circumstances, provided that they are used in the most effective way.

Although it is true that contemporary studies have not analyzed school samples with enormous variation in fiscal input, we can turn to historical data to answer the question of whether or not large changes in spending have led to impressive gains in overall student achievement. From the early years of the twentieth century until today, spending in constant per pupil dollars in the United States has, according to some sources, increased nearly eightfold, from approximately $1000/pupil in 1930 to nearly $8000/pupil today (Figure 1). Yet, this enormous increase (in inflation-controlled dollars) has not been accompanied by large gains in achievement of American students.

Of course there are difficulties concerning the interpretation of such historical data: $1000 in 1960 had a different buying power compared to $1000 today. Taking this into account, Rothstein and Miles (1995) of the
Economic Policy Institute, for example, assessed educational expenditures between 1967 and 1991 to measure inflation-adjusted growth of school spending. They determined that the real (inflation-adjusted) increase in per-pupil spending amounted to 61% during this 25-year period. Further, this increase in funding does not reach students directly, as it also reflects input for federally mandated programs that target small groups of students as well as indirect spending (e.g., building maintenance and improvements, lunch programs).

More recently (March, 2001), the National Center for Educational Statistics (NCES) reported on Revenues and Expenditures for Public Elementary and Secondary Education School Year 1998–1999. The NCES reports that for the 1998–1999 School Year, an average of $6508 (unadjusted) was spent per pupil, an increase from the $5566 per pupil expenditure reported by Rothstein and Miles in their 1991 data. Consistent with Rothstein and Miles’ data, the NCES reports that about 62% of total expenditures were targeted for instruction (i.e., direct spending). Even after adjustments are made for inflation and noninstructional spending, real spending per pupil in America has gone up substantially (Rothstein and Miles report a 28% increase for regular education for the time interval between 1967 and 1991, which equals an average annual rate of about 1%. 

A breakdown of total expenditures showed that spending on regular education decreased from 80% to 59% while expenditures for special education rose from 4% to 17%. What this means is that more and more students shifted from large regular educational settings to smaller, specialized settings. Thus, far more money is spent today on special educational needs of the lowest performing children.
per year, and we are assuming consistent growth for current spending). Yet, as reflected in the National Assessment of Educational Progress (NAEP), these changes in expenditures have not been accompanied by concomitant increases in test scores.

The National Assessment of Educational Progress (NAEP) is a national indicator of American students’ achievement in major academic areas. The results of the 2000 Science Assessment show that concurrent with the aforementioned increases in per pupil expenditures, students in grades 4 and 8 showed no significant changes in performance since 1996. Moreover, students in grade 12 showed a significant decline in performance, with a smaller percentage of students at or above the basic level of achievement. The corresponding 2000 Mathematics Assessment showed more promise. Results indicated gains for fourth- and eighth-graders. However, twelfth-graders, despite an increase compared to 1990 scores, showed a decline since 1996.

The literature on class size effects is growing rapidly. But the two camps remain unreconciled: One side (e.g., Greenwald, Hedges, and Laine, 1996) has argued from meta-analyses that smaller classes result in superior learning for some groups. Others (e.g., Hanushek, 1999) have raised a variety of methodological objections.

Recently, Ehrenberg, Brewer, Gamoran, and Willms (2001) provided a cogent analysis of the class size debate. They detailed the myriad interpretative snarls in all extant data sets, including the STAR project in Tennessee, the single best experimental study to date. Ehrenberg et al. concluded that smaller classes (16 rather than 25) boost poor children’s reading and arithmetic achievement significantly. However, this boost does not appear to be cumulative in the sense that children who attended smaller classes for 3 years were no better off than peers placed in small classes for only 1 year. Further, the class size effect seems to vary depending on where in the performance distribution one looks; for example, a needy child at the 40th percentile in reading might be elevated to the 50th percentile following a year in a small class. Five years after being in a small class, this child’s advantage is still apparent—between a 0.1 and 0.15 S.D. gain over control group children who were in larger classes throughout the same period (see Figure 2).

Employing multilevel modeling, Willms and Kerckhoff (1995) estimated the effects of pupil–teacher ratio on the reading and mathematics scores of 16-year-old pupils in 148 local education authorities in the U.K. Their set of controls included sex, SES, and age 12 achievement. They found a non-significant relationship between class size and reading ($\beta = -0.018$), but a significant relationship ($\beta = -0.032$) for mathematics. This latter result indicated that a decrease in pupil–teacher ratio from 25 to 16 could be expected to result in a .16 S.D. increase in achievement for reading, and a .29 S.D. increase for mathematics.
Rutter and Maughan view the findings on class size as ‘‘liberating in their implication that the main benefits of small size may derive from very small groups for especially needy younger children rather than a much smaller class size reduction across the school as a whole’’. While we agree with this sentiment, an additional consideration bears mentioning: From a policy perspective, we need to ask whether the resources used to reduce the size of classes could be allocated to greater effect in a different manner, say, by requiring teachers to be credentialed in their subjects, or requiring teachers to obtain higher educational degrees, or to require more pre-and in-service workshops. Or perhaps, the money could lead to greater gains with even more direct expenditures, for example, by supplying needy children with personal computers. With current research, we simply do not know if smaller classes represent the best strategy for narrowing the race/SES achievement gap. As long as resources are limited, policy makers will want to know which type of intervention will be the most successful strategy.

Proponents of increased inputs sometimes shift levels of discourse in their argument. Initially, they may focus on shortcomings of the very worst schools—citing outdated texts, shoddy physical conditions and high teacher turn-over. However, the reality is, as Rutter and Maughan correctly note, that most schools do not vary greatly in inputs, and even when they do, the effect of the disparity is not in the direction one might assume, as we describe below. Most schools appear to provide sufficient conditions for children’s learning and development, and numerous demonstrations in America show that expenditures are unrelated to outcomes. For example,
although a state such as Utah spends far less on its students than does the District of Columbia, the former far surpasses the latter in achievement as measured by national assessments. As can be seen in Table 1, Utah spends roughly half of what D.C. spends, and its average class size is approximately 22 students to 1 teacher, whereas the D.C. class size ratio is closer to 14 to 1. These are enormous differences even after adjustment for higher costs of living in D.C. And yet, at every grade level and in every subject area assessed by the National Assessment for Educational Progress, Utah students significantly outperform D.C. students.

In response to such data, proponents of increased inputs have pointed out that the SES/race mix of the communities that serve as the catchments for schools also exerts a significant influence on school effectiveness. For example, D.C. schools serve a greater proportion of poor and educationally needy students compared to Utah and most states; more of its students are programmed in small, special educational settings, as a result of their low levels of achievement upon entering school. However, Rutter and Maughan point out that in their early community study in London they found that variations among schools in student progress could not be fully attributable to variations in the types of students and communities the schools served, and that the correlations between pupil achievement and such factors as teacher

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Expenditures, Pupil–Teacher Ratio and Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>District of Columbia</td>
</tr>
<tr>
<td>Per-pupil expenditure(^a) (1998–1999)</td>
<td>$9650</td>
</tr>
<tr>
<td>Teacher salaries (1998–1999)(^b,c)</td>
<td>$47,150</td>
</tr>
<tr>
<td>Pupil–teacher ratio (1998)(^b,d)</td>
<td>13.9:1</td>
</tr>
<tr>
<td>N.A.E.P. scores(^e)</td>
<td></td>
</tr>
<tr>
<td>Math (grade 4/2000)</td>
<td>193</td>
</tr>
<tr>
<td>Math (grade 8/2000)</td>
<td>234</td>
</tr>
<tr>
<td>Reading (grade 4/1998)</td>
<td>182</td>
</tr>
<tr>
<td>Reading (grade 8/1998)</td>
<td>236</td>
</tr>
<tr>
<td>Science (grade 8/1996)</td>
<td>113</td>
</tr>
<tr>
<td>Writing (grade 8/1998)</td>
<td>126</td>
</tr>
</tbody>
</table>

\(^a\) From U.S. Department of Education, Office of Educational Research and Improvement (2001), Table 5.


\(^c\) Estimated average annual salaries of teachers in public elementary and secondary schools, 1998–1999.

\(^d\) Pupil–teacher ratios in public elementary and secondary schools, Fall 1998.

and student turn-over were greater at exit than at entry (Rutter, Maughan, Mortimore, Ouston, and Smith, 1979). This suggests that school characteristics, in fact, exert a significant effect on student progress, and not just the other way around. Since Rutter and Maughan’s study, newer studies have confirmed their wisdom. Most recently, Naomi Breslau and her colleagues have shown that growing up in a racially segregated and economically disadvantaged community was actually associated with a decline in cognitive ability scores between ages 6 and 11, over and above any effect of maternal IQ, educational attainment, marital status or birth weight (Breslau et al., 2001). Using General estimating equations and multilevel analysis, these researchers demonstrated that the gap between the poor urban students and their more affluent suburban peers increased over time, and the 4.9-point gap that was not due to maternal IQ, education, and marital status at age 6 increased to 10.1 points by age 11.² This is clear evidence of an effect of schools and communities on student outcomes, independent of the effect of students on school outcomes.

Finally, Rutter and Maughan ask whether it is realistic to suppose that the standards in the least effective schools could be raised to those of the most effective, and whether the least effective schools are functioning poorly for reasons entirely outside their control. Answers to these questions will be critical in guiding funding, of course. However, as Rutter and Maughan point out, and we agree, schools serving disproportionately disadvantaged students may need to be viewed differently from the typical econometric model in which all school and student outcomes are equated, and with which linear input–output relationships are assessed. Perhaps a better analogy is to a hospital in which the various units have differential inputs and outputs. Because of the nature of the patients and conditions that they treat, some units (e.g., emergency) require higher levels of inputs (e.g., funding, staff) in order to maintain a level of success consistent with other, less demanding departments. This requirement of disproportionately higher inputs, with an often less obvious success rate, does not necessarily indicate a failure in the system. Rather, it provides an impetus to re-evaluate not only the distribution of quantity of resources but also how to best allocate and balance these inputs to serve the entire hospital, or as it were, school system.

REFERENCES


² In an alternate model, these investigators reported that the age 6 IQ gap between urban and suburban children was 16.4 points, and this widened to 21.6 points at age 11; 7.4 points age 6 was unaccounted for by maternal education, IQ, and marital status, while 12.6 points were unaccounted for by age 11.


