Cost-Effective School Inputs

Is Class Size Reduction the Best Educational Expenditure for Florida?

Anthony H. Normore
Lynn Ilon
Florida International University

The current debate about class size is not centered on whether smaller class sizes are desirable. Rather, the debate is whether the costs involved are the best ways to spend taxpayers' monies. This analysis addresses this question for the state of Florida. Using the Florida Comprehensive Assessment Test as a measure of educational achievement, a state data set containing information on all elementary schools was used to examine which government-funded inputs were most cost effective. Using a three-step methodology leading to a cost effectiveness analysis, this article finds that reducing class sizes is the most expensive of state inputs that affect achievement scores. Varying the mix of school personnel (administrators, teachers, and teacher aides) and investments in teacher quality (training and experience) are shown to produce the same results (raising test scores) at a lower cost than the reduction of class sizes.

Keywords: cost effectiveness; class size reduction; Florida; educational expenditures

Class size reduction (CSR) has become a major public policy focal point in Florida. The discussion has been relatively subdued and was not considered a major issue prior to passing Amendment 9 to Florida's constitution in 2002. As a result, educators, policy makers, and politicians in Florida are currently debating whether the funding of CSR in public schools is a cost-effective use of Florida state tax dollars. Essentially, the class size issue has many facets and is not politically neutral. According to Kennedy (2003), the "direct and indirect costs of class-size reduction persuaded the Florida Association of District School Superintendents (FADSS)
to oppose the amendment" (p. 6). Some school district superintendents in Florida are elected. Consequently, it might not be surprising that the FADSS supported a white paper that claimed that smaller classes would be too costly for the state. The policy is estimated to cost billions of state dollars during the next 10 years. Beginning with 2003 to 2004, the legislature shall provide sufficient funds to reduce the average number of students in each classroom by at least two students per year until the number of students per classroom does not exceed the maximum. Funding needed to meet these requirements is the responsibility of the state and of local school districts. Assuming that improved test scores is a major goal of this policy, the question has been raised as to whether reduced classroom size is the state’s best use of educational dollars.

The notion of economic efficiency is central to the debate of which school inputs best ought to be funded. Funding of CSR is simply the latest in a range of educational policy reforms implicitly focused on cost-effective use of state funds for education. Such reforms have led to revamping of school accountability measures and the development of state-mandated tests to determine mastery of basics in the state of Florida. Rarely, though, have such reforms been driven by knowledge of the relative cost effectiveness of proposed reforms. A wide body of research focuses on the effectiveness of such reforms irrespective of costs (e.g., Okpala, 2002; Slavin, 1989, 1990; Weaver-Dunne, 2000; Witte, 1999), whereas more recent research focuses on the costs of implementing them and some notion of what the trade-offs are, given implied resource use (e.g., Borman & Hewes, 2002; Egelson, Harman, Hood, & Achilles, 2002; Finn & Achilles, 1999; Finn, Gerber, Achilles, & Boyd-Zaharias, 2001; Krueger, 2002; Reynolds, Temple, Robertson, & Mann, 2002). This research will add to the literature and debate by evaluating school inputs on the basis of the expenditures required to affect test scores. The debate about class size is not centered on whether smaller class sizes are desirable. Rather, the debate is whether the costs involved are the best ways to spend taxpayers’ monies. One question along this line is whether CSR is the best use of dollars targeted at education.

This analysis adds to the information needed for an informed discussion on this topic. We see an opportunity to raise the level of discussion about this important issue by raising awareness of a missing element in the current debate: the relative cost effectiveness of government inputs to schools for Florida. The purpose of this study is to investigate the issue of CSR and its effects with a combination of testing for effects of various government inputs to schooling and weighting those effects by their relative costs. Although the first step in this research follows a well-worn path in looking at government input effects on achievement, the weighting of these effects by relative cost should add some important information to the debate. Furthermore, with the focus on class size within the state of Florida, the Department of Education put considerable effort into constructing a valid, reliable, and policy-relevant measure of classroom size. Thus, this analysis provides an opportunity to use statewide data to analyze the cost effectiveness of class sizes using the considerably stronger measure of classroom size than has traditionally been available through statewide data sets in general and for Florida in particular. As such, this research contributes to the literature on effects of government inputs on achievement, on the financial implications of reducing class size (e.g., Correa, 1993; Ferguson & Ladd, 1996; Hanushek, 1986, 1997, 1999a, 1999b; Hazzard, 1989; Hedges, Laine, & Greenwald, 1994; Smith, 1971; Tomlinson, 1988, 1990) and on whether the reduction of class sizes is the best use of educational funds in the state of Florida.

Review of Literature

CSR Effects on Achievement

Studies on CSR began as early as the 1970s and quickly identified a link between class size and achievement scores. Glass and Smith (1978) conducted a meta-analysis of reduced class size studies. These authors concluded that reduced class size can be expected to produce increased academic achievement. This was followed by a landmark class size study in the 1980s, Tennessee’s Student Teacher Achievement Ratio (STAR) Project in 1985. Because of the positive student achievement results of Project STAR, states and districts began to engage in class size initiatives in earnest in the 1990s as part of public education reform. In discussing class size research, some authors have failed to distinguish between pupil-teacher ratios and the number of pupils in a particular classroom (e.g., Greenwald, Hedges, & Laine, 1996; Hanushek, 1998; Hedges et al., 1994, Wenglinsky, 1997). Finn et al. (2001) state that “pupil-teacher ratios, usually computed for entire school districts or states, include special education and Title I teachers with small classes, subject matter specialists with no full-time classes of their own, librarians, teaching assistants, and others” (p. 175). According to Wagner (2001).

It should not be confused with teacher/pupil ratio which counts all educators in a school. ... CSR is an educational, economic and political issue of interest
to all education players and public... Class size decisions significantly affect the amount that taxpayers contribute to public education and involve choices where to apply limited resources. The notion that smaller class sizes improve student learning has an intuitive feel, but the research is not conclusive. There are many variables that must be considered and implementation presents a complicated set of challenges. (p. 1) 1

Although the distinction is clear between pupils per teacher (PTR) and classroom size, the bulk of the literature uses measures of PTR as a proxy for classroom size. Actual classroom size is difficult to calculate whereas many state-level data sets have the data for PTR calculation. Recognizing this weakness as a major constraint in implementing a CSR policy, Florida undertook a major initiative to construct a valid and reliable measure of class size, at considerable costs and technical resource expenses (Haynes, 2003). The problems inherent in constructing a class size variable in large data sets are compounded by other data and modeling issues. It is not surprising that the resultant research has reached varying conclusions. Research studies have yielded contradictory findings (CSR Research Consortium, 1999; Hertling, Leonard, Lumdend, & Smith, 2000). Many studies have indicated that CSR does have a positive impact in several areas: more time to cover curriculum, higher levels of student-student and student-teacher engagement, and safer schools with fewer discipline issues (Schwartz, 2003, p. 1). Nevertheless, the influence of smaller class size on academic achievement continues to be debated. Some research (STAR Project Tennessee) indicates positive effects even beyond the treatment period, although other research (Project Prime Time Indiana) does not. Research on California’s Class Size Reduction Program finds achievement gains, but such gains cannot be directly attributed to CSR. Given the conflicting research findings and the high costs of reducing class size, states are gradually increasing class sizes.

Videro (2000a) reported that after following the performance of nearly 10,000 students for 3 years, first, second, and third graders in classes of no more than 15 students scored better on standardized reading and mathematics tests than their counterparts in bigger classes. Research on the STAR project concluded that the benefits of smaller classes were cumulative as long as students remained in small classes.

Whereas many states have begun CSR initiatives (Indiana in 1985; Nevada, 1989; Utah, 1990; Texas, 1985; Oklahoma, 1989; North Carolina, 1991) three states (i.e., Tennessee, 1985; Wisconsin, 1995; California, 1996) have engaged in CSR evaluation to determine costs and effects on students, teachers, and the system as a whole. Schwartz (2003, p. 1) reports that CSR by Project STAR in Tennessee was “the most effective strategy for increasing student achievement scores” (in math). Although the strategy was considered effective, it was also the second-most-expensive approach for high levels of reading achievement, rating second to adult tutoring, which was a more costly way of improving reading. In a follow-up study conducted by Pate-Bain, Fulton, & Boyd-Zaharias (1999). findings indicated that students who were in Project STAR (those who were tracked through high school) demonstrated lasting effects as compared with non-STARC students by taking more advanced courses and were less likely to be retained in a particular grade or to drop out. Essentially, this lower rate of grade retention enabled the state to save millions of dollars. Gerber, Finn, Achilles, & Boyd-Zaharias (2001) found that starting early and continuing in small classes for at least 3 years are necessary to assure long-term carryover effects. Additionally, as supported by other researchers (e.g., Brewer, Krop, Gill, & Reichardt, 1999; Witte, 1999), their investigation confirmed that attending a class with a full-time teaching assistant had little impact on academic performance in Grades K-3 and no significant effects on performance in later grades (p. 174). Grissmer, Flanagan, Kawata, & Williamson (2000) support the notion that teacher aides are a weak and expensive treatment for improvement in student achievement.

Findings on impact of reduced class size in Wisconsin’s Student Achievement Guarantee in Education (SAGE) are consistent with the Tennessee Project STAR. According to Egelson et al. (2002), in August 2001 the Wisconsin governor signed a budget that provided “$171 million for the 2001-2002 and 2002-2003 school years to fully fund a maximum class size of 15 students in grades K-3 in all SAGE schools” (p. 17). In the same report, for the Southeastern Regional Vision for Education, Egelson et al. (2002) provided an overview of several major state-level CSR initiatives, including the results across time of two North Carolina districts where CSR was implemented at no added cost by focusing on program design, type of classroom instruction, and professional development provided to teachers.

**Economic Benefits of CSR**

Educational expenditures are guided by some notion of optimal educational outcomes. Frequently, one specific outcome is chosen as a primary goal of education: test scores, purportedly measuring educational achievement. A variety of approaches (e.g., Bracey, 1995; Opkala, 2002) have
attempted to model the effects of school and contextual variables on achievement. Bracey and Steller (2003) point out that studies of costs are lopsided if there is no attempt to analyze or predict savings through benefits—that “the costs of a program must be compared against the benefits that the program generates.” (p. 783). These authors maintain that “benefit/cost rates that are greater than one indicate that a program is worthy of consideration regardless of the absolute level of program costs” (p. 783) and “provide strong evidence of long-term positive outcomes for high quality pre-school programs” (p. 780). Krueger (2002) summarizes some of these studies. In so doing, he makes a persuasive argument that investments in class size are likely justifiable on economic grounds. Along the same lines, Borman and Hewes (2002) and Viadero (2000b) concluded that long-term economic benefits of reducing class sizes outweigh the costs. This is supported by McRobbie, Finn, and Harman (2004), who argue that in the long run, potential benefits may offset the costs. Smaller classes in primary grades start students on a path that reduces the need for special education, grade retentions, or disciplinary measures and increases the likelihood of high school graduation. Such outcomes can translate into real savings.

Other programs that have garnered attention as to whether the benefits from the programs are worth the costs include Head Start (Meisels, 2003), Abecedarian Project (Masse & Barnett, 2002), and Perry Project (Schweinhart, Barnes, & Weikart, 1993). Cost-benefit analyses on all three conclude they are. Treatment-group students who had taken part in these projects in preschool (e.g., 123 in Perry Preschool, 109 in the Abecedarian study) were followed in adulthood and had lower crime rates, higher educational graduation rates, and fewer retention rates in grade (Molnar & Achilles, 2000).

The economic argument advanced here, although an important contribution to the CSR debate, falls short of answering the full question on whether CSR is a good investment for society. First, it assumes that societies generally make expenditure decisions on the basis of long-term net returns (benefits). This is clearly not the case. Worldwide, investments in education yield nearly universally positive social returns (Psacharopoulos, 1994), but virtually no society has increased its educational expenditures to the point where costs equal long-term benefits (a resting point of zero returns). Second, the findings do not answer the question of whether CSR is the best investment in raising achievement scores. That is, if reducing class sizes raises achievement scores that lead, in turn, to a more productive society, are there other school investments that would have accomplished this goal for less money?

### Financing CSR

Some economic analysis takes this research a step further. Hanushek (2002), a critic of studies favoring small classes, stated that schools might get a bigger payoff from making other changes, such as improving the quality of teachers. He reiterates that the issue is whether there exists any evidence that class size reduction ever matters... surely class size reductions are beneficial in specific circumstances—for specific groups of students, subject matters, and teachers...but that class size reductions necessarily involve hiring more teachers, and teacher quality is more important than class size in affecting student outcomes. (p. 27)

The correct summary, as suggested by Hanushek (2002) is that “just providing more resources—whether in the form of reduced class sizes or in other forms—is unlikely to lead to higher student achievement as long as future actions of schools are consistent with their past choices and behavior” (p. 28). Hanushek (1998) concluded, “We have extensive experience with class size reduction and it has NOT worked” (p. ii), and “extensive econometric investigation show NO relationship between class size and student performance” (p. iii).

Given that societies do not generally make educational expenditure decisions based solely on long-term economic returns, the parameters are narrowed to consideration of what can be funded under normal budgetary constraints. Hanushek (1999a) believes that the underlying contexts involve political decisions regarding school funding. Many states and school districts across the United States are seeing their budgets reduced and thus view commitments to smaller classes within this context. Limited federal aid has spoken to this need. According to Schwartz (2003).

In 1999 Congress began appropriating funds so that schools could hire additional teachers and invest in other CSR measures... Federal CSR funds for the 2001-02 school year totaled $1.6 billion, and allocations are now included in the No Child Left Behind Act of 2001. (p. 1)

A CSR study found that space problems were listed as the number one problem by school administrators in schools that were unable to implement reduced-size classes in the 1st year (CSR Research Consortium, 1999). In California, the size of the CSR efforts “has caused problems ranging from hiring thousands of new teachers and the addition of 18,000 classrooms” (Schwartz, 2003, p. 2). According to Bohnstedt and Stecher (2002), “some
schools declined to participate" in the CSR initiatives because "they were simply unable to meet its hiring and space requirements" (p. 2). These authors assert that more experienced teachers were transferring to less troubled schools, which left urban schools attracting the inexperienced and uncertified teachers. Finding spaces for additional classes was another problem reported. There was sufficient funding for increasing the size or number of schools but no available urban land on which to build these schools (Bolmstedt & Stecher, 2002, p. 2).

McRobbie et al. (2004) offer alternative ways of funding CSR that can help contain costs, such as (a) targeting the resources and (b) allowing local flexibility in attaining smaller class goals to include redistributing resources and creative scheduling. Odden and Archibald (2001) demonstrated that the financial burden of creating smaller classes could be minimized by careful planning and analysis of resources. These authors discuss how leaders in the Kenosha school district, Wisconsin, managed to significantly reduce class size in elementary schools by allocating resources creatively and by deploying all revenues made possible by student demographic characteristics and the state's school financial system (p. 10).

Other research indicates alternative means for improved student achievement. McIntyre and Scott (1989) indicated that smaller investments in other educational strategies may yield similar or greater achievement gains as reducing class sizes. Stern (1987) noted in his research that an increase in teachers' salaries would more likely have effects on student achievement scores and would be more cost effective than a reduction in class size. Harder (1990) asserted that instructional quality may be more important to student achievement as opposed to reduction in class size. Reynolds et al. (2002) showed that cost-effective outcomes of the Chicago child-parent centers far exceeded any costs or inputs (p. 294). The findings of the cost-benefit analysis show that a comprehensive child development program intervention has substantial long-term benefits to society through increased economic well-being and reduced expenditures for remediation and treatment.

**Conditions Specific to the State of Florida**

According to Canedy (2003), the state of Florida ranks near the bottom among states in per capita spending, graduation rates, and standardized achievement test scores. The state governor was quoted as stating that reducing class sizes in Florida "would cost more than $27 billion, in a state whose total annual budget is about twice that" (Canedy, 2003, p. 18). The state legislature introduced Amendment 9 (Senate Bill 30A). The bill addresses CSR and its implementation. According to Horne (2003) and Florida's Class Size Amendment (2004), the amendment to the state constitution in Section 1, Article IX establishes final goals as noted for the beginning of the 2010 to 2011 school year. It outlines the equity in funding per student; the responsibilities of the state, school, and school district; and the characteristics of an effective CSR model. In constitutional language (Florida's Class Size Amendment, 2004), by 2010, the maximum number of students assigned to each teacher who is teaching in public school classrooms (a) in pre-K through Grade 3, does not exceed 18; (b) in Grades 4 to 8, does not exceed 22; and (c) in Grades 9 to 12, does not exceed 25. Beginning in fiscal year 2003 to 2004, the legislature must provide sufficient funds to reduce the average number of students in each classroom by at least two students per year until the new maximum class sizes are met. Although the governor of Florida is willing to dedicate $628 million in the 2004 budget for smaller classes, he warned that responding to the new law will not be achieved without the possible shifting and expansion of private school vouchers, lifting of restrictions on the number of charter schools, or redrawing of attendance zones (Horne, 2003). The state will incur costs as the amendment is phased in. Estimated cumulative costs through 2010 range from $20 billion to $27.5 billion, depending on how classrooms are built. Once fully implemented, each year's operating costs are estimated to be $2.5 billion in today's dollars.

Florida, like California, has issues of overcrowding in schools, specifically in Miami-Dade County, where "about 41 percent of the county's schools are at 150 percent over capacity" (Weaver-Dunne, 2000). A school campus can have as many as 20 portables, giving the impression of a veritable minicity. The schooling population is rapidly expanding because of immigrant students, who make up about "22 percent of the system's 360,000 students" (Weaver-Dunne, 2000). Governor Bush's educational budget for 2003 to 2004 proposed $2.8 billion dollars to be committed above and beyond the funds already allocated for the school construction and renovation programs. Additionally, the governor proposed that $143.2 million dollars of the state's fund be allocated for recruitment, retention, and providing professional development for teachers. This proposal is because of the "project need of nearly 192,000 new teachers" (Florida's E-Budget, 2003). Along with this proposal, Florida has had the A+ Plan for education, which has increased student achievement and school accountability and has given parents the choice of which education forum is best for their child (Florida's E-Budget, 2003). Under the A+ Plan and No Child Left Behind programs, parents who have students in failing schools can take advantage
of a voucher system. The voucher program helps CSR while providing individual students with a choice in their education. Students are able to relocate to another public school or attend a private school of their choice. The private schools, however, do not abide by the accountability standards that are set forth for the public school sector, nor are all private schools open to voucher students (Nees & Keenan, 2003). In addition to the voucher system, the state of Florida has promoted a series of tools and options for each district to use as it tries to meet the requirements of CSR and Senate Bill 30A. Some of the options include year-round school and shifting school attendance zones, whereas other options include the use of joint facilities with postsecondary institutions to promote dual enrollment courses.

Research Methods

Data Collection

All data were taken from the online Florida Indicators Report, 2001 to 2002. This Florida state Department of Education database includes school-level data on all public (and some charter) schools in the state that were collected for the 2001-2002 academic year (the most recent complete year for which there were data). We included more than 1,700 schools in the elementary school portion of this database (Florida Department of Education, 2002b). In our cross-sectional analysis, all variables were either taken directly from this database or computed using variables from the database and were aggregated at the school level. The databases included results for the Florida Comprehensive Assessment Test (FCAT) scores, which were used as a measure of student achievement in this study.

Data Analysis

This cost analysis was conducted in three analytic stages. The first step was to run a regression that related FCAT scores to inputs that government pays for (along with some socioeconomic and school characteristic variables, as mentioned in the previous section). These results were used to measure the quantity of each input (for the average school) needed to raise test scores by a fixed amount. We chose a fixed increase of 2% in test scores for this simulation step. Finally, we calculated the cost for each of these quantities by first calculating the unit cost (measured in the same units as those input in the regression) and then weighted that unit cost by the quantity required for the 2% rise in test scores.

Variables

Our goal was to cull from the data set all variables that are indicators of public expenditures for schools. Nevertheless, we included some variables of socioeconomic status and school characteristics that have proven to be important in predicting student achievement in previous literature. The following are variable definitions and are sometimes direct quotes from the official state guide to the data (Florida Department of Education, 2002b). The methods for costing each variable are also described below where applicable.

- **Teacher-aides ratio** measures the ratio of the number of teachers per school (Grades K-3) to the number of teacher aides in those same grades. Costs are the average cost per school of changing the ratio of teachers to aides by converting (partial) aide positions to (partial) teacher positions.
- **Average FCAT score** is the weighted average of FCAT scores (reported in the database) for Grades 3, 4, and 5. Scores were weighted by the number of pupils taking each of these exams and then averaged for the school. Scores are reported as average school percentile rank.
- **Percentage low-income students** is the percentage of all students in the school who are low income.
- **Percentage non-White students** is the percentage of all students in the school who are not reported as White.
- **Percentage administrators** is the percentage of all staff in the school that are classified as administrators. Costs are the average cost per school of changing the percentage of total staff that are administrators (holding total staff size constant).
- **Percentage instructional staff** is the percentage of all staff in the school that are instructional. Costs are the average cost per school of changing the percentage of total staff that are instructional (holding total staff size constant).
- **Charter** indicates whether a school is a charter school (1) or not (0).
- **Expenditures per student** is the total school operating costs per unweighted full-time equivalent student. According to the Florida Department of Education, these costs are the sum of salaries, benefits, purchased services, materials, supplies, other direct expenses, capital outlay, and school indirect costs. This variable was reported in dollar amounts and was, therefore, not transformed into a unit cost measure.
- **School size** is the number of students attending the school in October.
- **Percentage teachers advanced degrees** is the percentage of teachers in the school who hold a master's degree or higher. The cost is the difference between salaries of teachers with advanced degrees and teachers without advanced degrees weighted by average number of teachers in a school.
Teachers average years experience is the average number of years of teaching experience for teachers (including in-state and out-of-state experience). The cost is the average salary increases per year for teachers with experience weighted by average years of experience needed to raise average teacher experience in a typical school.

Average class size is the count of students who are enrolled in classes for specified courses divided by the count of teachers for the classes in the specified range of courses. Unit cost is the cost of staffing additional (partial) classrooms for the average school.

Findings

The first step in our analysis produced a regression. The model used was

\[
\text{FCAT score} = \\
+ \% \text{ low-income students} \\
+ \% \text{ non-White students} \\
+ \% \text{ administrators} \\
+ \% \text{ instructional staff} \\
+ \text{charter} \\
+ \text{expenditures per student} \\
+ \text{school size} \\
+ \% \text{ teachers advanced degrees} \\
+ \text{teachers average years experience} \\
+ \text{average class size} \\
+ \text{teachers per aide K-3}
\]

The results were surprisingly robust with an $R^2$ of .88. All the variables (including the constant) were statistically significant at the .05 level. Table 1 shows the regression results.

Results generally pattern those found in much of the literature that relates school inputs and characteristics to student achievement. One surprising finding, however, is that expenditures per student are negatively related to FCAT scores. Although this finding could be counterintuitive, it is important to remember that in the context of a regression analysis (the first step in our analysis), the variable that captured expenditures per student would have been evaluated with the expenditures attributable to other variables partitioned out. That is, expenditures per student were statistically evaluated once the expenditures attributable to teacher qualifications, teacher experience, classroom size, and so on had been accounted for. In this regression context, expenditures per pupil were statistically significant (at the .01 level).

### Table 1
Regression Results and Variable Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Beta Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>82.519</td>
<td>-36.945</td>
</tr>
<tr>
<td>% Low-income students</td>
<td>55.7</td>
<td>-5.721</td>
</tr>
<tr>
<td>% Non-White</td>
<td>50.7</td>
<td>0.012</td>
</tr>
<tr>
<td>% Administrators</td>
<td>2.8</td>
<td>0.002</td>
</tr>
<tr>
<td>% Instructional staff</td>
<td>63.3</td>
<td>-14.523</td>
</tr>
<tr>
<td>% Charter</td>
<td>5.7</td>
<td>-0.002</td>
</tr>
<tr>
<td>Expenditures per student</td>
<td>5.609</td>
<td>-0.003</td>
</tr>
<tr>
<td>School size</td>
<td>681</td>
<td>0.000</td>
</tr>
<tr>
<td>% Teachers advanced degrees</td>
<td>27.4</td>
<td>0.274</td>
</tr>
<tr>
<td>Teachers average years experience</td>
<td>12.6</td>
<td>-0.451</td>
</tr>
<tr>
<td>Average class size</td>
<td>22.9</td>
<td>6.980</td>
</tr>
<tr>
<td>Teachers per aide K-3</td>
<td>0.37</td>
<td></td>
</tr>
</tbody>
</table>

Note: All variables significant at .05 level.

### Table 2
Costs of Effecting a 2% Gain in Florida Comprehensive Achievement Test Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cost per School</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Administrators</td>
<td>$16,284</td>
</tr>
<tr>
<td>% Instructional staff</td>
<td>$124,672</td>
</tr>
<tr>
<td>Expenditures per student</td>
<td>-$577</td>
</tr>
<tr>
<td>Teachers advanced degrees</td>
<td>$121,050</td>
</tr>
<tr>
<td>Teachers average years experience</td>
<td>$122,543</td>
</tr>
<tr>
<td>Average class size</td>
<td>$39,559</td>
</tr>
<tr>
<td>Teachers per aide K-3</td>
<td>$87,455</td>
</tr>
</tbody>
</table>

All other variables are related to FCAT scores in directions consistent with findings of most studies on student achievement. It is interesting to note, however, that changes in percentage of staff assigned to administration has a much more powerful influence on test scores than does the percentage of staff assigned to instruction, as evidenced by the beta coefficients. These regression results were used to weight the costs necessary to effect a 2% gain in FCAT scores. Table 2 shows these results.

The negative relationship between expenditures per student and test scores resulted in an apparent cost savings for that variable, though its interpretation is unclear and is discussed below. The next-cheapest change is to
increase the percentage of administrators in the school. This result is a function of the relative sensitivity of test score results to the percentage of administrators in the school and is further explored below. Raising the percentage of teachers in the teacher-aides mix in early grades (K-3) also appears to be a cost-effective means of raising test scores. Our results show that lower class size is the least cost-effective intervention for public monies that Florida elementary schools can make.\textsuperscript{11} The quality of instructors, as evidenced both by the mix of teachers to aides and by the percentage of teachers with advanced degrees, positively influences test scores and is more cost effective than lower class size.

Two caveats are important here. First, the model is linear, and as with all linear models, results are most reliable when predicting small changes. This is particularly important with these findings because some of the variables have limits to their ability to be changed. It is likely, for example, that there is a ceiling past which exchanging instructional staff for administrators either has no effect or begins to have a negative effect. Also, although hiring teachers with more-advanced degrees is apparently a cost-effective way of spending money on schools, one cannot hire more than 100\% of teachers with advanced degrees. Past this limit, one must find other investments.

Second, these changes are necessarily estimated on the average school.\textsuperscript{12} Not all schools will respond to changes identically, and not all schools can make the incremental changes suggested by our results. Thus, the results can guide basic investment decisions for the state but may not be a perfect guide to cost-effective investments for any given school. This, of course, can be said for all such models of school achievement but are, nonetheless, worth noting here.

**Discussion**

The data are rich and varied and drew us to ask further questions. Below is a discussion of three of the most intriguing. We asked what was behind the results of school administrators, expenditures per student, and the teacher-aides ratio.

**Administrators**

Increasing the administrative staff relative to other staff appears to be a cost-effective way of increasing FCAT scores. This finding is not inconsistent with a body of literature that suggests that good administrative support

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percentage of Administrators by School Size</strong></td>
</tr>
<tr>
<td><strong>Percentage of Staff Who Are Administrators</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>School size</td>
</tr>
<tr>
<td>% Low-income students</td>
</tr>
<tr>
<td>Expenditures per student</td>
</tr>
</tbody>
</table>

can have a substantial effect on student achievement. What is surprising here, though, is that increasing administrative staff is a substantially cheaper means of increasing FCAT scores than any other government investment we tested. Part of the answer lies in the sensitivity of FCAT scores to the percentage of staff that are administrators. According to our findings, FCAT scores are 6.5 times more sensitive to changes in the proportion of staff that are administrators than to similar changes in the proportion of teachers in schools (beta coefficients of 1.22 vs. 0.19). The other component driving this result is the fact that increasing administrative staff by a percentage (of total staffing) is relatively cheap. Because administrators earn only about $20,000 (on average) more than teachers, a shift in staffing of 1\% (from 2.8\% to 3.8\%) is relatively cheap (about $17,500). The combination of a relatively cheap input with a fairly high FCAT sensitivity makes for a cost-effective government input.

It seems clear that increasing the percentage of administrators can go only so far in raising FCAT scores. To begin investigating this relationship, schools were divided into five groups by their reported percentages of administrators. Table 3 shows these groupings, with Group 1 having about 1.7\% of staff as administrators, Group 2 having 2.3\%, and so on. Although the percentage of administrators is somewhat related to school size (Pearson’s $r = .24$), the relationship is weak; neither does there appear to be a strong relationship with the wealth of the school, as evidenced by the percentage of low-income students. Finally, it does not appear that a higher percentage of administrators is associated with higher expenditures per student (Pearson’s $r = .16$).

The relationship between the percentage of administrative staff and FCAT scores is complex. A simple bivariate analysis sheds very little light.
Table 4
Florida Comprehensive Achievement Test (FCAT) Mean Scores
by Percentage Administrators

<table>
<thead>
<tr>
<th>Percentage of Staff Who Are Administrators</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (1.7%)</td>
<td>56.4</td>
<td></td>
</tr>
<tr>
<td>Group 2 (2.5%)</td>
<td>56.8</td>
<td></td>
</tr>
<tr>
<td>Group 3 (2.6%)</td>
<td>57.8</td>
<td></td>
</tr>
<tr>
<td>Group 4 (3.1%)</td>
<td>57.9</td>
<td></td>
</tr>
<tr>
<td>Group 5 (4.1%)</td>
<td>56.2</td>
<td></td>
</tr>
</tbody>
</table>

The two factors are only weakly related when examined out of the context of a full model (Pearson’s r = .07). Table 4 shows that FCAT scores are not markedly different for the five groups of schools, ranked by percentage of administrators.

This leaves open the question of why FCAT scores are so sensitive to the percentage of administrators. It could be that administrators interact with other inputs in complex ways not easily revealed by descriptive analysis.

Teacher Aides

According to our findings, the ratio of teachers to teacher aides is the second most cost-effective investment that Florida can make. Both economic and policy implications depended on whether teacher aides are a supplement or a substitute for teachers. We calculated costs as if schools would keep staff size the same (number of teachers + number of aides) but would increase the number of teachers relative to the number of teacher aides. The underlying assumption is that if schools needed the results of this study, they would do so under some budgetary constraints that would, overall, across the state of Florida, cause the ratio of teachers to teacher aides to shift but would not increase overall average school staffing size.

There are currently about 6.3 teachers and 17.5 teacher aides in Grades K-3 in the typical school in Florida. This results in a ratio of teachers to aides of .37 (6.3:17.5). Our results indicate that it would take a 16% increase in this ratio to raise FCAT scores by the prescribed 2%. Doing so would shift the ratio from .37 to .53 (.37 + .16 = .53). Although keeping overall staff size the same, this would imply that teacher numbers increase by 3.8 teachers, while the number of aides would be reduced by the same amount. Given the differences in teacher and aide salaries, the typical schools would face an annual cost of $87,435 to fund these staffing changes. Although this cost calculation modeled behavior that might be typical if our research findings were used as a guide, the fact is it may not model actual practices of schools. Thus, the policy question remained: What, in actual fact, are schools really doing? Are aides adding to the total numbers of instruction-related staff, or are they being hired instead of more costly teachers?

Culling our data returned surprising results. We divided our data into five equal groups representing various levels of aides per teacher. We then asked whether teachers were either being supplemented or substituted through the hiring of aides. Figure 1 shows that teacher aides appear to be a substitute for teachers rather than a supplement.

Most striking in the figure is that the average staffing size is about the same across all groups. It is the composition of this staff that varies. In the first group (the group with the highest teacher-aides ratio), teachers almost equal aides. In the last group, aides outnumber teachers nearly 10:1. One possibility is that school size or class size was driving the hiring decisions. Yet class sizes for all groups varied little: between 22.1 and 23.7 with no apparent pattern vis-à-vis teacher-aides ratios. The same can be said for school size with a range of 605 to 754, again with no apparent pattern.
So school and class size do not seem to influence the relative mix of teachers and teacher aides.

The picture is muddied still further when we asked the question of whether the teacher-aides mixture was a function of socioeconomic status. Table 5 shows that there might be a relationship between teacher-aides ratios and socioeconomic status, although not the relationship many might expect to see. The schools with higher percentages of teachers relative to aides are associated with highest percentages of low-income students.

Although higher numbers of teachers relative to teacher aides was found in our regression results to be statistically significant in predicting FCAT scores, we also looked at a pattern of teacher-aides ratio to FCAT scores. Table 5 shows a patterned relationship, with FCAT scores rising modestly as more aides were present for every teacher. Again, this result is counterintuitive because it might be assumed that teacher-intensive staffing is preferable to high ratios of teacher aides.

The nexus of explanations concerning the ratio of teachers to teacher aides leaves open the question of why schools choose the hiring of one rather than another. Clearly, there is some trade-off being made between teachers and aides, with typical schools of about 650 students hiring about 23 of some combination of teachers and aides. The easy explanations escape our analysis. It is the schools in poorer areas that have the most teachers and fewest aides. FCAT scores are higher in schools with higher aides-to-teacher ratios. A more thorough investigation will be necessary to uncover what is going on in these schools.

### Table 5

<table>
<thead>
<tr>
<th>Teacher-Aides Ratio, Socioeconomic Status, and Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aides per teacher Grades K-3</td>
</tr>
<tr>
<td>% Low-income students</td>
</tr>
<tr>
<td>Average FCAT test scores (%)</td>
</tr>
</tbody>
</table>

### Expenditures per Student

Expenditures per student turned out to be a good predictor of FCAT scores but surprisingly in the opposite direction than we or previous research would have predicted. That is, the lower the expenditures per student, the higher the FCAT scores. This research does not attempt to answer the question of why such a contrary finding occurred. although it is possible to hypothesize that expenditures that go beyond teacher quality, staffing, and class size have little effect. Nevertheless, the finding was worth looking into as far as the data would allow.

We began by asking what the bivariate relationship was between FCAT scores and expenditures per pupil. Figure 2 shows this relationship. In this chart, schools were evenly divided into 10 groups representing average FCAT scores. We then asked what the average FCAT score was for each of these groups and compared it to the average expenditure per student. The chart shows that FCAT scores and per-pupil expenditures are inversely related. In the highest scoring schools, expenditures per pupil were about $4,750 per pupil per year whereas in the lowest scoring group of schools were roughly $5,860 per pupil.

We asked, then, whether expenditures per pupil were related to the socioeconomic status of the school area. Because the data set does not have community data, we looked at the percentage of low-income students in the school as a proxy for community socioeconomic status. Figure 3 shows the relationship we found.

In fact, schools that appear to expend the most per student are in relatively lower socioeconomic communities. Why this is so is not revealed by
many of its provisions are bold and well intentioned, many will need
further clarification if they are to achieve their intent. Our findings indicate
that CSR is likely not a cost-effective means of raising student achievement
as measured by test scores, at least in the state of Florida. Quality and mix
of staffing appears to yield the same results for substantially less cost.
Nevertheless, our results require careful examination if they are to be used
as a guide to policy.

Three of the seven inputs we measured required virtually the same level
of investment to cause a 2% change in FCAT scores. Percentage of staff
devoted to instruction, percentage of teachers with advanced degrees, and
average years of teacher experience each require an investment of about
$120,000 for the typical school to see similar gains in FCAT scores.
Economists might assume that the equilibration of these three investments
is not by chance. Rather, it is possible that policy makers collectively have
made investment decisions knowing where the next-best investment dollar
ought to be spent—whether these decisions have been made implicitly or
guided by research findings. Collectively, policy makers may have found
the right mix or balance of these three inputs. If this is so, then investment
decisions regarding any of these three inputs is now a matter of keeping
them balanced. Increasing one substantially versus the other may not have
the impact that keeping a balance between them would have.

Although there is little argument that smaller class sizes are desirable,
the debate is whether CSR is the most cost-effective use of taxpayers'
dollars. Based on the findings from this study, research in three areas is
necessary for further exploration. The issues requiring further investigation
are outlined as follows: (a) Why is the expenditure per student not posi-
tively affecting scores on the FCAT? Few would argue that expenditures per
student do not count, yet our results demonstrate otherwise; (b) it makes
sense that the quantity and quality of school administrators matters, but the
study raises the question as to why the scores on the FCAT are so sensitive
to the percentage of school administrators; and (c) considering the findings
from this study, one might question why Florida has chosen to support CSR
initiatives rather than other educational investments.

Reducing overall class sizes in the state of Florida is an enormous
financial undertaking. If the intention of voters was to put in place a public
policy that improved the quality of K-12 education in the state, the question
seems obvious as to why class size was the intervention proposed for an
amendment. No study we found addressed whether CSR was the most cost-
effective educational expenditure for raising achievement scores in Florida,
much less whether it was the most cost-effective way to spend taxpayer

the data. It is worth noting, however, that Title I provides monies for the
poorest schools, and Florida sometimes provides extra support for schools
that have scored particularly low on the FCAT exam.

**Conclusion and Implications for Educational Policy**

Attention to class size is a timely and appropriate focus for education
policy. Although CSR has enormous intuitive and political appeal, design-
ing a successful CSR policy is no simple matter. Educators and policy
makers may need to make choices between effective educational practices
because of financial, facility, or personnel issues. As knowledge from state
and local experiences continues to evolve, some of the lessons learned from
CSR include the need for adequate facilities, policies that allow flexibility
in the use of funds to maintain focus on learning and not just on getting
numbers down, excellent teaching, and directing resources, particularly
toward children in the minority and low-income children. With all of its
requirements and provisions, Florida's Senate Bill 30A on CSR is expected
to generate considerable controversy among educational stakeholders
(i.e., parents, teachers, school administrators, and policy makers). Although

---

**Figure 3**

**Expenditures per Pupil and Percentage of Low-Income Students**

- **Expenditures per Student**
- **% low-income students**

---

decreases in expenditures per pupil.
dollars across other sectors. The current governor made it clear that he
did not support the proposed amendment. The answer likely lies in the
confluence of two realities. First, a study such as this one is complex and not
easily translatable into results that the typical voter can digest easily.
Second, many people would like to see educational quality improve, and
class size appears to be a tangible way of influencing education, even lacking
hard data on various options. Given the strong will of the voters, the
politicized nature of the issue, and the large influence of the amendment on
Florida tax expenditures, the debate continues to center on ideological posi-
tions not well grounded in research findings.

Notes
1. See Note 7 for a brief discussion of these differences with respect to the Florida educa-
tional accountability data set.
2. Florida's data set does make this distinction. See Note 7.
3. Because the model is linear, the choice of any fixed amount would not affect the cost-
effectiveness results.
4. A more comprehensive discussion of the problems of costing this variable is contained
in a following section.
5. The data provided by the state does not make the distinction between administrators.
Rather, 100% of all personnel are classified in one of three categories: administrators, instruc-
tional staff, and all others. See Florida Department of Education (2002a) for the Internet
address explaining how this variable was constructed.
7. The Florida Department of Education attempted to differentiate between pupils per
teacher and class size in its 2001 to 2002 data set. In so doing, it made a clear departure from a
simple students-divided-by-teachers ratio. To determine class sizes, the department used number of
students who took certain basic courses and divided it by the number of teacher-periods that the
subjects were taught. This calculation is detailed in Florida Department of Education (2002a,
p. 3). A conversation with Martha Haynes, who supervised this effort, indicated that considerable
error went into the calculation of accurate class sizes. Her presentation on this detailed
methodology is available on the Web (Haynes, 2002). These authors also spoke personally with her
and thank her for her time and openness about the process.
8. This $R^2$ is unusually high. But alternative specifications also resulted in high $R^2$
(always greater than .80). Multicollinearity was checked as a possible mitigating factor, but
none was found.
9. A test for multicollinearity was negative.
10. Only those inputs that can be affected by changes in government spending were costed.
11. This is not to say that class size does not affect achievement. Our results (Table 1)
show that smaller class sizes have a significant positive association with higher Florida
Comprehensive Achievement Test scores. Rather, our question is whether it is the most cost
effective of the effective inputs.
12. This is a limitation of regression models that output beta coefficients based on values
from a "typical" (average) case.

References
Bohara, K. W., & Stecher, B. M. (2002). What have we learned about class size reduction
in California, Sacramento, CA: California Department of Education.
Bracey, G. (1995). Debunking the myths about money for schools. Educational Leadership,
53(3), 65-69.
costs. Phi Delta Kappan, 84, 780.
size reductions under different policy alternatives. Educational Evaluation and Policy
Analysis, 21(2), 1179-1192.
Comley, D. (2003, February 11). Florida struggles to find a way to achieve smaller classes. The
Education Economics, 1(2), 129-135.
summary/99-98/index.htm
ence. Washington, DC: U.S. Department of Education, Regional Educational Laboratory,
Office of Educational Research and Improvement.
schools. In H. F. Ladd (Ed.), Holding schools accountable: Performance-based reform in
Finn, C. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications,
Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of
small classes. Teachers College Record, 103(2), 145-183.
doe/arm/class-size.htm
Florida Department of Education (2002a). Guide to calculations for the Florida school indicators
elais/elaspubs/pdf/fsn0002.pdf
budget.state.fl.us
Gerber, S. B., Finn, J. D., Achilles, C. M., Boyd-Zaharias, J. (2001). Teacher aides and students'
Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class size and
achievement. San Francisco: Far West Laboratory for Educational Research and
Development.
ment: What state NCLB test scores tell us. Arlington, VA: RAND.


Anthony H. Normore is an assistant professor of educational administration in the Department of Educational Leadership and Policy Studies at Florida International University, Miami.
A former K-12 public school educator for 20 years (teacher, school administrator, district office program implementation specialist), he has presented at various local, state, provincial, national, and international conferences in Canada and the United States. He has facilitated in-service training and workshops for school and district personnel on school improvement planning, organizational change, organizational planning, and effective schools. He has represented the Canadian Teachers Federation to Nepal, where he worked with the National Nepalese Teachers Association. He is currently researching and writing in the areas of leadership development and preparation, planning, socialization, recruitment, selection, performance appraisal, and accountability of school leaders.

Lynn Ion is an educational economist specializing in educational policy and planning issues arising from the globalization of the world's economies. She is a director of the Intercultural Institute for Educational Initiatives at Florida International University. She also directs the International-Intercultural Studies in Education program. She has lectured and consulted for the World Bank, Harvard Institute for International Development, the United Nations, Educational Testing Service, the U.S. Agency for International Development, the Asian Development Bank, and several local and global nongovernmental organizations. She has lived and worked throughout the world, including 3 years in Micronesia, 2 years in Zimbabwe, and a year in Jordan. In addition, she has worked in or has been consulted on several other countries, including Jamaica, the former Soviet Union, Egypt, Kuwait, Nepal, Bangladesh, Vietnam, Malawi, Zambia, Uganda, and South Africa. Her reports, scholarly work, and invited lectures have dealt with several other countries.