



## **Public Sector Performance Measurement and Stakeholder Support**

**David N. Figlio**

Faculty Fellow, Institute for Policy Research  
Professor of Human Development and Social Policy and Economics  
Northwestern University

**Lawrence W. Kenny**

Professor of Economics  
University of Florida

Version: date: July 7, 2009

**DRAFT**

***Please do not quote or distribute without permission.***

## **Abstract**

Over the past several decades there has been dramatically increased attention paid to measuring the performance of public sector and nonprofit organizations in the United States and elsewhere. Recent research has indicated that public sector and nonprofit organizations are responsive to performance measurement in both productive and unproductive ways. However, it is not yet known how stakeholders respond to this measurement. This paper makes use of a unique panel survey data set of the population of elementary and middle schools in Florida to directly investigate this question. The authors exploit the fact that Florida changed its school grading system in 2002, and they study the degree to which private contributions to schools are responsive to the information contained in school grades. They find evidence that school grades can have substantial effects on a school's ability to obtain private contributions. They also observe that schools serving different clienteles are treated differently in response to changes in school grades.

The survey data used in this paper were collected with funding from the National Institutes of Child Health and Human Development, U.S. Department of Education, the Annie E. Casey, Smith Richardson and Spencer Foundations, and the Atlantic Philanthropic Society. The survey data were collected in conjunction with Cecilia Rouse, Dan Goldhaber and Jane Hannaway. We have benefitted from the suggestions made by the two referees, as well as seminar participants at Northwestern University and conference participants at the American Education Finance Association and Southern Economic Association annual meetings. We alone are responsible for all errors.

## Introduction

Over the past several decades there has been dramatically increased attention paid to measuring the performance of public sector and nonprofit organizations in the United States and elsewhere. These performance measures, ranging from report cards for Medicare HMOs to determinations of whether schools make "adequate yearly progress" as required by the federal No Child Left Behind Act to charity ratings provided by the American Institute of Philanthropy, are intended to induce health care providers, schools, charities and other organizations to provide their services more efficiently. Because public sector and nonprofit quality is multidimensional and costly to measure, stakeholders often have difficulty obtaining and processing information about these services.<sup>1</sup> Conveying information about public sector performance has the potential to improve stakeholders' monitoring abilities, and to the degree to which stakeholders may use their monitoring to effect change, this could improve the performance of service providers.

Recent research has indicated that public sector and nonprofit organizations are responsive to performance measurement in both productive and unproductive ways. Schools, for example, respond to accountability incentives by boosting overall performance and introducing substantive policy and practice changes aimed at improving performance (Rouse et al., 2007) but also by engaging in apparently strategic behavior (see, e.g., Figlio, 2006; Jacob, 2005; Neal and Schanzenbach, forthcoming) that makes it more difficult to know the extent to which accountability-induced improvements are genuine. These behavioral reactions are to be expected, given the high degree to which consumers use accountability information: Report cards are influential in determining housing prices (Figlio and Lucas, 2004), school choice

---

<sup>1</sup> Research on the frontier of economics and cognitive science (e.g., Gabaix, Laibson, Moloche and Weinberg, 2006; and Payne, Bettman and Johnson, 1993) describe the costs of cognition that individuals face when gathering and interpreting information about goods and services. See also DellaVigna (forthcoming) for examples of economic behavior in real-world informational settings.

(Hastings and Weinstein, 2007), and Medicare HMO enrollment decisions (Dafny and Dranove, 2008). Performance measurement influences consumer behavior regarding private-sector firms as well, in areas ranging from responses to restaurant hygiene grade cards (Jin and Leslie, 2003) to movie reviews (Reinstein and Snyder, 2005).

It is, therefore, clear that performance measurement influences the behavior of the measured, as well as overall market responses to the public sector and nonprofit organizations being measured. But it is not yet known how stakeholders respond to this measurement. These organizations frequently draw both from public support and from contributions by invested parties. Does receiving a good or bad rating influence the degree of support for the organization shown by these stakeholders?

This paper directly investigates this question. Making use of a unique panel survey dataset of the population of elementary and middle schools in the state of Florida, we study the degree to which private contributions to schools -- typically collected via parent-teacher organizations -- are responsive to the information contained in school grades. Beginning in 1999, Florida assigned letter grades to its public schools on the basis of measured school performance. We exploit the fact that in 2002 Florida dramatically changed its school grading system, generating an information "shock" that caused some schools to have better grades than they would have had under the previous system and other schools to have worse grades than would have otherwise occurred.

To our knowledge, this is the first paper of its type: The closest research of which we're aware is Jin and Whalley's (2007) study of the expansion of US News and World Report's ranking system to cover a larger number of universities, in which they measure the effect of attention per se on state financial support of universities. That study, however, is addressing a

fundamentally different question, as it does not identify the effect of placement in the ranking hierarchy. Brunner and Sonstelie (2003) utilize nonprofit tax return data to study the determination of voluntary contributions to schools in California in the aftermath of Proposition 13. They find greater total donations in larger schools and in schools serving richer and more educated families. However, their study describes the *presence* of voluntary contributions, rather than the *response* of contributions to performance measurement. And studying the response of contributions requires more detailed contributions data than tax return information, as the vast majority of schools receive sufficiently small amounts of donations that their parent-teacher organizations are not required to report contributions to the Internal Revenue Service. While this is not a serious issue for Brunner and Sonstelie's purposes, it renders tax return data useless for our purposes. One can only credibly study the effects of accountability on contributions using survey data, and the survey that we utilize is, to our knowledge, the first of its kind.

We find evidence that receiving a high school grade, conditional on past grading performance, does not increase the level of voluntary contributions to the school, but receiving a low grade is associated with considerable reductions in private financial support for the school. Indeed, we estimate that a school that receives a grade of "F" -- the lowest in the state's system -- will experience a drop in contributions of two-thirds or more, and a school that receives a grade of "D" also will experience a substantial reduction in contributions. In other words, stakeholder support, at least in the short run, is negatively impacted by receipt of a poor performance measurement score. We also observe that donations to schools serving poor or minority families are more sensitive to school grades than donations to schools serving more affluent families.

Our results provide empirical support to models (e.g., Vesterlund, 2003; Landry et al., 2006) that predict that donations respond to signals of charity quality such as well publicized

initial contributions. A related literature (e.g., Sloan, 2009; Chhaochharia and Ghosh, 2008) has found that charities that receive more favorable accountability ratings by groups such as the Better Business Bureau's Wise Giving Alliance get more contributions. A charity's overhead rate may provide evidence on how efficiently it is operated. Although the evidence is mixed (see Bowman, 2006, pp. 293-294), the preponderance of evidence suggests that less money is donated to charities with higher expense ratios. These results are consistent with our finding that less money is contributed to poorly run schools. Donors seem reluctant to throw good money at inefficient organizations. This withdrawal of support may provide another incentive for poorly-measured public and nonprofit organizations to improve their measured performance.

### **Conceptual framework**

Suppose that people care about their own private consumption ( $C$ ) and about how much learning is perceived to take place in the school ( $L$ ). Perceived learning in turn hinges on the perceived effectiveness of spending on education ( $\beta$ ) and on expenditures per pupil ( $E$ )

$$L = \beta \times f(E)$$

A lower school grade is hypothesized to decrease the perceived effectiveness of school spending ( $\beta$ ). This raises the cost of learning. The effect on preferred school expenditures, and thus on donations to the school, hinges on how responsive the amount of learning ( $L$ ) is to the drop in perceived school effectiveness.

If the price elasticity for learning is zero, then the percentage fall in  $\beta$  must be completely offset by the rise in  $f(E)$ . If resources from the school district or state are fixed, then donations to the school must rise. More generally, as long as the percentage fall in learning demanded ( $L$ ) is smaller than the percentage fall in school effectiveness ( $\beta$ ), a rise in school donations is needed

to bring about the desired small or nonexistent drop in  $L$ . In this scenario, the community rallies around the school, increasing donations to the less efficient school.

On the other hand, if the price elasticity for learning is sufficiently large, then the fall in desired learning ( $L$ ) is greater than the drop in school effectiveness ( $\beta$ ). A fall in school spending is needed to bring about a sufficiently large drop in learning. In this scenario, donations to schools diminish as the quantity of learning demanded is sharply curtailed.

In conclusion, whether school donations rise or fall when perceived school effectiveness drops depends on how sensitive desired learning is to an increase in its cost. In a very similar theoretical structure, Landry et al., (2006, pp. 750-751) conclude that an initial "seed money" donation has an ambiguous impact on subsequent individual contributions to a charity.<sup>2</sup>

The responsiveness of perceived school effectiveness ( $\beta$ ) to school report card grades should reflect how weight is placed on the new information provided by the school grades. Disadvantaged parents tend to be less involved in the school and thus are less informed about their school's effectiveness. Since they are less informed, they are expected to be more responsive to school grades than more affluent parents.

### **School grading in Florida**

Florida's 1999 A+ Plan for Education introduced a system of school accountability with a series of rewards and sanctions for high-performing and low-performing schools. The A+ Plan called for annual curriculum-based testing of all students in grades three through ten, and annual grading of all public and charter schools based on aggregate test performance. As noted above, the Florida accountability system assigns letter grades ("A," "B," etc.) to each school based on

---

<sup>2</sup> Similarly, neutral technological change results in a rise in the amount of labor and capital demanded in an industry only if the price elasticity of demand for the product is sufficiently large. See Blair and Kenny (1982).

students' achievement (measured in several ways). High-performing and improving schools receive rewards, while low-performing schools receive additional assistance as well as sanctions. In addition to the stigma associated with receiving low grades, schools that received a grade of "F" in two years out of four have their students become eligible for vouchers to attend a different (higher rated) public school, or an eligible private school. And while poorly performing schools receive additional assistance, they also are subject to additional scrutiny and oversight. All "D" and (especially) "F"-graded schools are subject to site visits and required to send regular progress reports to the state.<sup>3</sup>

School grading began in May 1999, immediately following passage into law of the A+ Plan. Between 1999 and summer 2001, schools were assessed primarily on the basis of aggregate test score *levels* (and also some additional non-test factors, such as attendance and suspension rates, for the higher grade levels) and only in the grades with existing statewide curriculum-based assessments<sup>4</sup>, rather than on the *progress* schools made toward higher levels of student achievement.

Starting in summer 2002, however, school grades began to incorporate test score data from all grades from three through ten and to evaluate schools not just on the level of student test performance but also on the year-to-year progress of individual students. However, while during the 2001-02 school year several things were known about the school grades that were to be assigned in summer 2002, the specifics of the formula that would put these components together to form the school grades were unknown until late in the school year. We take advantage of the fact that schools and their stakeholders could not necessarily anticipate their school grade in

---

<sup>3</sup> Details on oversight and reporting requirements can be found online at <http://www.bsi.fsu.edu/PerformanceUpdates/performanceupdates.aspx>.

<sup>4</sup> Students were tested in grade 4 in reading and writing, in grade 5 in mathematics, in grade 8 in reading, writing and math, and in 10 in reading, writing and math.

summer 2002 because the specific changes in the grading formula were not decided until well into the school year.

As can be seen in Table 1, there existed a considerable amount of change in the grade distribution between 2001 and 2002. Most notable is the fact that while no schools received a grade of "F" in 2001, 38 elementary and middle schools received an "F" grade in 2002.<sup>5</sup> Overall, the grade distribution shifted upward, with 836 "A" schools (at the elementary and middle school level) in 2002 as compared with 544 in 2001 and 484 "B" schools in 2002 as compared with 396 in 2001. Rouse et al. (2007) demonstrate that a substantial fraction of the change in school grades during this transition is due to changes in the grading formula, rather than changes in student demographics, student performance or institutional response, and nearly all of the newly "F"-graded schools received their bottom grade exclusively due to the change in the grading system. They estimate that 48 percent of elementary schools either received a higher or lower grade than they might have "expected" based on the prior grading system, and that 15 percent of schools that might have expected to receive a "D" under the old system received an "F" under the new one. That is, many schools were "shocked" by the change in the grading system.

### **Survey of public school principals**

The only administrative data on private donations to schools come from tax returns of nonprofits (e.g., parent-teacher associations) with revenues exceeding \$25,000 per year. As mentioned above, however, the vast majority of schools are not aided by nonprofits with

---

<sup>5</sup> We limit our analysis to elementary and middle schools because the nature of voluntary contributions to these schools is considerably less likely to be to help to purchase extracurricular services, through athletic booster fees, band booster fees, and so on. Brunner and Sonstelie (2003) made the same determination in their study of voluntary contributions.

revenues sufficiently large to require a tax return to be filed.<sup>6</sup> A second issue with using tax return data for this purpose is that revenues do not equal donations; these organizations are required to report basic financial statement data, and revenues vary considerably with the mode of fundraising employed. Some organizations raise revenues from sources such as catalog sales where they purchase the items for sale and keep typically one-third to one-half of the revenues as proceeds; other organizations raise revenues from sales of donated items or services (e.g., school carnivals, food sales). Two parent-teacher organizations making identical donations to their schools may appear to have dramatically different levels of revenues reported to the Internal Revenue Service.<sup>7</sup>

We take advantage of a unique survey conducted jointly by one of the authors of this study and colleagues at Princeton University and the Urban Institute in three waves – in the early portion of the spring during the 1999-2000, 2001-02, and 2003-04 academic years -- to the universe of "regular" Florida public school principals.<sup>8</sup> The survey team achieved response rates that were over 70 percent in each year. Rouse et al. (2007) demonstrate that while higher-graded schools were somewhat more likely to respond to the survey than were lower-graded schools, the characteristics of schools responding to the survey, by school grade, are quite similar to those of non-respondents.

---

<sup>6</sup> Note that these tax returns are for informational purposes only. These organizations are not required to pay taxes on their proceeds.

<sup>7</sup> As an entirely unscientific example, one of the authors has served as treasurer of parent-teacher associations at two different elementary schools that made very similar levels of donations -- approximately \$20,000 per year -- to their respective elementary schools. Both collected more than \$25,000 in annual revenues and were therefore required to file tax returns, but one organization averaged \$32,000 per year in revenues during the author's tenure as treasurer while the other organization averaged \$40,000 per year in revenues. The difference between the two is entirely due to the mix of fundraising selected.

<sup>8</sup> We excluded "alternative schools" such as adult schools, vocational/area voc-tech centers, schools administered by the Department of Juvenile Justice, and "other types" of schools. Note that we included charter schools serving "regular" students as well. The survey instruments are available on request.

The school surveys ask principals to identify a variety of policies and resource-use areas. Most important for the purposes of this study, principals are asked in each round of the survey:

Approximately how much additional revenue does this school raise annually through other sources of income (e.g., PTAs, community or business sponsorship, athletic or parking fees, etc.)?

In follow-up interviews with a subset of principals, it was determined that principals heavily weighted the most recent year of these revenues, so that each year's survey response can best be thought of as an approximation of the prior year's revenues from outside sources (e.g., the 2003-04 survey is reporting on 2002-03 additional revenues.) Nearly all responding schools answer the question on donations; for instance, in the 2003-04 survey, 69 percent of elementary and secondary schools statewide, and 98 percent of responding schools, answer this question.

Despite the fact that we have attained a very high response rate for the survey of school principals, it is important to gauge the degree to which schools that responded to the relevant question are similar to the population of schools as a whole. Table 2 reports descriptive statistics of elementary and middle schools in the 2003-04 survey that reported data on donations, compared with the universe of survey-eligible schools. As reported in Rouse et al. (2007), respondent schools are more likely to have received a grade of "A" in 2001 or 2002 and are less likely to have earned a grade of "D" in 2001 (though not in 2002). In addition, respondent schools have somewhat lower percentages of black and English language learner students, due to the fact that top-graded schools are most likely to respond to the survey. As Rouse et al. (2007) demonstrate, however, responding and non-responding schools at any given grade level are very similar along a variety of dimensions.

Figure 1 presents kernel density estimates of the distributions of log donations in each of the three rounds of the survey. The vast majority of schools report nonzero donations; in all three rounds of the survey, around three percent of the total number of schools have zero donations.<sup>9</sup> The distributions of log donations are roughly normal, and elementary and middle schools report similar distributions of donations. The typical level of donations has increased over time; median donations in the 1999-2000 and 2001-02 surveys are \$10,000, as compared with \$12,000 in the 2003-04 survey. It is apparent that a small number of schools in the most recent survey report very large levels of donations: The 95th percentile of reported donations in 2003-04 is \$100,000 versus \$60,000 in 1999-2000 and \$50,000 in 2001-02. This generates a large increase in mean donations from \$18,094 in 1999-2000 and \$18,925 in 2001-02 to \$36,649 in 2003-04 despite only modest increases in median donations in the 2003-04 survey. We therefore estimate a variety of models to investigate the degree to which the results are sensitive to outlier donation levels.

### **School grades and changes in donations**

*Kernel density estimates:* We now turn to the relationship between school grades and donations. Before estimating parametric models of the effects of grades on donation changes, it is instructive to evaluate the raw data on donations. Figure 2 presents kernel density estimates of the distribution of log donations in each of the three waves of the survey, for schools that would receive in 2002 grades of "C", "D" and "F". The top two figures can be thought of as pre-shock distributions, while the bottom figure reflects distributions of log donations observed after the grading shock occurred. One observes that in all three rounds of the survey, schools that would

---

<sup>9</sup> This fact underscores the importance of using survey data for these purposes. In Brunner and Sonstelie's (2003) application of tax return files only 20 percent of schools had a non-profit organization that raised at least \$25,000 and thus were counted as receiving contributions.

receive grades of "C" tend to garner higher levels of donations than those that would receive grades of "D" or "F", but the fraction of schools that received "D" or "F" grades with very low donations increased considerably after the shock. This is particularly true for the 2002 "F"-graded schools. Note, however, that a small fraction of "F"-graded schools received relatively high levels of post-shock donations, so the reduction in the donation levels is by no means uniform.

The change in donations from survey to survey can be more easily observed in Figure 3, which presents the change from the 2001-02 survey round to the 2003-04 survey round in log donations reported by a school. It can be seen that the density of "F" schools experiencing considerable reductions in donations post-grading is greater than that for "D" schools and especially "C" schools, although, as seen in Figure 2, some "F" schools experienced relatively large increases in donations as well. The mean change in log donations for "F" schools is -1.16, as compared with +0.11 for "D" schools and +0.22 for "C" schools. We revisit the distributional changes in donations later in the paper.

*Parametric models:* Schools vary along a number of different dimensions, so we next estimate a parametric model in which our dependent variable is the log of donations in the 2003-04 survey, and where we control for 2001-02 donations, prior school grades (in 2001, before the shock to the grading system), and a series of school-level variables collected from the Florida Department of Education. Specifically, we control for school size, which has an ambiguous estimated effect on donations because larger schools have a larger pool of potential donors, but also greater incentives for individuals to free ride (Brunner and Sonstelie, 2003). We also control for the racial and ethnic composition of the school, the rate of student absences, student stability, the rate of student suspensions, the percentage of students in the school who are gifted,

and the free-lunch-eligible percentage in the school, as all could influence the supply of donations to a school.

In Table 3 we report regressions based on the complete sample of elementary and middle schools (as well as elementary schools only) and on a truncated sample that excludes the top 1 percent of donations ( $> \$880,000$ ) to gauge how sensitive our results are to outliers. We observe that the results are substantively very similar regardless of whether we include or exclude the most extreme cases. We observe that schools earning grades of "D" in 2002, all else equal, experience reductions in donations relative to schools earning grades of "C". These are large estimated changes in donations, suggesting that donations would be 28.0 to 44.5 percent smaller in "D" schools than in "C" schools, all else equal. Still larger estimated effects are observed in the case of "F" grades. Donations are estimated to be 67.0 to 85.6 percent smaller in schools receiving an "F" than in schools that got a "C" in 2002.<sup>10</sup> These multivariate parametric results provide evidence that is consistent with the bivariate effects illustrated in the kernel density plots. Apparently, a "D" or an "F" results in a sharp fall in the amount of learning demanded and thus leads to a drop in contributions. Families may possibly believe that continued donations amount to "throwing good money after bad."<sup>11</sup> It is possible that, rather than withholding financial support from low-performing schools, families' contributions are being crowded out by increased state resources to "D" and "F"-graded schools, as the state increased its financial investment in these schools following the 2002 grading. That said, as reported by Rouse et al. (2007), the increases in state spending were virtually identical for "D" and "F" schools, so a crowd-out story cannot explain the relative reductions in stakeholder support for "F" schools as compared with "D" schools.

---

<sup>10</sup> The differences between "D" and "F" coefficients are statistically significant at conventional levels as well.

<sup>11</sup> One of the authors of this paper conducted a focus group with parents in a school that received an "F" in summer 2002, and this was a sentiment raised by several participants in the group.

The coefficients on other control variables are consistent with expectations as well. Reassuringly, log donations in 2002 are positively related to log donations in 2004. Larger schools receive no more voluntary contributions than smaller schools; the increase in potential donors is offset by greater free riding. As expected, schools that serve poorer families receive less voluntary contributions than schools whose students come from wealthier families. The percentage of students who are black is negatively related to donations, as is the percentage of students who have been suspended from school. Parents whose children are expected to stay in a school for more years have a greater benefit from improving a school and thus are expected to contribute more to a school. The student stability variable is, however, generally not significant.

It is possible that the responses to changes in school grades are more complex than the relationships presented in Table 3. For instance, it may be the case that schools whose grades fall from a "D" to an "F" may experience different responses than those whose grades fall from a "C" to an "F". Or those whose grades increase from a "B" to an "A" might experience different responses than those whose grades increase from a "C" to an "A". We therefore considered a model in which we estimated separate coefficients for each combination of 2001 grade and 2002 grade. We found no evidence that the magnitude of the grade change made a substantial difference at the top of the grade distribution, though it did make a modest qualitative difference at the bottom of the distribution. For instance, the estimated effect of going from a "C" to an "F", relative to remaining constant at a "C" grade, is -2.732 (with a p-value of 0.01) while the estimated effect of going from a "D" to an "F", relative to remaining constant at a "D" grade, is -2.251 (with a p-value of 0.01); these differences, however, are not statistically distinct from zero. Therefore, for ease of interpretation, throughout the remainder of the paper we concentrate on

models in which we estimate the effect of receiving a certain grade in 2002, holding constant 2001 grades, rather than the most flexible model possible.

*Regression discontinuity evidence:* An alternative approach to investigating the effects of school grades on changes in donations employs the fact that school grades are determined using a "grade points" formula, in which a school with 280 points earns a "D" and one with 279 points earns an "F". Therefore, we can employ a regression discontinuity approach to estimating the effects of receipt of the lowest school grade on log donations. Figure 4 presents graphical evidence of this relationship. The figure depicts mean residual log donations, generated from a regression of log donations from the 2003-04 survey on log donations from the 2001-02 survey, as well as past grade dummies and the set of school-level control variables included in Table 3. While in the regression that generated these residuals each school is a separate observation, for the purposes of illustration in Figure 4 each circle represents the mean residual value of all schools with the same number of grade points in the 2002 school grading system. It is important to note that given that we find that both "D" and "F"-graded schools experience reductions, on average, in donations, comparing schools at the margin of "D" and "F" grades will likely understate the overall effect of receipt of an "F" grade in a regression discontinuity framework.

Regression discontinuity model estimates are dependent upon the functional form of the model, and in the illustrated case, with the relationship between grade points and residual log donations estimated as a cubic function, the regression discontinuity estimated effect of receipt of an "F" grade versus a "D" is -1.523 with a standard error of 0.611. Other functional forms yield consistently large and negative estimated effects of "F" grade receipt: For instance, the regression discontinuity estimated effect when a quartic functional form is employed is -1.176 and the estimated effect when a quintic functional form is employed is -1.053, both of which are

somewhat smaller estimates than the cubic model but still statistically significant at conventional levels. A straight linear regression discontinuity model yields an estimated effect of "F" receipt of -0.868 with a standard error of 0.466. Therefore, the regression discontinuity models present additional evidence of the effects of school grades on a school's change in private donations. In summary, we observe consistent evidence, using a variety of graphical and parametric methods, that school grades limit the typical low-performing school's ability to collect donations to the school. We next turn to whether there exist any distributional differences in these effects.

### **Distributional effects of school grading**

Schools that are given a "D" or "F" are found to receive fewer donations. Are donations from some socio-economic groups more responsive to school report card grades than donations from other groups? Disadvantaged parents tend to be less involved in school activities. Figlio and Kenny (2007, p. 912) report that there is a strong positive relationship between parental income and various measures of parental activity in the school (PTA activity, parent-teacher contact, parental involvement as reported by principals). Thus parents in schools serving more affluent communities, schools with more gifted children, and schools with fewer minorities are expected to be better informed about their school's effectiveness ( $\beta$ ). Since they are more knowledgeable about the school, their estimate of perceived school effectiveness should be less responsive to the new information provided by the school report card grade. That is, getting a low grade should have a smaller effect on donations in schools serving rich families than in schools serving poor families.

The first two rows of Table 4 present estimated effects of receipt of various grades in 2002 (relative to a grade of "C") for schools stratified based on the percentage of students who

are eligible for free or reduced price lunch. We divide the sample into two sets of schools -- those above the median percentage low income for "D" or "F" schools (83.5 percent) and those below the median percentage low income for "D" or "F" schools. We find strong evidence that "D" and "F" schools serving relatively low income populations disproportionately experience reduced donation levels following the grading system change. While "F" schools that are comparatively high-income and comparatively low-income both face reduced donation levels, it is the schools serving the most disadvantaged students that see the largest reductions in donations. Among "D" schools, the estimated reduction in donations appears to be exclusively occurring in the relatively low-income schools. The contributions received by "A" schools and "B" schools were not significantly different than those received by "C" schools.

The next panel of Table 4 presents a similar comparison, but this time the schools are stratified on the basis of the percentage of students in a school labeled as gifted, according to the Florida Department of Education. We make this distinction because we suspect that the parents most likely to independently monitor school quality -- and therefore be least likely to rely on external ratings -- are those with the highest-ability children. We observe that the negative effect of low school grades on donations appears to be concentrated nearly entirely among the set of schools with below-median (for low-graded schools) rates of gifted program participation. Low-graded schools that serve relatively large numbers of gifted students do not appear to be affected by the school grades.

A third cut of the data involves stratifying schools based on whether they have relatively large fractions of minority students. We find that schools serving relative large fractions minority experience lower levels of donations when they receive grades of "D" or "F" and that schools with fewer minorities get fewer donations after receiving an "F". The differences

between the high minority and low minority "D" and "F" coefficients, however, are not statistically significant. On the other hand, high-minority schools that receive high grades of "A" or "B" tend to receive dramatically larger levels of donation than do high-minority schools receiving a "C" grade. Therefore, it appears that heavily minority schools are particularly responsive to receiving high grades. This makes sense given that the change in the grading system made it more likely that high-value-added schools serving disadvantaged populations would receive higher grades than before. In summary, donations are more responsive to school grades in schools serving disadvantaged students.

### **What motivates donors?**

We have laid out a simple, straight-forward theoretical framework to explain why potential donors might reduce their contributions to schools that have been given low marks by the state. That said, our framework does not take into consideration more complicated potential motivations by donors. One possible motivation is that changing schools is costly to parents, and so therefore, parents might increase their donations in an attempt to bolster the grade of the school in the next year. In such a situation, one might expect that schools that just missed a higher grade might see increased levels of donations as a consequence. It could also be the case that schools that just barely made their grade would experience relatively increased donations as well.

We directly test for whether this is the case by expanding our base specification reported in Table 3 to include variables for whether a school was within five points of the school grade threshold.<sup>12</sup> We estimate separate parameters for whether the school missed the threshold by

---

<sup>12</sup> The grade point thresholds used by the state of Florida in 2002 were 280 points for a "D", 320 points for a "C", 380 points for a "B", and 410 points for an "A".

between one and five points, or if the school barely exceeded the threshold by five or fewer points.<sup>13</sup> We find no evidence that schools near a grading threshold experience higher levels of donations; in fact, the point estimates are negative though far from statistical significance. The coefficient on being just above a grading threshold is  $-0.094$  ( $p=0.52$ ), while the coefficient on being just below a grading threshold is  $-0.173$  ( $p=0.33$ ).<sup>14</sup> Were we to instead increase the definition of "marginal schools" to be within ten points of the grading threshold, the fundamental points remain unchanged; the coefficient on being just above a grading threshold in that specification is  $-0.068$  ( $p=0.63$ ) and the coefficient on being just below a grading threshold is  $-0.132$  ( $p=0.39$ ). Therefore, it seems unlikely that stakeholders are deploying donations in an attempt to boost a school's grade in the future.

Another possible motivation for parental donation behavior is that parents may be withholding donations to schools that receive a grade of "F" because, if the school receives another "F" grade in the next three years families would become eligible for school vouchers under Florida's accountability system at the time. Such a story could explain the negative effect of a school receiving a grade of "F" (though not a grade of "D") but would involve different motivations from not wanting to "throw good money after bad." While rational parents might view this as a very low-probability event,<sup>15</sup> this story remains a possibility. We therefore estimate a variant of the Table 3 model in which we include a measure of the concentration of private schools in the county -- a private school Herfindahl index -- as a regressor.<sup>16</sup> We find no relationship between private school concentration and changes in donations to public schools; the

---

<sup>13</sup> In this model and the others in this section, we use the full set of elementary and middle schools, because our results in Table 3 suggest that there exists little difference in the results whether or not we exclude the schools with the highest donation levels.

<sup>14</sup> The latter coefficient is further decreased when excluding the "F" schools.

<sup>15</sup> Fewer than one-fifth of schools in Florida that had received grades of "F" prior to 2002 had ultimately received a second "F" grade, which would trigger school vouchers for students.

<sup>16</sup> We have also estimated models in which we limit private schools to be within ten miles of the public school, with effectively no change in the findings.

coefficient on the private school Herfindahl index is -0.174 ( $p=0.59$ ). However, such a model only would indicate that increased private school competition for public schools is unrelated to changes in public school donations. In order to investigate the potential that possible future school vouchers would lead to reduced donations today, one should interact private school competition variables with the 2002 school grade. When we estimate a model that includes a full set of interactions between the school grade and the private school Herfindahl index, we find that the interaction with an "F" grade is statistically insignificant. While the coefficient on the interaction is reasonably large in magnitude -- a one-standard-deviation increase in private school concentration is associated with a reduction in donations of 0.578 -- the point estimate on the interaction term is very imprecisely estimated ( $p=0.78$ ), but the sign is opposite of what would have been expected under the support-withdrawal-due-to-potential-future-vouchers story. Therefore, it seems unlikely that families are withholding support from "F" schools because they might receive vouchers in the future should the school continue to perform poorly.<sup>17</sup>

That said, the "F" result could be due to the most motivated families moving away from a school that had previously received a grade of "F" and was therefore "voucherized." We therefore estimate a variant of the Table 3 model that includes a variable reflecting whether the school had received a grade of "F" in 1999 or 2000.<sup>18</sup> We find that, indeed, schools that have received a second "F" grade experience further reduced donation support -- the relative reduction in donations for a second-time "F" school as compared with a first-time "F" school is -2.379 ( $p=0.00$ ). However, the coefficient on first-time "F" receipt remains large in magnitude and statistical significance: the coefficient is -1.548 ( $p=0.01$ ). So while a portion of the "F" grade

---

<sup>17</sup> We also estimated models in which we interacted an "F" grade with the fraction of public schools in the county earning grades of "C" or better. We found no evidence of differential contributions to "F" schools facing different degrees of public school competition either.

<sup>18</sup> No Florida schools received a grade of "F" in 2001.

result is apparently due to the fact that some of the "F" schools were recidivists -- and could reflect either a voucher-receipt effect or further evidence of "throwing good money after bad" -- the fact that the first-time "F" grade result remains large in magnitude and statistical significance indicates a general school-grading response is at work as well.

## **Summary**

This paper provides the first evidence of stakeholder financial reactions to changes in performance measurements in the education sector. We make use of rich population-based survey data on contributions to schools to measure how school contributions change after a major exogenous change in Florida's school grading system was introduced in 2002. We find that schools facing low grades ("D" and especially "F") experience substantial reductions in donations to the school. This negative reaction is particularly pronounced in relatively low-income schools and those with small gifted populations, and is present regardless of whether the school's students have become eligible for school vouchers as a consequence of the poor school grade. Similar to findings from the social psychology and marketing literatures, we find effects that apparently reflect a general aversion to "throwing good money after bad." In none of the subgroups do schools receiving a "D" or "F" receive significantly more donations, which would reflect desired learning levels that are relatively insensitive to perceived school productivity.

We observe little general increase in donations associated with high measured performance, except in the case of schools with relatively large fractions of minority students. Schools with very high rates of minority students that also receive grades of "A" or "B" tend to experience large increases in their level of donations after the change in grading system is realized. It could be that low-income or minority families rely more heavily on state school

grades for the purposes of school monitoring than do their higher-income or higher socio-economic status counterparts.

The results of this study have important implications for the practice of performance management in the public and nonprofit sectors. We find that stakeholders appear quick to withhold support for organizations with poor measured performance, and may reward organizations with high measured performance in circumstances where such measurements are less expected (e.g., schools with very large minority populations). To the degree to which stakeholder financial support is related to other levels of support as well (a correlation that is purely speculative), the results of this study indicate that stakeholder reactions may serve to reinforce the performance measurement. In light of these findings, the finding in the literature that schools improve considerably after receiving a low grade (see, e.g., Rouse et al., 2007) reflects even more favorably on the possibility for performance-improving effects of school accountability.

## References

- Blair, Roger D. and Lawrence W. Kenny. *Microeconomics for Managerial Decision Making*. New York: McGraw Hill, 1982.
- Bowman, Woods. Should Donors Care About Overhead Costs? Do They Care? *Nonprofit and Voluntary Sector Quarterly* 35 (June 2006): 288-310.
- Brunner, Eric and Jon Sonstelie. School Finance Reform and Voluntary Fiscal Federalism. *Journal of Public Economics* 87 (September 2003): 2157-2185.
- Chhaochharia, Vidhi and Suman Ghosh. Do Charity Ratings Matter? Working paper, University of Miami, February 2008.
- Dafny, Leemore and David Dranove. Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs. *RAND Journal of Economics* 39 (Autumn 2008): 790-821.
- DellaVigna, Stefano. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, forthcoming.
- Figlio, David. Testing, Crime and Punishment. *Journal of Public Economics* 90 (May 2006): 837-851.
- Figlio, David and Lawrence W. Kenny. Individual Teacher Incentives and Student Performance. *Journal of Public Economics* 91 (June 2007): 901-914.
- Figlio, David and Maurice Lucas. What's in a Grade? School Report Cards and the Housing Market. *American Economic Review* 94 (June 2004): 591-604.
- Gabaix, Xavier, David Laibson, Guillermo Moloche and Stephen Weinberg. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review* 96 (September 2006): 1043-1068.

- Hastings, Justine and Jeffrey Weinstein. Information, School Choice and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics* 123 (November 2008): 1373-1414.
- Jacob, Brian. Accountability, Incentives and Behavior: Evidence from School Reform in Chicago. *Journal of Public Economics* 89 (2005): 761-796.
- Jin, Ginger Zhe and Phillip Leslie. The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics* 118 (May 2003): 409-451.
- Jin, Ginger Zhe and Alex Whalley. The Power of Attention: Do Rankings Affect the Final Resources of Public Colleges? NBER working paper number 12941, February 2007.
- Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *Quarterly Journal of Economics* 121 (May 2006): 747-782.
- Neal, Derek and Diane Whitmore Schanzenbach. Left Behind by Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, forthcoming.
- Payne, John, James Bettman and Eric Johnson. *The Adaptive Decision Maker*. Cambridge, England: Cambridge University Press, 1993.
- Reinstein, David and Christopher Snyder. The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics. *Journal of Industrial Economics* 53 (March 2005): 27-51.
- Rouse, Cecilia, Jane Hannaway, Dan Goldhaber and David Figlio. Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure. NBER working paper 13681 (December 2007).

Sloan, Margaret F. The Effects of Nonprofit Accountability Ratings on Donor Behavior.

*Nonprofit and Voluntary Sector Quarterly* 38 (April 2009): 220-236.

Vesterlund, Lise. The Informational Value of Sequential Fundraising. *Journal of Public*

*Economics* 87 (March 2003): 627-657.

Table 1  
 Correlation between 2002 and 2001 School Grades  
 Elementary and Middle Schools

| 2001 School Grade | 2002 School Grade |          |          |          |          |      |
|-------------------|-------------------|----------|----------|----------|----------|------|
|                   | A                 | B        | C        | D        | F        |      |
| A                 | 370               | 129      | 45       | 0        | 0        | 544  |
| B                 | 257               | 93       | 46       | 0        | 0        | 396  |
| C                 | 203               | 244      | 371      | 48       | 5        | 871  |
| D                 | 6                 | 18       | 112      | 91       | 33       | 260  |
| F                 | <u>0</u>          | <u>0</u> | <u>0</u> | <u>0</u> | <u>0</u> | 0    |
|                   | 836               | 484      | 574      | 139      | 38       | 2071 |

Source: Authors' calculation from Florida Department of Education data.

Table 2  
 Comparison of Responding Primary and Middle Schools  
 With All Primary and Middle Schools

|                         | Responding<br>Schools<br><u>Mean</u><br>36649 | All<br>Schools<br><u>Mean</u> |
|-------------------------|---|-------------------------------|
| 2004 Donations          |   |                               |
| A in 2002               | 39.31*  | 38.19                         |
| B in 2002               | 22.95   | 22.37                         |
| C in 2002               | 26.02   | 26.01                         |
| D in 2002               | 5.71  | 6.54                          |
| F in 2002               | 1.69  | 1.87                          |
| A in 2001               | 26.51*  | 25.31                         |
| B in 2001               | 18.31   | 18.53                         |
| C in 2001               | 42.41   | 40.87                         |
| D in 2001               | 9.90*   | 12.08                         |
| % absent 21+ days       | 7.33  | 7.47                          |
| % gifted                | 5.02  | 4.91                          |
| % English lang learners | 7.20*   | 7.52                          |
| Student stability       | 93.30   | 92.80                         |
| % suspension in-school  | 5.36  | 5.27                          |
| %suspension out-school  | 5.73  | 5.65                          |
| Log # students          | 6.48  | 6.39                          |
| % free lunch            | 50.85   | 51.43                         |
| % black                 | 24.74*  | 26.91                         |
| % Asian                 | 1.75  | 1.67                          |
| % Hispanic              | 18.23   | 18.61                         |

Note: Means marked \* are statistically different from the all-schools mean at the five percent level.

Figure 1: Distributions of log donations, three rounds of surveys

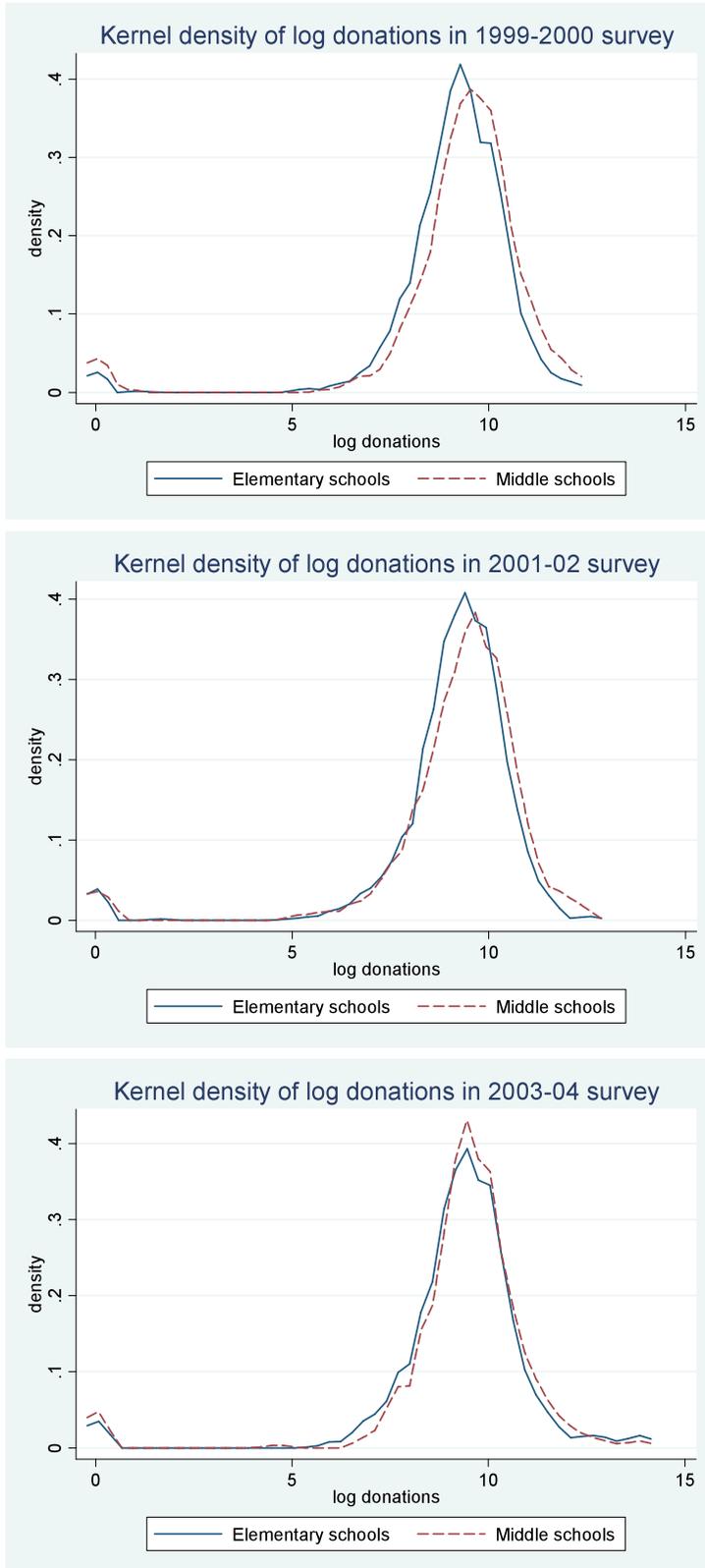


Figure 2: Kernel density estimates of pre/post donations of schools graded "C", "D" or "F" in 2002

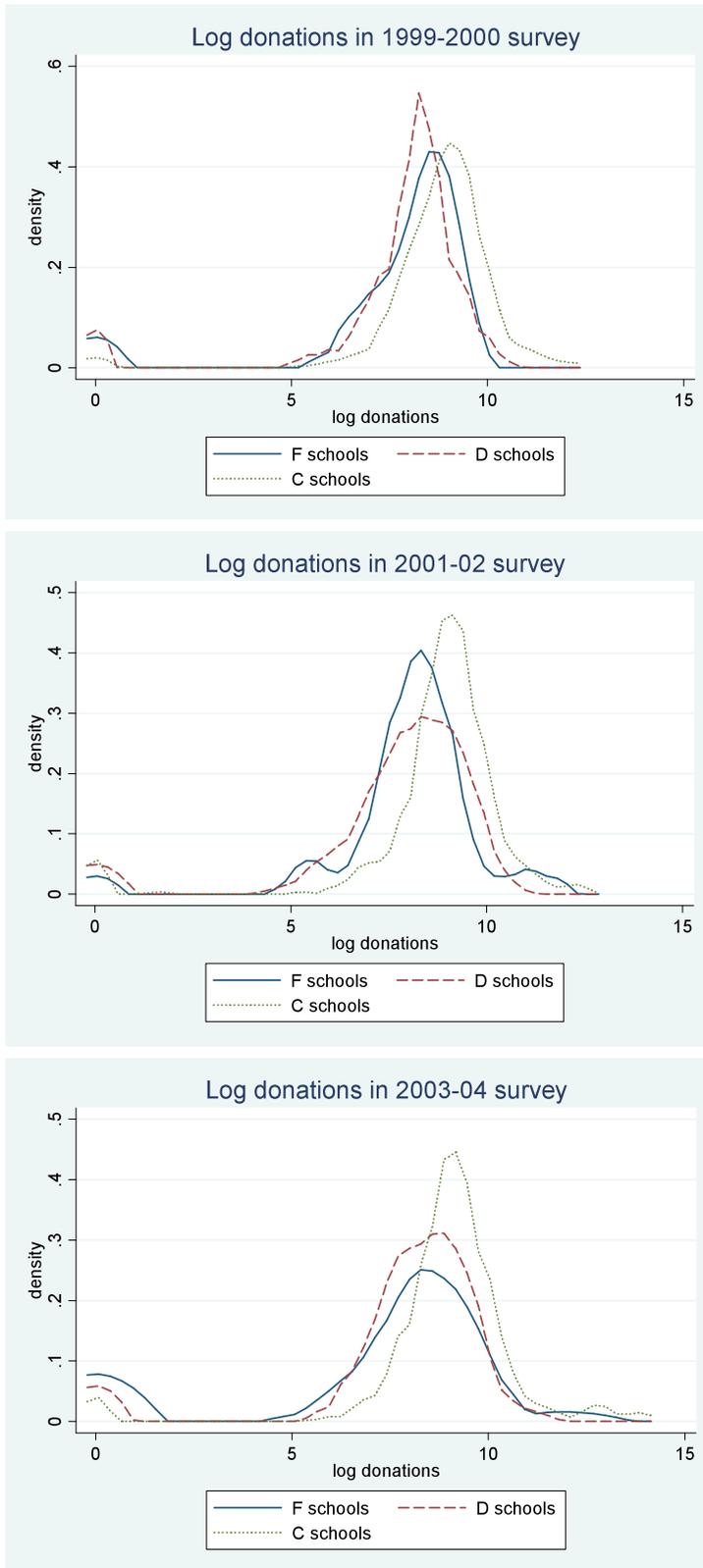


Figure 3: Change in log donations from 2001-02 survey to 2003-04 survey, by 2002 school grade

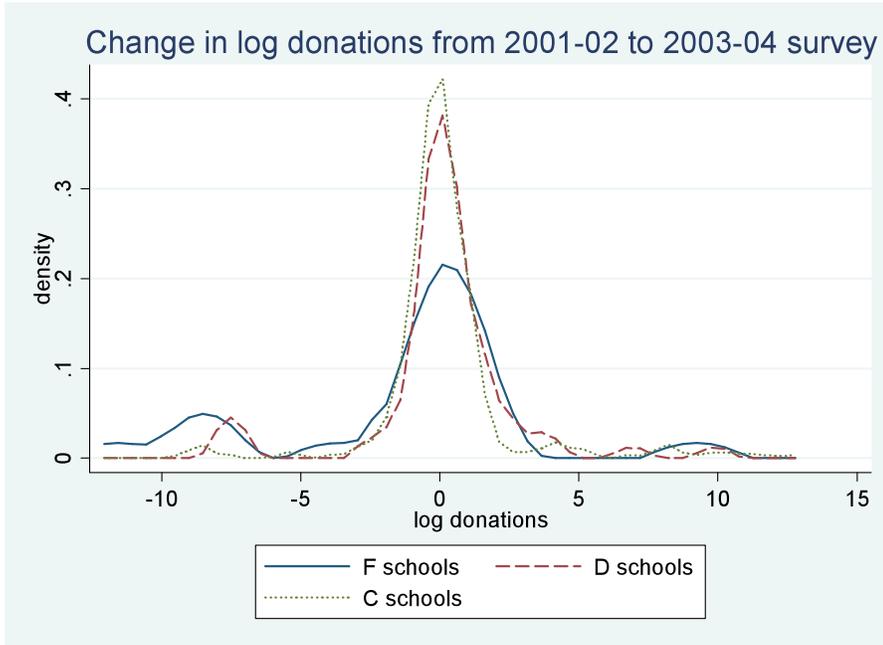


Table 3: Regressions Explaining Donations per School in 2004

(standard errors in parentheses)

|                         | Middle & Elementary |                     | Elementary          |                     |
|-------------------------|---------------------|---------------------|---------------------|---------------------|
|                         | Full Sample         | Top 1% Deleted      | Full Sample         | Top 1% Deleted      |
| A in 2002               | 0.397<br>(0.455)    | 0.504<br>(0.444)    | -0.195<br>(0.492)   | -0.071<br>(0.476)   |
| B in 2002               | 0.170<br>(0.267)    | 0.230<br>(0.261)    | -0.199<br>(0.290)   | -0.135<br>(0.281)   |
| D in 2002               | -0.679**<br>(0.306) | -0.666**<br>(0.298) | -0.574*<br>(0.312)  | -0.563*<br>(0.302)  |
| F in 2002               | -2.216**<br>(0.582) | -2.250**<br>(0.568) | -1.893**<br>(0.601) | -1.933**<br>(0.581) |
| No grade 2002           | -0.263<br>(0.192)   | -0.295<br>(0.187)   | 0.009<br>(0.210)    | -0.021<br>(0.203)   |
| Log 2002 donations      | 0.178**<br>(0.031)  | 0.176**<br>(0.031)  | 0.266**<br>(0.038)  | 0.266**<br>(0.036)  |
| Log # students          | 0.050<br>(0.109)    | 0.078<br>(0.107)    | -0.090<br>(0.148)   | -0.077<br>(0.146)   |
| Middle school           | -0.169<br>(0.194)   | -0.146<br>(0.190)   |                     |                     |
| % free lunch            | -0.018**<br>(0.004) | -0.017**<br>(0.004) | -0.017**<br>(0.004) | -0.015**<br>(0.004) |
| A in 2001               | 0.033<br>(0.138)    | 0.024<br>(0.135)    | 0.189<br>(0.152)    | 0.192<br>(0.148)    |
| B in 2001               | 0.117<br>(0.152)    | 0.109<br>(0.148)    | 0.187<br>(0.163)    | 0.201<br>(0.157)    |
| D in 2001               | -0.517**<br>(0.218) | -0.559**<br>(0.214) | -0.344<br>(0.220)   | -0.398*<br>(0.213)  |
| No grade 2001           | -0.841**<br>(0.310) | -0.824**<br>(0.302) | -0.237<br>(0.343)   | -0.204<br>(0.331)   |
| % absent 21+ days       | 0.043**<br>(0.016)  | 0.044**<br>(0.016)  | 0.064**<br>(0.022)  | 0.066**<br>(0.021)  |
| % gifted                | 0.009<br>(0.010)    | 0.004<br>(0.010)    | 0.016<br>(0.011)    | 0.010<br>(0.011)    |
| % English lang learners | 0.024**<br>(0.009)  | 0.021**<br>(0.009)  | 0.018*<br>(0.010)   | 0.013<br>(0.009)    |

Table 3 continued

|                        | Middle & Elementary |                     | Elementary          |                     |
|------------------------|---------------------|---------------------|---------------------|---------------------|
|                        | Full Sample         | Top 1% Deleted      | Full Sample         | Top 1% Deleted      |
| Student stability      | -0.009**<br>(0.002) | -0.003<br>(0.022)   | -0.010<br>(0.023)   | 0.000<br>(0.023)    |
| % suspension in-school | -0.001<br>(0.007)   | -0.002<br>(0.007)   | -0.015<br>(0.014)   | -0.017<br>(0.014)   |
| %suspension out-school | -0.021*<br>(0.012)  | -0.020*<br>(0.012)  | -0.055**<br>(0.023) | -0.052**<br>(0.022) |
| % black                | -0.009**<br>(0.003) | -0.009**<br>(0.003) | -0.007**<br>(0.003) | -0.005*<br>(0.003)  |
| % Asian                | 0.030<br>(0.028)    | 0.015<br>(0.028)    | 0.008<br>(0.029)    | -0.000<br>(0.028)   |
| % Hispanic             | -0.015**<br>(0.004) | -0.014**<br>(0.004) | -0.012**<br>(0.005) | -0.009*<br>(0.005)  |
| Constant               | 9.893**<br>(2.218)  | 9.091**<br>(2.170)  | 9.316**<br>(2.356)  | 8.200**<br>(2.287)  |
| # observations         | 1235                | 1222                | 916                 | 905                 |
| Adjusted R-squared     | 0.2184              | 0.2176              | 0.2855              | 0.2904              |
| Root MSE               | 1.694               | 1.651               | 1.571               | 1.516               |

\*Significant at 10% level under a two-tailed test.

\*\*Significant at 5% level under a two-tailed test.

Figure 4: Regression discontinuity estimates of the effects of "F" grades on residual log donations in 2003-04 survey, conditional on past grades and past donations and demographic/school variables

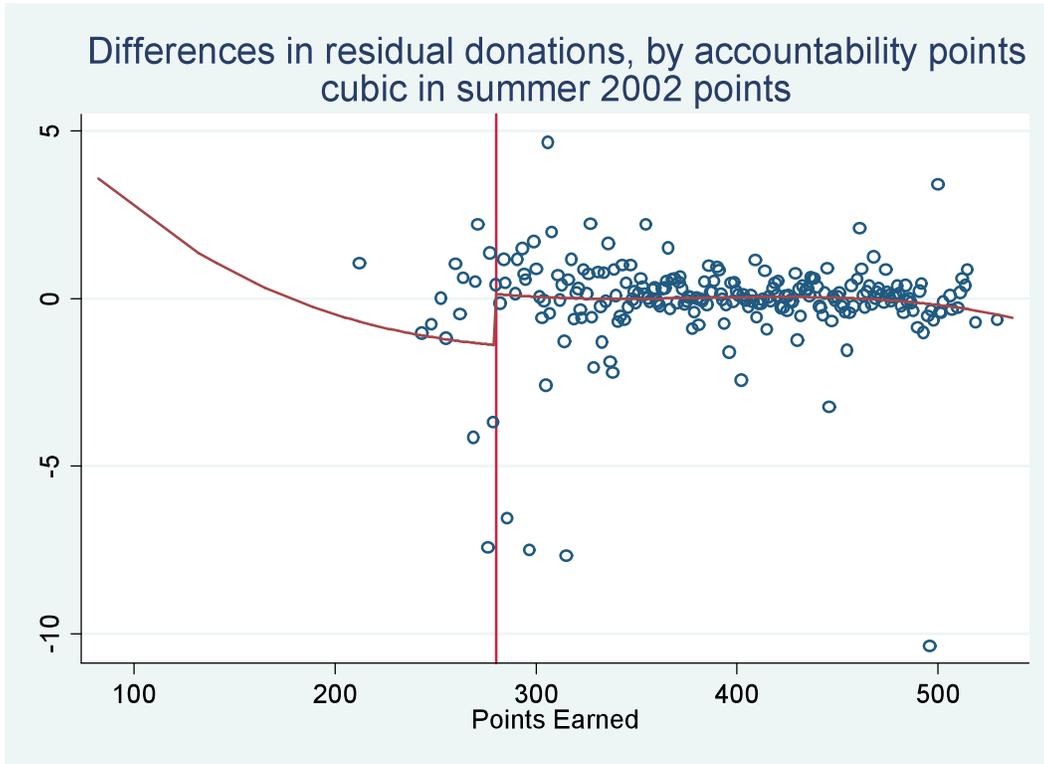


Table 4: Differential Effects of School Grades on Donations in 2003-04 Survey: Schools Stratified by School Attributes

(all estimates are relative to "C" grade in 2002; standard errors in parentheses; all covariates in Table 3 included)

|   | <u>"A" in 2002</u> | <u>"B" in 2002</u> | <u>"D" in 2002</u> | <u>"F" in 2002</u> |
|---|--------------------|--------------------|--------------------|--------------------|
| Schools with<br>% free lunch $\geq$ 83.5% | 1.388<br>(1.350)   | 0.926<br>(0.699)   | -1.753<br>(0.526)  | -3.709<br>(0.952)  |
| Schools with<br>% free lunch $<$ 83.5%    | 0.487<br>(0.467)   | 0.189<br>(0.275)   | 0.035<br>(0.363)   | -1.366<br>(0.738)  |
| p-value of difference                     | 0.508              | 0.303              | 0.003              | 0.034              |
| Schools with<br>% gifted $\geq$ 0.9%      | 0.107<br>(0.496)   | 0.034<br>(0.297)   | -0.338<br>(0.374)  | -0.195<br>(1.274)  |
| Schools with<br>% gifted $<$ 0.9%         | 1.223<br>(0.711)   | 0.491<br>(0.465)   | -1.278<br>(0.440)  | -3.119<br>(0.704)  |
| p-value of difference                     | 0.129              | 0.365              | 0.074              | 0.034              |
| Schools with<br>% minority $\geq$ 92.7%   | 2.496<br>(1.236)   | 1.683<br>(0.738)   | -1.204<br>(0.593)  | -2.955<br>(1.046)  |
| Schools with<br>% minority $<$ 92.7%      | 0.286<br>(0.464)   | 0.065<br>(0.275)   | -0.461<br>(0.355)  | -2.117<br>(0.861)  |
| p-value of difference                     | 0.073              | 0.032              | 0.256              | 0.509              |