2001

Does Charter School Attendance Improve Test Scores?: Comments and Reactions on the Arizona Achievement Study

Christopher Nelson
Western Michigan University

Kevin Hollenbeck
W.E. Upjohn Institute, hollenbeck@upjohn.org

Upjohn Institute Working Paper No. 01-70

“Does Charter School Attendance Improve Test Scores?”: Comments and Reactions on the Arizona Achievement Study

W.E. Upjohn Institute Staff Working Paper No. 01-70

Christopher Nelson
The Evaluation Center
Western Michigan University

and

Kevin Hollenbeck
W.E. Upjohn Institute for Employment Research

July 13, 2001

DRAFT: COMMENTS WELCOME

“Does Charter School Attendance Improve Test Scores?”: Comments and Reactions on the Arizona Achievement Study

ABSTRACT

In a recent report, Solmon, Paark, and Garcia (2001) seek to identify the impact of attending charter schools on student achievement using data from Arizona. Based on a sophisticated statistical analysis, these authors report that charter school attendance increases test score gains of students. This note raises some questions about the interpretation of the results reported and some questions about the empirical approach and underlying data. First, the report relies on a 2-x-2 evaluation design with type of school (charter or traditional) attended in a base year as the rows and type of school in the ensuing year as the columns. The report compares the observations in a cell of the design matrix to all other cells. This note questions the validity of that approach and suggests that the way that the data were constructed allows comparisons only across the rows. Second, the note questions whether grade level was used in the data matching procedure used to construct the comparison sample. Third, the note questions whether sex was used as a covariate in the outcomes equation and whether building or district fixed effects were used to control for unobservable factors at those aggregate levels. Finally, the note suggests that marginal costs are more appropriate for a cost-benefit or cost effectiveness analysis than average costs, which were used in the summary section of the report.

INTRODUCTION

In a recent report, Solmon, Paark, and Garcia (2001) (hereafter referred to as SPG) seek to identify the impact of attending charter schools on student achievement. Based on a sophisticated statistical analysis (rare in charter school research[1]), SPG report that charter school attendance increases test score gains of students. We applaud the rigorous approach that SPG have employed, but we believe that the empirical results presented do not warrant an unambiguous conclusion about superior results in charter schools. Furthermore, we are troubled by some data issues that are not addressed in the report. The purpose of our comments is to raise these issues in the spirit of collegial commentary, with the hope that our questions and concerns can be addressed and that we can achieve agreement about the impact of charter school attendance based on these data.

The next section briefly reviews the SPG methodological approach and the authors' key findings. That section is followed by a discussion of key questions and concerns about the data, methods, and reported findings. The concluding section summarizes our concerns.

THE ARIZONA STUDY

The SPG study aims at determining the impact of attending charter schools on student learning. The measures of student learning that the study uses are test scores on a norm-referenced examination, the SAT9. The study's sample includes student scores from the 1996–97, 1997–98, and 1998–99 academic years.[2] The dependent variable that SPG use in their analyses is the test score, but because observations are matched over time, they are able to use prior-year test scores as a control variable. By controlling for a prior-year test, they argue that they are modeling test score

---

[1]For exceptions see Eberts & Hollenbeck (2001) and Bettinger (1999).

[2]Since the tests are administered in the spring of each year, we refer to scores from the 1996–97 academic as "1997," scores from the 1997-98 academic year as "1998," and so on.

changes. Indeed, they note that "[t]he research question is whether or not differences in test score *changes* among students are related to the types of schools attended after other student characteristics are taken into account" (Solmon, Paark & Garcia, 2001: 11, emphasis added).

The sample of students used in the study seems to be a near-census of charter school students in Arizona supplemented by a "comparison group" sample of students from traditional public schools. SPG rely on a "blocking" design to select the latter sample.[3] We are not sure that we clearly understand the selection of the sample from the description in the report, but it appears as though the Arizona Department of Education matched all students who had 1997 and 1998 test scores and extracted all students who were enrolled in a charter school in either year. Then the "comparison group" was selected by matching each traditional school student who was in a traditional school in 1997 and a charter school in 1998 with any and all students who were in the same traditional school in 1997, in a traditional public school in 1998, and in the same reading and math quartiles.[4] Finally, 1999 test data were matched to the sample.

The models that SPG estimated were of two types. First, they estimated linear regressions with test scores as the dependent variable and with a prior test score included among the covariates. Second, they estimated panel data models with test scores as the dependent variable with fixed or random effects for individual students. In the panel data, models do not include prior test scores since

[3]We use the descriptive term "near-census" because presumably not all students in charter schools in 1997 have a matching test score in 1998. For one thing, the test was administered to students in grades 3–12 in the first year and 2–11 in the second. Furthermore, students may have moved out of state or been absent when the second test was administered. SPG report an 87 percent match rate on a statewide basis but do not report a rate for charter school students in 1997. We assume that it was approximately the same.

[4]Note that the study is inconsistent about whether grade level was used in the selection of the comparison group. The text on page 11 indicates that grade level was used, but footnote 6 on page 12 indicates that grade level was not used. As our comments below suggest, we believe that the footnote is correct, and this is more than just an esoteric point. It actually affects the results.

the fixed and random effect (dummy variables and variance components) allows the statistical analysis to examine changes over time for each student.

The key independent variable of interest in the models is the type of school attended (charter school or traditional public school). Simplifying greatly, the estimates derived from the models compare changes in mean test scores (year *t* minus year *t*–1) among charter school students against those in traditional public schools. Given the scaling of the variables, positive coefficient estimates imply that attending charter schools is advantageous to students.

SPG report several results. While readers are encouraged to consult the paper to see all of the results, the authors report that attendance is a charter school is positive, with relative gains increasing more the longer a student stays enrolled in a charter school. Moreover, the effect appears to be larger for reading than for mathematics. The authors summarize their main findings are as follows:

> For reading, … students enrolled in charter schools for two and three consecutive years have an advantage over students staying in the TPSs [traditional public schools] for the same periods of time. Students who enrolled in charter schools for two consecutive years show a 2.35–2.44 extra point advantage over students who stayed in TPSs for two consecutive years. Similarly, students in charter schools for three consecutive years show an additional 1.31 extra point advantage over students in TPSs [traditional public schools] for 3 consecutive years. … For mathematics, students in charter schools for two years show a slight advantage over TPS [traditional public schools] students who stayed for 2 consecutive years. However students in the charter sector for three consecutive years have insignificantly lower gain in math than the corresponding TPS [traditional public schools] students, on average" (Solmon, Paark & Garcia, 2001: 23).

QUESTIONS AND CONCERNS

If we consider charter schools as an educational intervention whose purpose is to improve the educational achievement of students, then the appropriate evaluation question is one of *net impact*:

how do the scores of charter school students compare with the scores those same students would have posted attending tradition public schools? Of course, the counterfactual is unobservable. In principle, the most rigorous way of attributing student learning impacts to charter schools would be through a random assignment experimental design.[5] The Arizona study uses the less rigorous approach of constructing a comparison group.

As we understand it, the study makes use of two quasi-experiments. The sample of traditional public school students was chosen to match the students who were in a traditional public school in 1997 and switched to a charter school in 1998. So for the years 1997 and 1998, the first quasi-experiment compares students who moved from a traditional public to a charter school (TC) to those who remained in a traditional school for two years (TT). There is little reason to expect that the TT group would be a good comparison group for the CC (those who were in a charter school both years) or CT (those who went from a charter to a traditional public school) groups.[6] However, the CC and CT groups might comprise the second quasi-experiment. If the "match rate" accomplished by the Arizona Department of Education was approximately the same for these two groups between the

[5]Inasmuch as many charter schools are oversubscribed and that most state laws require schools select students from their waiting lists at random, randomized experiments of the sort used to evaluate voucher programs ought to be possible, in principle. However, a number of logistical challenges must first be overcome. As voucher research has shown, there is often considerable attrition in control groups constructed using waiting lists (see e.g., Witte, 1997; Rouse, 1997). Moreover, field work conducted by the first author and his colleagues suggests that charter school waiting lists are often insufficient for the construction of a good randomized experiment. In many cases, such lists are well out of date or, in the most extreme cases, exist only in the minds of school administrators. Moreover, it is nearly impossible to assess whether students on the lists had subsequently enrolled in other charter schools or had been exposed to other educational reforms. In order to be convincing, any such analysis would have to include an audit of the waiting lists. While this is certainly possible, it would likely be very time consuming and costly. Finally, even assuming that researchers could overcome these logistical challenges, such a randomized experience would be, like many such experiments, of limited external validity. Indeed, the would be generalizable only to those students whose parents/guardians had attempted to enroll their children.

[6]Curiously, the authors clearly state in their discussion of sampling (p. 7) that the most appropriate comparison is between the TT and TC groups. Nonetheless, they appear to give equal weight to all comparisons when interpreting their regressions.

1997 and 1998 administrations of the SAT9, then the two groups are comparable.[7] Therefore, the effect of charter school attendance on student achievement is most appropriately determined by looking at these two differences (TC/TT and CC/CT). The results in the SPG table 4 (p. 12) show that students with the sequence TC clearly lag behind TT in both reading and math, with differences in reading ranging from –0.814 to –1.324 and differences in math ranging from –1.261 to –2.134 across various model specifications. That is, the results for the designed quasi-experiment shows a clear advantage to those students who stayed in traditional schools. Moreover, it appears as if there is little or no difference between CC and CT in reading, and that CC clearly lags behind CT in math. Table 4 estimates seem to clearly indicate that (some) traditional school attendance is advantageous to attendance in a charter school when we limit our purview to the strongest comparison groups.[8]

Another concern, given the well-known gender differences on reading and math tests, is that sex is not in the model or descriptive statistics. The discussion of the data (p. 5) indicates that sex is available as a covariate, yet it is not reported in any of the tables with descriptive statistics or in any of the model results. If it turns out that the sample of charter school students has a higher proportion of girls than the sample from the traditional public schools, then the charter school variables may be proxying for gender. If it were the case that charter schools have a higher proportion of girls, then we would expect charter school students to do better in reading and about the same or worse in math.

[7]A priori, we would probably expect the match rates to differ because the CT group involves a change in buildings. So the percentage of "matches" among students who stayed in charter schools for the two years probably exceeds the percentage of "matches" for those who changed school types. However, this also applies to the TC/TT comparison.

[8]As a more minor and technical matter, the authors fail to clearly describe how they derived claims that group differences were statistically discernible. We assume that the authors performed Wald tests of the null hypothesis that the difference between the coefficients in question were zero. Furthermore, the authors appear to have misinterpreted the interactions between time and the sector dummies in their first regression by ignoring both the fact that the coefficients on the interaction terms provide the differences from the base group and the fact that the standard errors on such terms must be adjusted to incorporate information from each of the two interacted variables (see, e.g., Aitken & West, 1985; Friedrich, 1982).

The descriptive statistics in tables 1, A-1, A-2, and A-3 indicate that there is a significant discrepancy between the charter school students and the traditional public school students in grade level. On average, students in the "comparison group" (traditional public schools) are a full grade level above those in charter schools. While the paper is not clear on this point (see footnote 3), this suggests to us that grade level was not used as a matching criterion. Why does this matter? Interestingly, grade level has a significant and large (in absolute value) negative effect on test scores, as reported in tables 4-6. Thus, if grade level had been controlled for in the matching process, then the traditional public school students would have had an even greater advantage in mean test scores over students in charter schools than is exhibited in the descriptive statistics tables, casting further doubt on SPG's conclusion that charter school attendance is associated with greater achievement gains. Furthermore, if the statistical relationship between grade level and test scores is nonlinear, then the coefficients on type of school may be picking up some of the grade-level effect.

SPG's paper clearly states that the test score means for traditional public school students are higher than the charter school students for all years for both tests. This raises the possibility that part of the difference in test score gains for the two types of schools may be explained by regression to the mean and by a relatively higher proportion of students "topping out" among the traditional school students.

One of the most salient advantages of the Arizona data is the fact that students are matched from year to year, and thus it is possible to analyze gains (or changes) in test scores. Generally, year-to-year changes are a better measure of learning than are test score levels.[9] But SPG forego the opportunity to analyze gains by using test score levels as their dependent variable (with a prior score

[9]See Kane and Staiger (2001) for some important limitations to change-score data.

as a control variable.) It would seem to us that a better specification would involve using the individual difference between the 1999–1998 test scores as the dependent variable and 1997 test score as a control variable for ability.[10]

Another question about model specification is the absence of building (and implicitly teacher) and district effects. Since SPG report that they are using administrative data, we wonder why they chose to ignore building-level or district-level variables in their specification. For example, charter school building enrollments tend to be much smaller than traditional school enrollments. Thus, the coefficients on charter school variables may be picking up building size effects. Similar arguments can be made about teacher characteristics such as experience or education. For that matter, the paper's concluding section discusses the importance of allocating resources appropriately, so it is curious that the authors have not included some measure of resources per pupil in the models. Building and district effects would be included in the error term of the ANCOVA (i.e., OLS regression) models. Thus, omission of these effects in the models jeopardizes the assumption of homoskedasticity. Such effects would be part of the fixed or random individual effects in the "dynamic" models presented by SPG, but would certainly be confounded with individual-specific effects such as family resources and parental educational backgrounds.

SPG discuss the possibility of selection bias in their analysis. They argue that because mean test score levels are lower for charter school students than traditional public school students, it is the case that charter schools are not cream skimming, but rather are "havens for students with special problems, returning former dropouts, and others 'referred' to them by TPSs" (Solmon, Paark &

[10]We assume that the authors have considered some fundamental issues about the dependent variable, the SAT9 score, but we think that the exposition would be improved if they were discussed in the paper. Such questions include, are scores equated across grades? Or do they include some sort of development metric? Are individual score levels comparable across time?

Garcia, 2001: 3). Lower test score levels are clearly not sufficient evidence to conclude an absence of positive selection. Even though test score levels are lower for charter school students than traditional public school students, the former may still systematically differ from the latter in unobserved ways. The fact that students and parents have chosen to attend charter schools suggests that the parents are interventionists: they have taken the initiative to transfer their students in the belief that they will get a better education. If such parent initiative is positively related to test scores, then the coefficient on charter schools will be biased upward.

Our final concern involves the "back of the envelope" cost effectiveness discussion in the conclusion of the paper. The study cites the average annual cost of educating a traditional public school students as $7000, and the average annual cost of educating a charter school student as $4500. Based on the analysis of test scores, SPG conclude that more learning is occurring in charter schools and that, therefore, charter schools are more cost effective and deserving of public support. These economists should surely know that cost effectiveness involves comparison of *marginal* costs, and that it is an extremely unlikely that the average costs in these two types of schools equal marginal costs. There are many fixed costs in traditional public schools that are not borne by charter schools, or may be borne by some, but not all, charter schools. These might include transportation, extracurricular programs, fine arts curricula, laboratory courses in science and career and technical education curricula, foreign language instruction, and so on. On the other hand, there may be facilities costs that charter schools must bear that traditional public schools can relegate to capital budgets. In short, without good, disaggregated cost data, any sort of cost-benefit analysis is fraught with uncertainty. We do not know what the cost components are of Arizona's educational systems, but it could well be the case that instead of charter schools having a $2500 cost advantage and learning gains, there is a \$200–\$500 (marginal) cost advantage with learning *dis*advantages (particularly, if we could factor in the selectivity bias).

SUMMARY AND CONCLUSIONS

The SPG study is one of the most sophisticated analyses of the impact of charter school attendance on student achievement to date. Having annual test score data and a rich set of covariates at the student level clearly allows for rigorous analysis. While we applaud the authors' attempt to bring such rigor to the study of student achievement in charter schools, we are not yet prepared to accept the conclusion that charter school attendance is associated with increased levels of student achievement. In particular, we believe that the construction of the comparison sample allows only a comparison of the TC and TT groups. (The sample construction would seem to allow also a comparison of the CC and CT groups). When the analysis is restricted to these groups, the authors' own reported findings suggest that charter school students' test scores and gains lag behind traditional school students. Furthermore, we question the robustness of the results to specification changes such as inclusion of gender, matching on grade level in the construction of the data set, and inclusion of building-level and district-level effects. Finally, we believe that policymakers will be ill-served by cost effectiveness analyses that rely on average cost rather than marginal costs. Clearly, in making a decision about investing a marginal dollar in order to raise student achievement in reading or mathematics, policymakers need to know the marginal return to dollars spent on those functions now. We hope that the authors will continue to refine and explicate their analysis so that we can achieve agreement about the impact of charter school attendance on student achievement.

REFERENCES

Aitken, L & West, S. (1985). *Multiple regression: Testing and interpreting interactions.* Beverly Hills, CA: Sage.

Bettinger, E. (1999). *The Effect of Charter Schools on Charter Students and Public Schools*. National Center for the Study of Privatization in Education, Occasional Paper No. 4.

Eberts, R. & Hollenbeck, K. (2001). *An Examination of Student Achievement in Michigan Charter Schools*. Kalamazoo, MI: Upjohn Institute for Employment Research.

Friedrich, R. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science* 26: 797-833.

Kane, T. & Staiger, D. (2001). "Improving school accountability measures." National Bureau of Economic Research Working Paper No. 8156, Washington DC.

Rouse, C.E. (1997). "Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program." National Bureau of Economic Research Working Paper No. 5964, Washington, DC.

Solmon, L., Paark, K. & Garcia, D. (2001). *Does charter school attendance improve test scores?: The Arizona results*. Phoenix, AZ: Goldwater Institute.

Witte, J. (1997). "Achievement effects of the Milwaukee Voucher Program." Paper delivered at the annual meeting of the American Economics Association. New Orleans, LA.