



DOCUMENT REVIEWED:	“The Effect of Milwaukee’s Parental Choice Program on Student Achievement in Milwaukee Public Schools”
AUTHORS:	Jay P. Greene & Ryan H. Marsh
PUBLISHER/THINK TANK:	School Choice Demonstration Project; Department of Education Reform, University of Arkansas
DOCUMENT RELEASE DATE:	March 2009
REVIEW DATE:	May 18, 2009
REVIEWER:	Gregory Camilli
E-MAIL ADDRESS:	<a href="mailto:camilli@rutgers.edu">camilli@rutgers.edu</a>
PHONE NUMBER:	(732) 932-7496

### **Summary of Review**

According to a new study of Milwaukee public schools, student achievement has benefited from voucher-based school competition. A novel method, using geocoding, was proposed for measuring the degree of competition within the city of Milwaukee and, in turn, for determining whether such competition has increased or decreased the achievement of public school students. Though a more traditional measurement of competition was eventually used in lieu of geocoding, the authors of the study determined that the overall effect of competition on student outcomes was positive over the seven-year span for which data were available. Specifically, it was argued that increased school choice improves the academic performance of students in traditional public schools who are voucher eligible by means of system-wide competitive pressures. Based on a review of several key issues—including statistical modeling and control, effect size interpretation, the role of explanation in causal inference, and the validity of reported conclusions—the practical effect of competition through vouchers appears to be small, if not negligible. It is also suggested that a number of methodological issues would benefit from greater clarity.

## Review

### **I. INTRODUCTION**

The School Choice Demonstration Project, a research center self-described as “devoted to the non-partisan study of the effects of school choice policy,” recently released *The Effect of Milwaukee’s Parent Choice Program [MPCP] on Student Achievement in Milwaukee Public Schools*, a study by Jay P. Greene and Ryan H. Marsh.<sup>1</sup> This report (referred to hereafter as EMPCP) describes the design and execution of a statistical analysis of student and school information obtained from 1999 to 2006. A novel method, using geocoding, is offered for measuring competition within the city of Milwaukee and for determining whether there is an association between this measure of competition and public school student achievement.<sup>2</sup>

Based on their analyses, Greene and Marsh conclude that “students in Milwaukee fare better academically when they have more free private options through the voucher program.” They further surmise that “the existence over the last two decades of MPCP” may have improved student scores substantially,” with a mild warning against extrapolating the findings beyond the 1999-2006 period for which data were collected.

### **II. FINDINGS AND CONCLUSIONS OF THE REPORT**

Notwithstanding the concerns described later in this review regarding methods and interpretation, the report is to be commended for a thorough and fair presentation of results. Reports such as this should make public the methods and statistical models used in a study and should link empirical results and interpretations in an explicit manner, and this is done well here.

The main statistical results are given in the report’s Table A, in the section labeled “No Controls” under the column labeled “Whole City.” This data set contains assessment information merged from two achievement tests (Wisconsin Knowledge Concepts Examination and the Terra Nova Achievement Examination). Importantly, the records in the analytic sample are for public school, not voucher school, students. The study therefore focuses on the potential benefits of competition for public school students only.

In Table A, statistically significant and positive regression coefficients (RC, for short) were obtained for Language Arts, Mathematics, and Reading. To help readers understand this somewhat arcane information, the authors provide a more accessible version of quantitative results in Table F, in terms of *effect size*.<sup>3</sup> In Language Arts, for example, an effect size of  $ES = .10$  is shown in the first row of Table F. This is intended to represent the effect of voucher competition over a seven-year period (the span of the data collection); the report does not annualize the reported effect sizes, a point given more attention below.

In the first four columns of Table A, results are provided for localized competition (using the geocoded variables)—especially the column indicating competition within about 1 mile of the public school. No statistically significant results are found for the model that controls for grade and year.

### **III. THE REPORT’S USE OF RESEARCH LITERATURE**

In the section “Prior Research,” a number of studies are cited that provide a helpful perspective on what is currently known regard-

ing the effects of school competition on student achievement. A number of studies having conflicting results (e.g., Carnoy *et al.* and Hoxby<sup>4</sup>) are briefly summarized, and the authors observe that evidence from studies conducted in Florida “gives us some reason to believe that the threat of voucher competition spurs significant improvement in traditional public schools” (p. 4).

The report includes the following quote from a comprehensive review of the literature by Belfield and Levin, who synthesized 41 empirical studies of competition, including studies of school achievement due to vouchers, charter schools, or other public schools:

A sizable majority [of the 41 studies] report beneficial effects of competition, and many report statistically significant correlations. For each study, the effect size of an increase of competition by one standard deviation [SD] is reported. The positive gains from competition are modest with respect to realistic changes in levels of competition. The review also notes several methodological challenges and recommends caution in reasoning from point estimates to public policy.<sup>5</sup>

The information in this quote suggests the following basic method for determining the effect of competition. First, an independent (or predictor) variable is required that quantifies competition. Belfield and Levin provided a brief review of the literature on *measuring* the degree of competition. Second, an outcome variable is needed, and this is typically student achievement. Finally, a statistical model is used to link independent and outcome variables, thereby obtaining quantitative measures of impact.

Given the novel method of geocoding em-

ployed in EMPCP (as explained below), some conceptual connection with the previous literature on measuring competition would have been helpful.<sup>6</sup> It also would have been helpful if the quantitative findings of Belfield and Levin for student outcomes had been mentioned; roughly speaking, the latter found an average benefit of  $ES = .10$ . This provides some context for interpreting the size of competition effects, though I could not determine whether Belfield and Levin computed annualized competition gains (I suspect not). Because different competition studies have different time horizons, it is difficult to compare benefits without some kind of time standardization.

#### IV. REVIEW OF THE REPORT'S METHODS

There are two important aspects of the method employed in EMPCP to obtain statistical estimates: (1) how analytic variables are constructed, and (2) model specification, in the sense of what variables should, and should not, be in the model.

##### Constructing Variables

Although a novel method involving geocoding was used to construct analytic variables, the key positive findings ultimately set forth in the report apply to the city as a whole (all of MPS), rather than to schools within smaller, competitive geographic zones. The geocoding was used because of an initial belief that competition effects would be felt more strongly by a public school when voucher-accepting private schools are located nearby. The empirical analyses, however, did not confirm this conjecture.

In the end, the main independent variable was the interaction term  $FRL*N$ , where  $FRL$  is the free-lunch status of students (representing a proxy for eligibility for a

voucher). The variable  $N$  is the total number of voucher-accepting schools in Milwaukee that served the grade level of the student. This interaction ( $FRL*N$ ) can then be described as measuring the combined influences of voucher-eligibility and system-wide competition of schools (rather than competition resulting from nearby schools), or as stated in the report (p. 1), “effects on the whole school system, not its effects on participating students.” Note that the coefficient estimated for  $FRL*N$  measures a competition effect limited to public school students who are voucher-eligible.

### Statistical Model Specification

A simple OLS (ordinary least squares) regression model was run, incorporating voucher, control, and student fixed effects. In addition to the no-controls model, the report presents models using “year,” “grade” and “year and grade” as controls. These controls essentially account for whether a given private school served the grade level of a given public school student and whether it actually accepted vouchers during the specific year observed. That is, these controls move the analysis from “there’s a private school that accepted vouchers at some point” to “there’s a private school that accepted vouchers for this student’s grade level in the year that we’re analyzing.”

Coefficients for the combined test-score data are given in the report’s Table A.<sup>7</sup> Greene and Marsh choose to interpret the analysis employing no controls, though the controlled coefficients are given in the bottom segment of the table. The issue of control is important, in part because—as explained later—the controls greatly reduce effect size (and statistical significance), and in part because it is not clear why  $FRL$  by itself was omitted from the analysis as a control variable (or covariate) at the student level.

Though not without construct validity issues,  $FRL$  is commonly used as an indicator of opportunity to learn (OTL), given that students eligible for free lunch generally have less access to educational resources.

Keep in mind that  $FRL$  is currently included in the model, but only as an interaction with  $N$ . Though it would seem plausible to include both  $FRL$  and  $FRL$  interactions in the same statistical model, this action would likely result in estimation problems that could be metaphorically explained as having two star quarterbacks on the same team: team performance may become erratic, and consequently it may be hard for the coach to judge the merits of either quarterback.<sup>8</sup> There is also the conceptual problem that  $FRL$  is required to play offense (as a measure of voucher eligibility) and defense (to control for OTL). Consequently, the interaction  $FRL*N$  could also be interpreted as a positive effect of voucher options for students with lower OTL. This is not necessarily the same thing as competition.

The statistical modeling aspect of this report could be more fully developed with a multilevel approach in which students are nested within schools. Multilevel modeling techniques are useful and often statistically necessary when observations have this structure.<sup>9</sup> From this perspective, the voucher-options variable is an interaction between the salience of competition at the student level and the availability of vouchers at the school level. It would be difficult, in my opinion, to construct a convincing argument for deleting  $FRL$  prior to examining empirical results.

In sum, the model specification in EMPCP is ultimately applied without including controls and does not include  $FRL$  as a main effect. The multilevel approach might lead to a more convincing account of model spe-

cification.<sup>10</sup> Standard errors in the multilevel framework are also likely to be more accurate than those obtained with OLS given nested educational data.<sup>11</sup>

## V. REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS

The report highlights uncontrolled results, and annualized effect sizes are not calculated. In general, the controlled coefficients (only for grade and year, since *FRL* was not included) in Table A are about one fifth the size of the uncontrolled coefficients. Translating this factor into Table F for the combined sample would result in effect sizes of .020, .016, and .020 for Language Arts, Mathematics, and Reading, respectively. If these gains were annualized over the seven years (the span of the data collection), it is clear they are very modest in absolute size.

This observation should not be read as an attempt to minimize the importance of the statistically significant results reported in this study. Even small gains with crude measures of competition may signal more powerful underlying effects. However, it is critical in any policy research to distinguish clearly between statistically significant estimates and practically significant results.

The report focuses on system-wide competition (under the Table A column “Whole City”). According to Greene and Marsh (p. 7):

It may be difficult to visualize exactly how MPS more effectively serves students with more voucher options given the difficulty of school officials to make fine distinctions between what is provided to each student. But school officials do not have to make fine distinctions for them to be more attentive to students with more options.

The authors appear to believe that if positive macro-level effects can be estimated reliably, then explanation or interpretation of the causal mechanism is secondary. A great deal of faith is thereby required in standard linear regression without controls as well as substantial forbearance for the puzzling finding that competition effects were unrelated to voucher-school proximity. The downside to the macro approach to evidence-based reasoning is the potential frailty of the assumptions of model specification and of construct validity for key independent variables. However, the suitability of these assumptions is most appropriately a topic for future studies, because the competition effects estimated for Milwaukee public school students are relatively small, and a good deal more robustness checking is required prior to interpreting the importance of the competition estimates with confidence.

Properly presented and understood, and setting aside the specification concerns raised earlier, the results in EMPCP are consistent with the conclusion of Belfield and Levin (p. 297): “the effects of competition on educational outcomes appear to be substantially modest.” The policy implications are thus appropriately refocused on whether there are (1) better programmatic competitors than voucher systems for increasing student achievement, and (2) whether there are unintended consequences requiring evaluation. For example, in a recent policy brief, Arsen & Ni wrote:

The research surveyed here suggests, rather than conclusively establishes, that competition from vouchers and charter schools is no more beneficial for TPS [traditional public school] performance than competition from nearby private or public schools in environments with no choice policy. Indeed, Belfield and Levin’s review

of studies of these traditional forms of school choice shows a higher proportion of findings indicating statistically significant positive effects on TPS outcomes than is evident among existing studies of voucher and charter school competition.<sup>12</sup>

Obviously, the issue of whether competition is most effectively implemented with voucher-versus non-voucher schools is an important question that involves economic analysis and, more importantly, careful consideration of alternative value systems regarding public education and long-term consequences.

#### **VII. USEFULNESS OF THE REPORT FOR GUIDANCE OF POLICY AND PRACTICE**

Before the results of this study can effectively inform the decision of whether to expand or reduce the size of the voucher program in Milwaukee, a number of issues need resolution. Most importantly, the use of uncontrolled estimates requires justification. After all, Greene and Marsh argue that statistical control is important: “we believe controlling for year effectively ‘de-trends’ test results without comparable loss of information” (p. 8). Why then, were uncontrolled estimates preferred?

Second, a more comprehensive approach to statistical modeling would allow important

distinctions between student and macro-level effects. Third, an annualized metric should be used to compare student achievement gains spurred by competition with gains from other efficient educational interventions. If the authors would counter that it is not important to report annualized gains, then a well-reasoned argument should be offered. This would clarify how the effects of competition can be compared with other potential interventions (or competition studies in other communities) for increasing student achievement.

Greene and Marsh (p. 10) may have anticipated such reactions with the caveat “No single study of a single program can be definitive...” However, the ensuing sentence in the closing paragraph of EMPCP is problematic: “but this research contributes to the finding that expanded choice and competition improve the academic performance of students who remain in traditional public schools.” I remain cautious, if not a bit skeptical, of the latter claim, and I would await further rationale and corroborating evidence along the lines suggested in this review.

To spur further research, it would be most helpful if the authors would make their data set freely available for secondary analyses. This is a great opportunity to challenge other researchers to explore the data from different perspectives, potentially leading to more robust estimates and interpretations of the effects of school competition.

## Notes and References

- <sup>1</sup> Greene, J.P. & Marsh, R.H. (2009, March). *The Effect of Milwaukee's Parent Choice Program [MPCP] on Student Achievement in Milwaukee Public Schools*. University of Arkansas, Department for Education Reform. Retrieved May 14, 2009, from [http://www.uark.edu/ua/der/SCDP/Milwaukee\\_Eval/Report\\_11.pdf](http://www.uark.edu/ua/der/SCDP/Milwaukee_Eval/Report_11.pdf)
- <sup>2</sup> Quotations in this paragraph are taken from the report's initial acknowledgements.
- <sup>3</sup> *Effect size* is now ubiquitous in educational research as a quantitative measure of outcome that can be compared or aggregated across studies, even if those studies use different outcome measures. In contrast to indices of statistical significance, effect sizes are often described as measures of practical significance. However, different research areas have different methods for calculating effect size and determining what constitutes a small, medium, or large value. These values reported in EMPCP were obtained by the equation
- $$\text{Effect size} = SD_n * RC / SD_{NCE},$$
- with the explanation, for example, that a 37-school (one standard deviation) increase in the number of voucher-accepting schools, multiplied by the regression coefficient for Language Arts ( $RC = .055$ ), and then divided by the standard deviation of the NCE (Norm Curve Equivalent) scale ( $SD_{NCE} = 21.06$ ) results in an effect size of .10 as shown in the first row of Table F. Greene and Marsh make the important qualification that the effect size computation is based in the assumption that a policy can result in a one standard deviation change in the number of competitive schools.
- <sup>4</sup> Martin Carnoy, et al., "Vouchers and Public School Performance," Economic Policy Institute, October 2007; Caroline Minter Hoxby, "The Rising Tide," *Education Next*, Winter, 2001.
- <sup>5</sup> Belfield, C.R., & Levin, H.M. (2002). The effects of competition on educational outcomes: A review of the U.S. evidence. *Review of Educational Research*, 72(2), 279-341, p. 279.
- <sup>6</sup> In terms of the report's audience, this information (e.g., the Herfindahl index) would certainly be no less complex than the description of the geocoded measure on pp. 5-6 of the report.
- <sup>7</sup> A number of geocoded variables were constructed by use of the voucher student data base, but interpreted results from the last column of Table A of the report did not make use of these variables. I assumed that the columns of Tables A-E were different regression models using alternative competition proxy variables; however, the report does not indicate this explicitly.
- <sup>8</sup> This is known as the problem of multicollinearity, in which correlation among independent variables results in higher standard errors and, in turn, lower levels of significance.
- <sup>9</sup> Namely, the structure of the data set used in EMPCP because the geocoded voucher variables are measured at the school-level. That is, all *FRL* students within a public school have the same values of the competition or voucher variables. Let the subscript  $i$  signify schools and  $j$  students within schools. The student level model can then be written as

$$(1) \quad y_{ij} = b_{0i} + b_{1i}FRL + b_2Z + r_{ij},$$

where for school  $i$ ,  $b_{0i}$  is the model intercept and  $b_{1i}$  is the slope of *FRL*. The  $i$  subscript denotes that these coefficients may vary across schools. For simplicity, let  $b_2$  be the fixed slopes  $Z$  for other potential covariates at the student level, and  $r_{ij}$  be the student-level residual. At the school level, two potential equations are

$$(2) \quad b_{0i} = b_0 + \gamma_{01}Voucher + v_{0i}, \text{ and}$$

$$(3) \quad b_{1i} = b_1 + \gamma_{11}Voucher + v_{1i}.$$

Substituting Equation 2 and 3 into 1 gives the reduced form:

$$(4) \quad y_{ij} = (b_0 + \gamma_{01}Voucher_j + v_{0i}) + (b_1 + \gamma_{11}Voucher_j + v_{1i})FRL_{ij} + b_2Z_{ij} + r_{ij} \\ = (b_0 + b_1FRL_{ij} + b_2Z_{ij}) + (\gamma_{01}Voucher_j + \gamma_{11}Voucher_j * FRL_{ij}) + \\ (v_{0i} + v_{1i}FRL_{ij} + r_{ij}).$$

It can then be seen that  $\gamma_{11}$ , which is the voucher-options variable in the EMPCP report, is a cross-level interaction in the framework of a multilevel model. In contrast, the coefficient  $\gamma_{01}$  would measure of the global effects of competition at the school level.

<sup>10</sup> The multilevel approach to analysis would appear to satisfy the methodological concern of Belfield and Levin (p. 293) with regard to studies of school competition:

Although many of the studies control for covariates [i.e., year or grade], there is still a possibility--at this aggregated level--the models are inappropriately specified [i.e., multilevel structures are ignored]. However, misspecification bias may inflate or deflate the point estimate on the measure of competition; there are also no strong theoretical grounds for the inclusion of particular covariates.

<sup>11</sup> Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications. Future researchers might also want to take spatial autocorrelation into account (given geocoded data), because schools that are closer in proximity may have more similar student outcomes.

<sup>12</sup> Arsen, D. & Ni, Y. (2008). *The Competitive Effect of School Choice Policies on Performance in Traditional Public Schools*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit, p. 15. Retrieved May 8, 2009, from <http://epicpolicy.org/files/CHOICE-09-Arsen2.pdf>.

---

The Think Tank Review Project is made possible by funding from the Great Lakes Center for Education Research and Practice.