



NEPC

NATIONAL EDUCATION
POLICY CENTER

NEPC REVIEW: *APPLES TO APPLES: THE DEFINITIVE LOOK AT SCHOOL TEST SCORES IN MILWAUKEE AND WISCONSIN*

Reviewed By:

Benjamin Shear
University of Colorado Boulder

April 2017

Summary of Review

The report reviewed here compares Wisconsin student test score performance for the 2015-16 school year across public schools, charter schools and private schools participating in one of the state's voucher programs. Comparing a single year's test scores across school sectors that serve different student populations is inherently problematic. The report uses linear regression models to attempt to adjust for these differences and make what the authors claim are "apples to apples" comparisons. Based on these comparisons, the report concludes that charter schools and private schools participating in the voucher programs are more effective than traditional public schools. Unfortunately, the limited nature of available data undermines any such causal conclusions. The inadequate and small number of school-level variables included in the regression models are not able to control for important confounding variables, most notably prior student achievement. Further, the use of aggregate percent proficient metrics masks variation in performance across grade levels and makes the results sensitive to the (arbitrary) location of the proficiency cut scores. The report's description of methods and results also includes some troubling inconsistencies. Thus, while the report does present important descriptive statistics about test score performance in Wisconsin, it cannot provide answers for those interested in determining which schools or school choice policies in Wisconsin are most effective.



Kevin Welner
NEPC Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

This review is one of a series made possible in part by funding from the Great Lakes Center for Education Research and Practice.



<http://www.greatlakescenter.org>
GreatLakesCenter@greatlakescenter.org

NEPC REVIEW: *APPLES TO APPLES: THE DEFINITIVE LOOK AT SCHOOL TEST SCORES IN MILWAUKEE AND WISCONSIN*

Benjamin Shear, University of Colorado Boulder

I. Introduction

The recent emphasis on voucher and choice programs by the federal government¹ brings new attention to the topic and makes a new empirical study of Wisconsin's choice programs especially timely. Wisconsin currently has three publicly-funded voucher (or "choice") programs: the Milwaukee Parental Choice Program (MPCP), the Racine Parental Choice Program (RPCP), and the Wisconsin Parental Choice Program (WPCP). The MPCP, begun in 1990, is the longest-running school choice program in the country. The programs are open to families with income below set cutoffs ranging from 185% to 300% of the federal poverty level. Eligible families may enroll their child in participating private schools (including religiously affiliated schools) and the state pays up to \$7,860 (for the 2015-16 school year) directly to the school on the family's behalf. Schools may only charge additional tuition for high school students whose family income exceeds a set limit. In the 2015-16 school year, there were approximately 27,000 students enrolled in the MPCP (across 117 schools), 2,000 enrolled in the RPCP (across 19 schools) and 2,500 enrolled in the WPCP (across 82 schools).² In 2015-16 there were also 242 public charter schools operating in Wisconsin that enrolled approximately 44,300 students.³

The state of Wisconsin enrolled approximately 870,000 public school students in 2015-16 (including charter school students); thus about 5% of Wisconsin students were enrolled in charter schools while a portion equivalent to about 3.5% were enrolled in private schools through one of three choice programs.

In 2012, Wisconsin began creating district and school report cards for all public schools (including charter schools) as part of its school accountability system. Beginning in the 2015-16 school year, Wisconsin also generated report cards for private schools participating in the three choice programs, albeit with more limited data. The report under review here, titled *Apples to Apples: The Definitive Look at School Test Scores in Milwaukee and Wisconsin*, authored by Will Flanders, and published by the Wisconsin Institute for Law & Liberty, uses these accountability data to make test score comparisons across these three school sectors.

II. Findings and Conclusions of the Report

The main conclusion of the report is summarized in the following excerpt from the report's executive summary:

“We find that private schools in the choice programs and public charter schools in Milwaukee and Wisconsin perform significantly better on the ACT and Forward Exams than traditional public schools when a proper apples-to-apples comparison is made” (bold in original)⁶

The report does not state precisely what is meant by an “apples to apples” comparison. However, the implication is that the resulting comparisons can be interpreted in a causal manner. This is implied by the next paragraph stating that, “we have to make the best use of the available data to provide parents with accurate information about what is working and what isn't.” Stating that this report indicates “what is working and what isn't” infers that the differences between school sectors can be interpreted as the causal effect of attending a school in one sector relative to the others. This interpretation is further supported by use of phrases such as “Effect of School Type on Forward Exam Performance,” elsewhere in the report.

The school-level test score data analyzed in the report are based on the 2015-16 Wisconsin state mathematics and English/Language Arts (ELA) exams administered in 3rd through 8th grades (the “Forward Exam”) and the ACT test, which all 11th graders in Wisconsin are required to take. The authors do not state exactly how the various grade level and ACT test scores were aggregated to the school level. All comparisons in the report summarized below adjust for a small number of school-level demographic characteristics.

Milwaukee Results

First, the report compares test scores by sector (public, private, or charter) in Milwaukee. The report finds that MPCP private schools and charter schools in Milwaukee have higher average proficiency rates than traditional public schools, while 11th graders in MPCP private schools and district operated non-instrumentality charter schools have higher average ACT scores than students in traditional public schools. The differences in proficiency rates range from 4% to almost 10%, while differences in average ACT scores range from 2.8 to 4.5 points.⁸ The report also highlights the exceptionally strong correlation between the percent of non-White students in a school and the average school proficiency rates.

Milwaukee School Sub-Types

Next, the report compares test score performance across different sub-types of schools within sectors in Milwaukee. The report finds that Catholic and Lutheran MPCP schools, district-operated non-instrumentality⁹ charter schools and independent charter schools authorized by the University of Wisconsin-Milwaukee (UW-M) have higher mathematics

proficiency rates than traditional public schools. Catholic, non-instrumentality and UW-M charters also have higher ELA proficiency rates than traditional public schools. The report finds no evidence of differences among non-charter public schools, once the school-level covariates are taken into account.

Non-Milwaukee School Comparisons

Finally, the report finds no statistically significant differences among public, private and charter schools outside of Milwaukee for mathematics and ELA proficiency rates. The report does find that private and charter schools have higher average ACT test scores relative to public schools. The report ends with a finding that public schools in urban, rural and town jurisdictions have lower proficiency rates in mathematics and ELA than schools in suburban jurisdictions.

III. The Report's Rationale for Its Findings and Conclusions

The primary rationale is that a comparison of school-level test score performance from a single year, adjusted for school demographics, is a valid indicator of the relative quality of schools across sectors. Specifically, the findings are based on statistical analyses of publicly available school-level test score data that control for a small number of school demographic characteristics using linear regression. The data come from various publicly available sources on the Wisconsin Department of Public Instruction (DPI) websites, although specific data sources and information that would be needed to replicate the analyses are not included. The regression models control for differences among the schools with respect to percent of non-White students, English language learners, and economically disadvantaged students, as well as school level (e.g., elementary versus middle) and total school enrollment. The report does not control for prior student achievement, which is a major shortcoming of the study.

IV. The Report's Use of Research Literature

The report cites little prior research about charter schools and school voucher programs. The report briefly mentions some of the prior studies evaluating the effectiveness of the MPCP carried out by the School Choice Demonstration Project , but the report does not discuss the findings or methods in depth. The report cites two studies providing descriptive comparisons of test score performance across charter and public schools in Milwaukee also using only a single, earlier year of data.

There is little reference made to the broader literature on school choice programs or charter

schools, including recent comprehensive reviews.¹¹ These recent reviews of studies evaluating school voucher programs and charter schools relative to public schools are mixed. Studies have found positive, negative and null effects across different locations and contexts. On average, there seem to be modest positive effects on test scores and school attainment from these policies, but the reviews highlight that there is much greater variability of the effectiveness of schools within sectors than between sectors. The report does not discuss how this variability or prior research might relate to the context in Wisconsin.

V. Review of the Report's Methods

The question the report is attempting to answer with an “apples to apples” comparison is an important one: whether one school sector (public, private, or charter) is more effective at educating students than the others. The fundamental concern for any study comparing student performance across school sectors is selection bias. In brief, the problem is that students select whether to attend a public, charter or private school. As a result, if we observe differences in test scores across schools there is no way to know (without additional information) whether the differences are due to some schools being more effective or due to some schools attracting better students (or, more likely, a combination of these and other factors). The report correctly acknowledges this problem, and attempts to use linear regression models to adjust for selection bias in order to make “apples to apples” comparisons. Unfortunately, although this is a reasonable approach, the limited nature of the publicly available data means the regression models are unlikely to overcome the selection bias problem. As a result, the comparisons made do not represent a “definitive” and “apples to apples” comparison of the test score performance across sectors.

In this section, I describe two fundamental shortcomings of the report's data and methods. First, the limited number of variables available and missing data for private schools make selection bias a major concern that prevents the results from being interpreted as causal estimates. Second, the use of the aggregate percent proficient metric is inherently problematic for making group comparisons.

Selection and Sample Bias

The use of regression in this study attempts to overcome selection bias by comparing school-level outcomes across sectors after adjusting for school-level demographic characteristics that are related to test score performance. In theory, if a regression model includes all relevant variables that determine whether a student will choose to attend a public, private or charter school (and that are related to test scores), then the adjusted differences estimated in a regression model can be interpreted as causal effects of attending one school sector relative to the others. But because this report relies on only a small number of school-level demographic control variables and cannot control for prior achievement, the resulting esti-

mates cannot be validly interpreted as causal effects.

As an example, consider the result in the report's Table 1 indicating that private schools in the MPCP program have higher average proficiency rates on mathematics tests by about 4.1% relative to public schools. One explanation for this difference is that private schools are better at educating students than public schools. Another explanation, however, is that students in private schools with similar racial, language and economic backgrounds differ in important ways from apparently similar students in public schools. Private school students, for example, may have greater parental support for their education, and may have already had higher levels of achievement before they entered the private schools. Both explanations are plausible, because students attending private schools and charter schools come from families that actively sought to send their children to such schools. More importantly, these alternative explanations cannot be ruled out by the analyses in this report due to the limited number of control variables.

Another important limitation involves missing data for private schools. Unlike public and charter schools, private schools are only required to administer the Forward Exam and ACT tests to students enrolled through the choice programs, not to all students. Results for private schools are then only reported when enough choice students are enrolled at the school. This has two consequences. First, not all participating private schools are included in the analyses in this report. While the report acknowledges that missing data is a problem, there is no description of what portion of private schools or private school students end up being excluded from the analyses. Second, while public and charter school test performance is based on nearly all students in the school, private school test performance is based on only a subsample of students in the school. This further undermines the claim to be making apples to apples comparisons.

Aggregate Percent Proficient

The percent proficient metrics in this report indicate the percent of students at each school scoring at or above the "proficient" cut score on the Wisconsin Forward Exam, and are inherently problematic for two reasons. First, the location of the cut scores is largely arbitrary, meaning there can be legitimate reasons for the cut score to be placed higher or lower. This is problematic because comparisons of group performance based on percent proficient metrics can change substantially when the cut score moves, and in some cases can even be reversed. The problem with such metrics has received considerable attention.¹³ A second problem is that the percent proficient metrics are aggregated to the school level. This means the percent of proficient students is averaged across all grades in the school. If the difficulty of reaching the proficiency cut score varies across grade levels or if schools in different sectors are more or less effective at educating students in different grades, the results could be affected in unpredictable ways. This makes it difficult to know how to best interpret the meaning of any differences in aggregate percent proficiency rates.

Technical Inconsistencies

In addition to the data limitations described above, there are also inconsistencies between the description of the methods and the results presented that make it impossible to fully understand the results, particularly for the comparisons of non-Milwaukee schools. These are described in more detail in an endnote for interested readers.¹⁴

VI. Review of the Validity of the Findings and Conclusions

The data and analyses presented do not support the claim to make “definitive” comparisons of a true “apples to apples” nature about “what is working and what isn’t” across Wisconsin public, charter and private schools. The report correctly worries about directly comparing test score performance across school sectors without considering differences between students enrolled in different sectors. Despite reasonable efforts, however, the report is unable to overcome these problems with the limited public data available. First, the limited number of demographic variables, lack of prior achievement data, and missing private school data mean that the reported differences in test scores should not be interpreted as direct estimates of the relative effectiveness of different school types. Second, the use of aggregate percent proficient metrics could mask variation in performance across grade levels and makes the results sensitive to the (arbitrary) location of the proficiency cut scores. Finally, the technical inconsistencies noted above make the conclusions about non-Milwaukee schools impossible to understand without additional information.

VII. Usefulness of the Report for Guidance of Policy and Practice

If policymakers or the public are interested in determining which schools or school choice policies in Wisconsin are most effective, this report cannot provide answers to such questions. The report does highlight important patterns in student test score performance, including the correlation between school demographics and test score performance. But the results should not be interpreted as estimating the relative effectiveness of school sectors. The publicly available data used in the analyses are simply too limited. Even as descriptive analyses, the results should be interpreted cautiously until additional details are provided clarifying the inconsistencies discussed above. Because the report focuses primarily on comparing average test score performance across sectors, there is no information provided about why some schools might have higher test score performance than others. The analyses in this report are most useful for providing descriptions of average test score performance and student demographics across schools in Wisconsin, rather than descriptions of what is working and what isn’t.

Notes and References

- 1 Maxwell, L.A. (2017). School choice a big winner in president Trump's budget. *Education Week*. Retrieved April 3, 2017, from http://blogs.edweek.org/edweek/charterschoice/2017/03/school_choice_a_big_winner_in_president_trumps_budget.html
- 2 Data obtained from <https://dpi.wi.gov/sms/choice-programs/data> on April 3, 2017.
- 3 Data obtained from <https://dpi.wi.gov/sms/charter-schools/current> on April 3, 2017.
- 4 Wisconsin Department of Public Instruction (2016). *District & School Report Cards Interpretive Guide*. Retrieved April 3, 2017, from https://dpi.wi.gov/sites/default/files/imce/accountability/pdf/Interpretive%20Guide%202016_Web.pdf
- 5 Flanders, W. (2017). *Apples to apples: The definitive look at school test scores in Milwaukee and Wisconsin*. Milwaukee, WI: Wisconsin Institute for Law and Liberty. Retrieved April 3, 2017, from <http://www.will-law.org/wp-content/uploads/2017/02/apples.pdf>
- 6 Flanders, W. (2017). *Apples to apples: The definitive look at school test scores in Milwaukee and Wisconsin*. Milwaukee, WI: Wisconsin Institute for Law and Liberty. Retrieved April 3, 2017, from <http://www.will-law.org/wp-content/uploads/2017/02/apples.pdf>
- 7 See the Table 1 heading on page 6.
- 8 Note, however, that the report uses a liberal statistical significance cutoff of 0.1 for some differences, which would not be considered statistically significant in many contexts.
- 9 Charter schools are classified based on their authorizers. There are three types of charter schools considered: independent, instrumentality, and non-instrumentality. Independent charters are authorized independently of the district, e.g. by the University of Wisconsin. Instrumentality and non-instrumentality charters are authorized by the district, but the non-instrumentality charter schools operate with more autonomy.
- 10 Wolf, P.J. (2012). *The comprehensive longitudinal evaluation of the Milwaukee Parental Choice Program: Summary of final reports*. Fayetteville, AR: University of Arkansas School Choice Demonstration Project. Retrieved April 3, 2017, from <http://www.uaedreform.org/downloads/2012/02/report-36-the-comprehensive-longitudinal-evaluation-of-the-milwaukee-parental-choice-program.pdf>
- 11 For a recent review of empirical studies on school choice programs, see for example: Egalite, A.J., & Wolf, P.J. (2016). A review of the empirical research on private school choice. *Peabody Journal of Education*, 91(4), 441-454. For recent reviews of the growing literature on charter schools, see: Berends, M. (2015). Sociology and school choice: What we know after two decades of charter schools. *Annual Review of Sociology*, 41, 159-180 or Epple, D., Romano, R., & Zimmer, R. (2015). *Charter schools: A survey of research on their characteristics and effectiveness*. Cambridge, MA: National Bureau of Economic Research. Retrieved on April 3, 2017, from <http://www.nber.org/papers/w21256>
- 12 Wisconsin Department of Public Instruction (n.d.). *School accountability for 2015-16: Ten things you need to know*. Retrieved April 3, 2017, from <https://dpi.wi.gov/sites/default/files/imce/accountability/pdf/Ten%20Things%20to%20Know%20about%202015-16%20Accountability-41316.pdf>
- 13 For discussions of the limitations of percent proficient metrics see Holland, P.W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, 27(1), 3-17 or Ho, A.D. (2008). The problem with "Proficiency": Limitations of statistics and policy

under No Child Left Behind. *Educational Researcher*, 37(6), 351–360.

- 14 To better understand the data source and results in this report I attempted to replicate the results in Table 1. I downloaded school report card data available from the DPI at <https://dpi.wi.gov/accountability/report-cards> (the file “2015-16_school_reportcard_data.xlsx”) and used information about charter school classifications from <https://dpi.wi.gov/sms/charter-schools/current> (the file “15-16 CS 3rd Fri Enrollment.xlsx”). I was able to replicate the results in Table 1 nearly exactly with the following differences. First, the coefficient for the variable “Non-White” in Table 1 is probably a typo and is actually the coefficient for the percent of White, not non-White, students. This appears to be a typo, given that the author claims this should be a negative coefficient on page 7. Second, the coefficients for “Middle School” and “High School” appear to be reversed. Finally, the sample of 241 schools that I used to replicate Table 1 includes 30 schools that do not have any students enrolled below 9th grade (according to the DPI data file), which suggests these schools do not administer the Forward Exam in ELA or mathematics. It is unclear whether these schools should be included in the analysis of Forward Exam performance or not. The problem with the results in Table 4 has to do with the report’s stated use of fixed effects. Specifically, the report claims that “private schools are assigned a single district fixed effect, as are independent charters” (footnote 2). If this is the case, however, it is no longer possible to compare private schools or independent charters to each other or to any of the public schools, which have their own respective district fixed effects. It is thus unclear how the comparisons in Tables 4 and 5 are being made.

DOCUMENT REVIEWED: Apples to Apples: The Definitive Look at School Test Scores in Milwaukee and Wisconsin

AUTHOR: Will Flanders

PUBLISHER/THINK TANK: Wisconsin Institute for Law and Liberty

DOCUMENT RELEASE DATE: March 2017

REVIEW DATE: April 25, 2017

REVIEWERS: Benjamin Shear, University of Colorado Boulder

E-MAIL ADDRESS: benjamin.shear@colorado.edu

PHONE NUMBER: (303) 492-8583

SUGGESTED CITATION:

Shear, B. (2017). *Review of "Apples to Apples: The Definitive Look at School Test Scores in Milwaukee and Wisconsin."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-milwaukee-vouchers>