The juice is worth the squeeze: A cost-effectiveness analysis of the experimental evidence on private school vouchers across the globe.

M. Danish Shakeel (mdshakee@uark.edu)
Kaitlin P. Anderson (kaitlina@uark.edu)
Patrick J. Wolf (pwolf@uark.edu)

Department of Education Reform
University of Arkansas
201 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-3172 Fax: (479) 575-3196

Prepared for the Association for Public Policy and Management’s International Conference in Brussels, Belgium, July 13-14, 2017. An earlier version of this paper was presented at the Society for Research on Educational Effectiveness conference in spring 2017. We thank Philip Gleason of Mathematica Policy Research and Robert Costrell of the Department of Education Reform for comments on earlier drafts. We also thank Albert Cheng from Harvard University for assistance with graphics. All remaining flaws are our own. Corresponding author is M. Danish Shakeel, mdshakee@uark.edu.
Abstract

School choice is a demand side intervention that seeks to improve educational effectiveness through market-oriented competition. Private school vouchers have emerged across the globe as a school choice reform. Twenty randomized control trials (RCTs) of school vouchers have been conducted on eleven different voucher programs for studying student achievement. A thorough cost-effectiveness analysis of the experimental studies of school vouchers across the globe would provide the foundation for a greater scholarly consensus regarding the efficiency of voucher interventions. This study adds to the literature on cost effectiveness by combining experimental estimates of the participant effects of private school vouchers with the estimated cost savings associated with these programs. We compare the efficiency of vouchers in terms of reading and math gains, for programs within and outside the U.S., and for publicly- versus privately-funded programs. The analysis shows that null test score findings from school voucher program evaluations should be viewed from a cost-effectiveness perspective. Generally, school vouchers are a more productive means of publicly funding education than traditional public schools in all cases but one. Voucher programs produce similar or better outcomes for an average student at a fraction of per student public school costs.

Keywords: Private school vouchers, school choice, public school, Randomized Control Trial, RCT, cost-effectiveness, productivity, fiscal effects
1. Introduction

Private school vouchers allow parents to choose any school for their children using government resources (Wolf, 2008). Parental choice and satisfaction make families active consumers of education where they can demand quality education in a school market. In the U.S., both government- and privately-sponsored voucher programs exist. School choice has also become a topic of high relevance since Donald Trump favored school choice during his recent successful presidential run and nominated Betsy DeVos—a voucher proponent—as his Secretary of Education. Experimental studies on school vouchers have generally found null to moderately positive achievement effects within the United States. Evaluations of the Louisiana Scholarship Program (LSP) and a recent second evaluation of the District of Columbia Opportunity Scholarship Program (DC-OSP II) are the only exceptions; three evaluations have found negative effects of these two school voucher programs on student achievement in the initial years (Abdulkadiroglu, Pathak, & Walters, 2015; Dynarski et al., 2017; Mills & Wolf, 2017). LSP and DC-OSP II voucher programs are relevant to our study given the current landscape of voucher programs.

Two definitions are central to our study. “Effectiveness” is the extent to which a program accomplishes its intended goals. “Cost-effectiveness” is “the efficacy of a program in achieving given intervention outcomes in relation to the program costs.” (Rossi, Lipsey & Freeman, 2004, p. 425) A program can be cost-effective because it generates better outcomes at a similar cost or because it produces similar outcomes at a lower cost. In the real world of scare resources, either result relatively benefits society.

The economic theory of vouchers is to increase educational effectiveness through choice and competition. The academic achievement effects of vouchers, though they tend to increase
COST EFFECTIVENESS OF PRIVATE SCHOOL VOUCHERS

with years of treatment, are generally modest in size (Shakeel, Anderson & Wolf, 2016). Outside the U.S., in developing countries, the achievement effects of school vouchers are generally larger. This differential could be due to a larger gap between public and private school quality in developing countries compared to the U.S. Some experimental studies on vouchers have found larger positive effects on graduation rates and college enrollment while having null to moderately positive achievement effects on participants (Cowen et al., 2013; Wolf et al., 2013; Chingos & Peterson, 2015). This has led some education researchers to study non-test score outcomes, which may be affected differently than cognitive (e.g. test score) outcomes in an education intervention.

Few education interventions produce large positive effects on test scores, according to experimental evidence. Lipsey et al. (2012) conducted a large-scale meta-analysis of all education randomized controlled trials (RCTs) since 1995, concluding that the average impact of an intervention on test scores for a broad-based standardized test was 0.08 standard deviations (SD) at the elementary level and 0.15 SD at the middle school level. Thus, the Shakeel, Anderson and Wolf (2016) meta-analysis indicates that school vouchers tend to produce global test score effects that are typical of other education interventions. To be promising, an education intervention must not only be effective but also efficiently use public resources. Hence, a better way to look at the cognitive outcomes of school vouchers is to also look at their cost-effectiveness.

First, we consider two basic issues concerning the effectiveness of vouchers at raising student test scores: 1) how to interpret null-effects of voucher interventions and 2) achievement effects that show a dichotomy between production function and policy parameters. Muralidharan (2015) lists five different interpretations of null-effects. A private school voucher intervention
may yield null-effects due to: a) lack of program fidelity, b) substitution effects as a result of pulling away of schooling inputs in response to the voucher, c) positive effects for the participants equaled by positive competitive effects on non-participants, d) lack of administrative reforms that hinder the effectiveness of the voucher intervention, or e) a true null-effect on all students.

Furthermore, not all voucher programs show a linear relationship between the effect size and years of treatment. While few voucher programs have been evaluated for more than a couple years, the existing evidence shows that the relation between effect size and years of treatment may be non-linear (Shakeel, Anderson & Wolf, 2016). This creates a dilemma for drawing relevant policy conclusions for the true effect of a voucher intervention, especially if there is a large variation, change in sign or statistical significance of effect size over the different years of treatment. A dichotomy exists if conclusions drawn from the early years of an evaluation are substantively different from conclusions drawn from later years of an evaluation, which is arguably more policy relevant, but more time- and resource-intensive to obtain.

This dichotomy can be understood mathematically by differentiating between the partial and total derivative of the outcome of interest with respect to the observables. In any voucher RCT, it is expected that by randomly assigning the observable explanatory variable $X_{it}$ the production parameter $\beta$ can be efficiently obtained. However, $\beta$ is a partial derivative of the outcome of interest with respect to the explanatory variable $\frac{\partial T_{it}}{\partial X_{it}}$. The partial derivative assumes that the observable explanatory variable $X_{it}$ remains constant. However, with increases in the length of treatment, the school and family-level characteristics may endogenously confound $X_{it}$, resulting in inaccurate estimates of $\beta$. Instead of the partial derivative, the total derivative $\frac{dT_{it}}{dX_{it}}$ may more appropriately account for any re-optimization by the school or family agents in
response to an exogenous change in the explanatory variables. Thus, an earlier year of an evaluation may yield the production function effect while the latest year of evaluation may yield the policy parameter effect which accounts for this re-optimization by agents and is more policy relevant (Das et al., 2013). Satisfaction of family is key to a market oriented reform like school vouchers and it is expected that changing schools would require some time for adjusting the household level inputs for the family. Similarly, the schools are also likely to re-optimize their inputs due to competition, entry and exit of students. The latest year of a voucher intervention may be more policy relevant, especially if a voucher intervention is to be scaled up based on results from a limited intervention. In some cases voucher may actually increase the amount of personal resources that families devote to education. This is clearly the case in the voucher program of Bogota, Colombia. It allowed families to top up and had incentives for students to get continued access to voucher if they passed the exams.

Results from Shakeel, Anderson, and Wolf (2016) suggest that, to draw more policy relevant conclusions concerning achievement effects from a voucher intervention, six issues must be raised in future studies: 1) long-term voucher interventions must establish the relation between effect size and years of treatment, 2) the interventions should address the five points raised above concerning null-effects, 3) the production function vs. policy parameters dilemma should be emphasized, 4) the last year’s cognitive effects should be the parameter to compare with non-cognitive outcomes of voucher interventions, 5) the details of the validation, reliability and construction process of the achievement test used in the interpretation should be provided in the papers, and 6) researchers should also study the outcomes from a cost-effectiveness perspective.
Muralidharan and Sundararaman (2015) and Wolf and McShane (2013) argue in favor of considering the cost-effectiveness of school vouchers as their findings indicate that null-effects in school voucher settings are achieved at a fraction of the cost of per pupil public school expenditures. A thorough analysis of the cost-effectiveness of school vouchers across the globe, conducted here, provides the foundation for a greater scholarly consensus regarding the ability of school vouchers to improve outcomes for students.

The objective of this analysis is to rigorously assess the cost-effectiveness of private school vouchers. This study adds to the literature on cost-effectiveness by combining experimental estimates of the participant effects of private school vouchers with the estimated cost savings associated with these programs. The experimental estimates come from the first systematic and meta-analytic review of all RCTs on private school vouchers internationally (Shakeel, Anderson, & Wolf, 2016). We focus our cost-effectiveness study on RCTs, because these are the “gold standard” of program evaluation in terms of assessing causal relationships (Mosteller & Boruch, 2002; Rossi, Lipsey & Freeman, 2005). RCTs essentially compare a treatment group (those receiving the offer of a voucher) relative to a control group (those who did not receive the offer of a voucher). In RCTs, the assignment of a voucher is random, and therefore the issue of selection bias is resolved, as treatment and control group units should be identical in expectation. While quasi-experimental design (QED) methods are often used to approximate the causal effect of a program, evidence from a within-study comparison (Anderson & Wolf, 2017) indicates that QEDs do not necessarily closely approximate causal estimates from an RCT, and that even the direction of the selection bias is not consistently predictable. For example, while it is often thought that more motivated or more able families self-sort into private schools and/or voucher programs, this is not always the case (Anderson & Wolf, 2017).
In addition to participant effects of vouchers, competitive effects of vouchers are also relevant for studying cost-effectiveness of vouchers. Positive competitive effects of voucher programs would strengthen the case for cost-effectiveness of vouchers while negative competitive effects would make an overall conclusion more difficult. In a review (Forster, 2016) of 33 studies, 31 indicate that school choice improves public schools, one finds a negative effect, and one finds a null effect. Given the generally positive competitive effects, we do not focus on competitive effects of voucher programs in the current analysis.

The article proceeds as follows. In section 2, we discuss earlier studies on school vouchers that have studied the outcomes from a productivity perspective. Section 3 describes the funding structure of the programs included in the cost-effectiveness analysis. In section 4, we describe the research methodology and assumptions made. Section 5 presents the results of the cost-effectiveness analysis. Section 6 concludes by discussing the policy implications of the cost-effectiveness analysis. We also discuss cautions when drawing policy implications from the study.

2. Literature Review

Earlier studies on the fiscal effects of publicly-funded school vouchers have found financial benefits for the funding body (i.e. the government). This is because the typical per-pupil voucher cost is less than the per-pupil cost when the same student would have attended a traditional public school. The Milwaukee Parental Choice Program (MPCP) is the longest running voucher program in the United States. Costrell (2010) estimated the net fiscal impacts of MPCP as approximately $52 million per year. With MPS denoting Milwaukee Public Schools, Costrell (2010, p. 4) relied upon the following:
Cost Effectiveness of Private School Vouchers

Net Impact = \((\text{MPS revenue/pupil} \times \text{reduction in MPS enrollment}) - (\text{voucher} \times \text{MPCP enrollment})\)

By comparing the new estimates with estimates in his earlier evaluation of the MPCP, Costrell found that the net fiscal impacts of the MPCP were positive and growing over time with program expansion.

The District of Columbia Opportunity Scholarship Program (OSP) is the only federally funded private school voucher program in the U.S. In a benefit/cost analysis of the DC OSP, Wolf and McShane (2013) took into account the increase in graduation rate induced by the program and estimated the economic returns to education attainment. Their estimates suggest that DC OSP’s impacts on educational attainment generated a return on investment (ROI) of approximately 162 percent. In other words for the low-income students in DC, for every dollar spent on the program, a return of $2.62 was estimated.

In a review of the fiscal benefits of 10 school voucher programs, Spalding (2014) estimated total savings of $1.7 billion as the lower bound since the inception of the MPCP in 1990-91 through 2010-2011. The calculations also showed an increase of over 230 times in the participation in school voucher programs in the analyzed timeframe. The increase in demand for school vouchers and the associated cost savings should draw the interest of policymakers for more experimental testing of vouchers programs in the U.S.

Generally, the students leaving the traditional public schools for a voucher-accepting school bring fiscal benefits to the school, as the leaving students cost more to educate than the revenue lost from the state (Scafidi 2012; Trivitt & DeAngelis 2016). As a result, when a student leaves a public school via a private school choice program, the remaining students in that school actually have increased financial resources, on a per-pupil basis. Hence, an average public school
district would generally receive fiscal benefits due to school choice interventions. Chakrabarti (2007) showed that when money follows the student, there are increasing incentives for the improvement in the traditional public school system.

Tuition tax credit programs, which operate like voucher programs but are funded through donations to non-profits, produce fiscal benefits, but these benefits may differ from the savings associated with vouchers for a variety of reasons. Publicly funded voucher programs require taxpayers to financially support the voucher amount while tuition tax credit programs allow taxpayers to receive partial or full tax credits for donating to nonprofits (Lueken, 2016). Some voucher programs involve fixed tuition amounts and regulations that affect participation by families and schools that prefer tuition top up. As tuition tax credit programs seemingly give more freedom to the taxpayers, they might produce different impacts than school vouchers in a market where school choice is in high demand. In addition, tuition tax credit programs have proven to be more politically palatable, as they do not require participation by all taxpayers, as full voucher programs due. On the other hand, however, tuition tax credit programs, if implemented, may have a lower take-up rate since individuals must choose to donate to nonprofits, as opposed to a universal voucher program, for example, in which taxpayers are required to contribute and all school-age children are able to participate at no additional private cost. Lueken (2016) estimated overall savings between $1,650 to $3,001 per student, on average, for ten tuition tax credit programs in the U.S. The cumulative savings per student for different programs ranged between $298 to $8,450.

Public charter schools also produce fiscal benefits. Wolf et al. (2014) used two measures for measuring the productivity of public charter schools. They estimated the gains in the student test scores on the 2010-11 National Assessment of Educational Progress (NAEP) per a $1,000
investment in education of a student in charter school in comparison to the traditional public school. In their second measure, they calculated a return on investment (ROI) by converting the learning gains in the charter and traditional public school sectors into an estimate of economic returns over a lifetime for the students. Thereafter, they compared the gains to the revenue amounts that had been invested in the student’s education. The authors conclude, “the analyses we present in this report indicate that charter schools are more productive than TPS, either because they produce higher student gains at a lower cost or because they produce similar or only slightly lower student gains at a significantly lower cost.” (Wolf et al., 2014, p. 9).

Positive fiscal impacts of private school vouchers are also expected in international contexts as the private schools charge less per-pupil tuition in comparison to the average per-pupil government expenditure in the public school system. Muralidharan and Sundararaman (2015) reported the funded voucher amount to be around 40% of the per-student costs in the public schools. Due to the larger quality gap between public and private schools in the developing world, the cost effectiveness of private school vouchers may be higher there in comparison to voucher interventions in the U.S.

3. Funding Structure of the Programs Included in the Analysis

The RCTs included in our analysis were located in three countries: the United States of America, Colombia and India. The majority of programs analyzed experimentally (8 out of 11) were administered within the U.S. The U.S. studies covered programs in Charlotte, NC; Dayton, OH; Milwaukee, WI; New York City; Toledo, OH; Washington, DC (two separate programs) and Louisiana. The participants in the RCTs were children who were randomized, through a lottery, to receive (or not) a voucher to attend a private school. The grades analyzed ranged from
K to 12, although most RCTs included a shorter grade range in their analysis. Most of the private schools that participate in voucher programs in the U.S. and other countries are relatively low-cost schools with per-student costs below the average amount spent in area public schools. The duration of studies analyzed ranged from one to seven years. The voucher interventions were targeted towards disadvantaged sections of the population through income limits and/or program location. Most voucher-accepting private schools were already serving disadvantaged students. Often they had a religious orientation, especially Catholic. With the exception of Louisiana, the U.S. programs were limited to particular cities.

Table 1 shows the summary of the funding structure for the publicly or privately funded school voucher or K-12 “scholarship” programs. The costs in Table 1 have been adjusted for inflation and cost-of-living/purchasing power to 2013 U.S. dollars and account for the variable proportion of the costs in traditional public schools. We discuss details in the next section. The publicly funded programs were in Bogota, Colombia; Washington, DC; Louisiana; and Milwaukee, WI; U.S.A. Generally, the publicly funded programs covered full tuition costs and the privately funded programs covered varying portions of the full tuition costs, with some combination of the parents or the schools making up the difference. For the purposes of calculating the cost savings of a voucher, relative to TPS expenditures, all costs and cost differences have been adjusted for inflation to 2013 USD. In addition, we adjust for regional cost-of-living within the U.S. using a comparable wage approach (Taylor & Fowler, 2006). The per-pupil cost differences between the voucher amount and public school variable cost for the publicly funded programs in the U.S. ranged from around -$1,322 in DC OSP II to -$2,842 in Louisiana; the original amounts of the voucher were $5,000 for Louisiana and $8,000 for grades K-8 and up to $12,000 for grades 9-12 for DC OSP II (Dynarski et al., 2017; Mills & Wolf,
2016). For the privately funded programs in the U.S., the proportion of support for the tuition cost varied vastly. While we are unable to determine the exact proportion and amount of per-pupil top-up (families contributing funds to make up the difference between the private school tuition and the voucher amount), the per-pupil cost difference between the voucher amounts (ignoring top-up) and the traditional public school variable cost ranged between -$8,116 in Toledo, OH to -$4,649 in Charlotte, NC.

[Table 1 about here]

Funding for the voucher programs in Colombia and India was extremely low. The two privately-funded voucher programs in Andhra Pradesh, India and Delhi, India were fully funded and the voucher program in Colombia allowed top-up and was partly funded by the World Bank. In nominal USD, the original voucher amounts ranged from about $117 in India (Wolf, Egalite & Dixon, 2015) to $190 in Colombia (Angrist et al., 2002). Both the privately- and publicly-funded voucher programs across the globe covered less than half of the per-pupil expenses in nearby public schools. The international voucher programs served students living in abject poverty (Angrist et al., 2002; Muralidharan & Sundararaman, 2015; Wolf, Egalite & Dixon, 2015). The private schools accepting voucher students have modest infrastructure, instructional facilities and special programs for differentiating instruction to students in comparison to the nearby public schools (Dixon, 2013; Wolf et al., 2013).

4. Research Methodology and Assumptions

The research design of the studies that inform the cost-effectiveness analysis was random assignment of children to treatment and control groups. Most studies had a one-stage randomization through administration of a lottery while one study in Andhra Pradesh, India
(Muralidharan & Sundararaman, 2015) was based on a two-stage randomization (randomly assigned students within randomly assigned villages). For this cost-effectiveness analysis, we relied on estimates of the participant effects of school vouchers from a meta-analysis by Shakeel, Anderson, and Wolf (2016). To graphically analyze the productivity of voucher programs, we compare these experimental estimates of program benefits, in Hedge’s $g$ effect sizes (Hedges, 1981), graphing each effect size within a circle sized according to the treatment sample size. The size of the treatment sample is the average of treatment samples for math and reading scores for a particular year of treatment (the treatment sizes do not differ vastly for math and reading scores). We inflation-adjust all costs (both the voucher costs and the applicable TPS costs) to September 2013 U.S. dollars. In particular, we use Consumer Price Index (CPI) data from the Bureau of Labor Statistics\(^1\) for the studies within the U.S., CPI data from India’s Open Government Data (OGD) Platform\(^2\) for the studies in India, and CPI data from Colombia’s Banco de la República\(^3\) for the study in Colombia. We inflation-adjust to September 2013 U.S. dollars. In addition, all U.S. costs are adjusted for regional differences in cost-of-living using a comparable wage approach (Taylor & Fowler, 2006), and costs outside of the U.S. are adjusted to the U.S. cost-of-living using Purchasing Power Parity (PPP) data from the Organisation for Economic Co-operation and Development (OECD).\(^4\) These adjustments allow for more careful cost comparisons, adjusted for both inflation over time, and the relative value of (purchasing power) of money in different localities. This allows us to analyze the relative productivity of the voucher in comparison to public school expenditures in the same locality (district or state).

---

1 Source: https://www.bls.gov/cpi/
2 Source: https://data.gov.in/catalog/state-level-consumer-price-index-ruralurban
4 Source: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
We focus on the treatment on the treated (TOT), but we also analyze the ITT estimates.\textsuperscript{5} The TOT includes the outcomes for voucher winners who actually used the voucher, and the Louisiana Scholarship Program was a placement lottery that only yielded TOT impacts, making the TOT impacts of greater relevance than the ITT impacts. The cost measures for the voucher programs are accumulated from a variety of sources including program websites, state-level websites, country-level websites, and research evaluations or other documents. In general, to be conservative, we use the maximum voucher amount available as an upper limit on the cost of the voucher, when available. Exceptions are the voucher amounts for the Milwaukee Parental Choice Program (MPCP) (Rouse, 1998; Greene, Peterson, & Du, 1999), the LSP (Abdulkadiroglu, Pathak, & Waters, 2015; Mills & Wolf, 2017), the Programa de Ampliación de Cobertura de la Educación Secundaria (PACES) program in Bogota, Colombia (Angrist, Bettinger, Bloom, King, & Kremer, 2002; Angrist, Bettinger, & Kremer, 2006) and the Andhra Pradesh (AP) School Choice Experiment (Muralidharan & Sundararaman, 2015) where the upper limit was not publicly available. In the first three cases, we use average voucher amounts instead. In the fourth case (Muralidharan & Sundararaman, 2015), we calculate the average voucher amount as 40 percent of the average public school expenditure per child (p. 1031, 1058). We obtain the per-pupil expenditure in public schools for Delhi, India from an Economic Survey of Delhi, 2016-2017.\textsuperscript{6} The per-pupil expenditures for public schools in Bogota, Colombia is obtained from Angrist et al. (2002, p. 1537).

For the U.S. studies, we subtract from the voucher costs the variable per pupil costs associated with public school education in the same locality. Only variable costs are subtracted, as the local public school system should theoretically be able to reduce costs by this amount,

\begin{itemize}
  \item \textsuperscript{5} Results are available from the authors on request.
  \item \textsuperscript{6} Source: \url{http://www.indiaenvironmentportal.org.in/category/15375/publisher/government-of-nct-of-delhi/}
\end{itemize}
while being unable to affect fixed costs in the short term. The cost measures for the traditional public schools system come from the per-pupil expenditures in a given locality in a given year, adjusted to September 2013 dollars. These expenditures are primarily obtained from the Census Bureau databases,\(^7\) supplemented with state-level databases from Ohio (Ohio Department of Education)\(^8\) and the National Center for Education Statistics Common Core of Data State Fiscal Reports,\(^9\) as needed.

We use conservative guidance from Spalding (2014) to determine which public school costs are considered variable. Specifically, we include (on a per-pupil basis) only current instructional expenses, current student support/pupil support, instructional support, and food service. Other fiscal effect studies have included enterprise operations (school bookstore, interscholastic activities, etc.) as variable costs as well (Scafidi, 2012), but Spalding assumes that as the costs associated with certain enterprise operations rise or fall, so would the associated revenues. The removal of these would have no net fiscal difference, and we exclude these from our lower bound estimates of the variables cost in the local TPS system. Therefore, we are assuming only a lower bound on the potential savings from the TPS system. In one study location and year combination (MPCP, 1990-1991), the corresponding traditional public school expenditures were not available, so we assumed that these costs were the same as the 1991-1992 Milwaukee TPS costs, adjusted to 2013 dollars.

The public school costs outside of the U.S. are not reported in the same level of detail\(^{10}\), so we are unable to differentiate between the variable and fixed components of these costs.

---

\(^7\) Source: [https://www.census.gov/programs-surveys/school-finances.html](https://www.census.gov/programs-surveys/school-finances.html)

\(^8\) Source: [http://education.ohio.gov/Topics/Finance-and-Funding/Finance-Related-Data/Expenditure-and-Revenue/Expenditure-Revenue-Data](http://education.ohio.gov/Topics/Finance-and-Funding/Finance-Related-Data/Expenditure-and-Revenue/Expenditure-Revenue-Data)


\(^{10}\) [http://ccs.in/sites/default/files/research/research-per-child-funding-model-for-schools-in-india.pdf](http://ccs.in/sites/default/files/research/research-per-child-funding-model-for-schools-in-india.pdf)

Therefore, when calculating the net “savings” from voucher programs outside the U.S., we make assumptions about the proportion of total TPS costs are variable, beginning with an assumption based on this proportion calculated from the U.S. studies. For example, we calculate, the percent of TPS expenditures that is considered variable for the last evaluation year for each U.S. voucher program. Across all U.S. programs, this percentage ranged from 64% to 79%, and on average was 69%. For LSP, we do not have public school expenditures for 2014-2015 (not published yet by NCES) so we use the 2013-14 expenditures instead. We also do the same thing for the recent DC OSP evaluation (i.e. DC OSP-II) by Dynarski et al. (2017). This yields an average of 69% as the typical percentage of total TPS expenditures that are considered variable in the U.S. localities with voucher evaluations included here. We use this number to estimate variable public school costs for the non-US studies. This decision is based on the assumption that non-U.S. governments allocate the same proportion of their per-student public expenditures to variable costs as in the U.S. voucher settings. This might be a strong assumption, so we report a breakeven point at which there would be no potential savings to the non-U.S. governments. The breakeven point is the percentage equivalent to the voucher amount divided by the total per-student public school expenditures in non-U.S. contexts. The breakeven points for Delhi, Andhra Pradesh and Bogota occur at 25, 40 and 54 percent, respectively. In other words, in Delhi, India, if 25% of the traditional public school costs are variable (and therefore represent savings from a voucher program), the program would break even, as the average per-pupil variable cost savings would equal the average voucher amount.

The National Center for Education Statistics (NCES) created a Comparable Wage Index (CWI) as a measure of “systematic, regional variations in the salaries of college graduates who are not educators.” The CWI is used to make financial comparisons in terms of cost of living, or
wages demanded across geographic regions within the U.S. (Bush School of Government and Public Service, 2016). For the U.S. studies, we utilize the CWI to adjust for financial comparisons related to cost of living. Due to the lack of a similar index for the two regions in India (urban Delhi and rural Andhra Pradesh), we are not able to adjust for cost of living differences across the two regions in India. For comparing the cost differences across countries we utilize the purchasing power parity (PPP) conversion factor\textsuperscript{11} from the World Bank. This allows us to adjust for the different in costs that are affected by the exchange rates of currencies across the countries.

Public school expenditures are generally reported as elementary and secondary expenditures combined. Since secondary education tends to be more expensive, we may be overstating the comparable TPS cost for the typical student in these voucher studies, as the voucher impacts are typically estimated for a younger population. Only one U.S. study, the DC OSP evaluation, included students through grade 12, and the second PACES study (Angrist, Bettinger, & Kremer, 2006) included students in grades 6 to 11.

If a voucher program accepts students who were not already attending public schools, then there is no corresponding savings for the public school system. However, few voucher/scholarship programs in the U.S. serve students already enrolled in private school, as most are limited to public school students or rising kindergarteners (EdChoice, 2017). In some developing countries, the public schools are often of low quality and teacher absenteeism is high (Chaudhury et al., 2006). It is possible that vouchers may induced kids who are not attending a school due to no learning to attend a private school. Three publicly funded U.S. programs in Table 1 have already been evaluated regarding their cost-effectiveness (Costrell, 2010; Wolf &

\textsuperscript{11} Source: \url{http://data.worldbank.org/indicator/PA.NUS.PPP}
McShane, 2013; Trivitt & DeAngelis, 2016). All three concluded that the voucher programs were more cost-effective than their TPS.

The total savings for a voucher program for the last year of treatment in each program can be calculated from Table 1 as:

\[
\text{Adjusted per – pupil cost difference}_{(\text{voucher} - \text{TP})} \times \text{Size of treatment sample}_{\text{TOT}}
\]

The cumulative savings associated with a voucher program can be calculated by multiplying the above equation by the number of years of treatment for these students in Table 1. This would generate a lower bound on the cumulative savings as the treatment sample for the last year is generally the lowest for any year of evaluation due to attrition. The savings can then be summed up across programs to generate net social savings due to these voucher programs. These calculations should not be viewed as net fiscal impacts or benefit/cost analysis of these voucher programs. Instead, they are a lower bound of the savings from these programs, as they represent only the savings for students who were randomly assigned and evaluated within an RCT, and only for certain years of the program. The overall savings from these programs accrue in part to the government and in part to society, although the savings that we estimate are limited in that we are unable to account for top-up in the privately-funded programs. In the privately funded programs, while savings accrue to the government as students leave public schools, private funds do represent a cost to society at large, and we are also ignoring additional private or social costs related to top-up amounts.
5. Results

We present the results graphically with the per-pupil cost difference between the voucher amount and the TPS variable cost on the horizontal-axis of our figures. These estimates have been adjusted for inflation and regional expense differences. A negative cost difference means that the voucher amount is less than the local TPS variable costs saved. In other words, negative cost differences indicate net savings from the voucher. On the vertical axis, we plot the effect size weighted by the size of the treatment group obtained from Shakeel, Anderson and Wolf (2016). Statistically significant (at the 95% confidence interval) effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above (below) the horizontal-axis represents a positive (negative) effect size which means that treatment (voucher) students had higher (lower) reading or math test scores, relative to their TPS peers in the same RCT study.

First, we present the overall results for reading and math (TOT) for the last year of evaluation of the programs. Then we show the results by nature of funding received by the program (public vs. private), and finally we graph the results by years of the treatment.

5-A. Overall Impacts

The overall results for the last year evaluation of the programs for reading in Figure 1 and for math in Figure 2 show that, generally, the productivity of vouchers is greater than that of traditional public schools. In fact, in all cases, the cost savings is positive (all centers of circles lie in the left two quadrants). The generally null effects come at a lower per-student cost in comparison to per-student variable cost in public schools. In only two programs (LSP and the second DC OSP evaluation), and only in math, were the effects of vouchers negative, and even
in these cases, the voucher programs were operating at savings relative to the TPS system. Thus, while vouchers tend to be at least as effective as public schools, their cost-effectiveness tends to be higher. Generally, the U.S. programs seem to save more money than the non-U.S. programs, however this is misleading given the stark differences in the level of spending between the U.S. and non-U.S. contexts. The variable costs for non-U.S. programs are calculated based on 69% as the typical percentage of total TPS expenditures that are considered variable in the U.S.

5-B. By Funding Type

In Figures 3 through 6 we plot the previous results by the type of funding received by the voucher program. The funding source is defined as public if a voucher program received any portion of the voucher amount from the government. Our results are limited in that they do not account for voucher top-up, which is quite common in the case of small, privately funded vouchers, and thus we are not able to determine which type of funding is more cost-effective among the publicly and privately funded programs. Generally, the privately funded vouchers in the U.S. cover only part of the private school tuition amount. The results show that even publicly-funded voucher programs are more cost-effective than the local public school system.
5-C. By Years of Treatment

In Figures 7 through 14 we plot the previous results by the duration of treatment (one, two, three and four or more years). Both the math and reading results show that treatment duration is positively related to voucher cost-effectiveness. While we cannot determine the exact form of this relationship, generally a positive relation is visible in the graphs.

[Figure 7 about here]
[Figure 8 about here]
[Figure 9 about here]
[Figure 10 about here]
[Figure 11 about here]
[Figure 12 about here]
[Figure 13 about here]
[Figure 14 about here]

There are two main takeaways from the preceding figures: the first is that school vouchers are consistently more cost-effective than the public schools, even if most of the effects are null. Secondly, the dilemma of the production function vs. policy parameter effect does not produce substantively different conclusions about voucher cost-effectiveness, as for all years of treatment, the voucher programs produced savings. While voucher effectiveness based on test scores may vary between worse and better (the effect size on test scores can change between positive and negative and between statistically significant to null effects) from the earliest year of an evaluation to the latest year, voucher cost-effectiveness generally remains positive (in DC OSP II and LSP negative impacts were obtained at lower costs). Hence, regardless of whether
one uses estimates from the early years of treatment or later years, the conclusion is still that voucher programs save money.

The total savings to the government from these voucher programs is high. Summing over the costs obtained from the voucher program yields savings worth $54,786,267 for the U.S. programs, $3,610,992 for the Colombia program, and $1,718,796 for the India programs. Overall the voucher interventions saved the governments $60,116,057 globally. This number does not adjust for student top-ups, per-student costs in elementary and secondary grades and cost of living differences across Delhi and Andhra Pradesh, India. This number is only for the students that were part of the RCTs and only for the number of years that an evaluation was conducted; hence it is a lower bound on the savings from the program. A program may have students that were not randomized and still participated. Furthermore, programs may not have been evaluated for all years of operation (e.g., the DC OSP). Nevertheless, aligning our analysis with previous voucher productivity studies of MPCP (Costrell, 2010), DC OSP (Wolf & McShane, 2013) and LSP (Trivitt & DeAngelis, 2016), we conclude voucher programs tend to save the government money. Further investment in voucher programs may be a cost-effective policy tool as these programs are generally at least as effective as the public schools in raising student outcomes on math and reading (Shakeel, Anderson, & Wolf, 2016).

6. Conclusions and Policy Implications

This cost-effectiveness analysis contributes to the field by combining rigorous evidence from RCT studies with actual public school expenditures. This study provides an estimate of the cost-effectiveness of voucher programs and indicates important policy implications about the effectiveness vs. cost-effectiveness aspect of voucher programs generally. While voucher
programs are growing across the globe, a cost-effectiveness analysis of the effect of vouchers internationally was lacking. There is variation across programs in both cost-effectiveness and impacts. No clear relationship has emerged between the cost and estimated impacts of a program from our analysis. As hypothesized, interpreting null student test-score impacts from a cost-effectiveness perspective reveals that private school vouchers generally produce student test scores similar to those in the local public school system, but at a lower cost. Our findings generally accord with Muralidharan and Sundararaman (2015) and Wolf and McShane (2013).

The evidence suggests that all programs pass the cost-effectiveness test for reading test score impacts, while only two evaluations had a more ambiguous overall effect (slightly negative test scores alongside lower costs in the LSP and DC OSP-II). A meta-analysis of the experimental studies on private school vouchers has shown generally higher impact of voucher programs on participants’ reading scores in comparison to math scores (Shakeel, Anderson, & Wolf, 2016). In the future, it would be interesting to see how the impacts and cost-effectiveness of LSP and DC OSP change with later years of evaluation of the program. Although, we distinguished between publicly- and privately-funded voucher programs, it is difficult to directly compare the cost-effectiveness of these two types of funding mechanisms, as they differ in terms of who pays, and generally only privately-funded programs allow top-up. This area requires further exploration. In terms of policy recommendations, the government could save money by investing in private school vouchers without generally losing student effectiveness. For an education intervention to be promising, it must not only improve student outcomes, but also be cost-effective. On the first measure, voucher programs, based on experimental evidence, are at least as effective as public schools, with the exception of two U.S. programs in LSP and Washington, D.C. On the second measure, they are substantially more cost-effective than public
schools. Before scaling up a voucher program, policymakers should consider funding more experimental evaluations of school vouchers. It would also be important to assess how instructional time is spent in public and private schools in voucher settings. Future studies should also use productivity measures based on unit of instructional time spent to address the education gap (Muralidharan, Singh, & Ganimian, 2016). Test scores and cost-effectiveness cover only part of the effects of voucher programs. For scaling up voucher programs, re-optimization of household and school level inputs deserve more exploration.
References


COST EFFECTIVENESS OF PRIVATE SCHOOL VOUCHERS


Muralidharan, Karthik. 2015. Field experiments in education in developing countries. *UC San Diego*.


Table 1: Description of 11 Voucher Programs in Cost-effectiveness Analysis

<table>
<thead>
<tr>
<th>Program Evaluated</th>
<th>Location</th>
<th>Funding Source</th>
<th>Funding Amount (Full or Partial)</th>
<th>Cost Difference (Voucher - TPS VC adj)</th>
<th>Size of treatment sample (TOT)</th>
<th>Grades</th>
<th>Years of Treatment</th>
<th>Studies Cited</th>
<th>Lower Bound Net Savings (Cost * Treatment Size * Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andhra Pradesh (AP) School Choice Experiment</td>
<td>Andhra Pradesh, India</td>
<td>Private Full</td>
<td>-$185</td>
<td>868</td>
<td>1 to 5</td>
<td>4</td>
<td></td>
<td>Muralidharan &amp; Sundararaman (2015)</td>
<td>$640,716</td>
</tr>
<tr>
<td>Charlotte Children’s Scholarship Fund</td>
<td>Charlotte, NC (USA)</td>
<td>Private Partial</td>
<td>-$4,649</td>
<td>152</td>
<td>2 to 8</td>
<td>1</td>
<td></td>
<td>Greene (2000); Cowen (2008)</td>
<td>$720,595</td>
</tr>
<tr>
<td>Children’s Scholarship Fund</td>
<td>Toledo, OH (USA)</td>
<td>Private</td>
<td>-$8,116</td>
<td>78</td>
<td>K to 8</td>
<td>3</td>
<td></td>
<td>Betringer &amp; Slonim (2006)</td>
<td>$1,899,144</td>
</tr>
<tr>
<td>District of Columbia Opportunity Scholarship Program (OSP)</td>
<td>Washington, DC (USA)</td>
<td>Public Full</td>
<td>-$1,322</td>
<td>635</td>
<td>K to 12</td>
<td>1</td>
<td></td>
<td>Dynarski, Rui, Webber, Gutmann, &amp; Bachman (2017)</td>
<td>$839,464</td>
</tr>
<tr>
<td>Ensure Access to Better Learning Experiences (ENABLE)</td>
<td>Delhi, India</td>
<td>Private Full</td>
<td>-$871</td>
<td>614</td>
<td>K to 2</td>
<td>2</td>
<td></td>
<td>Wolf, Egalite, &amp; Dixon (2015)</td>
<td>$1,078,081</td>
</tr>
<tr>
<td>Louisiana Scholarship Program (LSP)</td>
<td>Louisiana (USA)</td>
<td>Public Full</td>
<td>-$2,697</td>
<td>492</td>
<td>3 to 8</td>
<td>3</td>
<td></td>
<td>Abdulkadiroglu, Pathak, &amp; Walters (2015); Mills &amp; Wolf (2016); Mills &amp; Wolf (2017)</td>
<td>$3,981,152</td>
</tr>
<tr>
<td>Milwaukee Parental Choice Program (MPCP)</td>
<td>Milwaukee, WI (USA)</td>
<td>Public Full</td>
<td>-$3,538</td>
<td>87</td>
<td>K to 8</td>
<td>4</td>
<td></td>
<td>Rouge (1998); Greene, Peterson, &amp; Du (1999)</td>
<td>$1,231,224</td>
</tr>
<tr>
<td>School Choice Scholarships Foundation Program</td>
<td>New York, NY (USA)</td>
<td>Private Partial</td>
<td>-$6,833</td>
<td>890</td>
<td>1 to 4</td>
<td>3</td>
<td></td>
<td>Peterson, Howell, Wolf, &amp; CampbeI (2003); Barnard, Frangakis, Hill, &amp; Rubin (2003); Knaege &amp; Zhu (2004); Jin, Barnard, &amp; Rubin (2010); Bicker, Domina, Permer, &amp; Hoynes (2015)</td>
<td>$18,244,110</td>
</tr>
<tr>
<td>U.S. Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$54,786,267</td>
</tr>
<tr>
<td>Columbia Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$3,610,992</td>
</tr>
<tr>
<td>India Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1,718,797</td>
</tr>
<tr>
<td>TOTAL SAVINGS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$60,116,056</td>
</tr>
</tbody>
</table>

Note: Statistics are presented for the last year of each program. Please contact the authors for more details for all years of a program. DC OSP was subjected to two distinct evaluations for different student populations. There are less savings in DC OSP II than DC OSP I due to increases in the voucher amount. No ITT estimates exist for LSP as it was a placement lottery. The variable costs for non-U.S. programs are calculated based on 69% as the typical percentage of total TPS expenditures that are considered variable in the U.S. localities. Size of treatment sample is the average sample size for math and reading effect size estimates.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size, which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Delhi, India includes an overall estimate for English and Hindi. Reading estimate for Andhra Pradesh, India includes an overall estimate for English, Hindi, and Telugu. Reading estimate for Bogota, Colombia is for Spanish.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size, which means that vouchers produce better/worse outcomes than the local TPSs.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Delhi, India includes an overall estimate for English and Hindi. Reading estimate for Andhra Pradesh, India includes an overall estimate for English, Hindi, and Telugu.
Figure 4: TOT Reading (Publicly Funded)

Vouchers perform better than public schools

Vouchers perform worse than public schools

Note: On the vertical axis, we plot the effect sizes, Hedges’ g estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Bogota, Colombia is for Spanish.
**Figure 5: TOT Math (Privately Funded)**

<table>
<thead>
<tr>
<th>Location</th>
<th>Effect Size</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington, DC - WSF</td>
<td>0.9</td>
<td>Vouchers perform better than public schools</td>
</tr>
<tr>
<td>NYC</td>
<td>0.7</td>
<td>Vouchers perform better than public schools</td>
</tr>
<tr>
<td>Dayton, OH</td>
<td>0.5</td>
<td>Vouchers perform better than public schools</td>
</tr>
<tr>
<td>Charlotte, NC</td>
<td>0.3</td>
<td>Vouchers perform better than public schools</td>
</tr>
<tr>
<td>Toledo, OH</td>
<td>0.1</td>
<td>Vouchers perform better than public schools</td>
</tr>
<tr>
<td>Delhi, India</td>
<td>-0.3</td>
<td>Vouchers perform worse than public schools</td>
</tr>
<tr>
<td>Andhra Pradesh, India</td>
<td>-0.4</td>
<td>Vouchers perform worse than public schools</td>
</tr>
</tbody>
</table>

**Note:** On the vertical axis, we plot the effect sizes, Hedges’ \( g \) estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs.
Figure 6: TOT Math (Publicly Funded)

Vouchers perform better than public schools

Bogota, Colombia

Milwaukee, WI

Washington, DC - OSP I

Vouchers perform worse than public schools

Louisiana

Washington, DC - OSP II

Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes in comparison than the local TPSs.
Note: On the vertical axis, we plot the effect sizes, Hedges’ g estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Delhi, India includes an overall estimate for English and Hindi.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Delhi, India includes an overall estimate for English and Hindi. Reading estimate for Andhra Pradesh, India includes an overall estimate for English, Hindi, and Telugu.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Bogota, Colombia is for Spanish.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs. No reading estimates are reported for Toledo, OH as it had only math test outcomes. Reading estimate for Bogota, Colombia is for Spanish.
*Note:* On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs.
Note: On the vertical axis, we plot the effect sizes, Hedges’ $g$ estimates based on the final year for each study from Shakeel, Anderson, & Wolf (2016). The size of the circle corresponds to the number of treatment students in each study. Statistically significant effect sizes are in bold circles. The location of each program is connected to its effect size (center of the circle) with a line. The center of the circle above/below the horizontal axis represents a positive/negative effect size which means that vouchers produce better/worse outcomes than the local TPSs.