8-2018

The Private School Advantage: Evidence from School Vouchers and Educational Leadership

Mohammad Danish Shakeel
University of Arkansas, Fayetteville

The Private School Advantage: Evidence from School Vouchers and Educational Leadership

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Education Policy

by

Mohammad Danish Shakeel
Aligarh Muslim University
Bachelor of Science in Physics, 2004
Aligarh Muslim University
Master of Science in Physics, 2006
Youngstown State University
Master of Science in Mathematics, 2008

August 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Robert Maranto, Ph.D.
Dissertation Director

_____
Patrick Wolf, Ph.D.
Committee Member

_____
Jonathan Mills, Ph.D.
Committee Member

Abstract

School choice is becoming increasingly popular around the globe. Broadly the term 'school choice' is used to describe the options available for families to send children to school(s) other than the one they are residentially assigned to. Private school choice interventions known as 'school vouchers,' offer public or private funding to enable families to send their children to private school.

Research in 1970s and 80s by James Coleman and his colleagues showed a private school advantage in student achievement and graduation rates, in comparison to traditional public schools. Competing evidence was presented by Christopher Lubienski and Sarah Lubienski in 2013, claiming a public school advantage in student achievement. The debates surrounding a particular school sector advantage can be better addressed using causal evidence and using large datasets to understand possible mechanisms that differentiate the school sectors.

This dissertation reports on four analyses of the possibility of a private school advantage, using a variety of data. The first study looks at overall evidence on student achievement in math and reading scores from causal studies on private school vouchers around the globe. The second study offers a supplemental cost-effectiveness evaluation of the same set of voucher programs.

In the third study, nationally representative data on public and private school principals is analyzed to study principal autonomy over seven school-level activities across school sectors. Using the same dataset, the fourth study examines the determinants of principal attrition across school sectors. Principals' stated responses to stay in the profession in the baseline year are compared to their revealed status a year later.

Some contributions of this dissertation are evidence of vouchers increasing reading test scores more in comparison to math test scores and a larger test score impact in developing countries than in the U.S. The dissertation finds more autonomy over school-level activities and more likelihood to remain in the profession for the private school principal in comparison to the traditional public school principal. Hence, future studies may test the role of principal autonomy and principals' remaining in the profession as a mediator of school choice outcomes.

Acknowledgements

I would like to thank my committee: Robert Maranto, Patrick Wolf, and Jonathan Mills. Bob and Pat you introduced me to the concept of school choice and motivated me to think independently and work hard. I will miss your mentorship. Jon you have been an inspiration to think methodologically about empirical questions. I am highly indebted to you for the detailed comments that significantly improved this dissertation.

I also acknowledge inspiration and help from all other EDRE faculty – Jay Greene, Robert Costrell, Gary Ritter, Gema Zamarro and Julie Trivitt. The EDRE staff, alumni and current students in the PhD program made me feel at home. Dirk, discussions with you were very amusing. My cohort family – Kaitlin Anderson, Leesa Foreman and Yujie Sude, it was an amazing experience to be with you. I wish each of you the best of success with a bright future.

Kaitlin Anderson, Patrick Wolf and I thank former EDRE alumni Anna Egalite and Jonathan Mills for statistical details regarding studies included in our analysis in chapters 2 and 3 of the dissertation. We also thank Mark Lipsey, Philip Gleason, John Witte, Gary Ritter and our other colleagues in EDRE for comments on earlier drafts of chapter 2. We thank Philip Gleason of Mathematica Policy Research and Robert Costrell of EDRE for comments on earlier drafts of chapter 3. We also thank Albert Cheng from Harvard University for assistance with graphics. Corey DeAngelis and I thank Robert Maranto of EDRE and Albert Cheng from Harvard University for comments on earlier drafts of chapter 4.

My academic accomplishments would not have been possible without the support of friends, relatives and members of my family: my mother Asma Shakeel, sister Urooj Shakeel and brother-in-law Atiq Ahmad Khan. My nephew Ibrahim gives me a reason to smile. My faith

provides me a reason to look at the sky and realize that the sky is the limit. Thanks to all who helped me connect to my faith and emotions.

Finally, thanks to all who rejected me in life—it would not have been without your rejection that I learned to persist.

Anyone can do a dissertation—yours truly proved it.

Dedication

This dissertation is dedicated to my late father Shakeel Akhtar. He was my moral guide and was most happy for my accomplishments.

Table of Contents

List of Tables

Chapter 2 The participant effects of private school vouchers around the globe: A meta-analytic and systematic review

Chapter 3 The juice is worth the squeeze: A cost-effectiveness analysis of the experimental evidence on private school vouchers around the globe

Chapter 4 Who is more free? A comparison of the decision-making of private and public school principals

Chapter 5 Who stays, who leaves? Determinants of principal attrition across school sectors

List of Figures

Chapter 2 The participant effects of private school vouchers around the globe: A meta-analytic and systematic review

List of Abbreviations

ITT = Intent to Treat

TOT = Treatment on Treated

TPS = Traditional Public School

NCES = National Center for Education Statistics

SASS = School and Staffing Survey

PFS = Principal Follow-up Survey

List of Published Papers

Chapter 3:

Shakeel, M. D., & DeAngelis, C. A. (2017). Who is more free? A comparison of the decision-making of private and public school principals. *Journal of School Choice, 11*(3), 442-457.

Chapter 1

Introduction

> *My research in Kenya suggested that these poor families had always been able to afford private schools. Before free primary education, they were already in private schools. The real conundrum for me was why the development experts hadn't already figured this out.*
>
> **James Tooley, 2009, p. 125.**

In the U.S., the status quo in traditional public education is to assign students to schools based on where they reside. School choice is the practice of letting parents choose schools rather than residentially assign them. School voucher programs are scholarship initiatives – frequently government funded or incentivized – that pay for students to attend private schools of their choice. America's first private school voucher initiative was launched in Milwaukee in 1991 (Witte, 2000). Private school choice programs provide public or private funds for families within the jurisdiction of the program to send their children to private schools. In 2018, 54 private school choice programs exist in the United States. Such programs extend beyond vouchers and also comprise education savings accounts (ESAs), tax-credit scholarships and town-tuition options for families living in rural areas (EdChoice, 2018). Inside the U.S., disadvantaged families are the main participant and beneficiary of voucher programs (Wolf, 2018).

Outside the U.S., private schooling is on the rise in developing countries despite increased spending on public education and near-universal access to free of cost public primary schools (Glewwe & Muralidharan, 2016). Indeed, the private school share accounts for a fifth of primary school enrollment in the developing world (Baum et al., 2014). Research attributes this demand for fee charging low-cost private schools among poor families to the low academic quality of public education and high rate of teacher absenteeism in government operated schools (Dixon, 2013; Shakeel & Wolf, 2018; Tooley, 2009).

It is argued that school choice can improve education systems through academic competition among schools and also by providing families a better match between their needs and schools' quality (DeAngelis & Erickson, 2018). Studying school choice is becoming increasingly common and more data about choice programs are available (EdChoice, 2018). In the 1970s and 80s, sociologists James Coleman and Thomas Hoffer (1987) reported a private school advantage in student achievement and graduation rates relative to traditional public schools. Competing evidence claims a public school achievement advantage relative to private schools (Lubienski & Lubienski, 2013). Several studies argue for advantages in student success (over a variety of outcomes such as achievement, attainment, civic values and non-cognitive skills) for one form of schooling over the other. Arguments for advantages of private over public schools are purely philosophical (Mill, 1962[1869]; Paine, 1791), rely on qualitative evidence (Stewart & Wolf, 2016), rely on quantitative evidence (Coleman & Hoffer, 1987), are based on literature reviews (Ashley & Wales, 2015; Coulson, 2009; Morgan, Petrosino, & Fronius, 2015), are based on field work (Dixon, 2013; Tooley, 2009) and are meta-analytic (Anderson, Guzman, & Ringquist, 2013; Fryer, 2017). Other arguments for advantages of private over public schools are based on economics (Friedman, 1955), a theory of bureaucracy and autonomy of private institutions (Chubb & Moe, 1988, 1990), degree of family involvement in schools (Hiatt-Michael, 2017), civic values (Wolf, 2007) and religion (Bryk, Lee, & Holland, 1993). Competing arguments exist for advantages of public schools over private schools (Bowles & Gintis, 2002; Gutmann, 1987; Lubienski & Lubienski, 2013). Some commentators argue for accommodating private interests publicly through pluralism in public institutions (Berner, 2017) while other commentators highlight that for-profit private enterprise may provide public education (Hess & Horn, 2013).

This dissertation reports four analyses to determine if private schooling enjoys an advantage over public schools in student success. Chapter 2 presents a first analysis that relies on causal evidence around the globe of the effect of using a school voucher on student achievement in math and reading. Chapter 3 presents a cost-effectiveness analysis to evaluate the program impacts beyond the perspective of student achievement. However, merely looking at limited evidence from student achievement may understate the impact of private schools. Therefore, Chapter 4 looks at the issue of a private school advantage in student success over public schools from the perspective of educational leadership. Using a nationally representative dataset, chapter 4 examines in which sector principals enjoy relatively more autonomy – i.e. if school principals can run schools without taking orders from higher-level authorities. Increased autonomy for school principals may incentivize them to be more innovative, adapt to changing needs of students and staff within schools and influence school-level activities that matter for student success (Hess, 2013; Ouchi, 2009). Private school principals may have an advantage over public school principals, due to the former's more autonomy in influencing school-level activities such as hiring and firing teachers, shaping the schools' curriculum and setting their schools' discipline policies. Finally, chapter 5 examines the determinants of principals staying in or leaving their profession in both school sectors. The analysis tests if principals are more likely to stay in the profession in either school sector. Private school principals do not have tenure, have lower salaries on average and less opportunities for professional development in comparison to public school principals. Yet, the existence of an increased likelihood of principals in private schools remaining in the profession in comparison to principals in public schools may indicate principal satisfaction that may be correlated with student success.

The first study is a meta-analytic consolidation of the evidence from all Randomized Control Trials (RCTs) on school vouchers evaluating the student-level achievement test score effects in math and reading. Reading and math test scores are available and known with acceptable reliability in the analyzed voucher programs. English test score impacts are analyzed as a subcomponent of reading results for developing countries due to the importance parents in developing countries assign to English as a medium of instruction (Azam, Aimee, & Prakash, 2010; Mitra et al., 2003; Sen and Blatchford, 2001; Tooley & Dixon, 2002). The results of this meta-analysis indicate that voucher programs globally tend to increase test scores over time, particularly in developing countries like Colombia and India with a large private-public school quality gap. Generally, the achievement impacts are positive and small but larger for reading than for math and for programs outside the US relative to those within the US.

Inside the US, a major proportion of increase in the public school expenditures comprise the cost of staffing and infrastructure (Scaffidi, 2017). In developing nations, often the private schools operate with limited infrastructure and lower teacher salaries than comparison public schools (Dixon, 2013; Tooley, 2009). The participating private schools in the voucher programs generally have lower per-student expenses in comparison to public schools within the jurisdiction of the same voucher interventions. A global assessment of the cost-effectiveness of private school choice has never been done. Chapter 3 uses a variety of data from state, national and international-level sources to analyze the cost-effectiveness of the causal evidence on private school vouchers internationally. Results from the cost-effectiveness analysis of the experimental evidence on school vouchers around the globe establish that vouchers are generally cost-effective and even null impacts in a voucher program, obtained at a lower cost than education in a public

school setting, may have a net benefit for society. More research is needed to study the scalability of private school choice programs.

While substantial school choice research focuses on student achievement outcomes, little has explored the mechanisms involved in producing such outcomes. The roles of school principals, for example, may vary between public and private schools. This could be important because principals should have an effect on school environment/quality. School principals possess superior knowledge of and an ability to influence school culture and learning practices (Hess, 2013; Ouchi, 2009). Chubb and Moe (1988) theorize the existence of greater autonomy in private schools related to structure, goals and school operations. Increased autonomy over school level activities such as curriculum, budget and personnel may allow principals to be more innovative and influence student learning positively. Highly effective principals increase student learning by two to seven months within a single school year (Branch, Hanushek, & Rivkin, 2013).

The third study presents a comparative analysis of the autonomy of private and public school principals using data from the School and Staffing Survey (SASS) 2011–2012. Chapter 4 adds to the literature on educational leadership and school choice by examining the differences in private and public school principals' abilities to influence important decisions at their schools from a nationally representative sample of 9,230 school principals. The self-reported influence of principals on seven school-related activities are analyzed: setting performance for students, establishing curriculum, determining the content of in-service professional development programs for teachers, evaluating teachers, hiring new full time teachers, setting discipline policies and deciding how school budgets will be spent.

Results from logistic regressions indicate that private school principals exhibit more autonomy in influencing school level policies, perhaps explaining private school advantages. In particular, private school principals have a higher likelihood of reporting major influence over performance standards, curriculum, professional development, hiring teachers, discipline policy, and budget decisions. Conversely, private school principals have a lower likelihood of reporting having a major influence on the evaluation of teachers. Principals are those administrators most aware of the daily issues at the school level: more autonomy may help them address the issues faster. Perhaps, due to the heavy bureaucratization and centralized functioning of the public school system, principal autonomy is more robust in private schools.

The final study in my dissertation compares principal attrition rates between the public and private sectors. School principals play a key role in determining school quality and a school's academic outcomes. Principal attrition poses a challenge in maintaining a school's academic environment. Principal turnover likely generates short-term shocks into the school system which may affect the school environment and student learning negatively (Branch, Hanushek, & Rivkin, 2013; Burkhauser et al., 2012; Mascall & Leithwood, 2010; Miller, 2013).

Earlier research has explored the determinants of principal turnover in traditional public schools using nationally representative samples (Boyce, & Bowers, 2016; Mitani, 2017; Sun & Ni, 2016; Tekleselassie & Villarreal, 2010). A comparative analysis of the attrition patterns of public and private school principals using nationally representative samples does not exist (Rangel, 2018). A comparative analysis of attrition patterns of school principals may be informative of the underlying school-level mechanisms that differentiate public schools from private schools, thus offering possible explanations for intersectoral differences in student learning.

Using data from the School and Staffing Survey 2011-2012 and the Principal Follow-up Survey 2012, the final study in the dissertation compares principal attrition between public and private schools. Principals' stated preferences on a prior year baseline survey are compared to their revealed status a year later on a follow-up survey. The research presented in chapter 5 reports significant differences in principal attrition between public and private schools. Results show that private school principals are significantly more likely than their public school peers to intend to stay in the principal profession. The chapter reports that principals stated preferences a year before stand in contrast to their revealed status a year later. Although privately school principals are significantly more likely to intend to stay in the profession in comparison to their public counterparts at the baseline year survey, at the follow-up year, the results do not reveal a statistically significant difference. As most private school principals teach at their schools, they may be utilizing principalship to gain teaching experience and later return to teaching in the same schools (Chubb & Moe, 1988). Such a change from principalship to teaching is less likely for principals in public schools, probably because of bureaucratic rules and specialization (Gerth & Mills, 1946), and a greater social distance between teachers and principals in public schools in comparison to private schools (Ingersoll, 2003).

The contributions of this dissertation are: a) establishing that private school vouchers tend to increase reading score more in comparison to math test scores, b) establishing modest advantages of private schooling on student achievement, particularly outside the US, c) establishing that vouchers are generally cost-effective and even null impacts, if obtained at a lower per-student cost in comparison to public school settings, may yield a net benefit for society, d) presence of more autonomy for private school principals over school-level activities in comparison to public school principals and e) an increased likelihood to remain in the

profession for the private school principal in comparison to their public counterparts, with evidence of a disconnect between principals' stated preferences versus their revealed status. These findings should be taken with caution, however, as merely 11 voucher programs have been analyzed causally around the globe, and data from studies on school principals are not connected to student outcomes.

The next chapter investigates the experimental evidence on the use of private school vouchers around the globe. Chapter 3 explores the cost-effectiveness of the aforementioned experimental evidence on vouchers. Thereafter, comparative principal autonomy in the public and private school sectors is studied in chapter 4 using a nationally representative dataset. Chapter 5 analyzes the same nationally representative dataset to study comparative principal attrition from the principal profession in public and private schools. Lastly, chapter 6 concludes the dissertation by summarizing the overall findings.

References

Anderson, M. R., Guzman, T., & Ringquist, E. J. (2013). *Evaluating the effectiveness of educational vouchers*. In E. J. Ringquist (Ed.), Meta-analysis for public management and policy (pp. 311-351), San Francisco, CA: Jossey-Bass.

Ashley, D. L., and Wales, J. (2015). The impact of non-state schools in developing countries: A synthesis of the evidence from two rigorous reviews. EPPI-Centre, SSRU, UCL Institute of Education, London, UK. [DFID Education Rigorous Literature Review].

Azam, M., Aimee C., & Prakash, N. (2010). The returns to English-language skills in India. CReAM Discussion Papers, No. 02-10, University College London.

Baum, D., Lewis, L., Lusk-Stover, O., & Patrinos, H. (2014). What Matters Most for Engaging the Private Sector in Education: A Framework Paper. The World Bank, Washington, DC.

Berner, A. R. (2017). *Pluralism and American public education: No one way to school*. Springer.

Boyce, J., & Bowers, A. J. (2016). Principal Turnover: Are there different types of principals who move from or leave their schools? A latent class analysis of the 2007–2008 Schools and Staffing Survey and the 2008–2009 Principal Follow-up Survey. *Leadership and Policy in Schools, 15*(3), 237-272.

Bowles, S., & Gintis, H. (2002). Schooling In capitalist America revisited. *Sociology Of Education*, 1-18.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter. *Education Next, 13*(1).

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). Catholic schools and the common good. Harvard University Press, Cambridge: MA.

Burkhauser, S., Gates, S. M., Hamilton, L. S., & Ikemoto, G. S. (2012). *First-Year principals in urban school districts: How actions and working conditions relate to outcomes* (Technical Report). Santa Monica, CA: RAND Corporation.

Chubb, J. E., & Moe, T. M. (1988). Politics, markets, and the organization of schools. *American Political Science Review, 82*(04), 1065–1087.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities* (Vol. 41). New York: Basic Books.

Coulson, A. J. (2009). Comparing public, private, and market schools: The international evidence. *Journal of School Choice*, *3*(1), pp. 31-54.

DeAngelis, C. A., & Erickson, H. H. (2018). What leads to successful school choice programs? A review of the theories and evidence. *Cato Journal*, *38*(1), 247-263.

Dixon, P. (2013). *International aid and private schools for the poor: Smiles, miracles and markets*. Edward Elgar Publishing.

EdChoice (2018). *The ABCs of School Choice, 2018 Edition.* Indianapolis, IN: EdChoice.

Firestone, W. A., & Shipps, D. (2005). How do leaders interpret conflicting accountabilities to improve student learning. In William Firestone and Carolyn Riehl (Eds.) *A new agenda for research in educational leadership* (pp. 81–100).New York, NY: Teachers College Press, Columbia University.

Friedman, M. (1955). The role of government in education. In R. A. Solo (Ed.), *Economics and the public interest* (pp. 123–144). New Brunswick, NJ: Rutgers University Press.

Fryer Jr., R. G. (2017). *The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments*. In A. V. Banerjee & E. Duflo (Ed.) Handbook of Economic Field Experiments, 2, 95-322. Amsterdam, Netherlands: Elsevier.

Gerth, H. H., & Mills, C. W. (1946). *From Max Weber: Essays in sociology.* New York, NY: Oxford University Press.

Glewwe, P., & Muralidharan, K. (2016). *Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications*. Handbook of the Economics of Education, 5, 653-743. Amsterdam, Netherlands: Elsevier.

Gutmann, A. (1987). *Democratic Education.* Princeton, N.J.: Princeton University Press.

Hess, F. M. (2013). *Cage-busting leadership*. Cambridge, MA: Harvard Education Press.

Hess, F. M., & Horn, M. B. (Eds.). (2013). *Private enterprise and public education*. Teachers College Press.

Hiatt-Michael, D. B. (Ed.). (2017). *Family involvement in faith-based schools*. Charlotte, NC: IAP.

Ingersoll, R. M. (2003). *Who controls teachers' work?: Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Levenson, N. (2012). *Smarter budgets, smarter schools*. Cambridge: Cambridge, MA: Harvard Education Press.

Lubienski, C. A., & Lubienski, S. T. (2013). *The public school advantage: Why public schools outperform private schools*. University of Chicago Press.

Mascall, B. & Leithwood, K. (2010). Investing in leadership: The district's role in managing principal turnover. *Leadership and Policy in Schools*, *9*(4), 367-383.

Mill, J. S. (1962 [1869]) *Utilitarianism, on liberty, essay on Bentham.* (Warnock, M. ed.) Meridian.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review, 36*, 60-72.

Mitani, H. (2017). Principals' Working Conditions, Job Stress, and Turnover Behaviors under NCLB Accountability Pressure. Paper presented at the annual meeting of the Association for Education Finance and Policy.

Mitra, S., Tooley, J., Inamdar, P., & Dixon, P. (2003). Improving English pronunciation: An automated instructional approach. *Information Technologies and International Development, 1*, 75-84.

Morgan, C., Petrosino, A., and Fronius, T. (2015). The impact of school vouchers in developing countries: A systematic review. *International Journal of Educational Research*, *72*, pp.70-79.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Piscataway, NJ: Transaction Publishers.

Ouchi, W. G. (2009). *The secret of TSL: The revolutionary discovery that raises school performance*. New York, NY: Simon and Schuster.

Paine, T. (1791). *The rights of man: Answer to Mr. Burke's attack on the French revolution*. J. S. Jordan.

Rangel, V. S. (2018). A review of the literature on principal turnover. *Review of Educational Research*, *88*(1), 87-124.

Scaffidi, B. (2017). Back to the staffing surge. EdChoice.

Shakeel, M.D., & Wolf, P.J. (2018). *Poor parents are careful choosers: Dispelling the myth that school choice harms the poor*. In P. Dixon & S. Humble (Ed.), School choice around the world: What we can learn from national and international contexts. Amsterdam, The Netherlands: IEA Publications.

Shipps, D., & White, M. (2009). A new politics of the principalship? Accountability-driven change in New York City. *Peabody Journal of Education, 84*(3), 350–373.

Stewart, T., and Wolf, P. J. (2016). *The school choice journey: School vouchers and the empowerment of urban families*. Springer.

Sun, M., & Ni, Y. (2016). Work environments and labor markets: Explaining principal turnover gap between charter schools and traditional public schools. *Educational Administration Quarterly, 52(1),* 144–183.

Tekleselassie, A. A., & Villarreal, III, P. (2010). Career mobility and departure intentions among school principals in the United States: Incentives and disincentives. *Leadership and Policy in Schools, 10(3),* 251–293.

Tooley, J., & Dixon, P. (2002). Private schools for the poor: A case study from India. Reading, UK: CfBT.

Tooley, J. (2009). *The beautiful tree: A personal journey into how the world's poorest people are educating themselves*. Washington, DC: Cato Institute.

White, M. A. (2006). *A case of New York City high school principals' experiences with multiple accountability pressures and accompanying models for decision-making*. New York, NY: Columbia University Teacher's College.

Witte, J. F. (2000). *The market approach to education: An analysis of America's first voucher program*. Princeton University Press.

Wolf, P. J. (2018). Programs benefit disadvantaged students. In forum: Taking stock of private-school choice. *Education Next, 18*(2).

Chapter 2

The participant effects of private school vouchers around the globe: A meta-analytic and systematic review

M. Danish Shakeel (mdshakee@uark.edu)

Kaitlin P. Anderson (kaitlina@uark.edu)

Patrick J. Wolf (pwolf@uark.edu)

Earlier versions of this chapter were presented at conferences of the Society for Research on Educational Effectiveness, Association for Education Finance & Policy and Association for Public Policy Analysis & Management International.

1. Introduction

In the United States, families are typically residentially assigned to schools. School choice is the practice to let parents choose schools for their children rather than residentially assign them. School choice is a salient market-based school reform globally. School vouchers are a mechanism by which resources are provided to families that enable them to attend a private school of their choosing (Wolf, 2008a). Strictly speaking, a private school choice initiative is only a "voucher program" if the government funds the program directly out of an appropriation and the "voucher" only purchases educational services from a single provider. Other private school choice initiatives are funded indirectly, through tax credits provided to businesses or individuals who contribute to nonprofit scholarship-granting organizations. Such arrangements are commonly called tax-credit or opportunity scholarship programs. This study is the first meta-analytic consolidation of the evidence from all Randomized Control Trials (RCTs) evaluating the student-level achievement test score effects of school vouchers internationally. Our search process turned up 9,443 potential studies, 20 of which ultimately were included. These 20 studies represent 11 different voucher programs, eight in the U.S. and three in non-U.S. countries. We present math and reading outcomes when available, and present English results as a subcomponent of the reading effects in the international context to account for differences between English and the local language.

The funding structure of a voucher program in Colombia involved the use of student performance incentives, and the voucher program in Louisiana contained test-based accountability provisions not found in any of the other programs included in our study. We provide meta-analytic results that exclude these two outlier programs as a robustness check.

Economist Milton Friedman (1955) put forth the first concrete school voucher proposal. He argued that government should provide funds in support of compulsory education but need not deliver the schooling itself. Vouchers are a form of government outsourcing wherein the government provides funds to socioeconomically disadvantaged families to select private schools for educating their children. Supporters of vouchers claim that participating students will learn more, either by accessing higher-quality schools, or because the chosen private school will be a better match for the student's particular needs. It is expected that parents of these children are motivated to figure out their child's needs.

Whether or not students benefit from school vouchers is a fiercely contested empirical question (Doolittle & Connors, 2001). For example, education historian Diane Ravitch describes school vouchers as a "hoax" that has failed to benefit participants (Ravitch, 2014). Richard Murnane (2005, p. 181), in contrast, argues: "Providing families who lack resources with educational choices makes sense. The consequences of attempting to do this through a large-scale voucher…system are unknown. Carefully designed experiments could provide critical knowledge."

Experimental design is critical to evaluating school voucher programs because of concerns about selection bias. Some families may be more educationally motivated to find a better match between their child's needs and the chosen private school. Additionally, some families may seek quality education in accord with a religious nature and mission orientation of the chosen private school. Generally, we expect that more motivated and able families self-sort into private schools and/or voucher programs, though some within-study comparisons indicate this is not always the case (e.g. Anderson & Wolf, 2017). Fortunately, much of the research on school vouchers in the U.S., and some of the evaluations abroad, has involved random

assignment experiments that control for self-selection in expectation. Random assignment assumes that assignment of a private school to a family is no longer due to selection by a family but rather due to chance. Thus, any effect of private schooling is causal in theory. However, the caveat is that families who apply to such voucher programs may be motivated differently than families that do not apply for the voucher programs. Similarly private schools that participate in voucher programs may be different than private schools that choose not to participate in voucher programs. Both of these factors limit the external validity of experimental voucher findings.

In this meta-analysis we consolidate the evidence from 20 experimental evaluations of the achievement impacts of private school choice programs in the U.S., India, and Colombia. We primarily focus on estimates of the effect of using a voucher to attend a private school, referred to as the Treatment on the Treated (TOT) program effect, for several reasons. First, expanding access to private schools of choice is the key purpose of school voucher programs. The effects of private schooling specifically on the kinds of students who will access it with the support of vouchers is the most policy-relevant piece of information regarding school choice programs. Thus, the TOT effect of private school voucher use is considered by some analysts to be the policy-relevant effect parameter for school-choice interventions (Bifulco, 2012; Cowen, 2008; Howell, Wolf, Campbell, & Peterson, 2002). Second, the TOT effect is less sensitive to the rate at which voucher recipients use a voucher when offered and the rate at which control group students "crossover" to private schooling in the experiment. As such, the TOT estimate of the effect of private schooling is more consistent across programs. We also report in the appendices the experimental results for the mere offer of a voucher through winning a lottery, referred to as the Intent-to-Treat (ITT) effect, which is the pure experimental impact from the evaluations but lacks the advantages of the TOT effect described above.

We focus on reading and mathematics achievement for three reasons. First, they are measures of academic achievement. Schools are charged with enhancing student learning, so test-score outcomes are a proper measure of the effect of an intervention like school choice that expands the schooling options of students. Second, measures of reading and math achievement with known and acceptable reliability[1] are available around the globe. Third, the reading and math effects of private school choice programs are the only types of program effects that are sufficiently common to provide the foundation for a meta-analysis.

Very few voucher RCTs have systematically evaluated non-math and non-reading test score or even non-achievement outcomes. The evaluation of the PACES program in Colombia showed that lottery winners were ten percentage points more likely to finish 8th grade after three years. Lottery winners were also less likely to marry or cohabit as teenagers (Angrist, Bettinger, Bloom, King, & Kremer, 2002). Wolf et al. (2013) reported a positive impact on high school graduation in the evaluation of the D.C. Opportunity Scholarship Program (OSP). The Louisiana Scholarship Program (LSP) showed statistically significant negative impacts of voucher usage on student achievement in science and social studies in its first (Abdulkadiroglu, Pathak & Walters, 2015) and second (Mills & Wolf, 2016) years. By its third year, the LSP negative effect was statistically significant only for social studies (Mills & Wolf, 2017). Bettinger and Slonim (2006) reported positive and statistically significant effects of a voucher intervention on students' altruism towards charitable organizations but not towards their peers. The Andhra Pradesh school choice experiment showed no difference between test scores of lottery winners and losers on science and social studies after two and four years of the program (Muralidharan & Sundararaman, 2015). Although these non-reading and non-math experimental outcomes from

[1] Reliability of a test refers to the extent to which the test is stable and consistent in measuring the intended outcome(s).

voucher studies are interesting, there are too few of them to consolidate into our formal meta-analysis without introducing a substantial amount of statistical noise due to the diversity in distinctive outcomes and their associated measurement issues. Thus, we focus on reading and math test scores here.

Additionally, we examine English impacts as a subcomponent of reading results for developing countries. We do so because parents in developing countries assign high importance to English during the school selection process due to its likely association with increased academic opportunities and a higher economic return from the job market (Azam, Aimee, & Prakash, 2013; Mitra, Tooley, Inamdar, & Dixon, 2003; Sen and Blatchford, 2001; Tooley & Dixon, 2002).

2. Private school choice programs around the world

Government or philanthropic efforts providing greater access to private schools of choice are common around the world (e.g. Glenn, De Groof, & Candal, 2012a; 2012b; 2012c; 2012d; Wolf & Macedo, 2004). Voucher programs are either universal or targeted. Universal programs offer funding to all school-age children in a jurisdiction.

Universal private school choice programs operate in the Netherlands, Belgium, Denmark, Sweden, France, and other European and Commonwealth countries, mainly based on a constitutional right for parents to educate their children within a particular religious, philosophical, or pedagogical tradition (Glenn, 1989). A universal school voucher program has operated in Chile since the 1980s (Mizala & Romaguera, 2000). Of the 54 private school choice programs in 28 U.S. states plus the District of Columbia, nine are universal or nearly universal in design, only limited in scope due to funding amounts (EdChoice, 2017). The universal programs

in the U.S. are in the states of Arizona (2 programs), Georgia, Montana, and Nevada; the rural areas of Maine, New Hampshire and Vermont; and the urban area of Cleveland, Ohio.

Targeted programs have eligibility requirements that limit private school choice to certain disadvantaged populations of students, typically those with low family incomes or disabilities. Private school choice programs targeted to low-income students operate in Colombia, regions of India and Pakistan, and several developing countries in Africa. Many of these initiatives provide the equivalent of around $200/year to fund schooling at very low-cost private schools operated by education entrepreneurs (Dixon, 2013; Tooley, 2009). The U.S. is home to 45 targeted private school choice programs, of which 23 are means-tested, 17 are limited to students with disabilities, one is restricted to students attending failing public schools in Ohio, and four are doubly-targeted to low-income students in low-performing schools (EdChoice, 2017). The means-tested private school choice programs in the U.S. provide scholarships that range widely in size from around $1,000 to $13,000, with the lower-cost scholarship programs typically requiring families to contribute to the cost of private schooling. Vouchers for students with disabilities are typically larger, cover the full cost of educating the child, and in some cases are priced on a sliding scale based on the severity of the child's disability.

Private school choice is increasingly common throughout the U.S.[2] and the world. The research base on the effectiveness of voucher programs has been reviewed by multiple scholars over the past nine years, but those reviews do not render a clear judgment regarding whether students are helped or harmed academically by access to private school choice. In the next

[2] The newest form of private school choice in the U.S., education savings accounts (ESAs), permit parents to secure educational services from multiple private providers (Butcher & Burke, 2016). Other programs, in the U.S. and globally, use scholarships funded through private donations and philanthropy, with no extra government tax incentive involved. This study does not cover ESAs, because no experimental evaluations of their effects on student outcomes yet exist. Since privately-funded scholarships, tax-credit scholarships and school vouchers accomplish the same general purpose of expanding access to private schooling, we generally treat all three types of programs as equivalent within this study.

section, we review the systematic reviews of voucher effectiveness to evaluate if they reveal a clear, consistent and indisputable judgement on the achievement effect of vouchers.

3. A review of the systematic reviews of voucher effectiveness

From 2008 through 2017, 15 reviews of the achievement effects of private school choice in the U.S. have been published as reports, working papers, or journal articles. This plethora of school voucher research reviews underscores the salience of the topic. A meta-analysis is a statistical method to combine evidence from several studies on a chosen topic to develop an overall conclusion on the effectiveness of the evidence on the selected topic. Often, a meta-analysis combines the effect sizes on a selected outcome across several studies based on a similar methodology to evaluate if the overall evidence on the topic is statistically significant. Without methodological coherence, one cannot tell if the findings from different studies vary because the intervention truly had heterogeneous effects or simply because some of the research designs were biased in estimating the effect. The ideal meta-analysis is up-to-date, includes only studies with similar methodologies, is comprehensive, and provides a specific and verifiable determination of the average effect of an intervention on an important outcome (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rossi, Lipsey, & Freeman, 2004, pp. 324-328). Private schools account for at least a fifth of the share of total primary school enrollment in developing countries (Baum et al., 2014). The private school sector has witnessed continuous growth in the last two decades in the developing world, despite near-universal access to free public primary schools and increases in government spending on public education (Glewwe & Muralidharan, 2016). Thus, an ideal meta-analysis on the effectiveness of private school vouchers should incorporate studies from around the globe. We add to the standard set of four desirable meta-analytic features a fifth one: that the review be global if programs operate in multiple countries.

None of the 15 existing reviews, however, satisfies even three of the five criteria for an ideal meta-analysis, and seven satisfy none of them (Table 1).

Although some of these reviews are nearly current (Egalite & Wolf, 2016; Forster, 2016; Fryer, 2017; Lubienski & Brewer, 2016), none of them include the two most recent experimental studies of school vouchers (Dynarski, Rui, Webber, Gutmann, & Bachman, 2017; Mills & Wolf, 2017). Only seven of the reviews are methodologically coherent, restricting their scope to evaluations with experimental designs (Egalite & Wolf, 2016; Forster, 2011; 2013; 2016; Fryer, 2017; Morgan, Petrosino & Fronius, 2015; Wolf, 2008b). Only two reviews (Egalite & Wolf, 2016; Forster, 2016) include all studies available at the time of publication that fit the inclusion category of the authors of these reviews, described in Section 4 and Appendix A. Only two reviews (Anderson, Guzman & Ringquist, 2013; Fryer, 2017) are formal meta-analyses that include overall effect point estimates and confidence levels, with the other 13 studies being only literature reviews. The Anderson, Guzman, and Ringquist (2015) meta-analysis includes 17 U.S. school voucher studies of widely varying methodological designs. The Fryer (2017) meta-analysis is limited to school choice experiments but mixes RCTs of public school choice in with those of private school choice.

A final limitation of 13 of the existing reviews of voucher studies is their geographic focus. Eleven of them limit their consideration to the U.S. while Morgan, Petrosino and Fronius (2015) restrict their scope to developing countries and Fryer (2017) restricts its scope to developed countries. Epple, Romano, & Urquiola (2015) and Coulson (2009) are the only reviews to include both U.S. and non-U.S. studies.

Table 1. Extent to Which Previous Voucher Reviews Satisfy the Conditions for an Ideal Meta-Analysis (an "X" indicates it satisfies the condition)

Review	Up-to-Date	Methodologically Coherent	Comprehensive	Specifies Average Effect	Global	Conclusions
Miron, Evergreen & Urschel (2008)						Null to positive, encouraging
Lubienski & Weitzel (2008)						Null to positive, discouraging
Wolf (2008b)		X				Null to positive, encouraging
Rouse & Barrow (2009)						Null to positive, encouraging
Coulson (2009)					X	Null to positive, encouraging
Usher & Kober (2011)						No effect
Forster (2011)		X				Consistently positive
Forster (2013)		X				Consistently positive
Anderson, Guzman & Ringquist (2013)				X		Null to positive, discouraging
Epple, Romano & Urquiola (2015)					X	Null to positive, encouraging
Morgan, Petrosino & Fronius (2015)		X				Consistently positive
Forster (2016)		X	X			Consistently positive
Lubienski & Brewer (2016)						Null to positive, discouraging
Egalite & Wolf (2016)		X	X			Null to positive, encouraging
Fryer (2017)		X		X		Null to positive, encouraging

Previous selective reviews vary greatly in their conclusions regarding the effectiveness of vouchers. Based on these reviews, school vouchers have no effect on student achievement (Usher & Kober, 2011), consistently improve achievement (Forster, 2011; 2013; 2016; Morgan,

Petrosino & Fronius, 2015) or produce some mix of null to positive effects that are either encouraging to the authors (Egalite & Wolf, 2016; Epple, Romano, & Urquiola, 2015; Coulson, 2009; Fryer, 2017; Rouse & Barrow, 2009; Miron, Evergreen, & Urschel, 2008; Wolf, 2008b) or discouraging to them (Anderson, Guzman, & Ringquist, 2013; Lubienski & Brewer, 2016; Lubienski & Weitzel, 2008). Most of the individual studies covered in the reviews are only modestly powered to detect voucher effects, having analytic samples of less than 1,000 students in the final evaluation year. The many findings of "no significant effects" could be due to either limited power of studies to detect a significant effect, noisy data or the absence of a true school voucher effect. None of the reviews includes recent studies from 2017 that have generated extensive policy and media interest. Given the lack of any contemporary, complete, statistical meta-analysis of the effect of private school vouchers on student achievement around the world, this study offers a clear contribution to the literature on private school choice.

4. Method

4-A. Search strategy

For this meta-analysis, we identified publications from systematic computer and networked searches primarily through the EBSCO, JSTOR, and ProQuest databases, with Google Scholar used for additional checks. Lastly, we utilized subject matter experts in the field and snowballing techniques to find additional relevant studies missed by the systematic search. We identified 9,443 articles which ultimately produced 20 qualified RCTs meeting our key inclusion criteria. Specifically, to be included, a study had to be an experimental evaluation based on random assignment, include as an outcome individual student-level test scores in either math, reading, or both, and be available in English (see Section 4-B and Appendix A for details). Only studies of the participant effects of voucher programs were included; for example, we did

not include studies of the competitive effects of vouchers on traditional public school students or papers focusing on fiscal outcomes. Similarly, studies on participant effects of other private school choice programs like Tax Credit Scholarship programs were excluded, as those programs are not relevant to this review. Studies were not limited by publication date or publication status. The initial search process began in mid-2015 and yielded 16 qualified RCTs, after which we relied on network searches that identified four additional studies (Abdulkadiroglu, Pathak, & Walters, 2015; Dynarski, Rui, Webber, Gutmann, & Bachman, 2017; Mills & Wolf, 2017; Wolf, Egalite, & Dixon, 2015). The network searches involved contacting scholars who had produced experimental voucher studies previously and asking them if we were missing any unpublished or contemporary studies. We also monitored a daily Google alert list generated by the keywords "school choice."

We focused on identifying RCT (a.k.a. experimental) studies for several reasons. First, RCTs are the "gold standard" of program evaluation in terms of assessing causal relationships (e.g. Mosteller & Boruch, 2002; Pirog, Buffardi, Chrisinger, Singh, & Briney, 2009; Rossi, Lipsey, & Freeman, 2004). The random assignment of subjects in RCTs creates a treatment group (in this case, those receiving the offer of a voucher) and a control group (those who did not receive the offer of a voucher) that are similar to each other in expectation regarding all measurable and unmeasurable characteristics. This similarity is important when evaluating private school choice programs, since families who self-select into private schools likely differ from other families in unmeasurable ways that affect subsequent student achievement. In RCTs, access to private schooling through a voucher is random, solving the selection bias problem in

expectation. Thus, random assignment of a voucher generates strong internal validity, which is the confidence that any observed differences actually are due to the program.

Other voucher studies have been conducted using quasi-experimental methods such as Regression Discontinuity Design, propensity score matching, or use of control variables to try to minimize selection bias. The results from these non-experimental studies range from positive in some subject areas and years (Witte, Wolf, Cowen, Carlson, & Fleming, 2014) to consistently null (Witte, 2000) to negative in some subject areas (Figlio & Karbownik, 2016). While well-designed quasi-experimental school choice studies can approximate the results of experimental evaluations under certain conditions (Bifulco, 2012; Fortson, Verbitsky-Savitz, Kopa, & Gleason, 2012), at least sometimes they fail to reveal the true causal effects of attending private school (Anderson & Wolf, 2017; Betts, Tang, & Zau, 2010; Cowen, 2008). Moreover, it is often difficult to measure how much bias is present when relying on quasi-experimental approaches instead of experimental ones (Pirog, Buffardi, Chrisinger, Singh, & Briney, 2009). On the other hand, some quasi-experimental approaches tend to have greater external validity and generalizability than experimental ones. However, since the quasi-experimental approaches do not rely on distinctive randomly-assigned samples of students in a single location, our study compensates at least partially for the limited external validity of individual voucher experiments by systematically compiling the experimental results across varied student populations. The RCT approach generates the internal validity or causality of our study while the meta-analytic approach provides some external validity as the contexts vary across studies.

Compliance in RCTs refers to program participants who continue to remain in their originally assigned treatment or control groups. Control crossover refers to program participants who switch to the treatment group experience after being initially assigned to the control group.

The evidence from RCTs is more reliable under conditions with strong compliance, as control crossover or treatment non-compliance reintroduces selection bias if analysts merely compare the outcomes from voucher users with those from the subgroup of control group students who remained in public schools. Explicit TOT estimation methods eschew that simple but problematic approach and are designed to recover unbiased estimates of the effect of actually experiencing the intended school voucher treatment when some students randomly assigned to the treatment group fail to use their voucher and some students randomly assigned to the control group attend private school without a voucher. The three TOT strategies used in the studies that inform this meta-analysis were calculation of the Complier Average Causal Effect (CACE) (Cowen, 2008), Instrumental Variables (IV) analysis (Heckman, 1996), and the Bloom adjustment (Bloom, 1984). All three of these approaches seek to construct comparisons of the average outcomes for program participants compared to the outcomes those same students would have experienced if not for participation in the program. The RCT studies recover unbiased estimates by utilizing the underlying random assignment of access to the voucher. Original voucher RCT studies that did not provide the information required for us to judge the appropriateness of the use of the specific TOT method in that case (e.g. Greene, Peterson, & Du, 1999; Rouse, 1998) were excluded from the TOT calculations for the meta-analysis. Whenever the original voucher RCT report did not include TOT estimates (probably due to their emphasis on ITT impacts over TOT impacts of the voucher programs) but did include sufficient information for us to calculate the TOT, we did so using a Bloom adjustment because it requires fewer assumptions than CACE or IV approaches.[3]

[3] For example, CACE estimates of the TOT are only unbiased if the characteristics of treatment decliners are accurate predictors of which members of the control group similarly would have declined a voucher if they had won the lottery (Cowen, 2008). Instrumental variables estimates of the TOT are only unbiased if the effects of winning or losing a voucher lottery on student outcomes come solely through greater access to private schooling and not

Our second reason to focus on RCTs is because Anderson, Guzman, & Ringquist (2013) established that the conclusion one draws about the efficacy of vouchers is heavily influenced by which studies are selected for review. Quasi-experimental voucher studies have produced smaller voucher effect sizes and fewer statistically significant results, arguably because weaknesses in the research design and comparison groups biased the impact estimates towards zero. If one has to believe either the results from RCTs or the results from non-RCTs regarding the effects of a given intervention, then one should believe the results from RCTs because they have much stronger internal validity.

Third, we expected there would be a sufficient number of voucher effect estimates from RCTs to produce a reliable estimate of voucher impacts. Since the geographical scope of our search was global, and the temporal scope of our search was unrestricted, we assumed that we would identify many voucher achievement studies, even restricting our sample to gold standard experiments. Despite setting up a broad search, we only recovered 20 studies with 260 effect estimates. RCTs of private school choice outside of the U.S. are still relatively rare.

4-B. Selection Process

Each of the 9,443 collected sources was reviewed by two separate team members based on its title and abstract in order to determine whether it justified a full review. To be included in the meta-analysis, studies had to include student-level test score achievement effects of a private school voucher program in at least math or reading. Studies dealing with other impacts of vouchers such as competitive effects or fiscal impacts were excluded. We did not include

---

through an emotional response to the lottery outcome (e.g. Rhinesmith, 2017). Bloom adjustments generate their TOT simply by dividing the average difference in outcomes between all treatment and all control group members by the proportion of treatment students who ever used their private school voucher. This approach relies upon the simple assumption that any differences between the average outcomes of the treatment and control group students must have been produced solely by treatment uses, since non-users could have not been affected by the treatment (Bloom, 1984).

graduation rate, college attainment, or civic values outcomes in the current study as the focus of this study is only on achievement impacts in math and reading scores. Most of the articles excluded were theoretical discussions or opinion pieces without quantitative evidence, were focused on other issues such as competitive or fiscal rather than student achievement effects, or were merely quasi-experimental.[4] We only included studies published in English or with English translations.

A total of 6,549 sources were excluded based on title and/or abstract reviews. In some cases, the two coders initially disagreed over whether or not to include a particular study. When that happened, the two coders came to a consensual conclusion. Unless there was a clear reason to exclude a study, it was included in the full article review round, when more information would be available to judge its merits. Our full-article review process resulted in 16 studies remaining in the sample.

Our supplemental network search resulted in four additional articles added to the sample – two of the recently implemented Louisiana voucher program (Abdulkadiroglu, Pathak, & Walters, 2015; Mills & Wolf, 2017),[5] one of a philanthropic voucher program in Delhi, India (Wolf, Egalite, & Dixon, 2015), and one of a recent evaluation of the federal voucher program in Washington, D.C. (Dynarski, Rui, Webber, Gutmann, & Bachman, 2017). The four studies found in the network search did not show up in our computerized search as the four studies were either not in wide circulation on internet search engines or were not publicly available during the initial phase of our computerized search. In total, 20 RCT studies met the qualifications for

[4] A surprising number of education evaluations are described as "experimental" via keywords or in their abstracts but, upon a closer reading, actually do not create their comparison groups via random assignment and therefore are merely quasi-experimental.

[5] A third study of the LSP, Mills and Wolf (2016), is not reported separately, as Mills and Wolf (2017) includes the same effect sizes after years one and two, as well as an additional impact estimate after three years.

inclusion. Appendix A contains the details regarding the studies that were identified and eliminated at each stage. In Table 2 we summarize the studies, presenting attrition rates in terms of both sample attrition (the percent of study participants who are not observed in the final year) and program attrition (the percent of students offered a voucher who were not using the voucher in the final year). The program attrition rate is the additive inverse of the voucher take-up rate, which is also reported for other years, when available. The level of randomization of each study was at the child/family level.[6] Table 2 also indicates how the TOT effects were estimated or calculated, if applicable.

[6] One study (Muralidharan, & Sundararaman, 2015) involved randomization at two stages (randomly assigned students within randomly assigned villages); the two-stage randomization allows for higher external validity of the findings to the larger sample. As the randomization in the studies occurred at the child/family level, the studies can be combined for obtaining a meta-analytic estimate.

Table 2: Description of 20 RCT Studies included in Meta-Analysis

Program Evaluated	Publication	Years of Treatment	Study Duration	Grades	Sample Size (First Outcome Year)	Program Attrition (Final Year)	Sample Attrition (Final Year)	Control Crossover Rate	Voucher Take-up Rate	TOT method
Andhra Pradesh (AP) School Choice Experiment	Muralidharan & Sundararaman (2015)	4	2008-2012 (4 years)	1 to 5	4,620	49%	20.7% English; 68.1% Hindi; 17.5% Telugu; 17.5% Math	Not provided	61% (initial) or 51% (final year)	Original research team used Bloom (1984) adjustment
Charlotte Children's Scholarship Fund	Greene (2000)	1	1999-2000 (1 year)	2 to 8	357	51.60%	60%	Not provided	48.40%	N/A
	Cowen (2008)	1	1999-2000 (1 year)	2 to 8	347	25.50%	70%	Not provided	74.50%	Complier average causal effect (CACE), the mean treatment outcome across compliers
Children's Scholarship Fund (Toledo, OH)	Bettinger & Slonim (2006)	3	1998-2001 (4 years)	K to 8	186	43%	92%	21% (at time of survey) but 39% of control units attended private school at some point post-lottery	57% (at time of survey) but 65% of winners attended private school at some point post-lottery	Meta-analytic team estimated TOT using Bloom (1984) adjustment
District of Columbia Opportunity Scholarship Program (OSP)	Wolf, Kisida, Gutmann, Puma, Eissa, & Rizzo (2013)	4	2004-2009 (6 years)	K to 12	1,649	28.7%	37.8% Treatment, 48.5% Control	23.10%	74.3% (first year), 71.3% (final year)	Original research team used Bloom (1984) adjustment
	Dynarski, Rui, Webber, Gutmann, & Bachman (2017)	1	2012-2015 (3 cohorts, 1 outcome yr. each)	K to 12	1,077	30%	24% Reading; 25% Math	10% (first year)	70% (initial)	Original research team used Bloom (1984) adjustment
Ensure Access to Better Learning Experiences (ENABLE)	Wolf, Egalite, & Dixon (2015)	2	2011-2013 (2 years)	K to 2	1,306	11%	N/A	25% in total (or 30% of controls with outcome data)	89% (among students with second year outcome data)	Instrumental Variables analysis

Table 2 Cont'd: Description of 20 RCT Studies included in Meta-Analysis

Program Evaluated	Publication	Years of Treatment	Study Duration	Grades	Sample Size (First Outcome Year)	Program Attrition (Final Year)	Sample Attrition (Final Year)	Control Crossover Rate	Voucher Take-up Rate	TOT Method
Louisiana Scholarship Program (LSP)	Abdulkadiroglu, Pathak, & Walters (2015)	1	2012-2013 (1 year)	3 to 8	N/A	27%	N/A	5%	73%	Instrumental Variables analysis
	Mills & Wolf (2017)	3	2012-2015 (3 years)	3 to 8	1,184	46%	10%	6% (Year 1), 15% (Year 2), 14% (Year 3)	80% (Year 1), 64% (Year 2), 54% (Year 3)	Instrumental Variables analysis
Milwaukee Parental Choice Program (MPCP)	Rouse (1998)	4	1990-1994 (5 years)	K to 8	1,343	64.7%	N/A	Not provided	86.5% (initial fall), 35.3% (final spring)	Insufficient Information
	Greene, Peterson, & Du (1999)	4	1990-1994 (5 years)	K to 8	816	N/A	60% Treatment, 52% Control	Not provided	Not provided	N/A (TOT in paper was non-experimental)
Parents Advancing Choice in Education (Dayton, OH)	Peterson, Howell, Wolf, & Campbell (2003)	2	1998-2000 (2 years)	K to 12	404	N/A	51%	18% (first year) or 10% (first and second years)	78% (first year), 60% (second year)	Instrumental Variables analysis
Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES)	Angrist, Bettinger, Bloom, King, & Kremer (2002)	3	1995-1999 (4 years)	6 to 9	283	43%	75.30%	Majority of controls went to private school at least one year*	By survey year, about 57% still using voucher**	Meta-analytic team Bloom (1984) adjusted
	Angrist, Bettinger, & Kremer (2006)	7	1994-2001 (8 years)	6 to 11	3,541	50% (after three years)	12.40%	Majority of controls went to private school at least one year*	50% after three years	Meta-analytic team Bloom (1984) adjusted

Table 2 Cont'd: Description of 20 RCT Studies included in Meta-Analysis

Program Evaluated	Publication	Years of Treatment	Study Duration	Grades	Sample Size (First Outcome Year)	Program Attrition (Final Year)	Sample Attrition (Final Year)	Control Crossover Rate	Voucher Take-up Rate	TOT Method
School Choice Scholarships Foundation Program (NYC)	Barnard, Frangakis, Hill, & Rubin (2003)	1	1997-2000 (4 years)	1 to 4	525	20% to 27% (overall, not final year)	22.30%	6% to 10%***	73% to 80%***	CACE
	Peterson, Howell, Wolf, & Campbell (2003)	3	1997-2000 (4 years)	1 to 4	1,434	30%	33%	5% (first) or 3% (both first and second years)	82% (first year), 79% (both years), 70% (final year)	Instrumental Variables analysis
	Krueger & Zhu (2004)	3	1997-2000 (4 years)	K to 4	2,080	41.3% (from Bitler et al.)	36.20%	11% of control students attended private school at least one year	77% used voucher at least one year	Instrumental Variables analysis
	Jin, Barnard, & Rubin (2010)	1	1997-2000 (4 years)	1 to 4	525	20% (overall, not final)	22.30%	10%	About 80% overall	CACE
	Bitler, Domina, Penner, & Hoynes (2015)	3	1997-2000 (4 years)	K to 4	2,080	41.3%	34.6% Reading; 35.0% Math	5% (first year), 6% (second), 8% (third)	74% (first year), 65% (second), 55% (third)	Meta-analytic team Bloom (1984) adjusted
Washington Scholarship Fund (WSF)	Peterson, Howell, Wolf, & Campbell (2003)	3	1998-2001 (3 years)	K to 8	930	71%	40%	11% (first year), 8% (both first and second years)	68% (first year), 47% (two years), 29% (final year)	Instrumental Variables analysis

Notes: The sample size and attrition rates are based on the estimates from ITT Reading with the exception of Bettinger & Slonim (2006) which had only math impacts. The actual sample sizes for calculating the ITT and TOT Reading and Math impacts may differ slightly.
** Voucher eligibility was conditional on admission to a participating school)*
*** The 57% not using the voucher is based on 15% not in school, 16% in public school, and 12% that lost the voucher for other reasons.*
****Control cross-over and take-up rates from Barnard et al. (2003) are reported for single-child families, depending on time of application and background strata.*

Table 2 also includes the control cross-over rate, generally defined as the percent of voucher applicants randomly assigned to the control group that still attended a private school. In the U.S. voucher programs in our meta-analysis, students who lost the voucher lotteries often found other ways to access school choices. In the experimental evaluation in Dayton, Ohio, 18% of the control group students enrolled in a private school even without the assistance of a voucher (Howell & Peterson, 2006, p. 44). In the first evaluation of the D.C. OSP, 12% of the students that lost the lottery subsequently enrolled in a private school and 35% attended an independent public charter school, leaving just 53% of the control group students in traditional public schools (Wolf et al., 2013, p. 257). In the more recent D.C. OSP study, after one year, 10% and 42% of control group students attended private schools or charter schools, respectively (Dynarski, Rui, Webber, Gutmann, & Bachman, 2017). The New York City program achieved the clearest treatment-control contrast in type of school attended, as only 4% of the students that lost the lottery attended a private school and public charter schools were uncommon in NYC at the time; thus, almost all of the control group attended traditional public schools (Howell & Peterson, 2006, p. 44). In the experimental studies included in this meta-analysis, students remained in the control group and their outcomes counted towards the control group average for the ITT impact estimates even if they attended a private school. The rates at which control-group students crossed over to private schooling factored into the TOT effect calculations. Appendix D provides assumptions and calculations by study.

Our search recovered several studies unexamined by prior systematic reviews and meta-analyses. The global scope of our search added important data to our meta-analysis. Two studies of a large voucher program in Bogota, Colombia (Angrist, Bettinger, Bloom, King, & Kremer, 2002, Angrist, Bettinger, & Kremer, 2006), and two studies of different programs in separate

regions of India (Muralidharan & Sundararaman, 2015; Wolf, Egalite, & Dixon, 2015) were included. Our combined computerized and networked search also identified an RCT of a small privately-funded voucher program in Toledo (Bettinger & Slonim, 2006) that had been missed by all previous systematic reviews. Finally, we included four recent experimental evaluations of voucher programs (Abdulkadiroglu, Pathak, & Walters, 2015; Bitler, Domina, Penner, & Hoynes, 2015; Dynarski, Rui, Webber, Gutmann, & Bachman, 2017; Mills & Wolf, 2017). This meta-analysis represents a new look at a more comprehensive body of rigorous research on private school vouchers than ever before.

Many of the published reports of experimental evaluations of school voucher programs were nested in ways that affected how much independent information they contributed to the meta-analysis. In particular, at least six different research teams have published more than two dozen reports or articles analyzing the experimental data from the New York City Children's Scholarship Fund evaluation, 1998-2002. Including all 24 of those reports would generate substantial spatial auto-correlation due to repeated counting of the same finding. To ensure that the meta-analysis accounts for this nesting, we treated as a single study any group of publications of the same results, using the same methodology, by essentially the same research team. Any variation on that, such as publication of different results, using the same methodology, by a different research team (e.g. a failed replication), represented a different study even though it drew upon the same data. That determination reduced the number of New York City studies to five (Barnard, Frangakis, Hill, & Rubin, 2003; Bitler, Domina, Penner, & Hoynes, 2015; Jin, Barnard, & Rubin, 2010; Krueger, & Zhu, 2004; Peterson, Howell, Wolf, & Campbell, 2003). We then extracted data primarily from the final publication, unless an earlier publication contained more complete information, and supplemented those data with descriptive information

from other studies in the "nest" as needed. A study in our meta-analysis is the final and most complete presentation of a specific set of findings from a specific research team using a particular analytic method.

4-C. Programs included in the meta-analysis

The 20 RCTs identified by our search represent 11 separate school voucher programs (Table 3). Five programs – in Andhra Pradesh and Delhi, India; Toledo and Dayton, Ohio; and the D.C. Washington Scholarship Fund[7] – were each subject to a single experimental evaluation. Four programs – in Charlotte, NC; Louisiana; Milwaukee, WI; and Bogota, Colombia – were each the focus of both an original experimental study and one replication study. The New York City program was the subject of five different experimental analyses. The D.C. OSP was the subject of two different evaluations involving different samples of students.

In Table 3 each program is categorized as either privately or publicly funded, with either fully or partially funded vouchers. Fully funded vouchers must be accepted by participating private schools as the full cost of educating the student while partially funded vouchers require an additional payment from the student's family. In general, the fully funded vouchers were publicly funded, and the partially funded vouchers were privately funded. Funding for the programs in India and Colombia, whether "full" or "partial," was extremely low in nominal U.S. Dollars, ranging from about $117 in India to $190 in Colombia (Angrist, Bettinger, Bloom, King, & Kremer, 2002; Wolf, Egalite & Dixon, 2015). These low voucher amounts in India and Colombia are also reflective of lower educational costs in these areas. Comparable average costs of education in the public schools is $350 for Colombia (Angrist, Bettinger, Bloom, King, &

[7] The Washington Scholarship Fund (WSF) was a privately-funded scholarship program that preceded the government-funded Opportunity Scholarship (voucher) Program (OSP). The WSF was phased out in 2009, five years after the OSP was launched.

Kremer, 2002, pp. 1537-1538) and $1,963 for Delhi, India.[8] For Andhra Pradesh, Muralidharan and Sundararaman (2015, pp. 1031, 1058) report that the voucher amount was 40 percent of the average public school expenditure per child. The fully funded programs in the U.S. provided vouchers with maximum values that ranged from around $5,000 in Louisiana to over $13,000 for high school students in D.C. (Dynarski, Rui, Webber, Gutmann, & Bachman, 2017; Mills & Wolf, 2016). Fully funded programs provide all or nearly all of the state government funding that normally would go to the child's public school but do not include any of the federal or local education formula funding for the child. Partially funded programs in the U.S. provided about $2,000 in tuition support to families (Peterson, Howell, Wolf, & Campbell, 2003). Regardless of jurisdiction and full or partial funding, the maximum voucher amount allotted for all the programs in this meta-analysis generally represented less than half of the amount being spent per-pupil on students in area public schools.

All programs were targeted to low-income students through either income limits or program location, but usually both. The voucher initiatives in India and Colombia served students living in abject poverty either in cities (Angrist, Bettinger, Bloom, King, & Kremer, 2002; Wolf, Egalite, & Dixon, 2015) or villages (Muralidharan & Sundararaman, 2015). The U.S. programs were limited to students with family incomes near or below the cut-off for the federal lunch program,[9] almost entirely in cities. Almost all U.S. voucher participants were either African American or Hispanic. As a result, this meta-analysis is a study of the achievement effects of low-cost private school vouchers on low-income, primarily urban minority children.

[8] Per-pupil expenditure in public schools for Delhi, India was obtained from an Economic Survey of Delhi, 2016-2017. Source: http://www.indiaenvironmentportal.org.in/category/15375/publisher/government-of-nct-of-delhi/. We converted the per-pupil expenditure for public schools (rupees 29,641) to U.S. dollars for the year 2013 for comparison with the voucher amount (rupees 7,300). The comparable inflation adjusted voucher amount is $483.
[9] The federal lunch program provides low-cost or free lunches to children from low-income families in public and nonprofit private schools for each school day.

The private schools participating in some of these voucher programs (D.C. and Louisiana) were found to charge modest tuition and have experience serving disadvantaged student populations (Sude, DeAngelis & Wolf, 2017). Religious schools in general, and Catholic schools in particular, were the main participants in voucher programs in the U.S. In the first evaluation of the D.C. OSP, 80% of the participating students attended a religious school with their voucher, and 53% of them specifically enrolled in a Catholic school (Wolf et al. 2013, p. 257). Across programs, the private schools serving students with vouchers tended to have modest school facilities and few special programs for differentiating instruction to students (e.g. Dixon, 2013; Wolf et al., 2013). School-level quality measures based on test scores generally are not available for private schools participating in voucher programs because private schools are seldom required to report school-wide test scores publicly.

The counterfactual condition for control group students varied across the programs. In India and Colombia, almost all the students who lost the voucher lotteries attended local government-run schools. In India especially, public schools have many more resources than low-cost private schools but are plagued by teacher daily absenteeism rates of around 30% (Probe Team, 1999). When public school teachers in developing countries fail to show up for work, typically the children are on their own and are not supervised by substitute teachers. The counterfactual condition for students participating in voucher programs in the U.S. are traditional public schools that are free to attend and residentially assigned. Unlike their counterparts in developing countries, the public schools in the U.S. do not face the problems of teacher absenteeism or lack funds for school infrastructure and staff salaries.

Table 3: Description of 11 Voucher Programs included in Meta-Analysis

Program Evaluated	Location	Funding Source	Funding Amount (Full or Partial)	Grades	Studies Cited
Andhra Pradesh (AP) School Choice Experiment	Andhra Pradesh, India	Private	Full	1 to 5	Muralidharan & Sundararaman (2015)
Charlotte Children's Scholarship Fund	Charlotte, NC (USA)	Private	Partial	2 to 8	Greene (2000); Cowen (2008)
Children's Scholarship Fund	Toledo, OH (USA)	Private	Partial	K to 8	Bettinger & Slonim (2006)
District of Columbia Opportunity Scholarship Program (OSP)	Washington, DC (USA)	Public	Full	K to 12	Wolf, Kisida, Gutmann, Puma, Eissa, & Rizzo (2013); Dynarski, Rui, Webber, Gutmann, & Bachman (2017)
Ensure Access to Better Learning Experiences (ENABLE)	Delhi, India	Private	Full	K to 2	Wolf, Egalite & Dixon (2015)
Louisiana Scholarship Program (LSP)	Louisiana (USA)	Public	Full	3 to 8	Abdulkadiroglu, Pathak, & Walters (2015); Mills & Wolf (2016); Mills & Wolf (2017)
Milwaukee Parental Choice Program (MPCP)	Milwaukee, WI (USA)	Public	Full	K to 8	Rouse (1998); Greene, Peterson, & Du (1999)
Parents Advancing Choice in Education	Dayton, OH (USA)	Private	Partial	K to 12	Peterson, Howell, Wolf, & Campbell (2003)
Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES)	Bogota, Colombia	(partly funded by World Bank)	Partial	6 to 9 (2002 paper) and 6 to 11 (2006 paper)	Angrist, Bettinger, Bloom, King, & Kremer (2002); Angrist, Bettinger, & Kremer (2006)
School Choice Scholarships Foundation Program	New York, NY (USA)	Private	Partial	1 to 4	Peterson, Howell, Wolf, & Campbell (2003); Barnard, Frangakis, Hill, & Rubin (2003); Krueger & Zhu (2004); Jin, Barnard, & Rubin (2010); Bitler, Domina, Penner, & Hoynes (2015)
Washington Scholarship Fund	Washington, DC (USA)	Private	Partial	K to 8	Peterson, Howell, Wolf, & Campbell (2003)

Note: Studies do not necessarilly contain all years of a program. See Table 2 for more details at the study level.

4-D. Data extraction

The 20 included studies were coded in Microsoft (MS) Excel for details on author, publication year, location, funding type (public/private), years of evaluation, duration of study, grades analyzed, outcome (math and reading in English or local language), size of treatment and control group and overall sample size. Some studies had multiple evaluation years. Each evaluation year, type of impact estimate (TOT or ITT), and subject was treated as a separate observation in the database. A study that reported results in each of three years, in both reading and math, that included both TOT and ITT estimates, contributed 12 observations to the database

(3 x 2 x 2). The 12 observations that a given study might produce were only analyzed within a specific meta-analytic estimate of effect, such as the TOT estimate of the voucher effect in math in Year 2 after random assignment. When the authors provided results from multiple estimation models or from robustness checks, we only extracted the estimates from the final or most preferred model as signaled by the authors.

The extracted data filled 260 rows of an MS Excel spreadsheet, meaning 260 distinct effect estimates informed our meta-analysis (67 reading TOT estimates, 59 math TOT estimates, 71 reading ITT estimates, and 63 math ITT estimates). The extraction process was performed by two team members to minimize human error.[10] When necessary, we made assumptions to derive accurate sample sizes for the treatment and control groups. Appendix D provides assumptions and calculations by study.

4-E. Data synthesis

Meta-analysis combines results from several studies which individually have relatively small sample sizes and low precision. The fixed effects meta-analysis of the RCTs created an overall effect size by combining the effect sizes extracted from each study in standard deviation (SD) units. Effect sizes were analyzed separately for math and reading/English outcomes. All reading impacts were measured in the native language of the students and also some additional languages. The non-English languages specified in the reading results are presented in the figure notes. Both TOT and ITT effects were calculated, when possible. We focus on reporting the TOT impacts of voucher programs as the kinds of students who will actually use the vouchers to attend private schools may provide the most policy relevant information regarding these

[10] Inter-rater reliability is the extent to which researchers agree on the aspects of ratings and decisions. For the current study, the two members fully agreed on the data extraction and assumptions.

programs. The ITT effects are reported in the Appendix C. The overall effect size was a weighted average of the individual effect sizes extracted from the studies. Each observation's weight was the inverse of the variance around the effect size, so more precise effects were weighted more heavily in the meta-analysis (Borenstein, Hedges, & Rothstein, 2007).

The effect size and standard errors were extracted directly from the source if available. If necessary, they were calculated by the team using available data and the formulas in Appendix B. We calculated[11] the effect size and pooled standard deviation using Hedges' g (Hedges & Olkin, 1985). Lastly, the grand effect size and lower and upper bound of the overall 95% confidence interval were determined. Bettinger & Slonim (2006) had only math test outcomes while the other 20 studies included both math and reading/English.

The entire analysis was performed in two meta-analytic steps using a fixed effects meta-analysis. In the first step, we estimated an overall TOT and ITT effect for each year of the outcomes available for reading/English and math for each program by combining estimates reported across different studies for the same program in the same year. The program/year estimates vary slightly across the programs studied by multiple research teams due to methodological differences such as which baseline control variables are included in the estimation model, how missing data challenges are addressed, and even how students are classified by race. In the cases of the New York City studies, what might seem to be minor methodological variations produced substantively different interpretations of the results of hypothesis tests (Krueger & Zhu, 2004; Peterson & Howell, 2004), thereby underscoring the

[11] The calculated Hedges' g effect sizes and associated standard errors may differ from the effect sizes and standard errors reported in the studies. For comparability across studies, we estimate Hedges' g, relying on all available evidence and the formulas in Appendix B, instead of merely relying on the reported statistics in the papers.

importance of our meta-analytic approach to move beyond the peculiarities of individual voucher studies.

This mini "meta-analysis" of findings by site-year reduced the total number of effect size estimates to 96: 23 reading and English TOT estimates, 24 math TOT estimates, 24 reading and English ITT estimates, and 25 math ITT estimates. These 96 estimates represent the independent observations that inform our meta-analysis, as the actual set of student data used for each of the 96 estimates was unique in terms of the set of students involved, the year of evaluation, or the outcomes estimated (reading or math, ITT or TOT). We report all 96 effect sizes and their standard errors in Appendix E.

In the second step, we estimated overall voucher effects for 11 programs using a fixed effects meta-analysis. Use of random effects would not result in precise estimates[12] as the between-studies variance cannot be estimated with precision with our limited number of studies (in this case 11 programs) (Borenstein, Hedges, & Rothstein, 2007). The voucher effects in reading/English and math were estimated overall and separately based on geography (U.S. vs. non-U.S.), funding type (publicly vs. privately funded programs) and years of treatment (one, two, three, and four or more). The analysis for years of treatment used all 96 effect size estimates (which themselves represent a consolidation of the 260 extracted estimates) and all other analyses used the 46 effect size estimates for the last year covered by each study (12 estimates for reading/English TOT and 11 estimates for math TOT). We highlight the effects from the last study year because they represent the cumulative effects of the school voucher intervention across all years of the evaluations. We also report the overall variation in impacts across studies as well as the impacts across the subgroups analyzed. In addition to describing the statistical

[12] A caveat with the use of fixed effects meta-analysis is that it assumes that the estimates for all voucher programs are drawn from the same population. Such an assumption is likely unrealistic.

significance of the findings from each individual study, we present the p-value of the I^2 statistic, which measures whether there is sufficient study heterogeneity to eliminate sampling error as a likely cause of the observed results. All analyses have been carried out using a pre-coded worksheet in MS-Excel that implements the meta-analysis effect formula in Appendix B. The forest plots were generated using STATA.

5. Results

First, we present the global results for reading and math with English results as a subcomponent of the reading effects where English was taught in schools and tested. For each effect, we compare U.S. and non-U.S. programs. Next, we split the findings into effects from publicly funded versus privately funded programs. For those comparisons, the results are restricted to the final year of the evaluations. Finally, we present the results by years of treatment, which relaxes that condition.

To present our results we use forest plots which show the effect size and confidence interval for each study, for the U.S. and non-U.S. components, and overall. Individual studies are represented by box and whisker plots where the size of the gray square represents the relative weighting for the study and the whiskers represent the 95% confidence interval around a point estimate. The diamonds represent composite effects across all observations. Any confidence interval that includes zero indicates that an effect is not statistically significant. In the discussion that follows, we only present the forest plots for the TOT analysis. The forest plots for the ITT analysis are available in Appendix C. In general, the TOT results mirror the ITT results except

that the TOT effects, whether positive or negative, tend to be larger in magnitude because they adjust for noncompliance with random assignment.[13]

5-A. Overall impacts

Figures 1 and 2 present forest plots for the effects for reading and English globally. In addition, composites of the U.S. and non-U.S. effects are provided. The overall effects of the voucher programs are gains of 0.28 SD in reading and 0.08 SD in English, when English was not the language of the reading test and was assessed separately. The effect in reading is dominated by a large positive effect in the PACES program in Bogota, Colombia (1.4 SD). Comparing the seven U.S. and three non-U.S. programs with reading impacts, we see that the U.S. programs had an overall effect in reading (all of which was in English) of 0.03 SD that is small in size and barely missed statistical significance [95% CI: -0.00, 0.07]. The programs outside of the U.S. had a more definitive positive impact on reading scores of 0.51 SD (0.25 SD excluding Bogota, Colombia) and on English scores of 0.23 SD.

[13] Milwaukee, WI does not contribute to TOT effects as Greene's (1999) TOT in paper was non-experimental and Rouse's (1998) paper does not provide sufficient details to compute TOT effects. Louisiana was a placement lottery and it only contributes to TOT estimates.

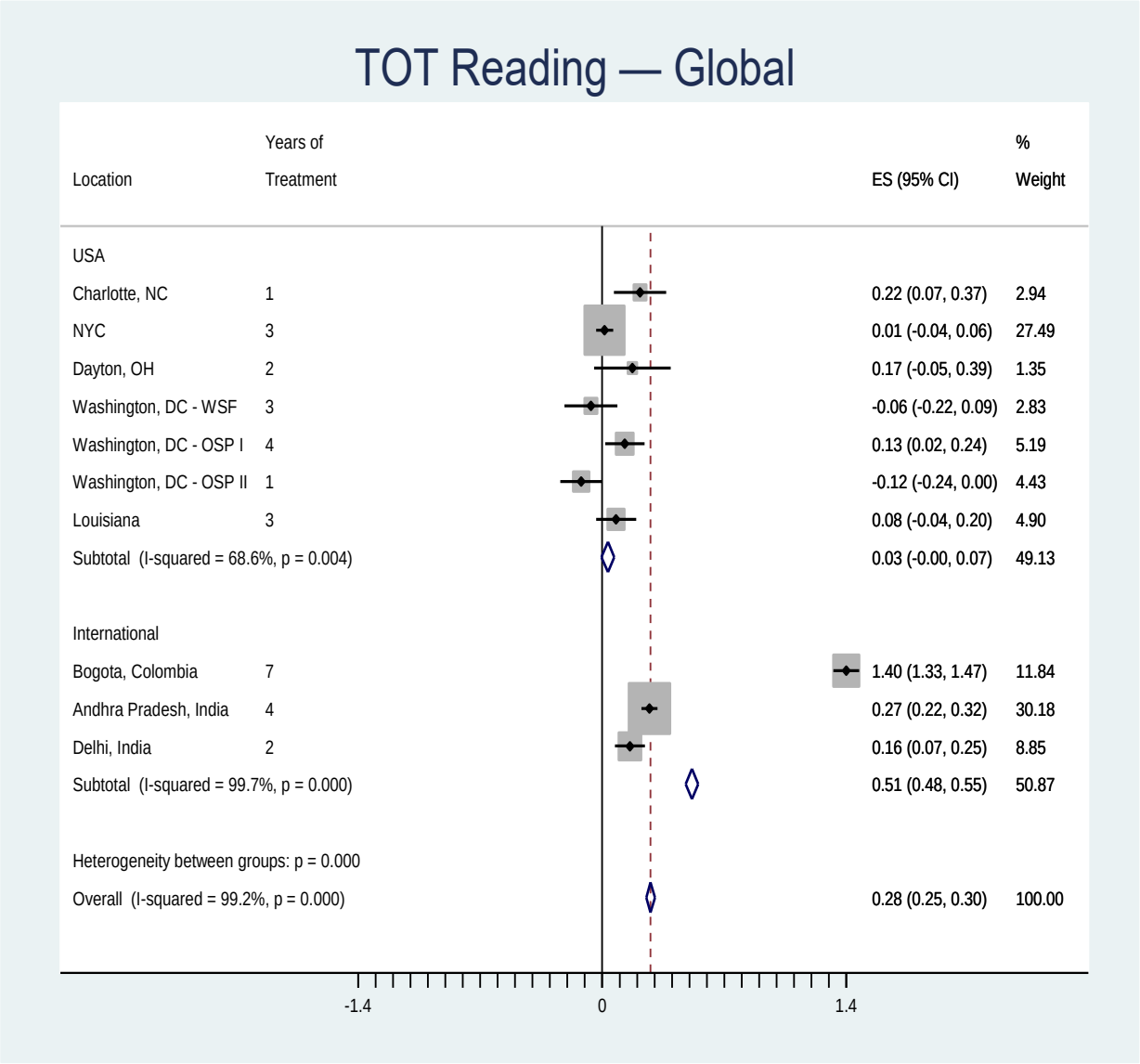

Figure 1: Overall Global impacts – TOT Reading.

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Overall effect size for US studies excluding Louisiana is 0.03 (-0.01, 0.07). Overall effect size for non-U.S. studies excluding Bogota, Colombia is 0.25 (0.20, 0.29) and overall global average excluding Louisiana and Bogota, Colombia is 0.13 (0.10, 0.16).

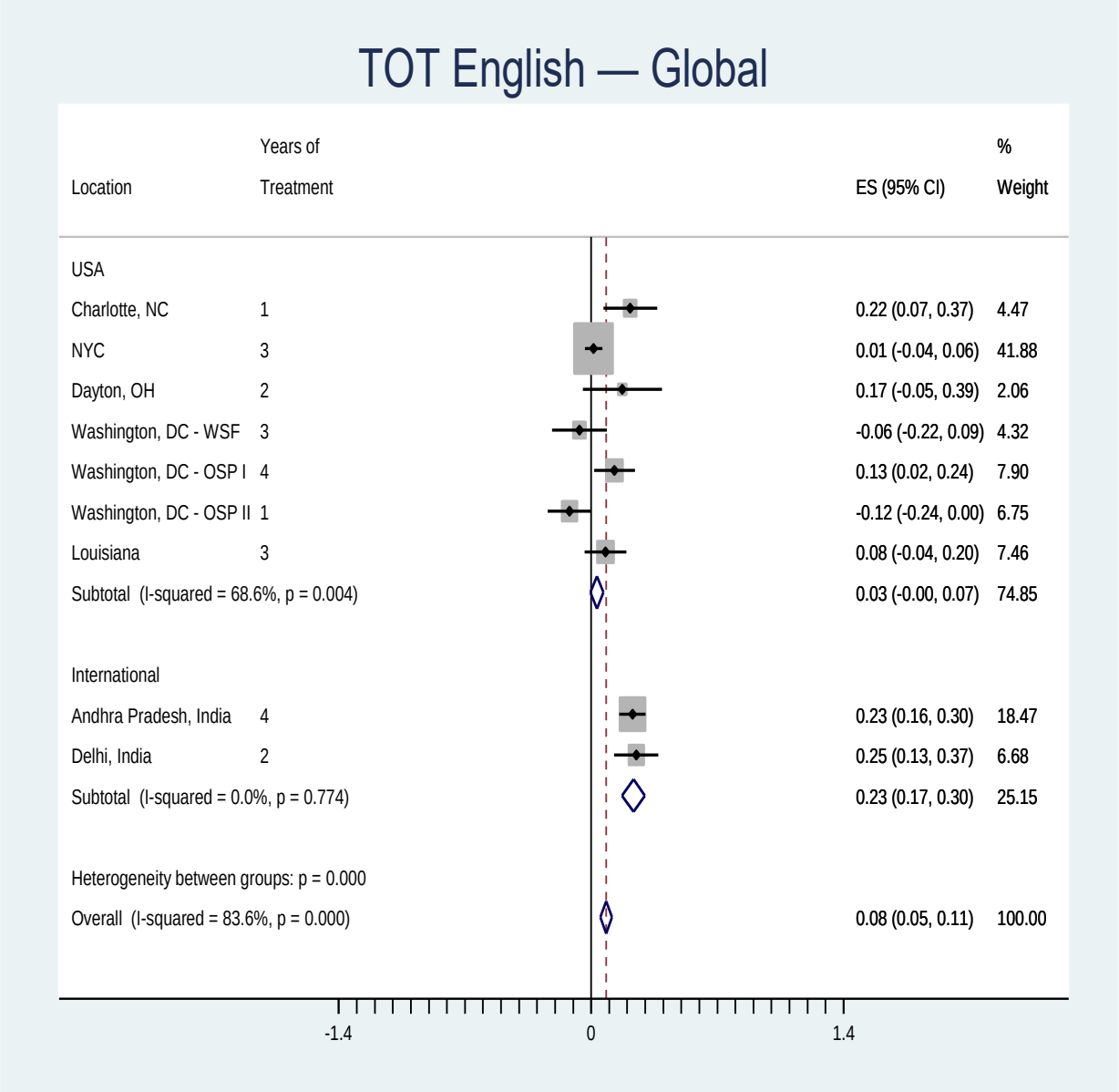

Figure 2: Overall Global impacts – TOT English.

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). No reading estimates are reported for Toledo, OH as it had only math test outcomes. Bogota, Colombia did not have an English estimate as the tests were administered in Spanish. Overall effect size for U.S. studies excluding Louisiana is 0.03 (-0.01, 0.07) and overall global average excluding Louisiana is 0.08 (0.05, 0.12).

The global effects in math are in Figure 3. Using a voucher improves math scores by 0.15 SD, on average [95% CI: 0.12, 0.18]. The U.S. programs, overall, have an insignificant effect of 0.01 [95% CI: -0.03, 0.05]. The non-U.S. programs have a positive effect of about 0.35 SD [95% CI: 0.30, 0.39]. The large effects for the non-U.S. programs are driven by the Bogota, Colombia study. Excluding Bogota, the non-U.S. studies had an overall insignificant effect on math test scores.

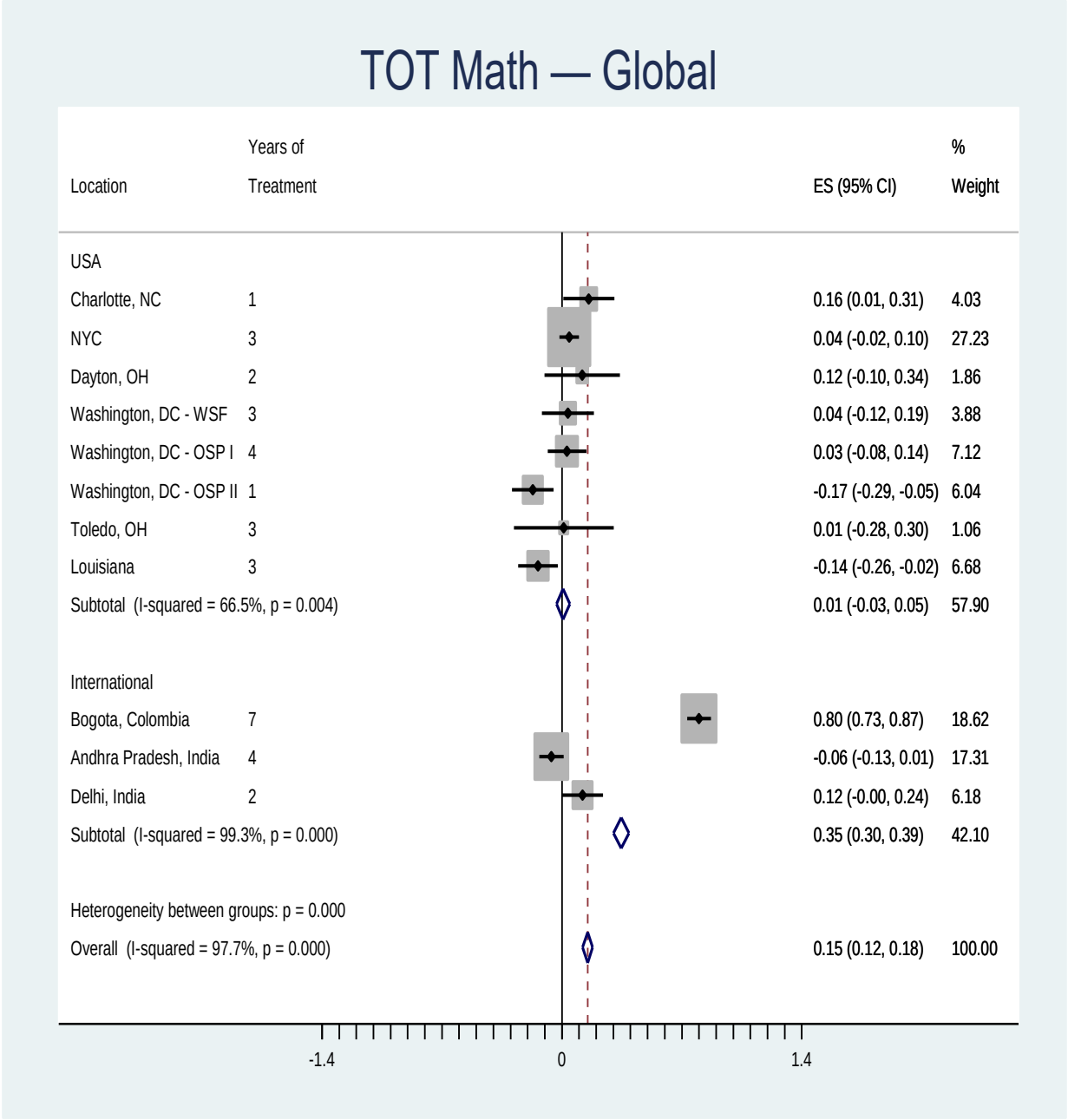

Figure 3: Overall Global impacts – TOT Math.

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). Overall effect size for US studies excluding Louisiana is 0.03 (-0.02, 0.07). Overall effect size for non-U.S. studies excluding Bogota, Colombia is -0.01 (-0.08, 0.05) and overall global average excluding Louisiana and Bogota, Colombia is 0.01 (-0.02, 0.05).

So far we have presented the results globally, as well as specifically inside and outside of the U.S. School vouchers have positive effects in reading, but these impacts are largest outside of the U.S., while the U.S. programs, as a subgroup, had insignificant impacts on reading. Overall, there is a positive impact on math test scores, but this result is entirely driven by one program in Bogota, Colombia. Next, we separate the effects by funding type (private or public).

5-B. Overall impacts by funding type

Figure 4 presents the results in reading, by funding type. For the purposes of this distinction, we define publicly funded programs as those with *any* amount of public funding, and privately funded programs as those that are *exclusively* privately funded through development or philanthropic funds. Both the publicly and privately funded voucher programs have positive effects on reading, overall. The average impact of using a voucher in privately funded programs is a gain in reading of 0.14 SD. The impact of voucher use in publicly funded programs is much larger, averaging reading gains of 0.65 SD. This large effect of publicly funded programs is primarily driven by the PACES program in Bogota, Colombia (1.4 SD).[14] In fact, excluding this Colombia based program indicates that publicly funded programs had a smaller (null) effect than privately funded programs. Due to the sensitivity of this result to the inclusion of PACES, we do not place much emphasis on this result.

[14] ITT results are provided in the Appendix C.

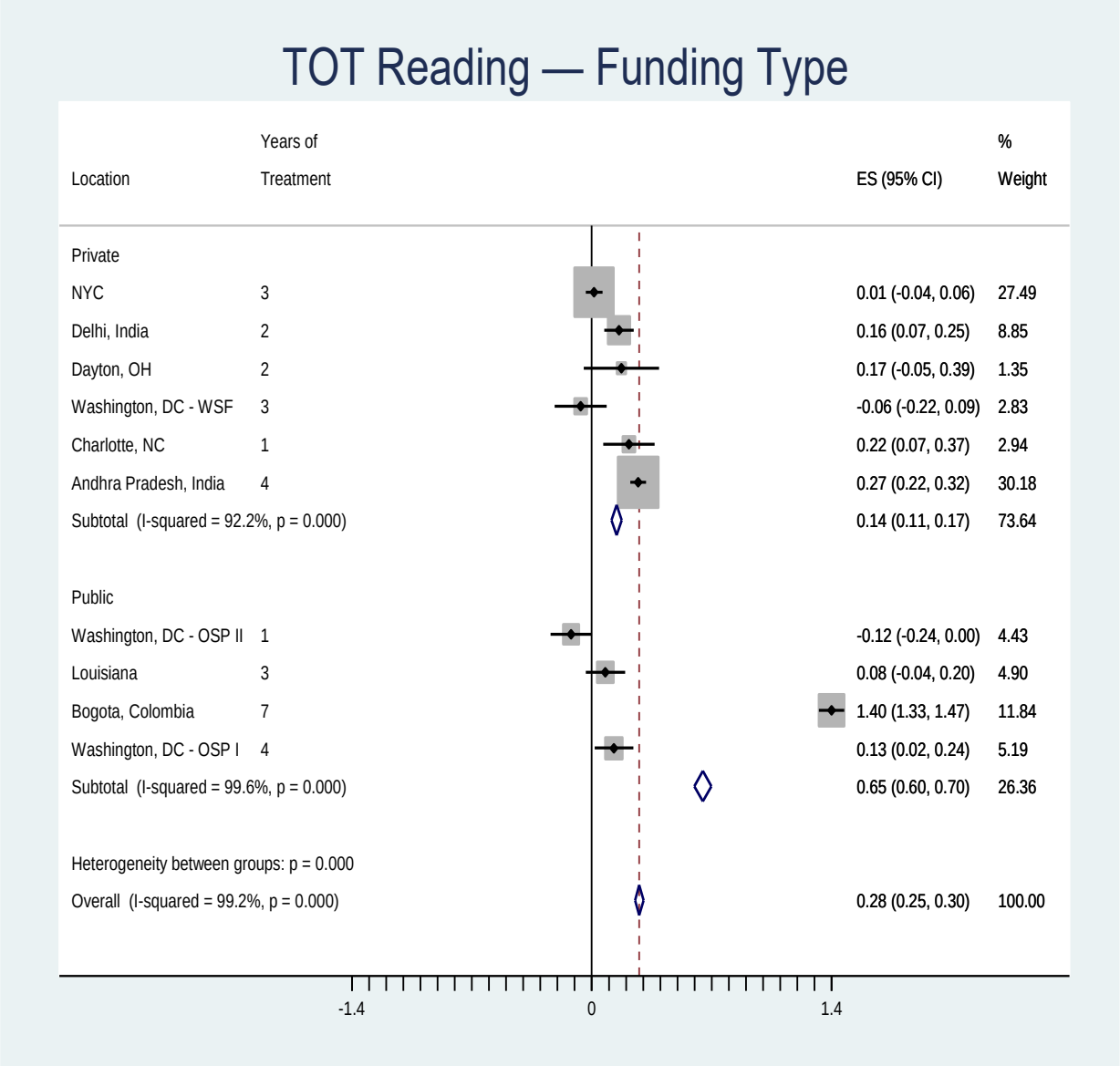

Figure 4: Overall impacts by Funding Type – TOT Reading.

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for privately and publicly (having received any public funds) funded programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Overall effect size for publicly funded programs excluding Louisiana and Bogota, Colombia is 0.01 (-0.07, 0.10).

Figure 5 presents the results in math, by funding type. The impact in math for privately funded programs, on average, is null, but the impact for publicly funded programs is an increase of 0.34 SD [95% CI: 0.29, 0.18], or null when excluding PACES in Bogota, Colombia.

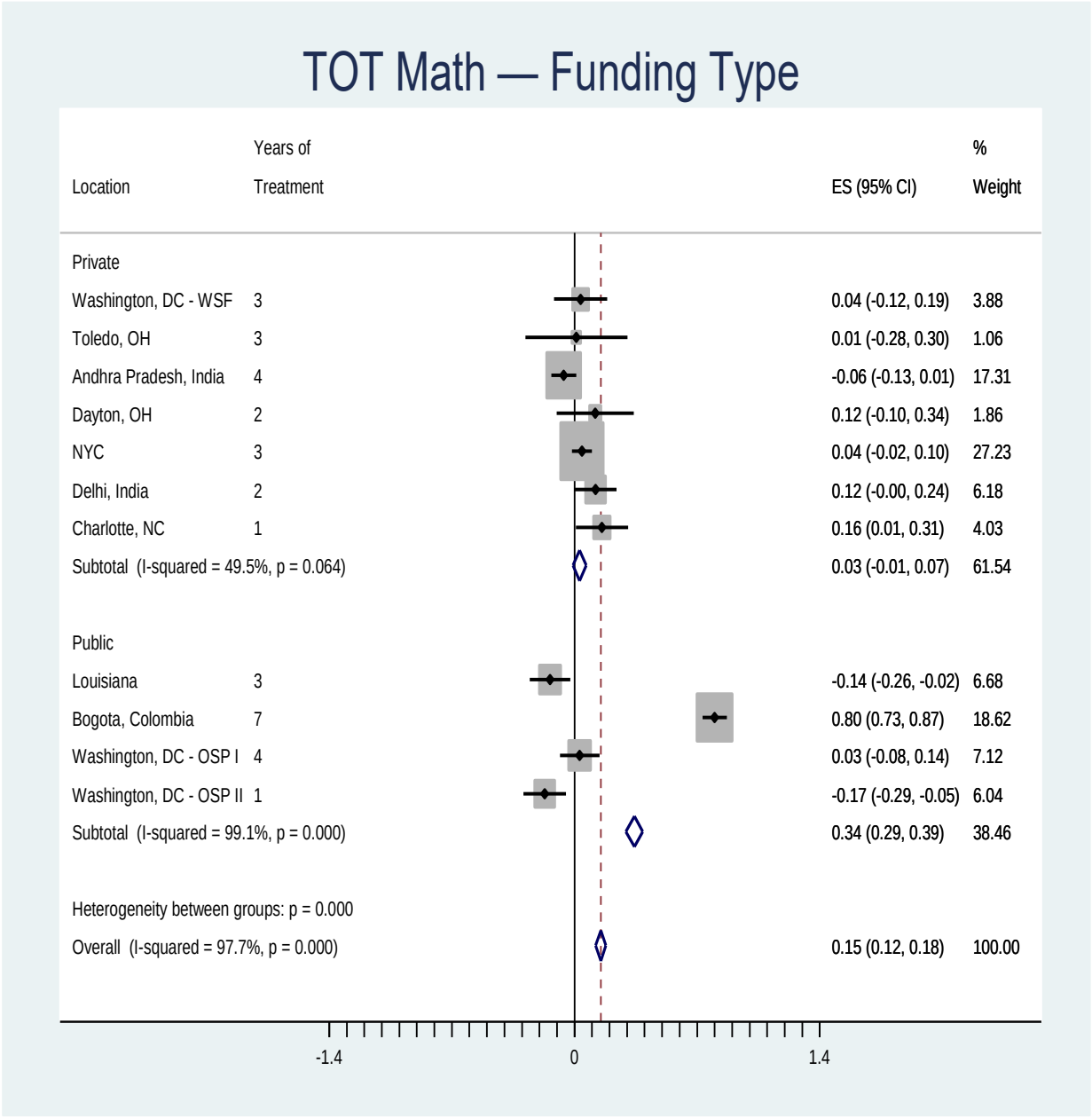

Figure 5: Overall impacts by Funding Type – TOT Math.

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for privately and publicly (having received any public funds) funded programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). Overall effect size for publicly funded programs excluding Louisiana and Bogota, Colombia is -0.06 (-0.14, 0.02).

5-C. Overall impacts by years of treatment

The last set of effects we present are the impacts on reading and math by years of treatment. If there is a cumulative positive effect of voucher treatment over time, we would expect impacts to increase[15] with the number of years of voucher use. The results are presented for eight programs with effects after one year, seven programs with effects after two years, six programs with effects after three years in math (five in reading), and three programs with effects of four or more years of treatment. All the estimates for years of treatment presented in this study are cumulative effects and do not represent annual estimates of the program effects.

Figure 6 shows the reading impacts by years of treatment. As expected, treatment effects increase with time of exposure. There is a null effect associated with one year of treatment, small positive effects for two and three years (0.07 SD and 0.06 SD, respectively), and a large positive effect (0.54 SD, [95% CI: 0.50, 0.58]) for four or more years.

[15] If compliance is changing over time, with the lowest performers dropping out, ITT analysis may be preferred over TOT analysis.

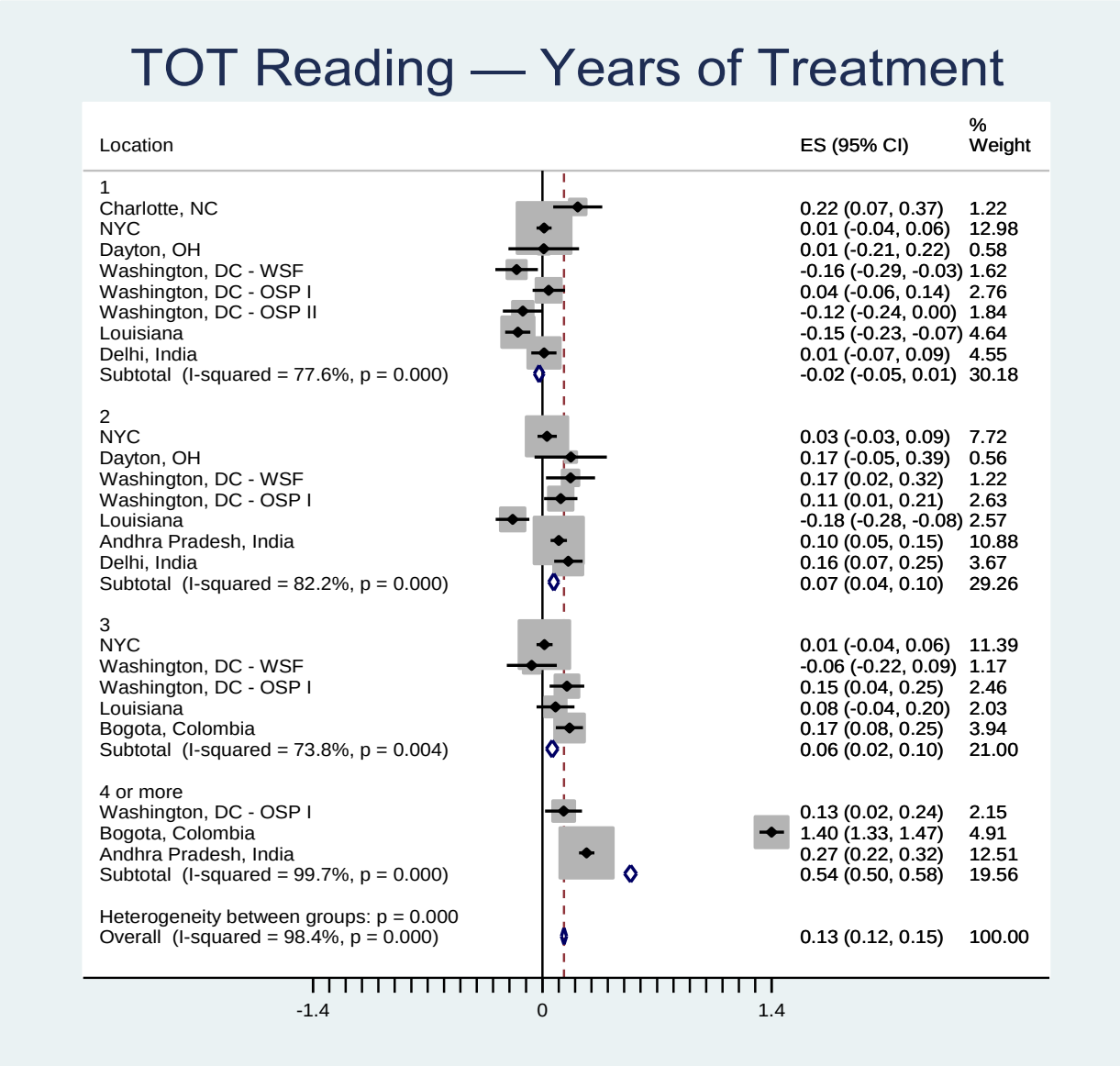

Figure 6: Overall impacts by Years of Treatment – TOT Reading.

Note: The Hedges' *g* estimates are effect sizes for one, two, three, and four or more years. The diamonds show overall yearly (dosage) estimates. The gray area around each point (effect size) is the study weight (inverse of variance). Toledo, OH only had math 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. Overall effect size with Louisiana removed is 0.00 (-0.03, 0.04) for one year and 0.09 (0.06, 0.13) for two years. Overall effect size with Louisiana and Bogota, Colombia removed is 0.03 (-0.01, 0.07) for three years and with Bogota, Columbia removed is 0.25 (0.21, 0.29) for four or more years. The overall estimate treats each effect size as independent and we do not focus on it due to spatial auto-correlation among different years of treatment within the same program.

The math results in Figure 7 show null effects after the first year and after three years, a negative effect after two years, and positive effects after four or more years. The average negative effect of two years of treatment is small (-0.05 SD), and primarily driven by the Louisiana Scholarship Program (-0.34 SD) and the program in Andhra Pradesh, India (-0.10 SD). The positive math effect of four or more years of treatment is large (0.33 SD) and precisely estimated [95% CI: 0.28, 0.37]). This longer-term outcome is primarily driven by the PACES program in Bogota, Colombia (0.80 SD).[16]

[16] The large jump in treatment effects in Bogota, Columbia between three years of treatment and seven years of treatment may be partly related to a change in examination and data-collection methods. The outcomes after three years are based on *La Prueba de Realización*, a grade-specific multiple-choice achievement test, and the outcomes after seven years are based on the ICFES, Colombia's centralized college entrance examinations.

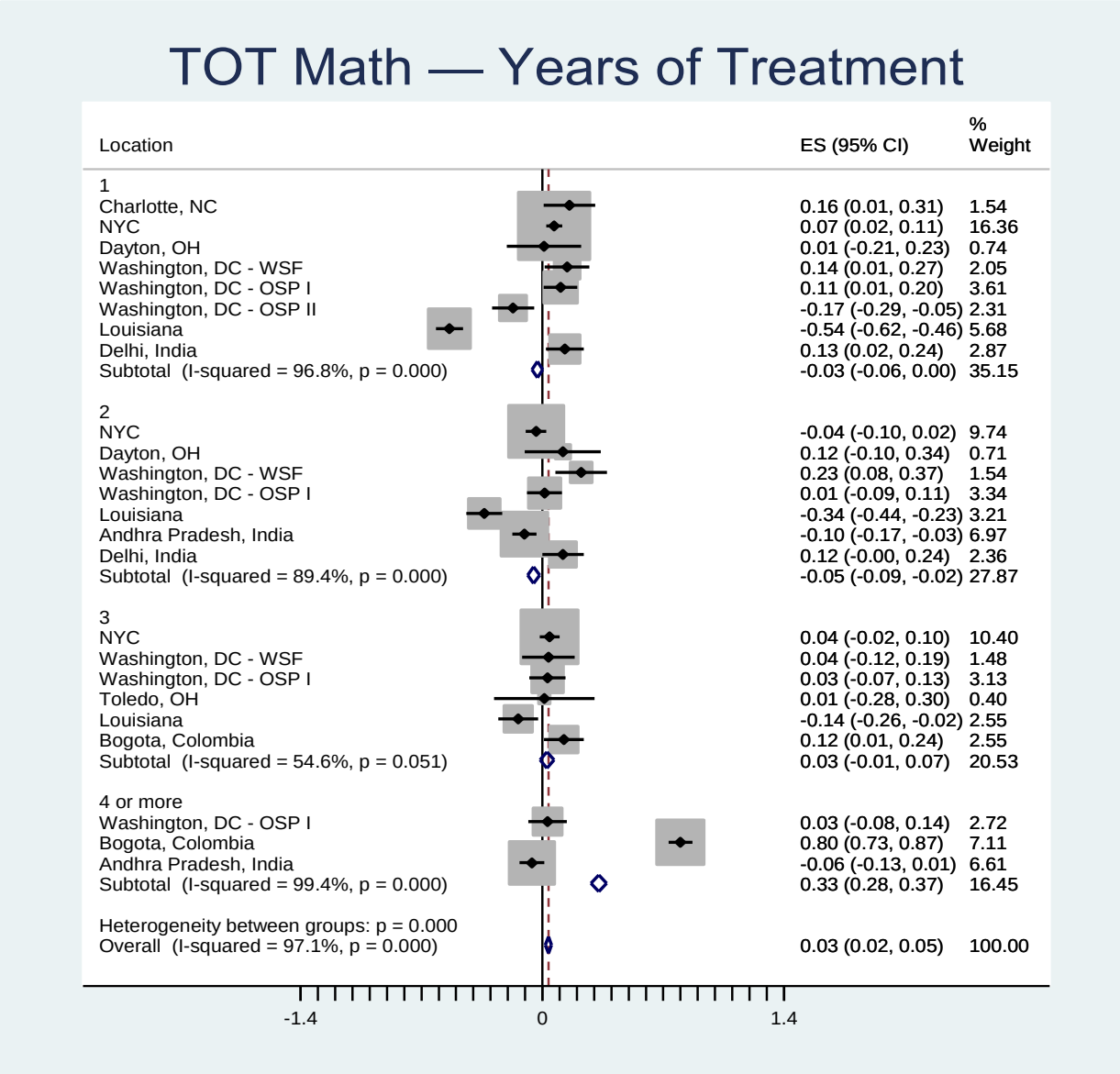

Figure 7: Overall impacts by Years of Treatment – TOT Math.

Note: The Hedges' *g* estimates are based on one year, two year, three year, and four or more year effect sizes for each study. The diamonds show overall estimates for yearly (dosage) effect of programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). Overall effect size with Louisiana removed is 0.07 (0.03, 0.10) for one year and -0.01 (-0.05, 0.02) for two years. Overall effect size for programs with Louisiana and Bogota, Colombia removed is 0.04 (-0.01, 0.09) for three years and with Bogota, Columbia removed is -0.03 (-0.10, 0.03) for four or more years. The overall estimate treats each effect size as independent and we do not focus on it due to spatial auto-correlation among different years of treatment within the same program.

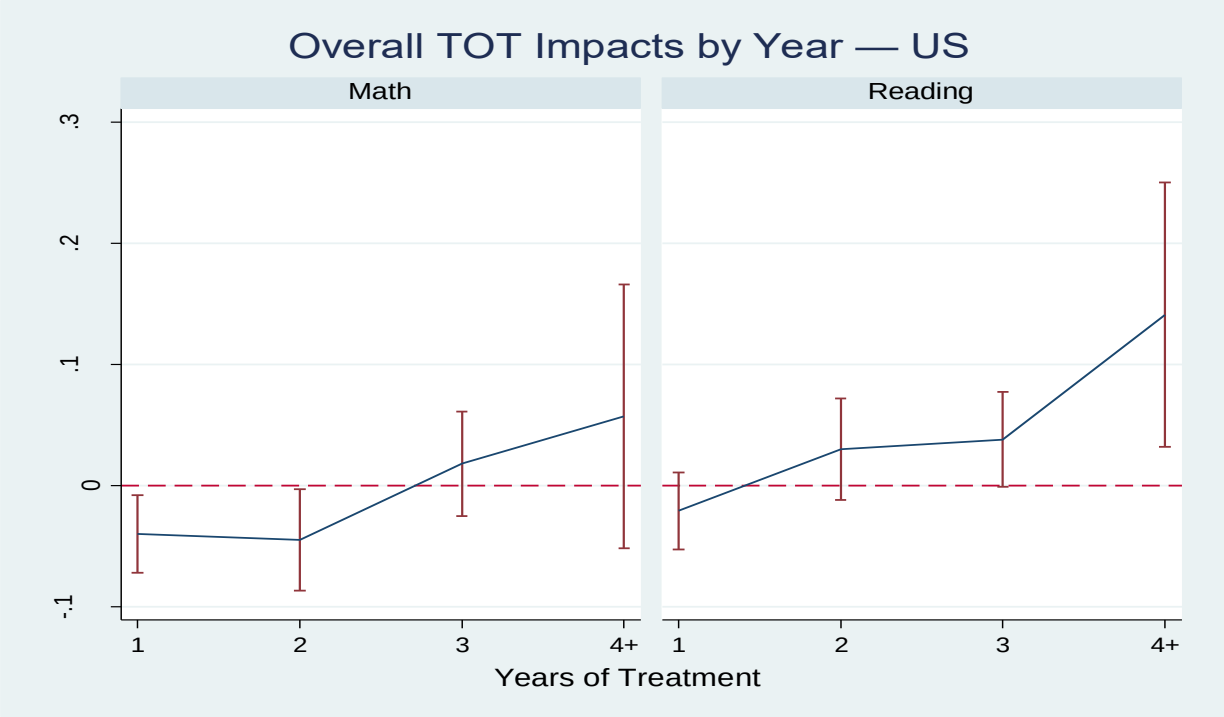

Figure 8: Overall TOT impacts by Year – US.

Note: Hedges' *g* estimates are based on one year effect, two year effect, three year effect and four or more year effect size calculated for each study. The effect size and confidence interval for each year are plotted vertically.

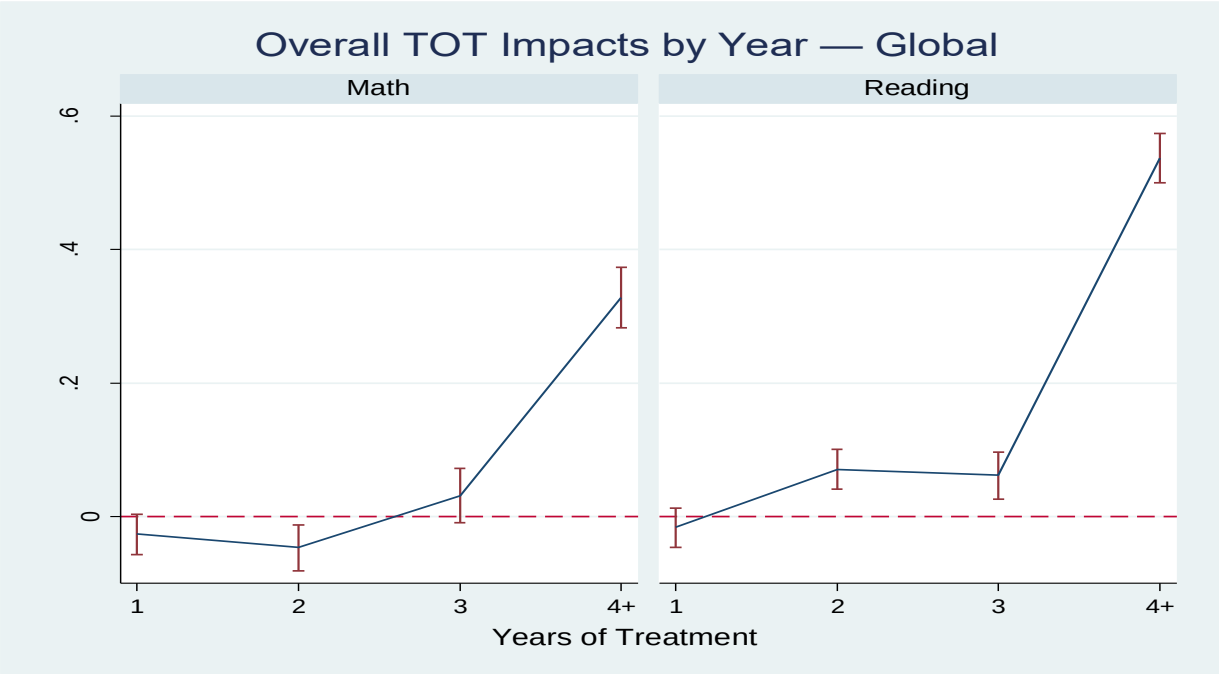

Figure 9: Overall TOT impacts by Year – Global.

Note: Hedges' *g* estimates are based on one year effect, two year effect, three year effect and four or more year effect size calculated for each study. The effect size and confidence interval for each year are plotted vertically.

These results indicate generally neutral to positive effects of school vouchers that vary by subject (math or reading), location (U.S. vs. non-U.S.), and dosage. The impacts of private school vouchers tend to be larger for reading than for math. Impacts generally are larger for programs outside the U.S. relative to those within the U.S. Impacts tend to grow from small to moderately large after four or more years. Unfortunately, a meta-regression to analyze the *ceteris paribus* relationships of each of these variables to voucher impacts from only 20 studies (11 programs) would be underpowered.

5-D. Robustness of the results

Sample attrition poses a challenge to our confidence in the true impacts of the 11 programs. Hence, we categorized the program impacts as either "low" or "high" risk of attrition bias in accordance with Cochrane's risk of bias tool (Cochrane Handbook, Chapter 8:

http://handbook-5-1.cochrane.org/). We classified individual studies as "low" risk for bias if they either met the What Works Clearinghouse (WWC) standards for acceptable and non-differential sample attrition levels or passed an appropriate robustness test for sample attrition bias. The checks for robustness to attrition used by the authors of the original studies included testing for baseline equivalence between treatment and control group students after program attrition had occurred and analyzing the sensitivity of the impact estimates to artificial truncations in the respondent samples for the treatment and control groups. In some cases, the authors conducted various bounding analyses such as inverse probability weighting or Lee bounds to check for the robustness of the results to differential attrition. For multiple studies that analyzed the same sample (such as the studies for NYC; Milwaukee, WI; and Charlotte, NC), if one of the studies was labelled as "low" risk, the other studies were read to check if they contradicted this classification. If no discrepancy was found, the overall impacts for a particular program were labelled "low" risk for attrition bias. With this categorization, only the program impacts for Delhi, India were labelled "high" risk for attrition bias, as the authors did not carry out any robustness checks for sensitivity of the results to sample attrition. Excluding the results for Delhi, India yields overall TOT global reading impacts of 0.29 (0.26, 0.32) and overall TOT math impacts of 0.15 (0.12, 0.18) which are essentially identical to the estimates that include the Delhi study. Thus, the meta-analytic estimates are robust to possible attrition bias.

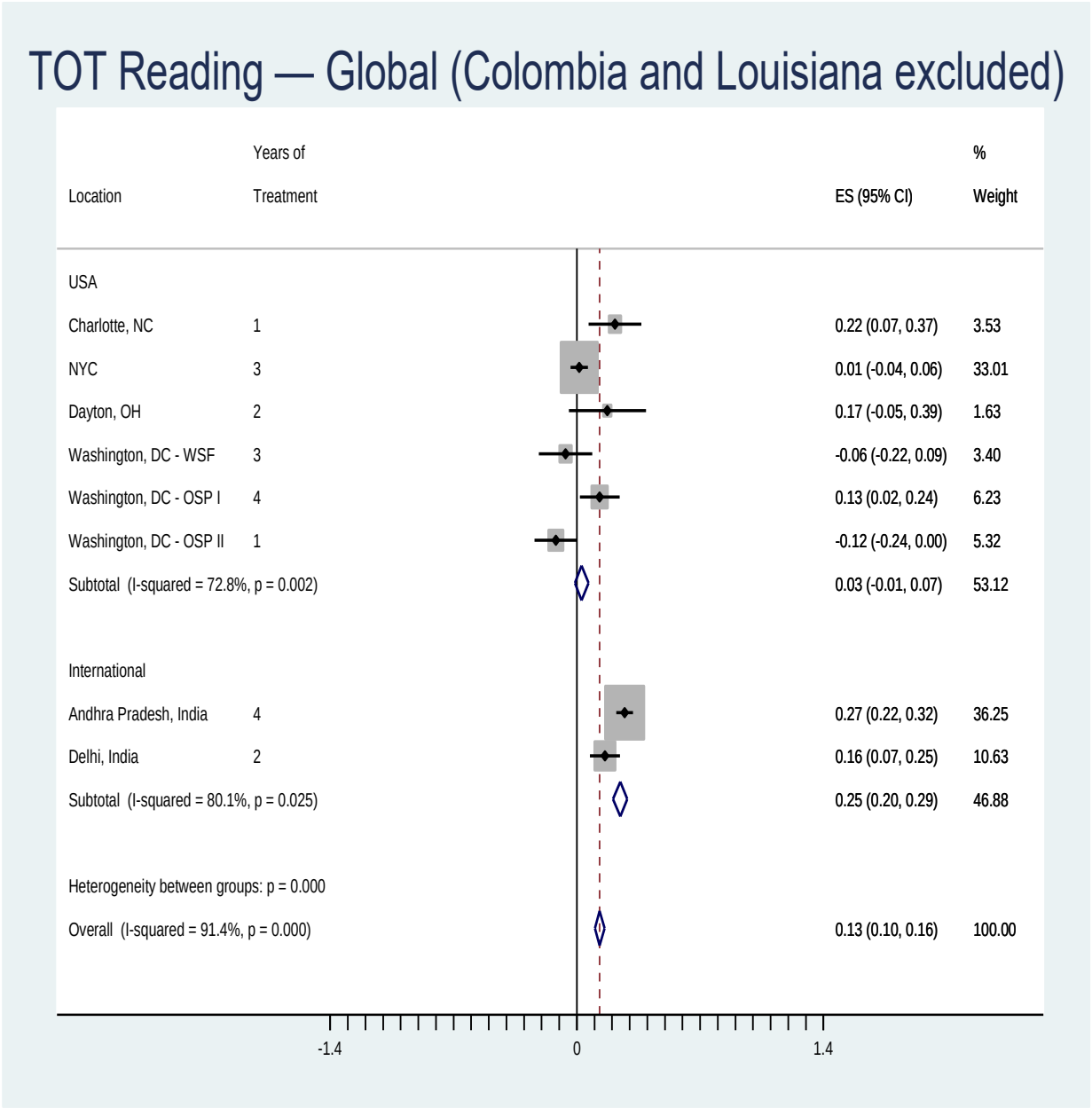

Figure 10: Overall Global impacts – TOT Reading (Colombia and Louisiana excluded).

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Estimates for Colombia and Louisiana are excluded.

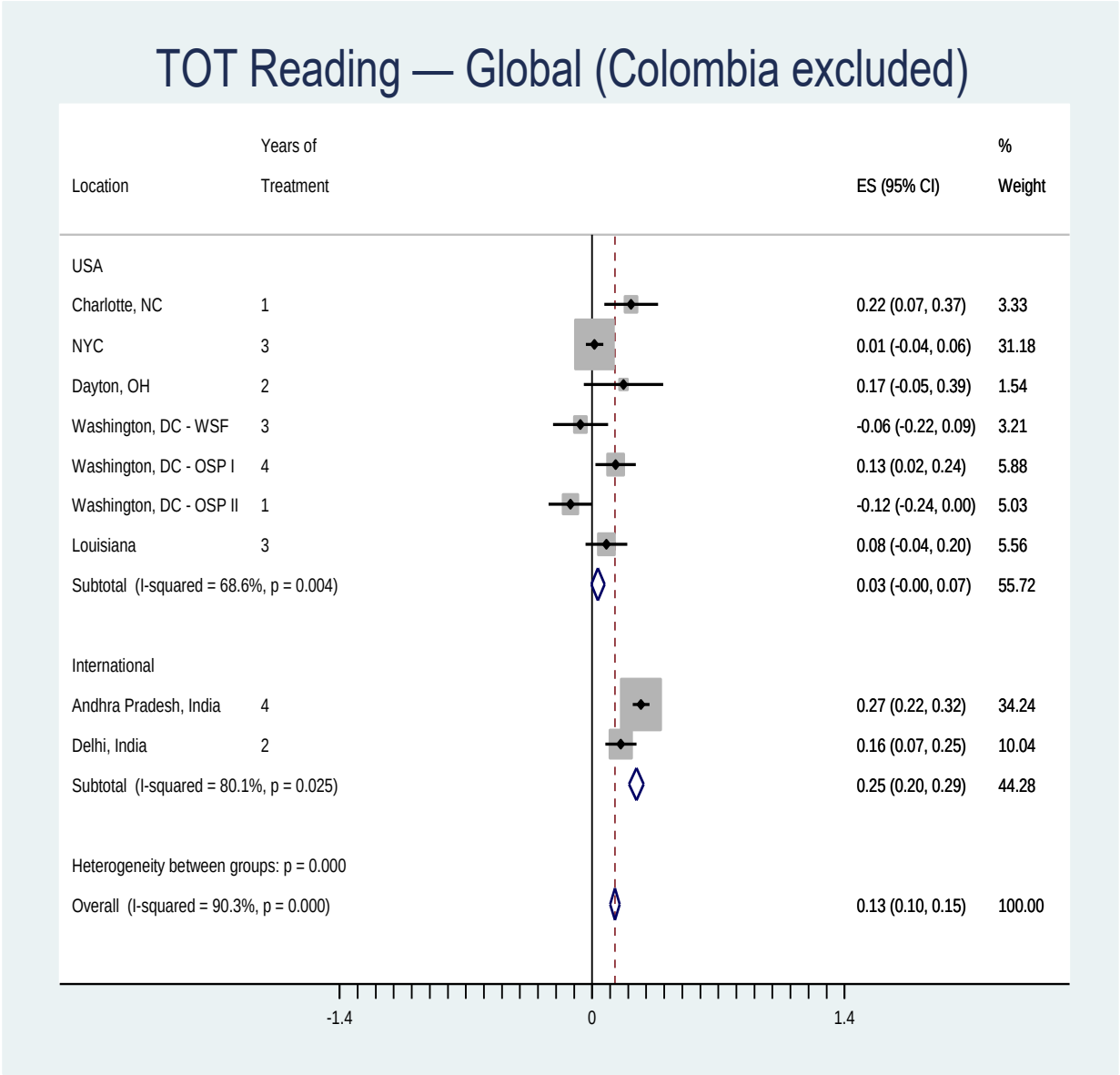

Figure 11: Overall Global impacts – TOT Reading (Colombia excluded).

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Estimate for Colombia is excluded.

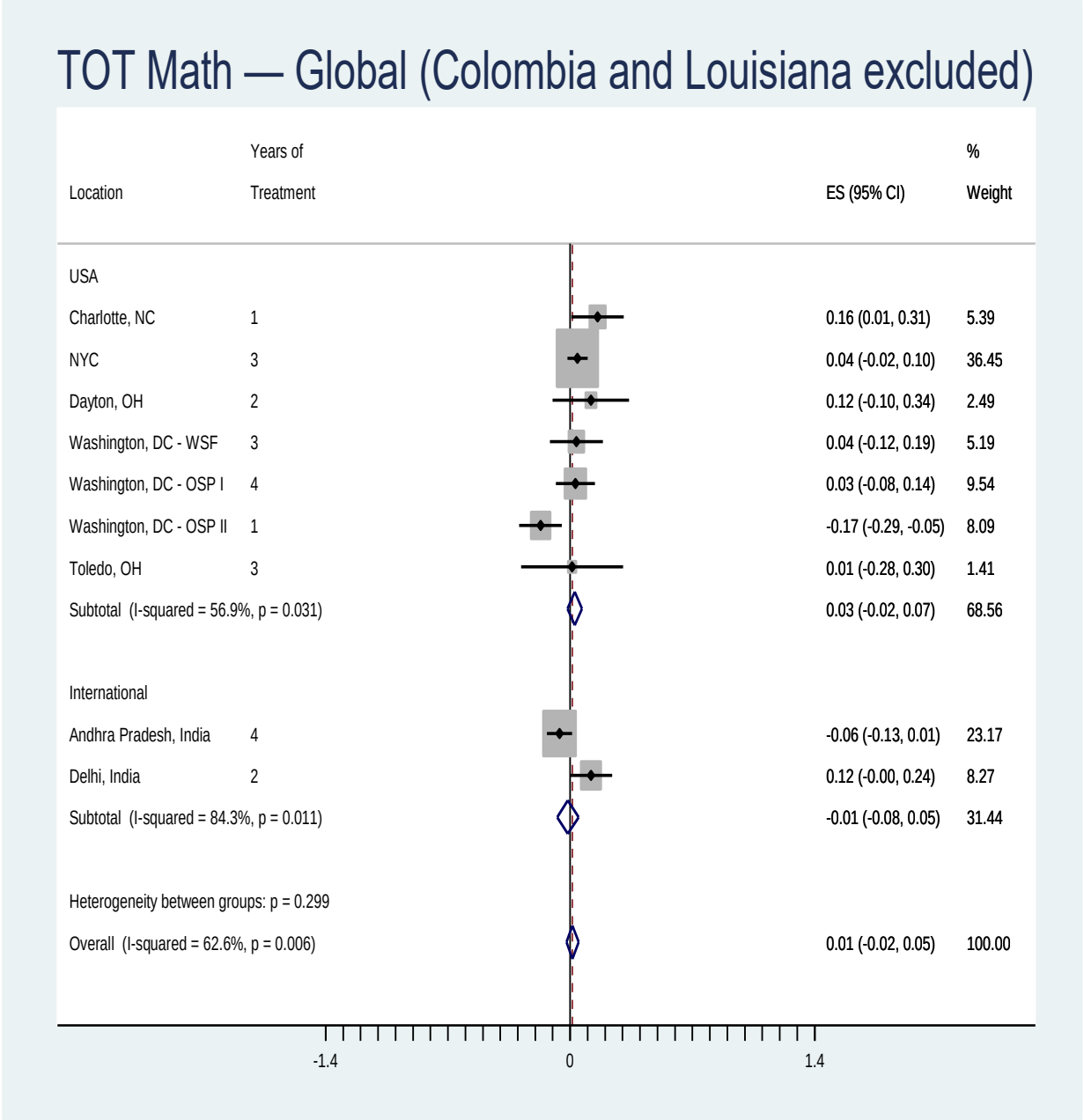

Figure 12: Overall Global impacts – TOT Math (Colombia and Louisiana excluded).

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). Estimates for Colombia and Louisiana are excluded.

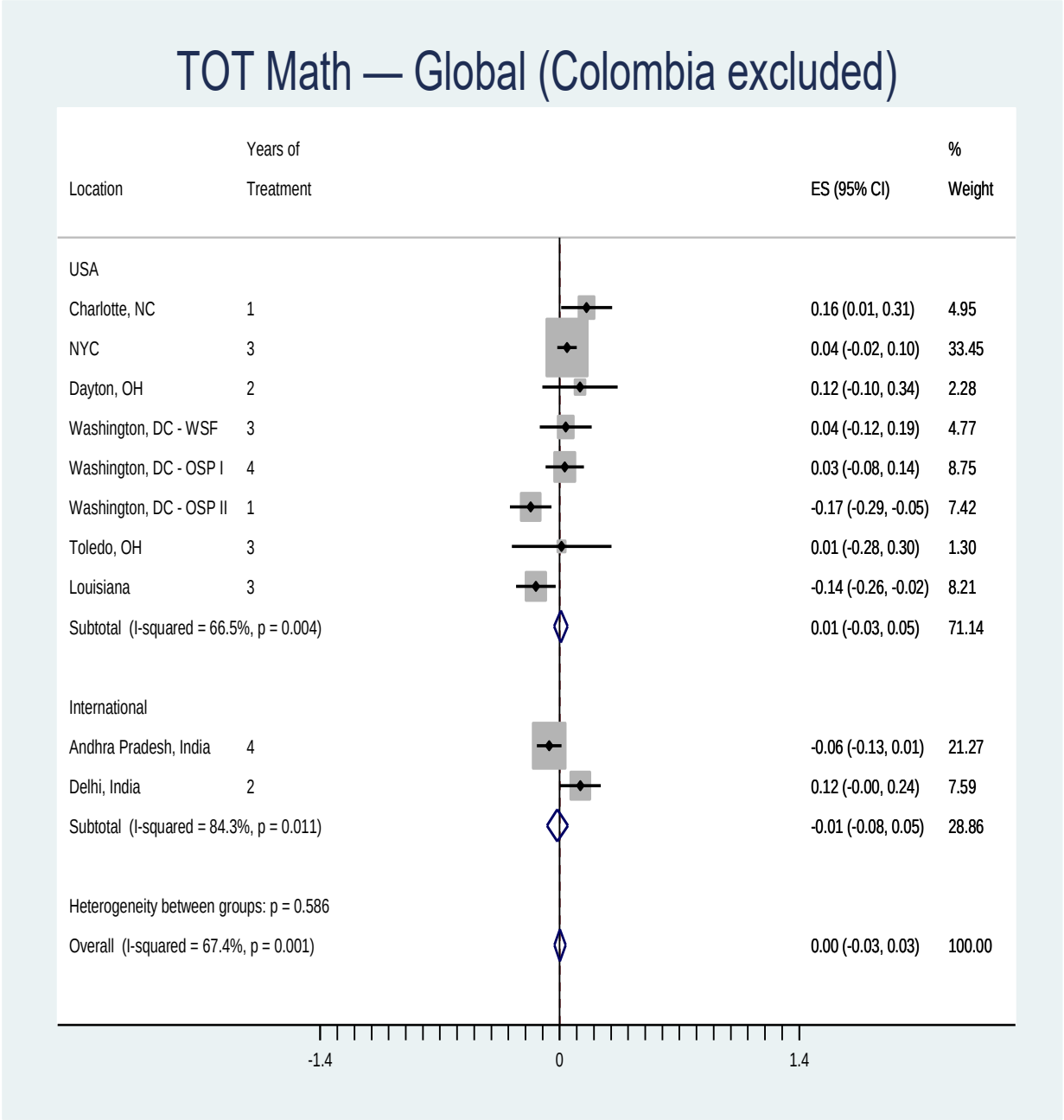

Figure 13: Overall Global impacts – TOT Math (Colombia excluded).

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). Estimate for Colombia is excluded.

The effect size estimates for Louisiana[17] and Bogota, Colombia regularly appear as outliers in our forest plots. We repeat the meta-analysis (Figures 10, 11, 12 and 13), excluding the results from these two potential outliers which contribute the most extreme negative and positive effect estimates in our study. The resulting overall estimates shrink in size but some of them retain statistical significance. We provide these results in the footnotes to each figure. With Louisiana and Bogota excluded, the math estimates have an overall null effect globally. The reading estimates are smaller but still positive and statistically significant even with these outlier findings removed.

We also conduct this same robustness test (removing outlier effects for Louisiana and Bogota) and assess whether the effect sizes for various years of treatment change. The reading estimates remain null for one year of treatment and increase slightly (from 0.07 SD to 0.09 SD) for two years of treatment. However, the estimates shrink from 0.06 SD to null for three years of treatment and from 0.54 SD to 0.25 SD for four or more years of treatment. In all of these cases, the confidence intervals still overlap, so there is no conclusive evidence that the results from the robustness check themselves are significantly different than the results from the main analysis. With the outliers omitted, the math estimates of positive effects increase substantially for one year of treatment, increase slightly but remain substantively similar for two and three years of treatment, and shrink dramatically for four or more years (from 0.33 SD to null).

From this robustness check, it seems that the magnitude of the overall reading impacts is reduced, but the overall conclusion of positive reading impacts is not affected by the Louisiana and Bogota outliers, as the reading impacts remain positive and significant. Math impacts are affected negatively by the exclusion of the outlier cases, becoming null overall. The LSP

[17] Louisiana was the only program that used criterion references tests. The treatment effects may be sensitive to the sampling of questions and test format.

evaluation was the only voucher RCT to use state criterion-referenced tests to measure program impact. Since the curricula used in the private schools in Louisiana are not necessarily aligned to the state test the way that they are in Louisiana public schools, it is possible that using the state test biased the Louisiana voucher impacts negatively, especially in math, which relies upon a specific sequencing of topics and skills. The results for Bogota, Colombia represent a blend of student incentives (to continue receiving the voucher, the students had to maintain minimum academic standards), additional education spending (through top up by parents) and private school productivity (Glewwe & Muralidharan, 2016). The fact that the Bogota program was a blend of school vouchers and other reforms could explain its status as an outlier in our meta-analysis.

5-E. Heterogeneity of the results

Testing for heterogeneity of effect sizes in a meta-analytic estimate allows for determination if the studies reasonably share a common effect size (Hedges & Olkin, 1985). We are interested in testing if our meta-analytic estimates have more variation than would be expected simply due to sampling error. The existence of significant heterogeneity between study subgroups (U.S. vs. non-U.S. programs, publicly vs. privately funded programs, and various years of treatment) would indicate true differences between the effects of vouchers on distinct subgroups that are not due to sampling error.

The forest plots provide an estimate of the percentage of variation in the meta-analytic effect sizes derived from within and across subgroup comparisons. The I^2 statistic in the forest plots provides a p-value. The null hypothesis for the underlying test assumes homogeneity across

the analyzed studies. A higher value of the I^2 statistic[18] resulting in a significant p-value suggests that the variation in the results of the studies are due to underlying heterogeneity rather than due to chance alone (Higgins, Thompson, Deeks, & Altman, 2003). The tests for heterogeneity in the figures show that they are statistically significant in both math and reading for the total sample of studies. Generally, there is statistically significant heterogeneity between subgroups (U.S. vs. non-U.S. programs, publicly vs. privately funded programs and between years of treatment). However, within the privately funded programs there is a lack of significant heterogeneity for math impacts. With that one exception, the heterogeneity analysis indicates that variation in effect sizes across subgroups reflects a true underlying variation and not statistical noise. The within subgroup analysis also reflects statistically significant heterogeneity in most cases—we do not observe significant heterogeneity in some cases probably due to small subgroup sample sizes.

6. Discussion

This meta-analysis contributes to the field of private school choice by combining and systematically evaluating rigorous evidence from all RCT studies of the effects of private school vouchers on student achievement. This review provides an up-to-date, methodologically coherent and comprehensive overview of all the rigorous experimental findings and yields important policy implications about the effectiveness of voucher programs.

Our search process turned up 9,443 potential studies, 20 of which ultimately were included. These 20 studies represent 11 different voucher programs, eight in the U.S. and three in

[18] In a fixed effects meta-analysis, the studies are assumed to have been carried out under similar conditions with similar study participants. Thus, one true effect size is shared by all the studies included in the meta-analysis. In contrast, in a random effects meta-analysis, the true effect size could vary across studies. Hence, the random effects meta-analysis estimates the mean of a distribution of true effects of all the studies. Thus, I^2 statistic has a lower value under a fixed-effects meta-analysis in comparison to a random effects meta-analysis.

non-U.S. countries. A total of 260 effect sizes are included, with a two-stage consolidation of those estimates yielding a total of 46 average findings drawn from the last year of the studies. We report 10 meta-analytic TOT effect sizes for reading (seven in the U.S. and three outside of the U.S.). In reading, we find an overall positive effect of about 0.28 SD with null effects in the U.S. and large positive effects (0.51 SD) outside of the U.S., primarily driven by PACES, in Bogota, Colombia. A much larger gap in the quality of public and private schools in countries like Colombia than in the U.S. may explain this finding (Angrist, Bettinger, & Kremer, 2006). In addition, the PACES program was distinctive in providing individual student incentives for academic achievement. Excluding this outlier, the overall impact on reading was still significant (0.13 SD). The overall impact on reading remains same (0.13 SD) when both Colombia and Louisiana outliers are excluded.

For math scores, we report 11 meta-analytic TOT effect sizes (eight in the U.S. and three outside of the U.S.). The math effects (0.15 SD overall) are large for the non-U.S. studies (0.35 SD) and null for the U.S. studies. Again, this is driven primarily by the PACES program – excluding this outlier results in a null overall impact on math test scores. Excluding both Colombia and Louisiana outliers yields a null overall impact on math test scores. A large-scale meta-analysis of all education RCTs from 1995 to 2010 has shown that the average impact of an intervention on test scores is 0.08 SD at the elementary level and 0.15 SD at the middle school level (Lipsey et al. 2012). Hence, our meta-analysis shows that voucher interventions produce positive test scores outcomes chiefly in reading, that are comparable in size to outcomes from other education interventions, but that there is heterogeneity within the set of programs.

The overall results just described are for the final year of data in each study. It could be that these effects do not represent the initial effects one might expect from a new program. In

fact, our analysis of the effects by years of treatment indicates that the effects of private school voucher programs often start out null initially and then turn positive. Longer-term achievement effects are much more salient than immediate achievement effects whenever longer-term effects are available (Das et al., 2013; Glewwe & Muralidharan, 2016). The general pattern of results also indicates that voucher interventions tend to increase reading scores more than math scores.

The results of this meta-analysis indicate that voucher programs globally tend to moderately increase test scores, particularly in countries like Colombia with a large private-public school quality gap. Although the scope of our search process was global, more international RCTs are needed to reach definitive conclusions about the impacts of voucher programs around the globe. Our search process yielded RCTs on private school vouchers only in U.S., India and Colombia. In addition, many of the programs included here are still relatively small-scale, and more experimental work should be done on larger programs to understand whether the potential benefits of private school vouchers would replicate at scale. We cannot learn much from even large-scale programs if they are not implemented alongside an experimental evaluation. Further, more experimental evaluations that consider the impacts of vouchers on key non-cognitive outcomes such as educational attainment and civic values (e.g. Angrist, Bettinger, & Kremer, 2006; Wolf, 2007; Wolf et al., 2013) would be of great value to the field. We hope that our study motivates researchers to pursue experimental evaluations of voucher impacts whenever feasible.

Additionally, it is critical to consider the cost-benefit tradeoffs associated with voucher programs. Numerous studies find that vouchers are cost effective, since they tend to generate achievement outcomes that are as good or better than traditional public schools but at a fraction of the cost (Muralidharan & Sundararaman, 2015; Wolf & McShane, 2013). Therefore, even null

impacts in a voucher program, if obtained at a lower cost than education in a public school setting, may have a net benefit for society. The average per-pupil spending in the public school jurisdictions covered in our meta-analysis was higher than the voucher amounts. For programs where the voucher covered the full amount of private school tuition, the average school funding in private vs. public sectors in the same geographic location[19] was $255 vs. $636 for Andhra Pradesh, India, $483 vs. $1,963 for Delhi, India, $4,817 vs. $11,846 for Milwaukee, WI, $5,456 vs. $10,853 for Louisiana, $7,761 vs. $21,081 for D.C. OSP I, and $12,306 vs. $20,577 for D.C. OSP II. The partially funded programs required parents to top up the voucher amounts. The data indicate that the partially funded programs in Charlotte, NC, Dayton and Toledo, OH, NYC, Washington, D.C. and Bogota, Colombia had much lower voucher amounts – less than half of the per-pupil expenditures in the neighborhood public schools. A cost-effectiveness analysis of the 20 studies included in this study may inform us about the savings from the experimental interventions of 11 voucher programs.

When taking into account the total costs and benefits of these types of programs, it is important to also study the impacts on the students who remain in the traditional public school (TPS) system. A voucher program in place may generate competitive effects for comparison TPS system as the school systems would compete for academic efficiency and for drawing children and financial resources towards the schools. Four of the voucher programs in our meta-analysis of participant test-score effects also have been evaluated regarding their systemic effects on the test scores of students who remained in public schools. The Louisiana (Egalite, 2014) and Milwaukee (Carnoy, Adamson, Chudgar, Luschei, & Witte, 2007; Chakrabarti, 2008; Greene &

[19] We calculated the numbers from the data provided in the papers, department of education websites and reports from sources such as EdChoice. For India, the rupees have been converted to dollars. All amounts are based off the last year of evaluation of a program and weighted averages are taken across cohorts in D.C. OSP II. The numbers have been inflation adjusted to 2013 dollars for comparison.

Forster, 2002; Hoxby, 2003) voucher programs have been found to have a positive effect on the subsequent test scores of affected public schools. The Washington, DC (Greene & Winters, 2007) and Andhra Pradesh (Muralidharan & Sundararaman, 2015) programs have been found to have no significant effects on the achievement of the students who remained in public schools. Thus, the research to date suggests that the modest average achievement benefits from private school choice for participants are not coming at the expense of achievement declines for non-choosing students. It is practically difficult to randomly assign school systems to a treatment group comprising of choice-based competition and a control group that is isolated from the choice-based competitive forces. Due to the nature of the question concerning how expanding private school choice and competition affects the performance of traditional public schools, the systemic effects of private school choice have been evaluated almost exclusively using merely quasi-experimental methods. Still, such studies in the U.S. consistently report effects that range from null to positive (Egalite & Wolf, 2016).

This meta-analysis provides a systematic summary of the generally modest positive effects of private school choice programs around the globe on the test scores of participating students. With time and more years of outcome data from voucher interventions, a better understanding of voucher impacts may be obtained.

References

Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2015). School vouchers and student achievement: First-year evidence from the Louisiana Scholarship Program (NBER Working Paper No. 21839). Cambridge, MA: National Bureau of Economic Research.

Anderson, M. R., Guzman, T., & Ringquist, E. J. (2013). Evaluating the effectiveness of educational vouchers. In E. J. Ringquist (Ed.), *Meta-analysis for public management and policy* (pp. 311-351), San Francisco, CA: Jossey-Bass.

Anderson, K. P. & Wolf, P. J. (2017). *Evaluating school vouchers: Evidence from a within-study comparison*. EDRE Working Paper 2017-10. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2952967

Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review, 92(5),* 1535-1558.

Angrist, J., Bettinger, E., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *The American Economic Review, 96*(3), 847-862.

Azam, M., Chin, A., & Prakash, N. (2013). The returns to English-language skills in India. *Economic Development and Cultural Change*, *61*(2), 335-367.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association*, *98*(462), 299-323.

Baum, D., Lewis, L., Lusk-Stover, O., & Patrinos, H. (2014). What matters most for engaging the private sector in education: A framework paper. The World Bank, Washington, DC.

Bettinger, E., & Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics*, *90*(8-9), 1625-1648.

Betts, J. R., Tang, Y. E., & Zau, A. C. (2010). Madness in the method? A critical analysis of popular methods of estimating the effect of charter schools on student achievement. In Hill, P. T. & Betts, J. R. (Eds). *Taking measure of charter schools: Better assessments, better policy making, better schools*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Bifulco, R. (2012). Can nonexperimental estimates replicate estimates based on random assignment in evaluations of school choice? *A within-study comparison. Journal of Policy Analysis and Management*, *31*(3), 729-751.

Bitler, M., Domina, T., Penner, E. K., & Hoynes, H. (2015). Distributional effects of a school voucher program: Evidence from New York City. *Journal of Research on Education Effectiveness, 8*(3), 419-450.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225-246.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis fixed effect vs. random effects. Retrieved June 1, 2016, from: https://www.meta-analysis.com/downloads/Meta-analysis%20fixed%20effect%20vs%20random%20effects.pdf

Butcher, J., & Burke, L. M. (2016). The education debit card II: What Arizona parents purchase with education savings accounts. Friedman Foundation for Education Choice. Retrieved June 1, 2016, from: https://www.edchoice.org/research/the-education-debit-card-ii/

Carnoy, M., Adamson, F., Chudgar, A., Luschei, T. F., & Witte, J. F. (2007). Vouchers and public school performance: A case study of the Milwaukee parental choice program. Washington, D.C.: Economic Policy Institute.

Chakrabarti, R. (2008). Can increasing private school participation and monetary loss in a voucher program affect public school performance? Evidence from Milwaukee. *Journal of Public Economics, 92*(5-6), 1371–1393.

Cowen, J. M. (2008). School choice as a latent variable: Estimating the "complier average causal effect" of vouchers in Charlotte. *Policy Studies Journal, 36*(2), 301-315.

Coulson, A. J. (2009). Comparing public, private, and market schools: The international evidence. *Journal of School Choice, 3*(1), 31-54.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics, 5*(2), 29-57.

Dixon, P. (2013). *International aid and private schools for the poor: Smiles, miracles and markets*. Northampton, MA: Edward Elgar.

Doolittle, F., & Connors, W. (2001). Designing education voucher experiments: Recommendations for researchers, funders, and users. In Levin, H. M. (Ed.) *Privatizing education: Can the marketplace deliver choice, efficiency, equity, and social cohesion?* Boulder, CO: Westview Press, Perseus Books Group.

Dynarski, M., Rui, N., Webber, A., Gutmann, B., & Bachman, M. (2017). Evaluation of the DC Opportunity Scholarship Program: Impacts after one year. NCEE 2017-4022. Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance. U. S. Department of Education.

EdChoice (2017). The ABCs of school choice: The comprehensive guide to every private school choice program in America. EdChoice.

Egalite, A. J. (2014). Competitive impacts of means-tested vouchers on public school performance: Evidence from Louisiana and Indiana. PEPG Working Paper No.14-05. Cambridge, MA: Harvard University.

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.

Epple, D., Romano, R. E., & Urquiola, M. (2015). School vouchers: A survey of the economics literature (NBER Working Paper No. 21523). Cambridge, MA: National Bureau of Economic Research.

Figlio, D., & Karbownik, K. (2016). Evaluation of Ohio's EdChoice Scholarship Program: Selection, competition, and performance effects. Washington, DC: Thomas B. Fordham Institute.

Forster, G. (2011). A win-win solution: The empirical evidence on school vouchers. Indianapolis: Friedman Foundation for Educational Choice.

Forster, G. (2013). A win-win solution: The empirical evidence on school choice. Indianapolis: Friedman Foundation for Educational Choice.

Forster, G. (2016). A win-win solution: The empirical evidence on school choice. Indianapolis: Friedman Foundation for Educational Choice.

Fortson, K., Verbitsky-Savitz, N., Kopa, E., Gleason, P. (2012). Using an experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates. U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, NCEE 2012-4019.

Friedman, M. (1955). The role of government in education. In Solo, R. A. (Ed.). *Economics and the public interest* (pp. 123–144). New Brunswick, NJ: Rutgers University Press.

Fryer Jr., R. G. (2017). The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments. In Banerjee, A. V. & Duflo, E. (Ed.). *Handbook of Economic Field Experiments*, (Vol. 2, p. 95-322). Amsterdam, Netherlands: Elsevier.

Glenn, C. L. (1989). Choice of schools in six nations: France, Netherlands, Belgium, Britain, Canada, West Germany. University of Michigan.

Glenn, C. L., De Groof, J., & Candal, C. S. (2012a). *Balancing freedom, autonomy and accountability in education*. Volume 1. Wolf Legal Publishers.

Glenn, C. L., De Groof, J., & Candal, C. S. (2012b). *Balancing freedom, autonomy and accountability in education*. Volume 2. Wolf Legal Publishers.

Glenn, C. L., De Groof, J., & Candal, C. S. (2012c). *Balancing freedom, autonomy and accountability in education*. Volume 3. Wolf Legal Publishers.

Glenn, C. L., De Groof, J., & Candal, C. S. (2012d). *Balancing freedom, autonomy and accountability in education*. Volume 4. Wolf Legal Publishers.

Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In Hanushek, E. A., Machin, S. J., & Woessmann, L. (Eds.). *Handbook of the Economics of Education*, (Vol. 5, pp. 653-743). Amsterdam, Netherlands: Elsevier.

Greene, J. P. (2000). The effect of school choice: An evaluation of the Charlotte children's scholarship fund program. Manhattan Institute for Policy Research Civic Report, 12.

Greene, J. P., & Forster, G. (2002). Rising to the challenge: The effect of school choice on public schools in Milwaukee and San Antonio. New York City: Manhattan Institute for Policy Research.

Greene, J. P., Peterson, P. E., & Du, J. (1999). Effectiveness of school choice: The Milwaukee experiment. *Education and Urban Society, 31*(2), 190-213.

Greene, J. P., & Winters, M., A. (2007). An evaluation of the effect of DC's voucher program on public school achievement and racial integration after one year. *Catholic Education: A Journal of Inquiry and Practice, 11*(1), 83–101.

Heckman, J. J. (1996). Randomization as an Instrumental Variable. *Review of Economics and Statistics, 78*(2), 336-341.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Academic Press.

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560.

Howell, W. G., & Peterson, P. E. (2006). *The education gap: Vouchers and urban schools*. Washington, DC: Brookings Institution Press.

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School vouchers and academic performance: Results from three randomized field trials. *Journal of Policy Analysis and Management, 21*(2), 191-217.

Hoxby, C., M. (2003). School choice and school productivity. In Hoxby, C. M. (Ed.). *The Economics of School Choice*. Chicago, IL: University of Chicago Press.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Washington, DC: Sage.

Jin, H., Barnard, J., & Rubin, D. B. (2010). A modified general location model for noncompliance with missing data revisiting the New York City school choice scholarship program using principal stratification. *Journal of Educational and Behavioral Statistics, 35*(2), 154-173.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist, 47*(5), 658-698.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. NCSER 2013-3000. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Lubienski, C., & Brewer, T. J. (2016). An analysis of voucher advocacy: Taking a closer look at the uses and limitations of "gold standard" research. *Peabody Journal of Education, 91*(4), 455-472.

Lubienski, C., & Weitzel, P. (2008). The effects of vouchers and private schools in improving academic achievement: A critique of advocacy research. *Brigham Young University Law Review*, 447-485.

Mills, J. N., & Wolf, P. J. (2016). The effects of the Louisiana scholarship program on student achievement after two years. Available at SSRN: https://ssrn.com/abstract=2738805

Mills, J. N., & Wolf, P. J. (2017). The effects of the Louisiana Scholarship Program on student achievement after three years (Louisiana Scholarship Program Evaluation Report #5). New Orleans, Louisiana: Education Research Alliance for New Orleans & School Choice Demonstration Project, Department of Education Reform, University of Arkansas.

Miron, G., Evergreen, S., Urschel, J. (2008). The impact of school choice reforms on student achievement. The Great Lakes Center for Education Research & Practice.

Mitra, S., Tooley, J., Inamdar, P., & Dixon, P. (2003). Improving English pronunciation: An automated instructional approach. *Information Technologies and International Development, 1*(1), 75-84.

Mizala, A., & Romaguera, P. (2000). School performance and choice: The Chilean experience. *The Journal of Human Resources, 35*(2), 392-417.

Morgan, C., Petrosino, A., & Fronius, T. (2015). The impact of school vouchers in developing countries: A systematic review. *International Journal of Educational Research, 72*, 70-79.

Mosteller, F. & Boruch, F. (Eds.) (2002). *Evidence matters: Randomized trials in education research*. Washington, D.C.: The Brookings Institution.

Muralidharan, K., & Sundararaman. V. (2015). The aggregate effect of school choice evidence from a two-stage experiment in India. *The Quarterly Journal of Economics, 130*(3), 1011-1066.

Murnane, R.J. (2005). The role of markets in American K-12 education. In Nelson, R. R. (Ed.). *The limits of market organization.* New York: Russell Sage Foundation.

Peterson, P. E., & Howell, W. G. (2004). Efficiency, bias, and classification schemes: A response to Alan B. Krueger and Pei Zhu. *American Behavioral Scientist, 47*(5), 699-717.

Peterson, P. E., Howell, W. G., Wolf, P. J., & Campbell, D. E. (2003). School vouchers: Results from randomized experiments. In Hoxby, C. M. (Ed.). *The economics of school choice* (pp. 107-144). University of Chicago Press.

Pirog, M. A., Buffardi, A. L., Chrisinger, C. K., Singh, P., & Briney, J. (2009). Are alternatives to randomized assignment nearly as good? Statistical corrections to nonrandomized evaluations. *Journal of Policy Analysis and Management*, *28*(1), 169–172.

PROBE Team. (1999). Public report on basic education in India. Oxford, Eng.: Oxford University Press.

Ravitch, D. (2014). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. Vintage.

Rhinesmith, E. (2017). A review of the research on parent satisfaction in private school choice programs. *Journal of School Choice, 11*(4), 585-603.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Seventh edition. Sage.

Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics, 113*(2), 553-602.

Rouse, C. E., & Barrow, L. (2009). School vouchers and student achievement: Recent evidence and remaining questions. *Annual Review of Economics*, *1*(1), 17–42.

Sen, R., & Blatchford, P. (2001). Reading in a second language: Factors associated with progress in young children. *Educational Psychology 21*(2), 189-202.

Sude, Y., DeAngelis, C. A., & Wolf, P. J. (2018). Supplying choice: An analysis of school participation decisions in voucher programs in Washington, DC, Indiana, and Louisiana. *Journal of School Choice*, *12*(1), 8-33.

Tooley, J., & Dixon, P. (2002). Private schools for the poor: A case study from India. Reading, UK: CfBT.

Tooley, J. (2009). *The beautiful tree: A personal journey into how the world's poorest people are educating themselves*. Washington, D. C.: Cato Institute.

Usher, A., & Kober, N. (2011). Keeping informed about school vouchers: A review of major developments and research. Center on Education Policy.

Witte, J. F. (2000). *The market approach to education: An analysis of America's first voucher program*. Princeton, NJ: Princeton University Press.

Witte, J. F., Wolf, P. J., Cowen, J. M., Carlson, D. E., & Fleming, D. J. (2014). High-stakes choice: Achievement and accountability in the nation's oldest urban voucher program. *Education Evaluation and Policy Analysis, 36*(4), 437-456.

Wolf, P. J. (2007). Civics exam: Schools of choice boost civic values. *Education Next, 7*(3), 66-72.

Wolf, P. J. (2008a). Vouchers. In The International Encyclopedia of Education, (pp. 635-36). Routledge.

Wolf, P. J. (2008b). School voucher programs: What the research says about parental school choice. *Brigham Young University Law Review*, 415-446.

Wolf, P. J., Egalite, A. J., & Dixon, P. (2015). Private school choice in developing countries: Experimental results from Delhi, India. *Handbook of International Development and Education*, 456-471.

Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management, 32*(2), 246-270.

Wolf, P. J., & Macedo, S., eds. (2004). *Educating citizens: International perspectives on civic values and school choice*. Brookings.

Wolf, P. J. & McShane, M. (2013). Is the juice worth the squeeze? A benefit/cost analysis of the District of Columbia opportunity scholarship program. *Education Finance and Policy, 8*(1), 74–99.

Appendix A. Details on Search and Exclusion Process

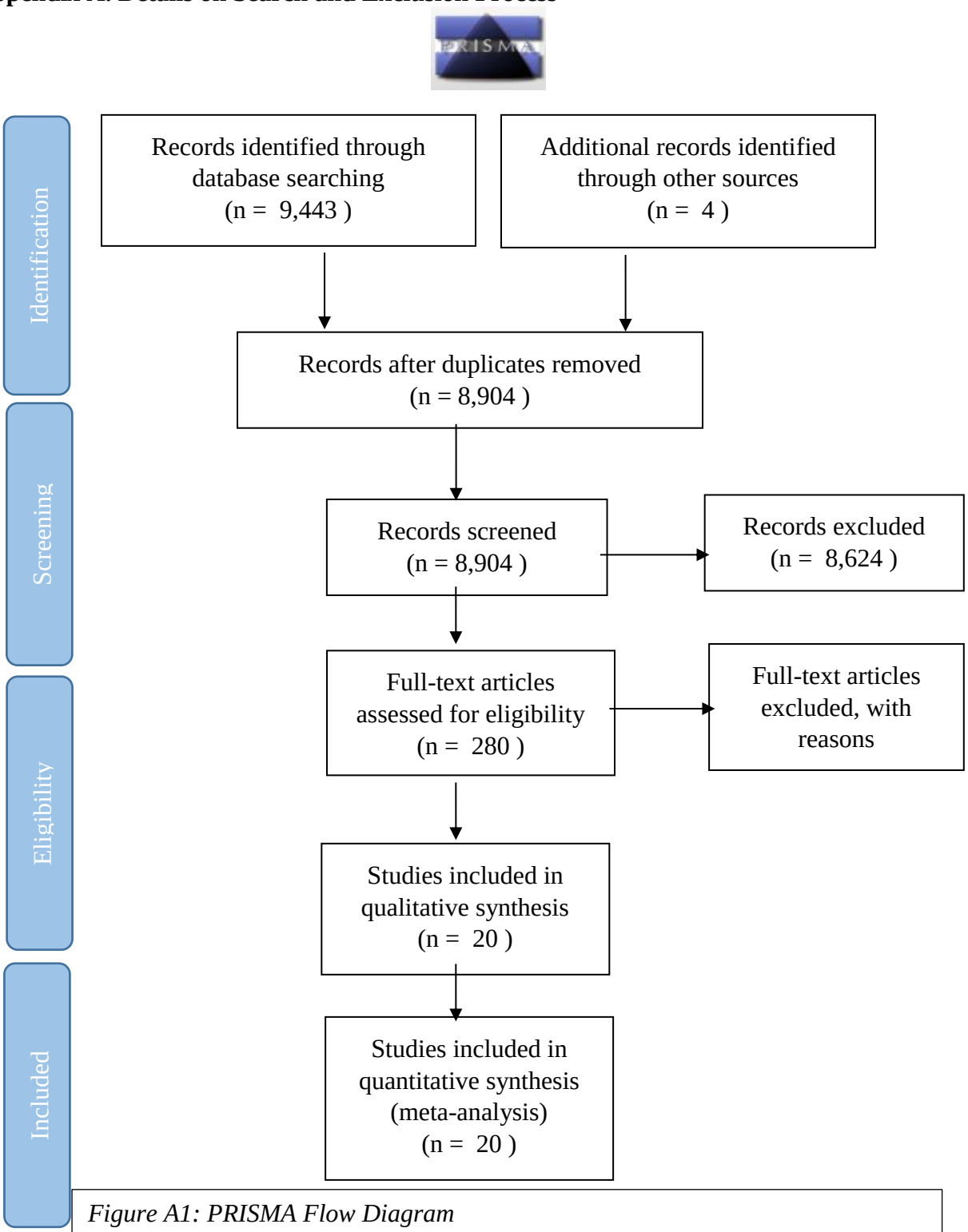

Figure A1: PRISMA Flow Diagram
From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. For more information, visit www.prisma-statement.org.

Table A1: Overview of Article Sources and Exclusions

	Number of Articles
Search 1 (University Library)	
Three library sources (EBSCO, JSTOR, ProQuest)	2,737
Duplicates Removed	-534
Unique articles (EBSCO, JSTOR, ProQuest)	2,203
Excluded Based on Title and/or Abstract	-2,075
Remaining Articles (EBSCO, JSTOR, ProQuest)	**128**
Search 2 (Google Scholar)	
Number of Google Scholar Sources Initially Found	6,706
Excluded Based on Title and Abstract	-6,549
Remaining Google Articles	157
Duplicates Removed	-9
Remaining Articles (Google Scholar)	**148**
Sum of Remaining Articles (Both Searches)	**276**
Excluded Based on Full Article	-260
Studies added through networked search	+4
Total search results (RCTs)	**20**

Table A2: Reason for Exclusion - for 260 papers were excluded at full article phase

Not a quantitative analysis	69
Not related to educational vouchers for K-12 students (housing vouchers, etc.)	30
Private schooling in general (not voucher specific)	7
Religious schooling in general (not voucher specific)	1
Different question related to vouchers (e.g. competitive effects, cost efficiency, segregation/stratification, school participation, parental preferences, etc.)	63
Earlier version of an included study	24
Not randomly assigned	63
Methodology fails to utilize available lottery to conduct RCT	3
Lack of information necessary to be included*	1
Total Excluded	**260**

*Attempted to find author to obtain this information, but was unable to.

Details on Search Strategy

Our search process was comprised of two stages. Our initial search focused on only the studies published since 2005 or later, but due to a lack of RCT studies identified during this process, we added a second search, including all years, but narrowing the search criteria to only include studies that included text related to randomization. The study selection was based on systematic search procedures. Keywords and phrases were chosen to be as inclusive as possible for our preliminary search.

The search criteria were as follows:

Initial Search: 2005 or later

EBSCO Search 1

Search terms: school voucher* OR education* voucher*

Time period: 2005 or later

Types of sources included: Academic Journals, Journals, and Reports

Total number of results: 765

EBSCO Search 2

Search terms: opportunity scholarship

Time period: 2005 or later

Types of sources included: Academic Journals, Journals, and Reports

Total number of results: 48

JSTOR Search 1

Search terms: voucher* AND education* or school AND research AND experiment* or "randomized controlled trial"

Time period: 2005 or later

Language: English

Included only Articles related to: Business and Economics, Economics, Education, Political Science, Public Policy & Administration, Social Sciences

Total number of results: 853 search results

JSTOR Search 2

Search terms: "opportunity scholarship"

Time period: 2005 or later

Language: English

Included only Articles related to: Business and Economics, Economics, Education, Political Science, Public Policy & Administration, Social Sciences.

Total number of results: 30 search results

ProQuest Search 1

Search terms: all(voucher) AND all(school*) AND all(research*)

Time period: 2005 or later

Excludes: Wire Feeds, Magazines, and Newspapers

Total number of results: 603 results

ProQuest Search 2

Search terms: all("opportunity scholarship")

Time period: 2005 or later

Excludes: Wire Feeds, Magazines, and Newspapers

Total number of results: 122 results

The searches of the three library databases (EBSCO, JSTOR, and ProQuest) resulted in a total of 1,934 unique papers, after removing duplicates.

<u>Secondary Search: All RCTs (including prior to 2005)</u>

Since RCTs or experiments are especially prized as education evaluations, we decided to extend our meta-analysis to any RCTs we could find on the topic, regardless of when they were conducted or published. In order to find these, a secondary search was conducted.

EBSCO Search 3 (for all RCTs)

Search terms: school voucher* OR education* voucher* AND AB: random*

Time period: No restriction

Types of sources included: Academic Journals, Journals, and Reports

Total number of results: 85

Note: AB: random means that the abstract had to include a stem of the word random**

EBSCO Search 4 (for all RCTs)

Search terms: opportunity scholarship AND AB: random*

Time period: No restriction

Types of sources included: Academic Journals, Journals, and Reports

Total number of results: 9

Note: AB: random means that the abstract had to include a stem of the word random**

JSTOR Search 3

Search terms: voucher* AND education* or school AND ab(random*)

Time period: No restriction

Language: English

Included only Articles related to: Business and Economics, Economics, Education, Political Science, Public Policy & Administration, Social Sciences

Total number of results: 116 search results

JSTOR Search 4

Search terms: "opportunity scholarship" AND ab(random*)

Time period: No restriction

Language: English

Included only Articles related to: Business and Economics, Economics, Education, Political Science, Public Policy & Administration, Social Sciences.

Total number of results: 2 search results

ProQuest Search 3

Search terms: all(voucher) AND all(school*) AND all(research*) AND ab(random*)

Time period: No restriction

Excludes: Wire Feeds, Magazines, and Newspapers

Total number of results: 95 results

Note: ab(random) means that the abstract had to include a stem of the word random**

ProQuest Search 4

Search terms: all("opportunity scholarship") AND ab(random*)

Time period: No restriction

Excludes: Wire Feeds, Magazines, and Newspapers

Total number of results: 9 results

Note: ab(random) means that the abstract had to include a stem of the word random.*

This secondary search of the three library databases (EBSCO, JSTOR, and ProQuest) resulted in a total of 269 additional unique papers, after removing duplicates.

Google Scholar and Other Website Searches

In addition to the three main library databases, we searched a variety of other sources. First, using the first search criteria, we searched Google Scholar for articles from 2005 or later using the search terms "school voucher" OR "voucher school" to find the maximum number of results. The search returned approximately 4,000 results including patents and citations. Other places we searched, due to their interest in school vouchers, were the websites of the National Bureau of Economic Research (NBER), University of Chile, Uppsala University in Sweden, and the Poverty Action Lab at MIT.

Using the second search criteria in Google Scholar: (("opportunity scholarship" OR "education* voucher*" OR "school voucher*") AND random*), we found 2,570 results including citations. Apart from importing the references in Refworks, we also did individual Google Scholar searches of the imported references whose titles did not end up in Refworks.

Details on Selection Process and Coding

After our title/abstract review, 148 sources remained from the Google and snowball search along with 128 sources from the library searches. Two members of our team reviewed each of these 276 sources in their entirety to determine if they met our inclusion criteria. In some cases, the researchers initially disagreed on the inclusion decision, in which case they met to discuss and come to a consensus. 260 papers were excluded at the full article phase. The reasons for exclusion of the 260 papers have been provided above in Appendix A, Table A2. Additional efforts were made to ensure our results are not affected by publication bias (described in the following section).

The remaining 20 studies were coded using a predesigned format in MS-Excel. The coding format included information on authors, publication year, years of treatment, program evaluated by the study, location of the study, source of funding, duration of study, grades analyzed, sample sizes of treatment and control groups, total sample sizes and information on program and sample attrition. All information was collected separately for different years of treatment reported in the studies and separately by TOT/ITT and separately by math/reading (information for effect sizes in English and local languages were coded for international studies).

Statistical details such as mean test scores (treatment/control), differences in means (treatment – control), standard deviation (treatment/control), treatment effect sizes (in standard deviations), standard error of Cohen's *d*, treatment effect and standard error in other units (such as National Percentile Rank (NPR)), t-statistic, p-values, upper/lower 95% confidence intervals were coded into the excel sheets. Every detail entered into the excel sheet had to be finally agreed upon by two of the coauthors to reduce human error. Lastly the information was used to arrive at Hedge's *g* and standard error of Hedge's *g* for carrying out the meta-analysis. The excel

sheet also contained information on comments/assumptions made for each row as well as mini meta-analytic estimates for multiple studies that reported an effect of the same program for the same year. For example, a mini meta-analysis was carried out to obtain an overall ITT effect for reading estimates for one year of treatment from two studies related to the Charlotte, NC program: Greene (2000) and Cowen (2008). The mini-meta analysis relied on a fixed effects strategy (Borenstein Hedges, & Rothstein, 2007) due to small sample size in all cases, and a pre-coded excel sheet was used to arrive at these mini meta-analytic estimates using the formula described in Appendix B. As a robustness check, STATA software was also used to check the accuracy of the estimates obtained using the pre-coded excel sheet in Excel.

Robustness check for unpublished literature

Tables in Appendix A include details of the literature identified through the computerized and network search described above. The final 20 studies included for meta-analysis are either a published study (when a study actually got published; this forms the majority of studies included), a working paper (when the paper did not yet get published; an example is Abdulkadiroglu, Pathak, & Walters, 2015) and a book chapter (Wolf, Egalite, & Dixon, 2015). Wherever necessary, supplementary details were added through other supporting documents available including direct contact with the authors if required.

To ensure our findings were not affected by publication bias, we conducted a variety of additional searches for unpublished documents such as working papers, conference drafts, and technical reports. Also, multiple versions of the same study published elsewhere in book chapters and different formats of publication were read for details. We manually searched known websites of universities and research institutes such as University of Chile, Uppsala University in Sweden, Poverty Action Lab at MIT, national Bureau of Economic Research (NBER), Department For

International Development (DFID) in UK, Center for Civil Society in New Delhi, India etc. Our team members also utilized Google translator in some cases where it appeared to the team member that a document in foreign language may tell us something about a voucher RCT. To look for possible ongoing evaluations of school vouchers, a hand search of journals that publish studies on school choice such as the *Journal of School Choice* were carried out.

As a last robustness check, a separate search was carried out for master's and doctoral theses in EBSCO, JSTOR, and ProQuest databases using the same search terms as described in above sections of Appendix A. Finally, one of the coauthors of this study is an internationally known expert in school choice and has a well-connected network of leading researchers in the area. The coauthor independently did a search in his contacts to find past or ongoing projects without publicly available results. No study or unpublished document found in the robustness checks was an additional experimental evaluation of school vouchers. Thus, no further studies contribute to the meta-analysis.

Appendix B: Formula used during meta-analysis

1. **Mean differences:** $\bar{X}_T - \bar{X}_C$
2. **SD Pooled :** $Std_{(pool)} = \sqrt{\frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2}}$
3. **Cohen's D:** $d = \frac{\bar{X}_T - \bar{X}_C}{Std_{(pool)}}$
4. **Lower bound ES (95%):** LB = $ES - (SE_d * 1.96)$
5. **Upper bound ES (95%)**: UB= $ES + (SE_d * 1.96)$
6. **Effect Size by correlation:** $ES = \frac{2r}{\sqrt{1-r^2}}$
7. **Effect Size by t ratio:** $d = t\sqrt{\frac{n_1+n_2}{n_1 n_2}}$
8. **Hedges' g (Unbiased D):** $ES(d') = \left[1 - \frac{3}{4N-9}\right] d$
9. **Standard error for effect size:** $SE_{d'} = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{d'^2}{2(n_1+n_2)}}$
10. **Inverse Variance (w)** $w = \frac{1}{(SE)^2}$
11. **Grand Effect size:** $\overline{ES} = \frac{\sum(w \times ES)}{\sum w}$

Where *ES* is effect size of each study, and *w* is the inverse variance weight.

Appendix C: Intent to treat (ITT) analysis

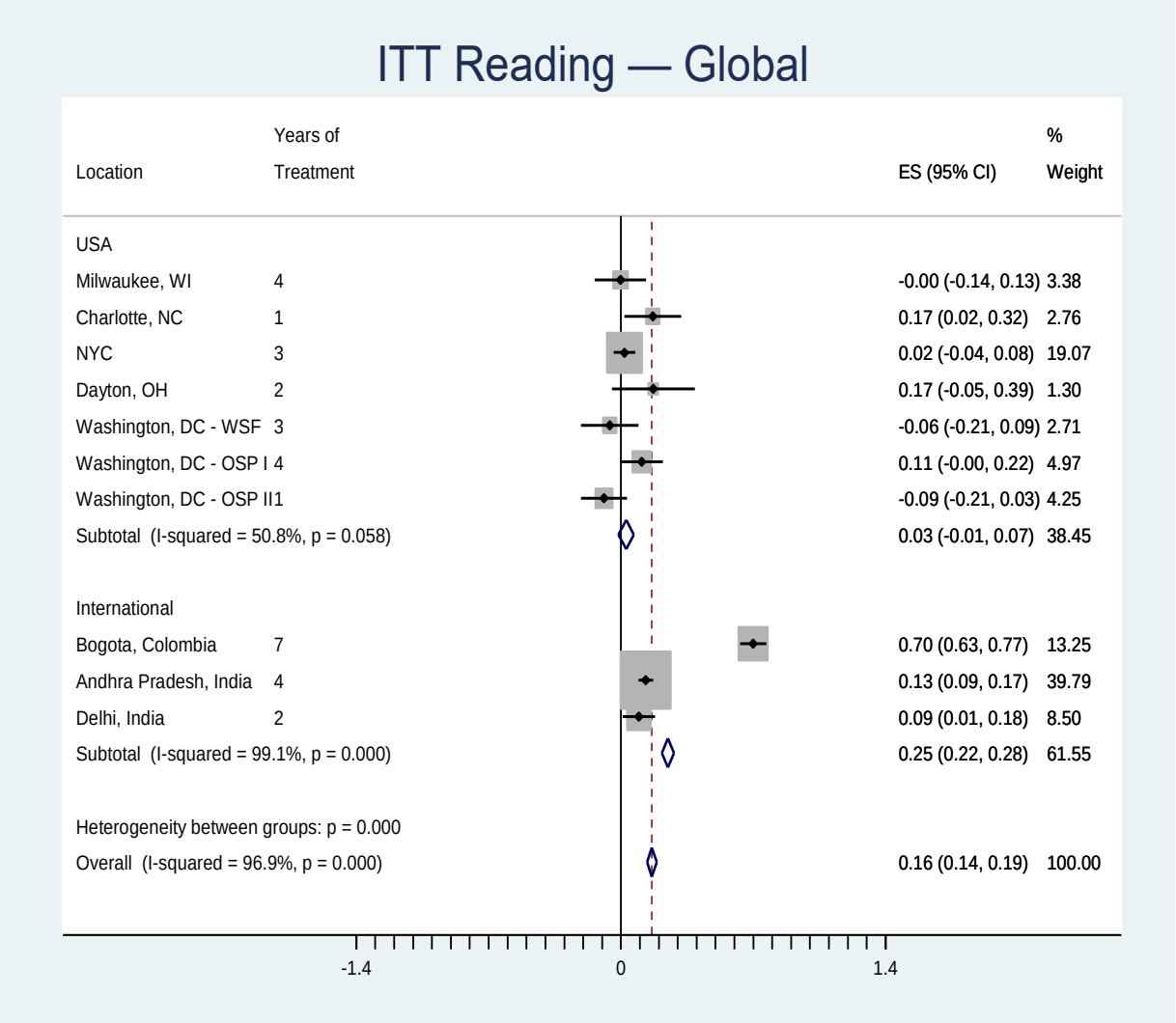

Figure C1: Overall Global impacts – ITT Reading.

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Louisiana voucher program did not have ITT estimates as it was a placement lottery. Overall effect size for non-U.S. studies with Bogota, Colombia removed is 0.12 (0.09, 0.16) and overall global average is 0.08 (0.05, 0.11).

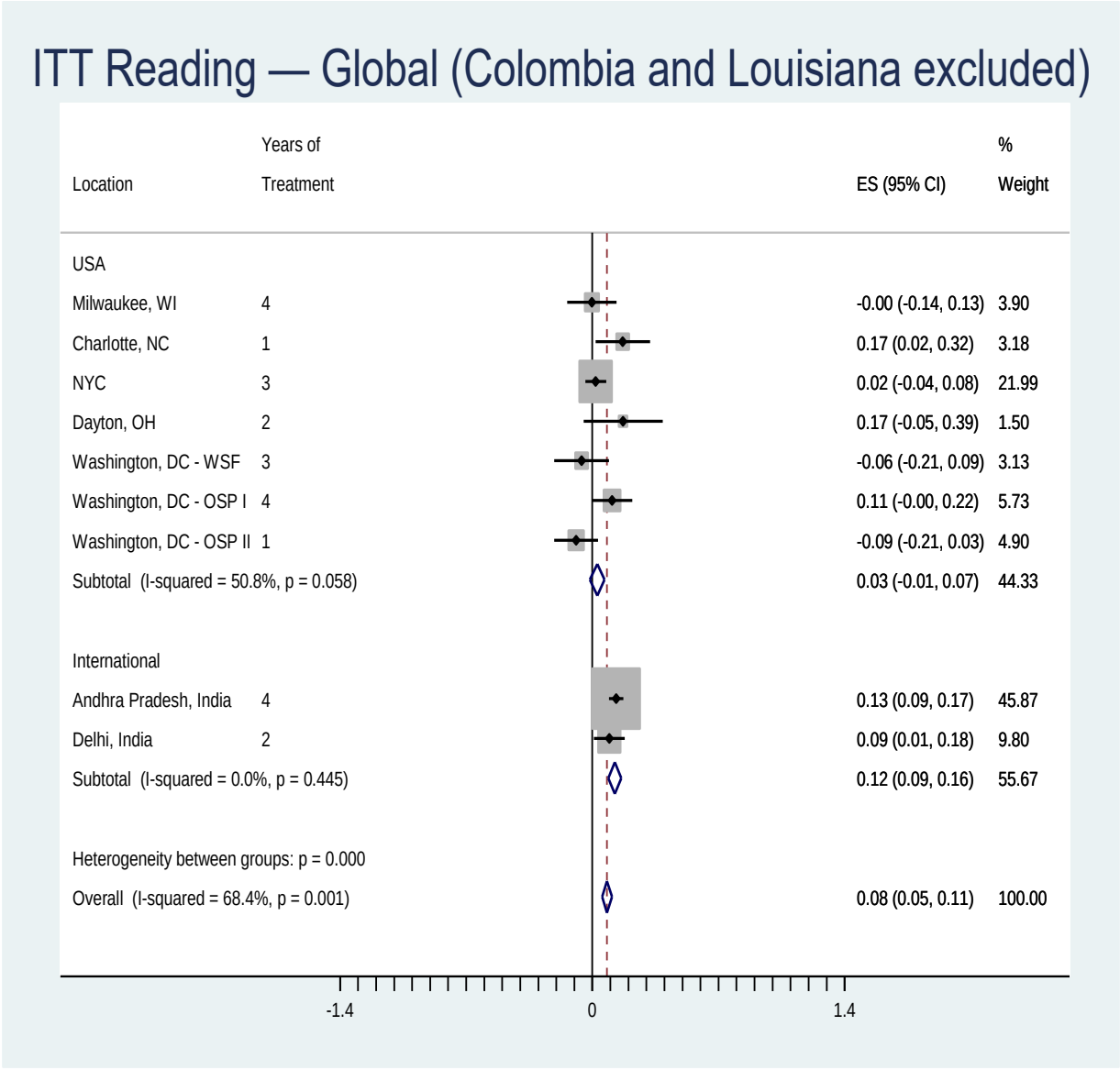

Figure C1.1: Overall Global impacts – ITT Reading (Colombia and Louisiana excluded).

Note: The Hedges' *g* estimates are based on the final year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Louisiana voucher program did not have ITT estimates as it was a placement lottery. Estimates for Colombia and Louisiana are excluded.

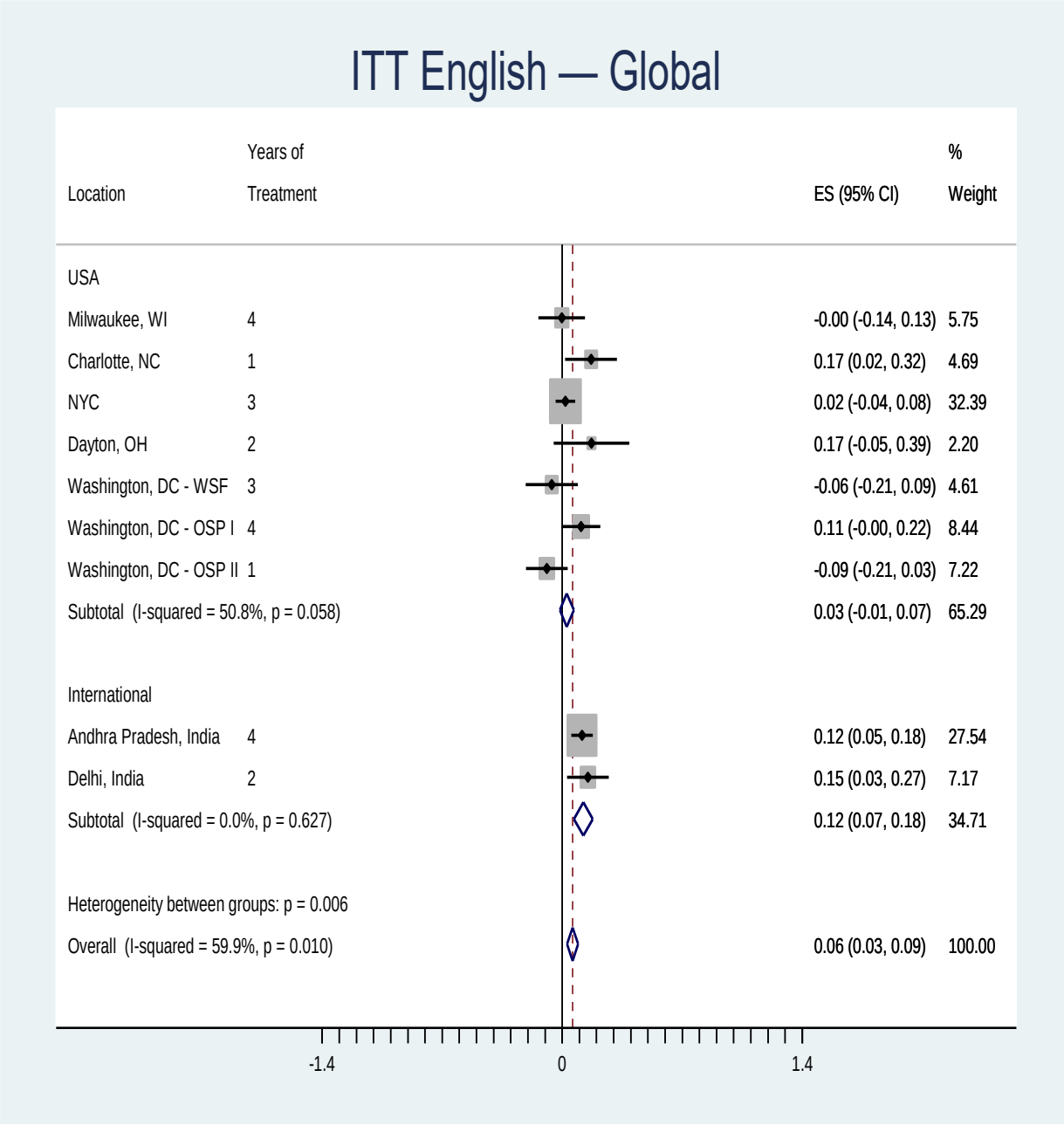

Figure C2: Overall Global impacts – ITT English.

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). No reading estimates are reported for Toledo, OH as it had only math test outcomes. Bogota, Colombia did not have an English estimate as the tests were administered in Spanish. Louisiana voucher program did not have ITT estimates as it was a placement lottery.

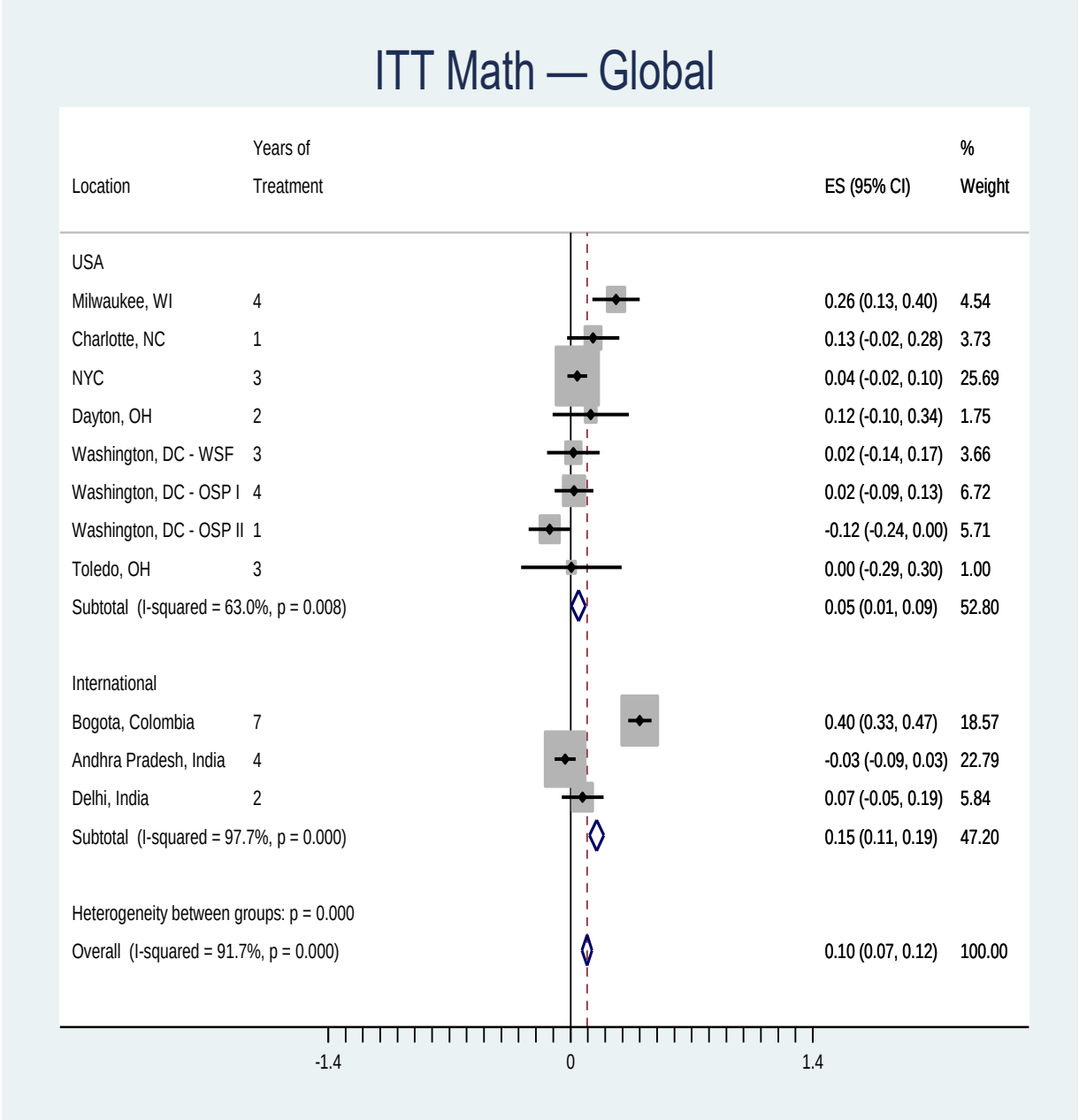

Figure C3: Overall Global impacts – ITT Math.

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). Louisiana voucher program did not have ITT estimates as it was a placement lottery. Overall effect size for non-U.S. studies excluding Bogota, Colombia is -0.01 (-0.06, 0.04) and overall global average is 0.03 (-0.01, 0.06).

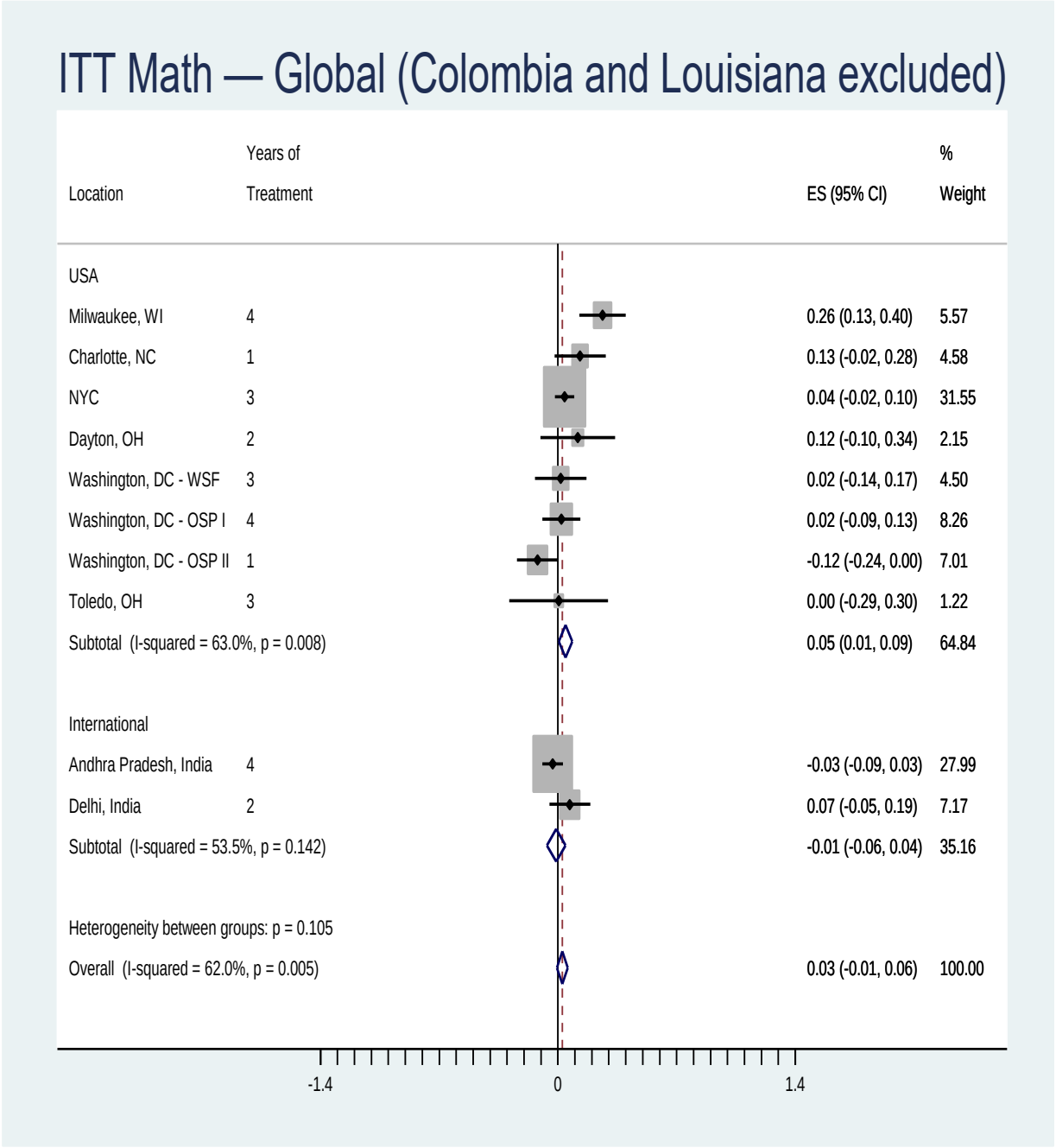

Figure C3.1: Overall Global impacts – ITT Math (Colombia and Louisiana excluded).

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for USA, International (outside USA) and Global (red dotted line). The gray area around each point (effect size) is the weight of each study (inverse of variance). Louisiana voucher program did not have ITT estimates as it was a placement lottery. Estimates for Colombia and Louisiana are excluded.

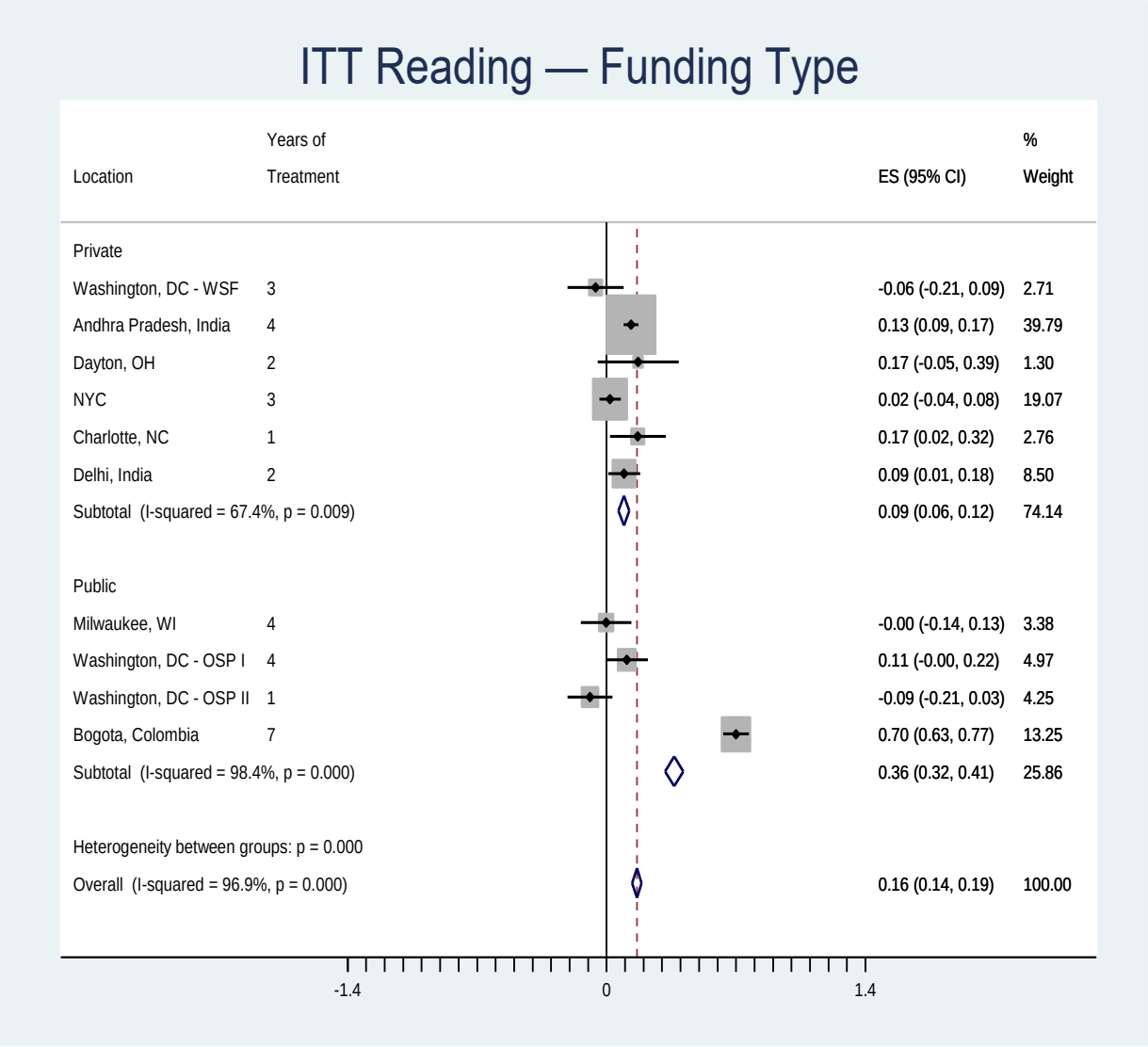

Figure C4: Overall impacts by Funding Type – ITT Reading.

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for privately and publicly (having received any public funds) funded programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). 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. Louisiana voucher program did not have ITT estimates as it was a placement lottery. Overall effect size for publicly funded programs excluding Bogota, Colombia is 0.01 (-0.06, 0.08).

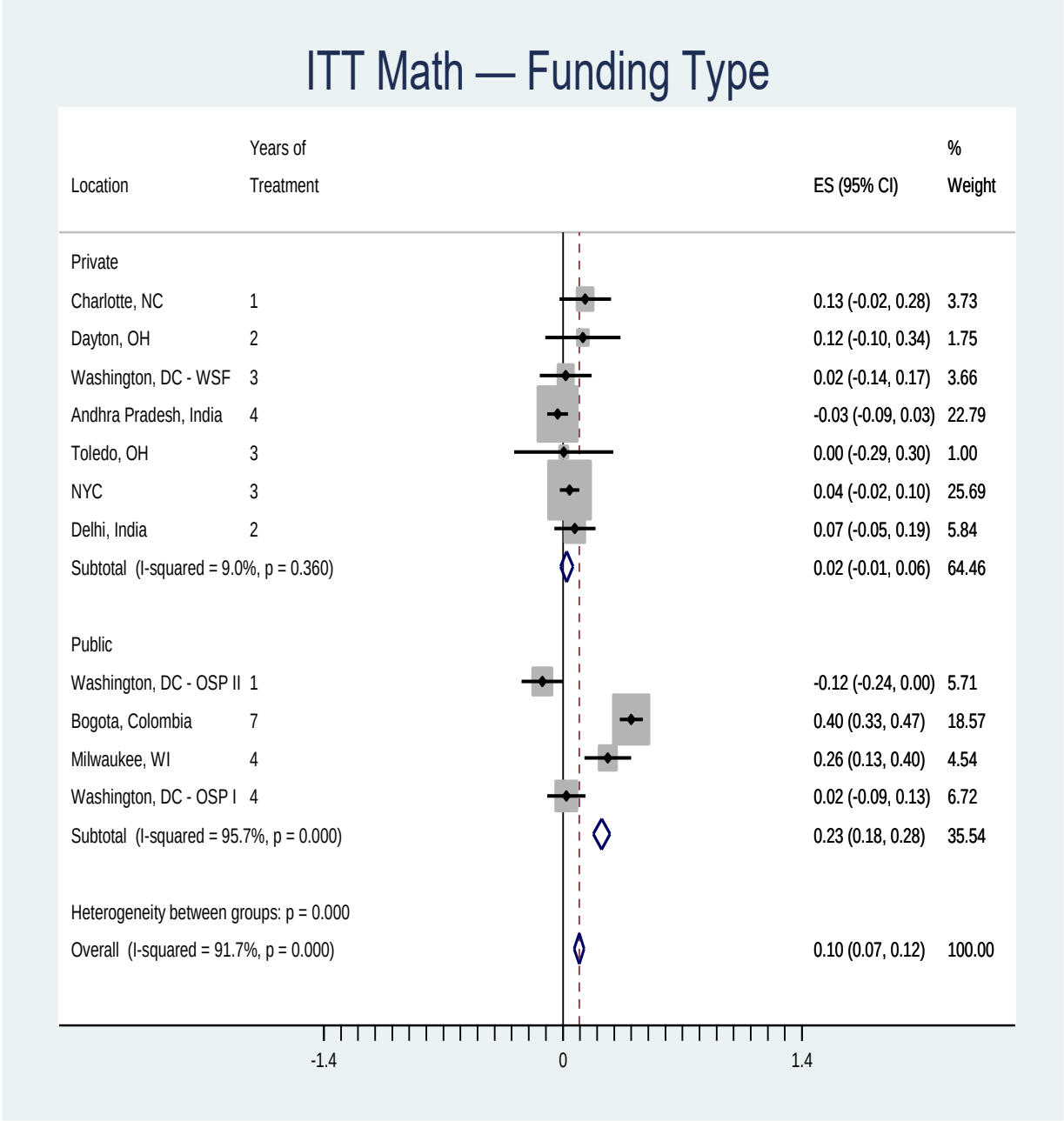

Figure C5: Overall impacts by Funding Type – ITT Math.

Note: The Hedges' *g* estimates are based on last year effect size calculated for each study. The diamonds show overall estimates for privately and publicly (having received any public funds) funded programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). Louisiana voucher program did not have ITT estimates as it was a placement lottery. Overall effect size for publicly funded programs excluding Bogota, Colombia is 0.04 (-0.03, 0.11).

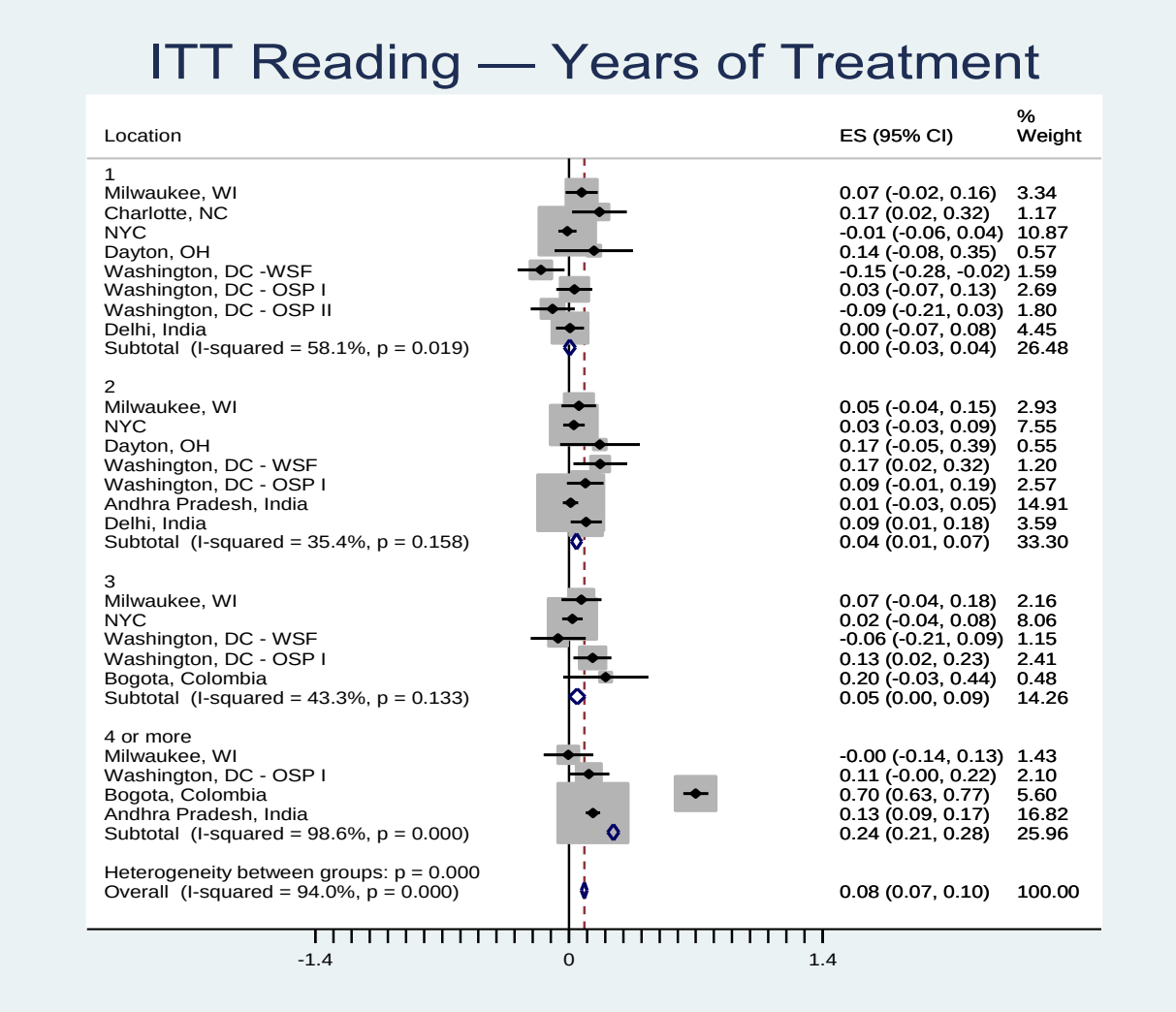

Figure C6: Overall impacts by Years of Treatment – ITT Reading.

Note: The Hedges' *g* estimates are one year, two year, three year, and four or more year effect sizes for each study. The diamonds show overall yearly (dosage) effect estimates. The gray area around each point (effect size) is the weight of each study (inverse of variance). No reading estimates were reported for Toledo, OH. Reading estimate for Delhi is an overall estimate for English and Hindi. Reading estimate for Andhra Pradesh is an overall estimate for English, Hindi, and Telugu. Reading estimate for Bogota is for Spanish. Louisiana uses a placement lottery and thus does not have ITT estimates. Overall effect sizes with Bogota, Colombia removed are 0.04 (-0.00, 0.08) for three years and 0.12 (0.08, 0.16) for four or more years. The overall estimate treats each effect size as independent and we do not focus on it due to spatial auto-correlation among different years of treatment within the same program.

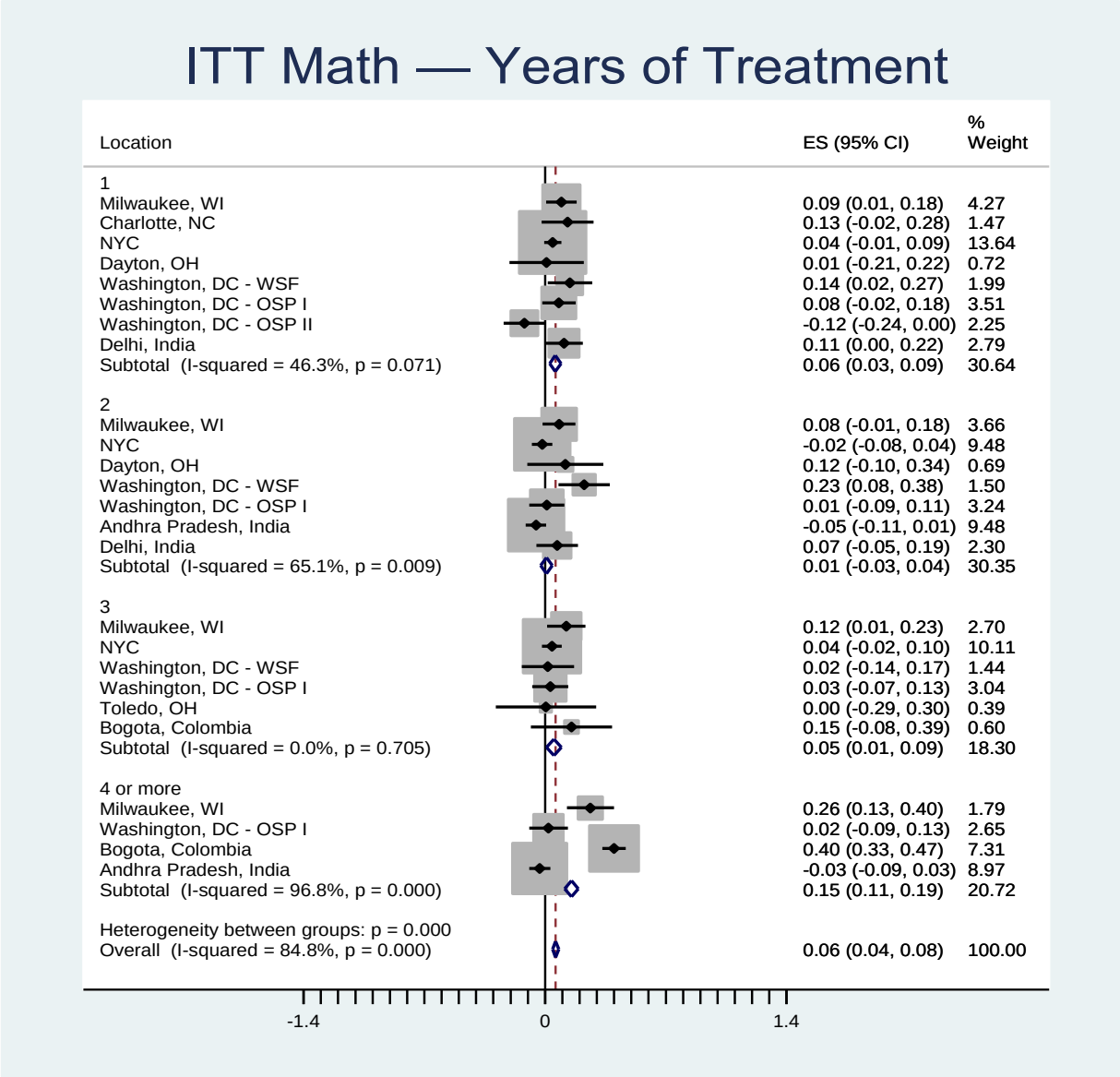

Figure C7: Overall impacts by Years of Treatment – ITT Math.

Note: The Hedges' *g* estimates are based on one year, two year, three year, and four or more year effect sizes calculated for each study. The diamonds show overall estimates for yearly (dosage) effect of programs. The gray area around each point (effect size) is the weight of each study (inverse of variance). Louisiana voucher program did not have ITT estimates as it was a placement lottery. Overall effect size for programs with Bogota, Colombia removed is 0.05 (0.00, 0.09) for three years of treatment and 0.02 (-0.03, 0.07) for four or more years of treatment. The overall estimate treats each effect size as independent and we do not focus on it due to spatial auto-correlation among different years of treatment within the same program.

Appendix D: Assumptions and Calculations for Studies, by Program

Andhra Pradesh (AP) School Choice Experiment

Muralidharan & Sundararaman (2015)

- ITT effects from Table VI, Panel A. Two languages impacts were meta-analyzed into one overall for two years, and three impacts for three years.
- TOT effects from Table VI, Panel B. Two languages impacts were meta-analyzed into one overall for two years, and three impacts for three years.
- 2 year program attrition: 39%: 39% of those offered did not use the voucher (p.10).
- 4 year program attrition: 49.2%: 39% of those offered did not use the voucher (p.10), but at the end of four years only 1,005 out of the 1,980 original treatment group were still using it. (1,980-1,005)/1980 = 49.2%.
- Sample attrition rates differ by year and test but are based on Table A.2 and Table VI. For example, the year 2 English sample attrition is 14.9%: (5,316 – 4,525/5,316) where 5,316 is the sum of the 1,980 + 3,336 in Table A.2 and 4,525 is the sample size in Table VI.
- Control crossover rates not provided.
- Voucher take-up rates were initially 61% (p. 1026), and 51% at the end of the project (p. 1038).

Charlotte Children's Scholarship Fund

Greene (2000)

- Program attrition calculated as the percent of students who were offered a voucher but did not attend divided by the total who were offered a voucher (413/ (413+388)) = 51.6% (p. 3).

- Sample attrition: Overall sample attrition 60% (p. 3).
- TOT estimates are IV results from Table 3. T-statistic was calculated using a p-value of 0.05 and degrees of freedom of 350 (N=357 – 7 variables including constant).
- Treatment/control split was based on the ratio of Choice students to Public students in Table 2 (Choice = 145, Public is 197), applied to the total N of 357.
- ITT estimates were calculated from the TOT estimates using the following Bloom adjustment:

  $$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

  Usage rate for was 48.4% (1-program attrition rate of 51.6%).
- Control crossover rate not provided.
- Voucher take up rate was based on statistics from p. 56 (388/(388+413)) where 388 was the number of students who were offered a voucher that attended private school, and 413 was the number of students offered a voucher who did not use it to attend a private school.

Cowen (2008)

- Program attrition: 25.5% (54/212 of those offered voucher declined it), Table 1 (p. 307).
- Sample attrition: 70% based on 30% of participants with outcome testing (Table 1, p. 307).
- ITT sample sizes from Table 1.
- ITT effects from Table 2.
- TOT in this case is the Complier Average Casual Effect (CACE), the mean treatment outcome across the subpopulation of compliers.
- TOT treatment group sample size (N = 212, number of users, p. 307).

- TOT control group sample size (From Table 1: N = "Total" minus "Choice" = 347 – 158).
- Control crossover rate not provided.
- Voucher take up rate was based on statistics in Table 1.

Children's Scholarship Fund (Toledo, OH)

Bettinger & Slonim (2006)

- Math effects only. ITT effect size from Table 3.
- Used some information from Bettinger & Slonim (2003) as needed.
- Sample size reported in Table 3 (N=349) was based on stacking two sets of math test scores, but this overstates the actual number of students. The footnote indicated 163 students who took both parts of the test, and 23 who took one part of the test, so we used a total sample size of 163 + 23 = 186.
- Control group is calculated as 58% of the 186 total sample, where 58% is the number of lottery losers (1,416 from p. 30), divided by the difference between the number of applicants (2,424) from p. 7 of Bettinger & Slonim (2003) and 39 "mystery winner" students who were excluded from the analysis. 58% = 1,416/(2,424-39).
- Program attrition: N/A. Table 1 on p. 30 indicates that the total number of winners was 2,385 (1,126 + 1,259). The number of losers was 1,416 (331 + 1085), but no indication of how many lottery winners actually used the vouchers.
- Sample attrition: 186 tested out of 2,385, indicates sample attrition of 92% (Table 1).
- TOT math effect was calculated from the ITT estimate using the following Bloom adjustment:

$$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

where usage rate is 43% (p. 12).

- Voucher take-up rates and control crossover rates from p. 12.

District of Columbia Opportunity Scholarship Program (OSP)

- This is the only program having two different evaluations for two distinct samples.

Wolf, Kisida, Gutmann, Puma, Eissa, & Rizzo (2013)

- ITT reading effects from Table 3-2 and Figure 3-1.
- ITT math effects from Tables 3-2 and 4-1 and Figure 3-2.
- Program attrition: Based on p. 67-67 (year 1), p. A-34 (year 2), p. A-32 (year 3), and p. A-41 (year 4).
- TOT effects after year one and two were calculated from the ITT estimates using the following Bloom adjustment:

$$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

where usage rates for year one and two rare based on p. 67-68.

- TOT effects after year three and four were based on percent of "never users."
- Control crossover rates from p. 225.
- Voucher take-up rate from Table 1 (using just cohorts 1 and 2, which represent the analytic sample).

Dynarski, Rui, Webber, Gutmann, & Bachman (2017)

- Program attrition rate from Table 1 (p. 4).
- Sample attrition rates from p. 8.
- Control crossover rates from p. 5 and voucher-take up rate from p. 4.

Ensure Access to Better Learning Experiences (ENABLE)

Wolf, Egalite & Dixon (2015)

- Year 1 ITT: treatment-control means, difference, effect size, and p-value taken from Tables 1, 2, and 3.
- Year 2 ITT: treatment-control means, difference, effect size, and p-value taken from Table 25.2. All other statistics acquired from data output obtained directly from the authors.
- TOT effects were calculated from the ITT estimate using the following Bloom adjustment:

$$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

  where the usage rate is 0.8678.
- Control crossover rate from p. 12-13 and voucher take up rate from p. 13.

Louisiana Scholarship Program (LSP)

Abdulkadiroglu, Pathak, & Walters (2015)

- No ITT effects because it was a placement lottery.
- Sample attrition was 17% for lottery losers (p. 13), and Table 10 indicates the probability of observing a score is about 8 percentage points higher for lottery winners than lottery losers, so we assume 9% sample attrition rate for lottery winners. Overall sample attrition is calculated as the number of attriters divided by the assumed beginning N (1,456) where the assumed beginning N = (treatment N/(1-attrition rate of treatment group) + (control N/(1-attrition rate of control group)). Overall sample attrition, therefore, is (1,456-1,248)/1,456 = 14.3%.
- Treatment and control splits is based the following: Control group sample size is equal to the total sample size from Table 4 (1,247 in Math or 1,248 in Reading) times the loser

rate from Table 10 (903/1412 or about 64%). Then the treatment group size is the Total N – Control N.

- Control crossover rate and voucher take up rate from Table 10.

Mills & Wolf (2017)

- Control crossover rate is based on control "non-compliers" in Table 8.
- Voucher take up rates are the "complier" rates for the treatment group from Table 8

Milwaukee Parental Choice Program (MPCP)

Rouse (1998)

- Treatment and control group sample sizes are based on Table 1, p. 555. Assumption is that reading analytic samples are identical to math analytic samples.
- TOT effect not calculated as the necessary information to compute TOT effects for each year is not available.
- Control crossover rates not available.
- Voucher take up rates are based on the "ever in fall" rates in Table II.

Greene, Peterson, & Du (1999)

- Sample attrition was calculated as the 1 – proportion of each group with test outcomes available. For example, 40% of the treatment group had test data available by the third and fourth year, so sample attrition was 60%. 48% of the control group had test data available by the third and fourth year, so sample attrition was 52%.
- Table 6 was used to calculate treatment/control splits for the ITT estimates. For example, for Reading ITT, Control N= 48/(48+63) or 43.2% of the total sample.
- TOT estimates are not calculated and the method used to derive at TOT estimates is non-experimental.

- Control crossover and voucher take-up rates not available.

Parents Advancing Choice in Education (Dayton, OH)

Peterson, Howell, Wolf & Campbell (2003)

- ITT effects: combined African-American and Other Ethnic Group results from Table 4B.3 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.
- TOT effects: combined African-American and Other Ethnic Group results from Table 4.3 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.
- ITT treatment and Control group sample sizes after years 1 and 2 based on response rate in each year times number of vouchers offered. For example: first year treatment group sample size is the total number of offers times the response rate (515 x 56%) from p. 195 of Howell, Wolf, Campbell, & Peterson (2002).
- Response rates between treatment and control assumed to be the same according to Peterson, Howell, Wolf & Campbell (2003), p. 197.
- Voucher take up rate from Peterson, Howell, Wolf & Campbell (2003).
- Control crossover rate from Table 6. (p. 2014) in Howell, Wolf, Campbell, & Peterson (2002).

Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES)

Angrist, Bettinger, Bloom, King, & Kremer (2002)

- ITT reading effect from Table 5.
- Control group sample size from Table 2, total ITT sample size from Table 5.
- TOT sample sizes from Table 7 (Control = 562, N of "Loser Means'; Total = 1,147)
- Sample attrition (year 3) is based on 283 students who took the test (Table 2) out of the total 1,147 (Table 3).

- Program attrition estimated to be 43% from p. 1543, which stated: "Not all winners were using their PACES vouchers in the survey year. This is because 15 percent of winners were not in school at all, and another 16 percent were in public schools, and therefore ineligible for scholarships. Some lottery winners also lost their voucher after repeating a grade (7 percent), while 5 percent switched to nonparticipating private schools or failed to complete the paperwork for a transfer. Others attended schools that stopped accepting vouchers or lost their vouchers for unreported reasons." The sum of 15%, 16%, 7%, and 5% is 43%.
- TOT effects were bloom adjusted using the voucher take up rate of 57%:

$$TOT\ Reading = \frac{ITT\ Reading\ Impact}{Voucher\ take\ up\ rate}$$

$$TOT\ Math = \frac{ITT\ Math\ Impact}{Voucher\ take\ up\ rate}$$

- Voucher take-up rate from p. 1543 (see program attrition calculation above).
- Control crossover rate is somewhat different from other studies, because what is reported is the percent of control group that received scholarships from other sources (p. 1536).

Angrist, Bettinger, & Kremer (2006)

- ITT effects for year 7 (ICFES exam scores) are the Tobit 10% results on p. 853.
- Total sample size (3,541) from footnote in Table 3. Treatment group was 58.5% of total sample size (Table 1, p. 850)
- Program attrition: 50% within three years (p. 854). Voucher take-up rate based on this as well.
- Sample attrition Table 1 as:

$$Sample\ attrition = \frac{Full\ Sample - Obsereved\ (Valid\ ID\ and\ Age)}{Full\ Sample} = \frac{4{,}044 - 3{,}542}{4{,}044} = 12.4\%$$

- TOT effects were calculated from the ITT estimate using the following Bloom adjustment:

$$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

where the usage rate is 1- program attrition = 1 - .5 = .5

School Choice Scholarships Foundation Program (NYC)

Barnard, Frangakis, Hill, & Rubin (2003)

- Sample attrition: Utilized Table 1 for total number at randomization (676+676 = 1352), and 1,050 as the observed sample, to calculate attrition rate of 22%: (1,352-1,050)/1,352
- Program attrition: Midpoint of 20% and 27%, the percentage of children who won scholarships and did not use them (p. 301).
- ITT effects: overall estimate based on a meta-analytic average of the "Low School" and "High School" impacts presented in Table 4. "Overall" impacts (combination of different grades at application) were used.
- There was a lack of detail on sample sizes, so treatment and control group sample sizes were based on a 50/50 split of the total number of single-child families included in the analysis (p. 301).
- Voucher take-up rates and control crossover rates from p. 301.

Peterson, Howell, Wolf & Campbell (2003)

- ITT effects: combined African-American and Other Ethnic Group results from Table 4B.1 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.

- TOT effects: combined African-American and Other Ethnic Group results from Table 4.2 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.
- ITT treatment and Control group sample sizes after years 1 and 2 based on response rate in each year times number of vouchers offered. For example: first year treatment group sample size is the total number of offers times the response rate (1,300 x 82% = 1,066) from p. 195 of Howell, Wolf, Campbell, & Peterson (2002). first year control group sample size is total N from Peterson, Howell, Wolf & Campbell (2003) of 1,434 minus the 1,066 treatment units.
- Response rates between treatment and control assumed to be the same according to Peterson, Howell, Wolf & Campbell (2003), p. 197, with the exception of in year 2. In year 2, the response rate was 7 percentage points higher in the treatment group than in the control group. Treatment and control split in year 2 was generated so that this differential was approximately 7 percentage points (912/1300 = 70.2% is the treatment group response rate and 284/449 = 63.3% is the control group response rate).
- Control crossover rate from Table 6, p. 2014 oh Howell et al. (2002). No data reported for third year control crossover rate.
- Voucher take-up rate from p. 110 of Peterson et al. (2003).

Krueger & Zhu (2004)

- Assumed to be same data as Bitler et al. (2015) so if statistics were not available in Krueger & Zhu (2004), we referenced Bitler et al. (2015).
- ITT treatment effects from Table 3b (with revised weights and without controls for baseline scores).

- For year three sample sizes, 2,770 is assumed to be the original all inclusive sample, because 1,801 was reportedly left after roughly 35% attrition. Half each of 2,770 is assumed to be treatment and control (1,385 each). Treatment and Control attrition rates (p. 638) were then used to calculate the number of treatment and control units in the analytic sample. For example 35.4% of the control group attrite, so the remaining is 895, and the remaining 906 in the total sample size are assumed to be treatment units.
- Year 1 and 2 treatment and control splits were assumed to be in the same ratio in year three.
- Sample attrition rates for each year were then calculated based on the observed sample size in a given year and the original sample size (2,770).
- Program attrition rates in each year are assumed to be the same as Bitler et al. (2015), from Table A2, Panel B.
- TOT effects from Table 6 2SLS results.
- TOT samples sizes: assumed to be the same as ITT, because not enough information.
- Control crossover rate from p. 695 (percent of students in control group who attended private school in at least one year).
- Voucher-take up rate from Table 1.

Jin, Barnard, & Rubin (2010)

- No ITT effects, because this is just using a different TOT-methodology with the same Barnard et al. (2003) and Krueger & Zhu (2004) sample.
- TOT effects from Table 7. Same assumptions made as Barnard et al. (2003).
- Control crossover and voucher take-up rates from p. 156.

Bitler, Domina, Penner, & Hoynes (2015)

- Sample sizes all assumed to be the same as Krueger & Zhu (2004).
- Sample attrition from Panel A of Table A1 (Bitler et al., 2015). For example, year 1 math attrition was calculated as the difference between the number of students randomized and the number of students with valid test scores (2,666 – 1,977), divided by the number of students randomized (2,666).
- Program attrition: From Panel B of Table A2 (Bitler et al., 2015). For example, in year 1, 1,022 of the 1,292 students randomized were attending a private school, indicating a first year usage rate of 79.1% and program attrition in the first year of 20.9%.
- ITT effects from Table 3, last column.
- TOT effects were calculated from the ITT estimates using the following Bloom adjustment:

$$TOT\ estimate\ = \frac{ITT\ estimate}{usage\ rate}$$

where usage rates were based on Table A2, Panel B. For example, in year 1, 1,022 of the 1,292 students randomized were attending a private school, indicating a first year usage rate of 79.1%.

- Control crossover rate calculated from Table A2, p. 446.
- Voucher take-up rate from Table A2, p. 446 (assuming that if in private, they are using voucher). However, there appears to be an error in Table A2, because the sum of the private and public students among the treatment group does not add to the 1292 (e.g. 352 + 1022 does not equal 1292). Table A1 shows the correct total treatment number (1374) so we use this as the denominator to calculate the take up rates.

<u>Washington Scholarship Fund (WSF)</u>

Peterson, Howell, Wolf & Campbell (2003)

- ITT effects: combined African-American and Other Ethnic Group results from Table 4B.2 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.
- TOT effects: combined African-American and Other Ethnic Group results from Table 4.4 in Peterson, Howell, Wolf & Campbell (2003) using meta-analytic average.
- ITT treatment and control group sizes after years 1 and 2 based on response rate in each year times number of vouchers offered. For example: first year treatment group sample size is the total number of offers times the response rate (809 x 63% = 510) from p. 195 of Howell, Wolf, Campbell, & Peterson (2002)
- Response rates between treatment and control assumed to be the same according to Peterson, Howell, Wolf & Campbell (2003), p. 197.
- The standard error on the three year reading impact for Other Ethnic Groups was not reported in Peterson, Howell, Wolf & Campbell (2003), but due to uniformity of standard error patterns across years within each subject, we calculated an average. For example, the standard errors for DC reading ITT impacts for African-American students were 1.5, 1.4, and 1.5 standard deviations for years 1, 2, and 3). The three year reading ITT standard error is the average of the one and two year standard errors (8.0 and 9.1).
- Control crossover rate from Table 6, p. 204 of Howell et al. (2002). No data reported for third year control crossover rate.
- Voucher take-up rate from p. 111 of Peterson et al. (2003).

Appendix E:

Table E1: Effect Sizes (Hedge's g) and Standard Errors, by Study

			ITT Reading		ITT English		ITT Math		TOT Reading		TOT English		TOT Math	
Program Evaluated	**Studies included**	**Years of Treatment**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**
Andhra Pradesh School Choice Experiment	Muralidharan & Sundararaman (2015)	2	0.009	0.022	0.185	0.031	-0.053	0.030	0.101	0.025	0.364	0.036	-0.104	0.036
		4	0.131	0.020	0.116	0.032	-0.031	0.031	0.271	0.024	0.229	0.038	-0.061	0.037
Charlotte Children's Scholarship Fund	Greene (2000); Cowen (2008)	1	0.168	0.077	0.168	0.077	0.131	0.077	0.217	0.076	0.217	0.076	0.157	0.076
Children's Scholarship Fund (Toledo, OH)	Bettinger & Slonim (2006)	3	N/A	N/A	N/A	N/A	0.005	0.149	N/A	N/A	N/A	N/A	0.011	0.149
District of Columbia Opportunity Scholarship Program (OSP)	Wolf, Kisida, Gutmann, Puma, Eissa, & Rizzo (2013)	1	0.030	0.051	0.030	0.051	0.080	0.050	0.039	0.051	0.039	0.051	0.105	0.050
	Wolf et al. (2013)	2	0.090	0.052	0.090	0.052	0.010	0.052	0.113	0.052	0.113	0.052	0.013	0.052
	Wolf et al. (2013)	3	0.130	0.054	0.130	0.054	0.030	0.053	0.150	0.054	0.150	0.054	0.030	0.053
	Wolf et al. (2013)	4	0.110	0.057	0.110	0.057	0.020	0.057	0.130	0.057	0.130	0.057	0.030	0.057
	Dynarski, Rui, Webber, Gutmann, & Bachman (2017)	1	-0.090	0.062	-0.090	0.062	-0.120	0.062	-0.120	0.062	-0.120	0.062	-0.170	0.062
Ensure Access to Better Learning Experiences (ENABLE)	Wolf, Egalite, & Dixon (2015)	1	0.005	0.039	0.010	0.056	0.110	0.056	0.010	0.039	0.020	0.056	0.130	0.056
		2	0.095	0.044	0.150	0.062	0.070	0.061	0.159	0.044	0.250	0.062	0.120	0.061
Louisiana Scholarship Program (LSP)	Abdulkadiroglu, Pathak & Walters (2015); Mills & Wolf (2017)	1	N/A	N/A	N/A	N/A	N/A	N/A	-0.150	0.039	-0.150	0.039	-0.538	0.040
	Mills & Wolf (2017)	2	N/A	N/A	N/A	N/A	N/A	N/A	-0.181	0.052	-0.181	0.052	-0.337	0.053
	Mills & Wolf (2017)	3	N/A	N/A	N/A	N/A	N/A	N/A	0.080	0.059	0.080	0.059	-0.140	0.059

Table E1 Cont'd: Effect Sizes (Hedge's g) and Standard Errors, by Study

			ITT Reading		ITT English		ITT Math		TOT Reading		TOT English		TOT Math	
Program Evaluated	**Studies included**	**Years of Treatment**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**
Milwaukee Parental Choice Program (MPCP)	Rouse (1998); Greene, Peterson & Du (1999)	1	0.070	0.045	0.070	0.045	0.094	0.045	N/A	N/A	N/A	N/A	N/A	N/A
		2	0.054	0.049	0.054	0.049	0.081	0.049	N/A	N/A	N/A	N/A	N/A	N/A
		3	0.069	0.057	0.069	0.057	0.123	0.057	N/A	N/A	N/A	N/A	N/A	N/A
		4	-0.003	0.069	-0.003	0.069	0.263	0.070	N/A	N/A	N/A	N/A	N/A	N/A
Parents Advancing Choice in Education (Dayton, OH)	Peterson, Howell, Wolf & Campbell (2003)	1	0.137	0.110	0.137	0.110	0.008	0.110	0.008	0.110	0.008	0.110	0.009	0.110
		2	0.171	0.112	0.171	0.112	0.117	0.112	0.173	0.112	0.173	0.112	0.118	0.112
Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES)	Angrist, Bettinger, Bloom, King & Kremer (2002)	3	0.202	0.120	N/A	N/A	0.153	0.120	0.166	0.042	N/A	N/A	0.125	0.059
	Angrist, Bettinger, & Kremer (2006)	7	0.700	0.035	N/A	N/A	0.400	0.034	1.400	0.038	N/A	N/A	0.800	0.035

Table E1 Cont'd: Effect Sizes (Hedge's g) and Standard Errors, by Study

			ITT Reading		ITT English		ITT Math		TOT Reading		TOT English		TOT Math	
Program Evaluated	**Studies included**	**Years of Treatment**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**	**Hedge's g**	**SE**
School Choice Scholarships Foundation Program (NYC)	Peterson, Howell, Wolf & Campbell (2003); Barnard, Frangakis, Hill & Rubin (2003); Jin, Barnard & Rubin (2010)*; Krueger & Zhu (2004); Bitler, Domina, Penner & Hoynes (2015)	1	-0.008	0.025	-0.008	0.025	0.044	0.025	0.011	0.023	0.011	0.023	0.069	0.023
	Peterson et al. (2003); Krueger & Zhu (2004); Bitler et al. (2015)	2	0.027	0.030	0.027	0.030	-0.017	0.030	0.029	0.030	0.029	0.030	-0.037	0.030
	Peterson et al. (2003); Krueger & Zhu (2004); Bitler et al. (2015)	3	0.019	0.029	0.019	0.029	0.039	0.029	0.013	0.025	0.013	0.025	0.042	0.029
Washington Scholarship Fund (WSF)	Peterson, Howell, Wolf & Campbell (2003)	1	-0.154	0.066	-0.154	0.066	0.144	0.066	-0.156	0.066	-0.156	0.066	0.143	0.066
		2	0.173	0.076	0.173	0.076	0.226	0.076	0.172	0.076	0.172	0.076	0.225	0.076
		3	-0.060	0.078	-0.060	0.078	0.016	0.078	-0.064	0.078	-0.064	0.078	0.035	0.078

*Jin, Barnard & Rubin (2010) contributes only to TOT effects.

Chapter 3

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

M. Danish Shakeel (mdshakee@uark.edu)

Kaitlin P. Anderson (kaitlina@uark.edu)

Patrick J. Wolf (pwolf@uark.edu)

Earlier versions of this chapter were presented at the conferences of the Society for Research on Educational Effectiveness in spring 2017 and the Association for Public Policy and Management's International in Brussels, Belgium, July 13-14, 2017.

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. Chapter 2 of this dissertation shows that experimental studies on school voucher programs 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).

Chapter 3 studies the 11 school voucher programs around the globe (also studied in chapter 2 of this dissertation) from a cost-effectiveness perspective. 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. We argue that null to positive test score findings from school voucher program evaluations should be viewed from a cost-effectiveness perspective.

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. Chapter 2 of this dissertation showed that the academic achievement effects of vouchers, though they tend to increase with years of treatment, are generally modest in size. 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 studies on vouchers have found larger positive effects on graduation rates (Cowen et al., 2013; Wolf et al., 2013; Chingos & Peterson, 2015) and college enrollment (Cowen et al., 2013; Chingos & Peterson, 2015) while having null to moderately positive achievement effects on participants. This pattern 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 meta-analysis in chapter 2 of this dissertation indicates that school vouchers tend to produce global test score

effects that are typical of other education interventions. Nevertheless, promising educational interventions must not only be effective but also efficiently use public resources. Hence, we study the cost-effectiveness of the experimental evidence on school vouchers.

This study suffers from several concerns. 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 change in magnitude and statistical significance over time. Concerning the first issue, 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. Howell et al. (2002) found positive achievement gains for the African-American subgroup of participants in their evaluation of the voucher program in Dayton, OH. The same evaluation, however, yielded an overall null achievement impact on all participants. In case of an overall null impact, a targeted voucher program may meet its objectives by raising student achievement for the disadvantaged families. Thus, overall null impacts in a voucher program warrant further analysis.

Second, 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 in chapter 2 of this dissertation shows that the relation between effect size and years of treatment may be non-linear. This pattern creates a dilemma for drawing relevant policy conclusions for the true effect of a voucher intervention, especially if there is large variation, change in sign or statistical significance of effect size over the different years of

treatment. Thus, conclusions drawn from the early years of an evaluation may be substantively different from conclusions drawn from later years of an evaluation, which are arguably more policy relevant, but more time- and resource-intensive to obtain.

This issue 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 β can be efficiently obtained. However, β 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 β. 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. Family satisfaction 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 vouchers 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 their voucher if they passed the exams. 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; Glewwe & Muralidharan, 2016).

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.

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). 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 (quasi-experimental designs allow estimation of the causal impact of an intervention on a targeted population but lack the feature of

random assignment of program participants to treatment and control groups) 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 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 the participant effects of vouchers, the competitive effects of vouchers (the effect of choice-based competition on the performance of affected public schools) are also relevant for studying the cost-effectiveness of vouchers. Positive competitive effects of voucher programs would strengthen the case for the cost-effectiveness of vouchers while negative competitive effects would make an overall conclusion more difficult. Chakrabarti (2008) showed that when money follows the student, there are increasing incentives for the improvement in the traditional public school system. Reviews on competitive effect of vouchers generally indicate positive or neutral to positive impacts of vouchers on the performance of affected public schools (Egalite & Wolf, 2016; Forster, 2016). Given the generally positive competitive effects, the current analysis on the cost-effectiveness of vouchers provides a lower bound on the societal benefit of vouchers.

The cost-effectiveness analysis 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 voucher funding body (i.e. the government). This conclusion is because the typical per-pupil voucher cost is less than the per-pupil cost had the same student 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:

$$Net\ Impact = (MPS\ revenue\ perpupil \times reduction\ in\ MPS\ enrollment) - (voucher \times MPCP\ enrollment)$$

The MPCP program expanded over time. Thus, by comparing the newer estimates of the net fiscal impact of MPCP 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 enrollment 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 voucher programs in the U.S. The contribution of our study relative to Spalding (2014) is the cost-effectiveness analysis of 11 voucher program around the globe that have been experimentally evaluated for achievement effects.

Generally, the students leaving the traditional public schools for a voucher-accepting school generate fiscal benefits for the public school, as the leaving students cost more to educate than the revenue lost from the state (Scafidi 2012; Trivitt & DeAngelis 2016). This is so as the cost of educating a child in the school constitutes both variable (such as costs associated with current instructional expenses, current student support/pupil support, instructional support, and food service) and fixed costs (such as facilities costs). When a child leaves a school, only the variable portion of the costs associated with educating the child is lost but there is no effect on the fixed portion of the cost in the short term. 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.

Tax-credit scholarship programs, which operate like voucher programs but are funded through donations to non-profits, also 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 tax-credit scholarship

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 tax-credit scholarship programs seemingly give more freedom to the taxpayers, they might produce different impacts than school vouchers due to choice differentiation, in a market where school choice is in high demand. In addition, tax-credit scholarship programs have proven to be more politically palatable, as they do not require participation by all taxpayers, as full voucher programs do. On the other hand, tax-credit scholarship programs, if implemented, may have a lower take-up rate by donors, since individuals must choose to donate to nonprofits, as opposed to a 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 and $3,001 per student, on average, for ten tax-credit scholarship programs in the U.S. The cumulative savings per student to the taxpayers for different tax-credit scholarship programs ranged between $298 and $8,450.

Fiscal benefits are not restricted to private school choice. Public charter schools also produce fiscal benefits. Wolf et al. (2014) used two measures for 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 may also be expected in international contexts where 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 of private school vouchers included in our analysis were located in three countries: the United States of America (U.S.), Colombia and India. Eight out of these eleven voucher programs analyzed 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 (Sude, DeAngelis & Wolf, 2018). 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 (as 2013 was the first year of inception of the LSP program) 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. In addition, we adjust for regional cost-of-living within the U.S. using a comparable wage approach (Taylor & Fowler, 2006). Our calculations show that 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 unadjusted amounts of the voucher were approximately $5,000 (median award amount) 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: Description of 11 Voucher Programs in Cost-effectiveness Analysis

Program Evaluated	Location	Funding Source	Funding Amount (Full or Partial)	Cost Difference (Voucher - TPS VC adj)	Size of treatment sample (TOT)	Grades	Years of Treatment	Studies Cited	Lower Bound Net Savings (Cost * Treatment Size * Years)
Andhra Pradesh (AP) School Choice Experiment	Andhra Pradesh, India	Private	Full	-$185	868	1 to 5	4	Muralidharan & Sundararaman (2015)	$ 640,716
Charlotte Children's Scholarship Fund	Charlotte, NC (USA)	Private	Partial	-$4,649	155	2 to 8	1	Greene (2000); Cowen (2008)	$ 720,595
Children's Scholarship Fund	Toledo, OH (USA)	Private	Partial	-$8,116	78	K to 8	3	Bettinger & Slonim (2006)	$ 1,899,144
District of Columbia Opportunity Scholarship Program (OSP)	Washington, DC (USA)	Public	Full	-$4,738	855	K to 12	4	Wolf, Kisida, Gutmann, Puma, Eissa, & Rizzo (2013)	$ 16,203,960
District of Columbia Opportunity Scholarship Program (OSP)	Washington, DC (USA)	Public	Full	-$1,322	635	K to 12	1	Dynarski, Rui, Webber, Gutmann, & Bachman (2017)	$ 839,464
Ensure Access to Better Learning Experiences (ENABLE)	Delhi, India	Private	Full	-$871	619	K to 2	2	Wolf, Egalite, & Dixon (2015)	$ 1,078,081
Louisiana Scholarship Program (LSP)	Louisiana (USA)	Public	Full	-$2,697	492	3 to 8	3	Abdulkadiroglu, Pathak, & Walters (2015); Mills & Wolf (2016); Mills & Wolf (2017)	$ 3,981,152
Milwaukee Parental Choice Program (MPCP)	Milwaukee, WI (USA)	Public	Full	-$3,538	87	K to 8	4	Rouse (1998); Greene, Peterson, & Du (1999)	$ 1,231,224
Parents Advancing Choice in Education	Dayton, OH (USA)	Private	Partial	-$6,396	252	K to 12	2	Peterson, Howell, Wolf, & Campbell (2003)	$ 3,223,584
Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES)	Bogota, Colombia	Public (partly funded by World Bank)	Partial	-$249	2071	6-9 (2002) and 6-11 (2006)	7	Angrist, Bettinger, Bloom, King, & Kremer (2002); Angrist, Bettinger, & Kremer (2006)	$ 3,610,992
School Choice Scholarships Foundation Program	New York, NY (USA)	Private	Partial	-$6,833	890	1 to 4	3	Peterson, Howell, Wolf, & Campbell (2003); Barnard, Frangakis, Hill, & Rubin (2003); Krueger & Zhu (2004); Jin, Barnard, & Rubin (2010); Bitler, Domina, Penner, & Hoynes (2015)	$ 18,244,110
Washington Scholarship Fund	Washington, DC (USA)	Private	Partial	-$6,949	405	K to 8	3	Peterson, Howell, Wolf, & Campbell (2003)	$ 8,443,035
U.S. Programs									$ 54,786,267
Colombia Program									$ 3,610,992
India Programs									$ 1,718,797
TOTAL SAVINGS									$ 60,116,056

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 student population. 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. TOT impacts for Milwaukee are non-experimental. 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 avearge sample size for math and reading effect size estimates.

Funding for the voucher programs in Colombia and India was extremely low. The two privately-funded voucher programs in 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.[20] 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 the meta-analysis in chapter 2 of this dissertation. To graphically analyze the productivity of voucher programs, we compare these experimental estimates of program benefits, in Hedge's g effect sizes (Hedges, 1981),

[20] Chapter 2 of this dissertation excluded TOT impacts for Milwaukee as an experimental estimate was not available. However, ITT impacts for Milwaukee are based on experimental estimates. For allowing comparison of cost-effectiveness across all voucher interventions, the non-experimental TOT estimates for Milwaukee have been maintained in this study.

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). In particular, we use Consumer Price Index (CPI) data from the Bureau of Labor Statistics[21] for the studies within the U.S., CPI data from India's Open Government Data (OGD) Platform[22] for the studies in India, and CPI data from Colombia's Banco de la República[23] for the study in Colombia. 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).[24] 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 approach allows us to analyze the relative productivity of the voucher in comparison to public school expenditures in the same locality (district or state).

We focus on the treatment on the treated (TOT), but we also analyze the ITT estimates.[25] The TOT informs about the impact of the voucher on voucher winners who actually used it, and the Louisiana Scholarship Program was a placement lottery that only yielded TOT 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

[21] Source: https://www.bls.gov/cpi/
[22] Source: https://data.gov.in/catalog/state-level-consumer-price-index-ruralurban
[23] Source: http://www.banrep.gov.co/en/prices/consumer-price-index
[24] Source: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
[25] Results are available from the authors on request.

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.[26] 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 expenditures are subtracted, as the local public school system should theoretically be able to reduce costs by this amount, while being unable to affect fixed costs in the short term. The cost measures for the traditional public school 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,[27] supplemented with state-level databases from Ohio (Ohio

[26] Source: http://www.indiaenvironmentportal.org.in/category/15375/publisher/government-of-nct-of-delhi/
[27] Source: https://www.census.gov/programs-surveys/school-finances.html

Department of Education)[28] and the National Center for Education Statistics Common Core of Data State Fiscal Reports,[29] 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,[30] so we are unable to differentiate between the variable and fixed components of these costs. Therefore, when calculating the net "savings" from voucher programs outside the U.S., we make assumptions about the proportion of total TPS costs that 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

[28] Source: http://education.ohio.gov/Topics/Finance-and-Funding/Finance-Related-Data/Expenditure-and-Revenue/Expenditure-Revenue-Data
[29] Sources: https://nces.ed.gov/pubs2016/2016301.pdf and https://nces.ed.gov/pubs2015/2015301.pdf
[30] http://ccs.in/sites/default/files/research/research-per-child-funding-model-for-schools-in-india.pdf
http://www.accountabilityindia.in/sites/default/files/state-report-cards/paisa_report_2012.pdf

U.S. voucher program.[31] Across all U.S. programs, this percentage ranged from 64% to 79%, and on average was 69%. This average is 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 regional differences 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 those areas. For comparing the cost differences across countries we utilize the purchasing power

[31] For LSP and DC OSP II, we do not have public school expenditures for 2014-2015 (not published yet by NCES) so we use the 2013-14 expenditures instead.

parity (PPP) conversion factor[32] from the World Bank. This tool allows us to adjust for the difference 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, who tends to be younger than the average K-12 student. 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 (i.e., already attending private 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 in developing countries may induce kids who are not attending a public 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 & 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[33] from Table 1 as:

$$Adjusted\ perpupil\ cost\ difference_{(voucher-TPS)} \times Size\ of\ treatment\ sample_{TOT}$$

[32] Source: http://data.worldbank.org/indicator/PA.NUS.PPP
[33] Size of treatment sample may differ from the number of compliers.

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 formula 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. We could think of the privately funded program as providing seed money with the later intention that funds would be public.

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. 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 the meta-analysis in chapter 2 of this dissertation. 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 the type 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 are 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 consistently 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.

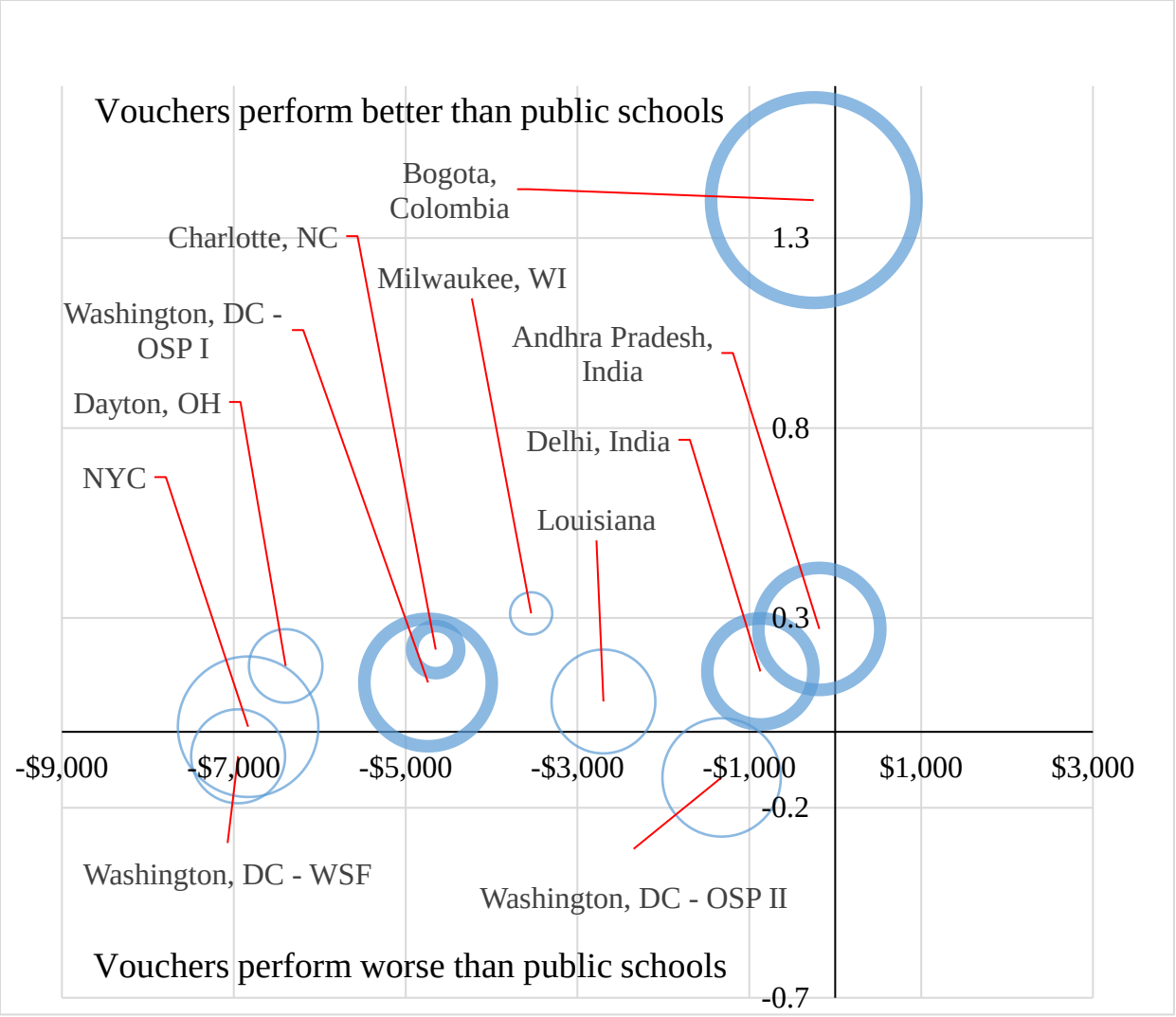

Figure 1: Overall impacts – TOT Reading.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

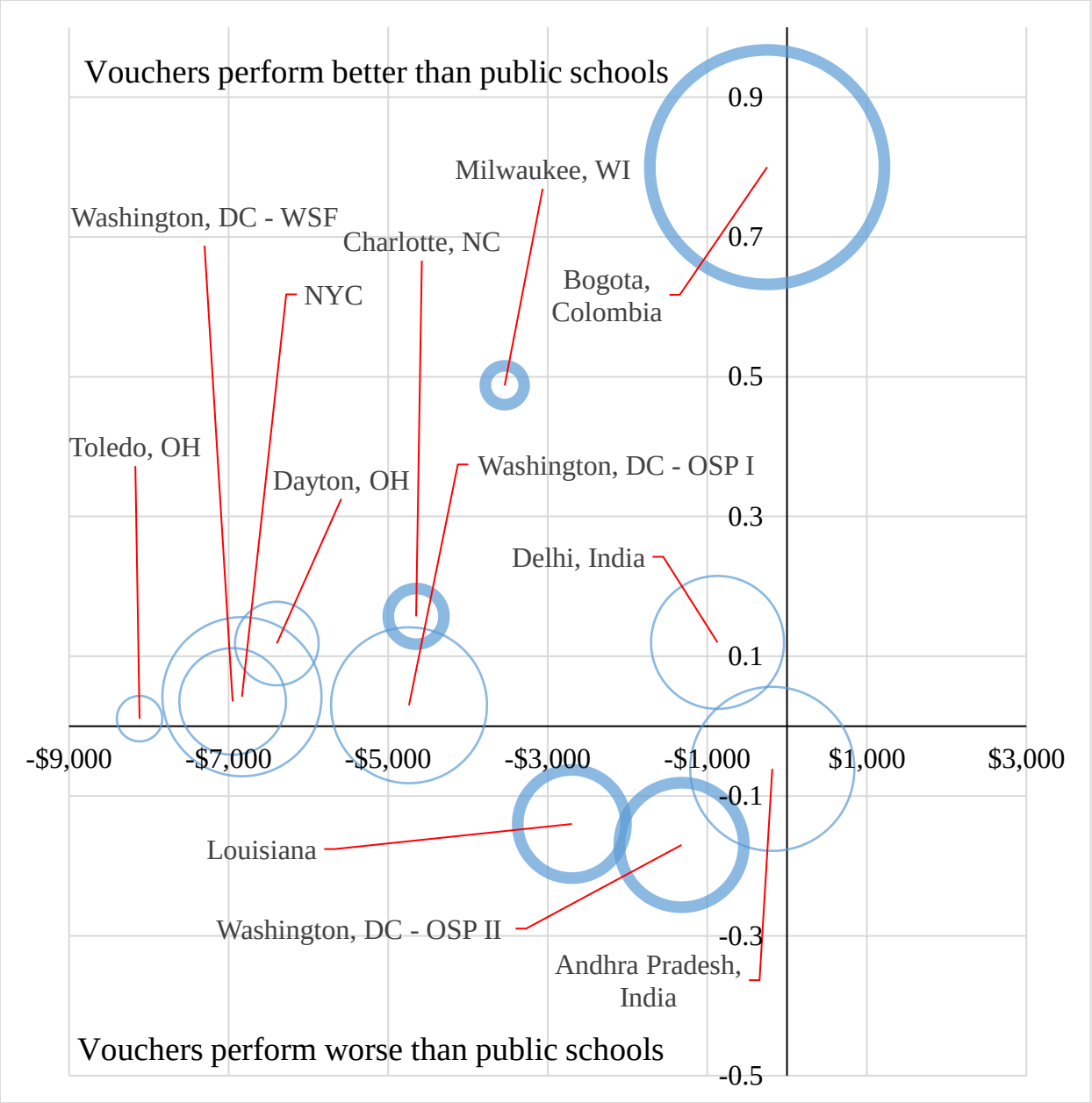

Figure 2: Overall impacts – TOT Math.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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. Overall impacts by funding type

In Figures 3 through 6 we plot 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.

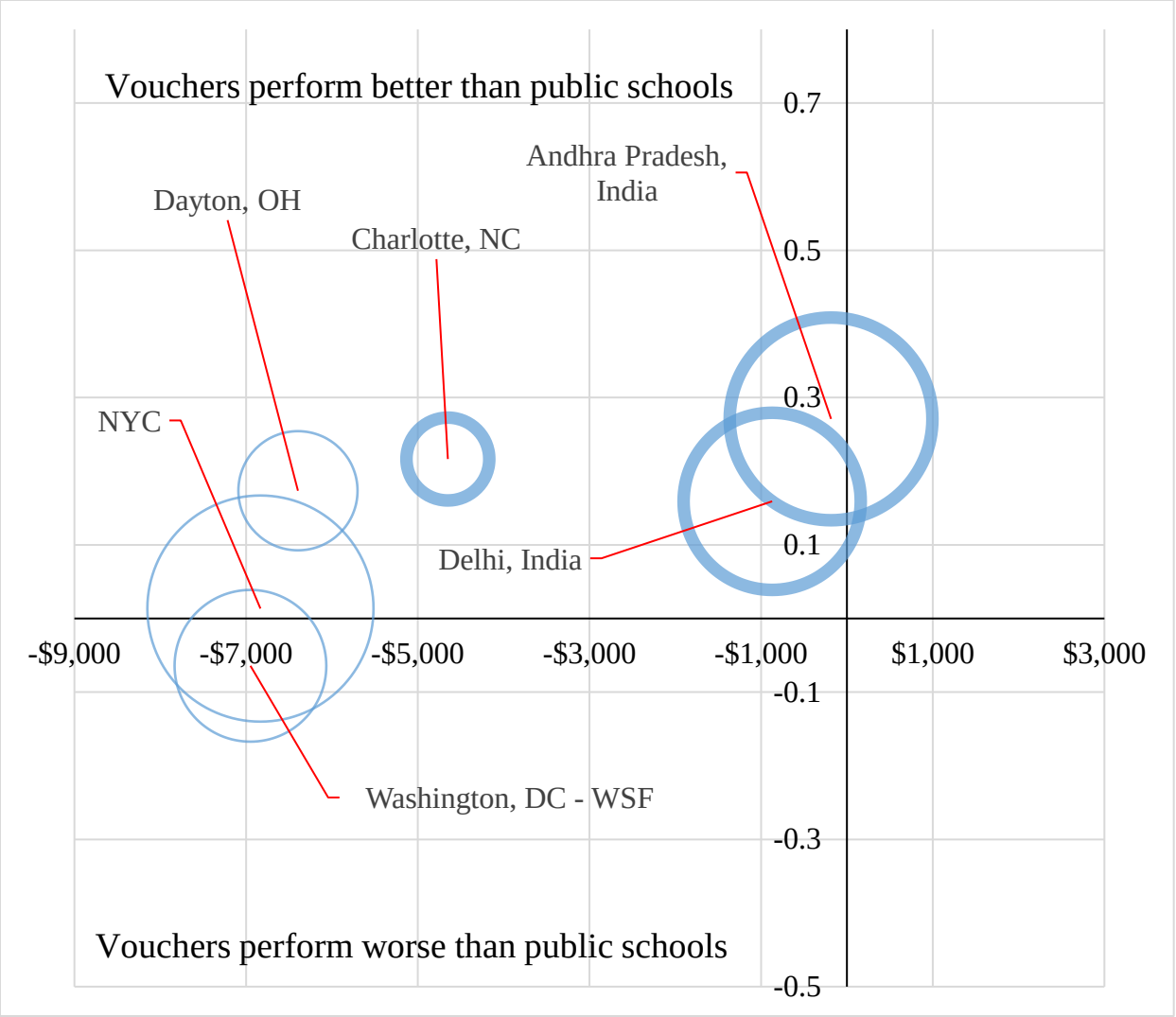

Figure 3: Overall impacts – TOT Reading (Privately Funded).

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

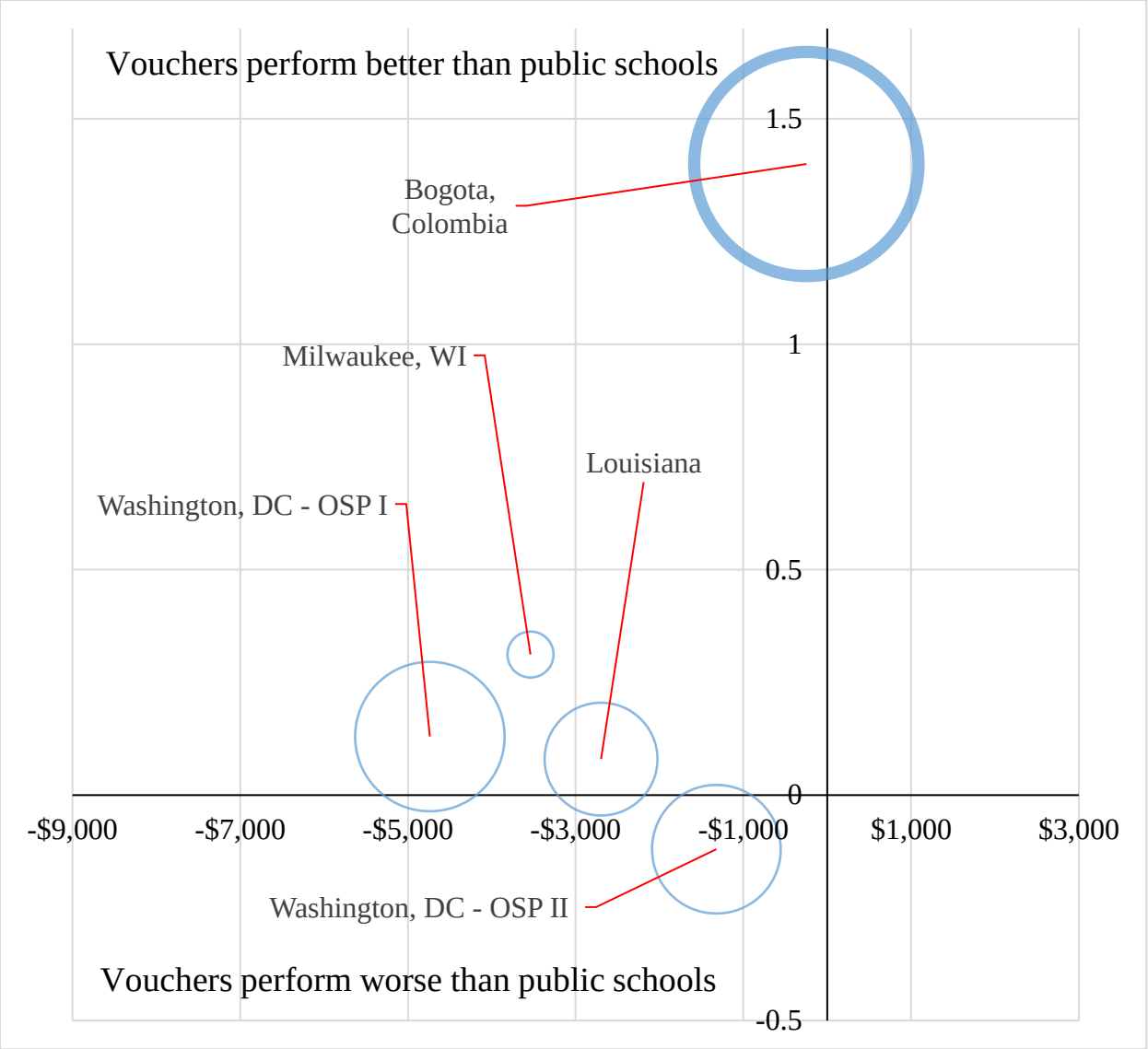

Figure 4: Overall impacts – TOT Reading (Publicly Funded).

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

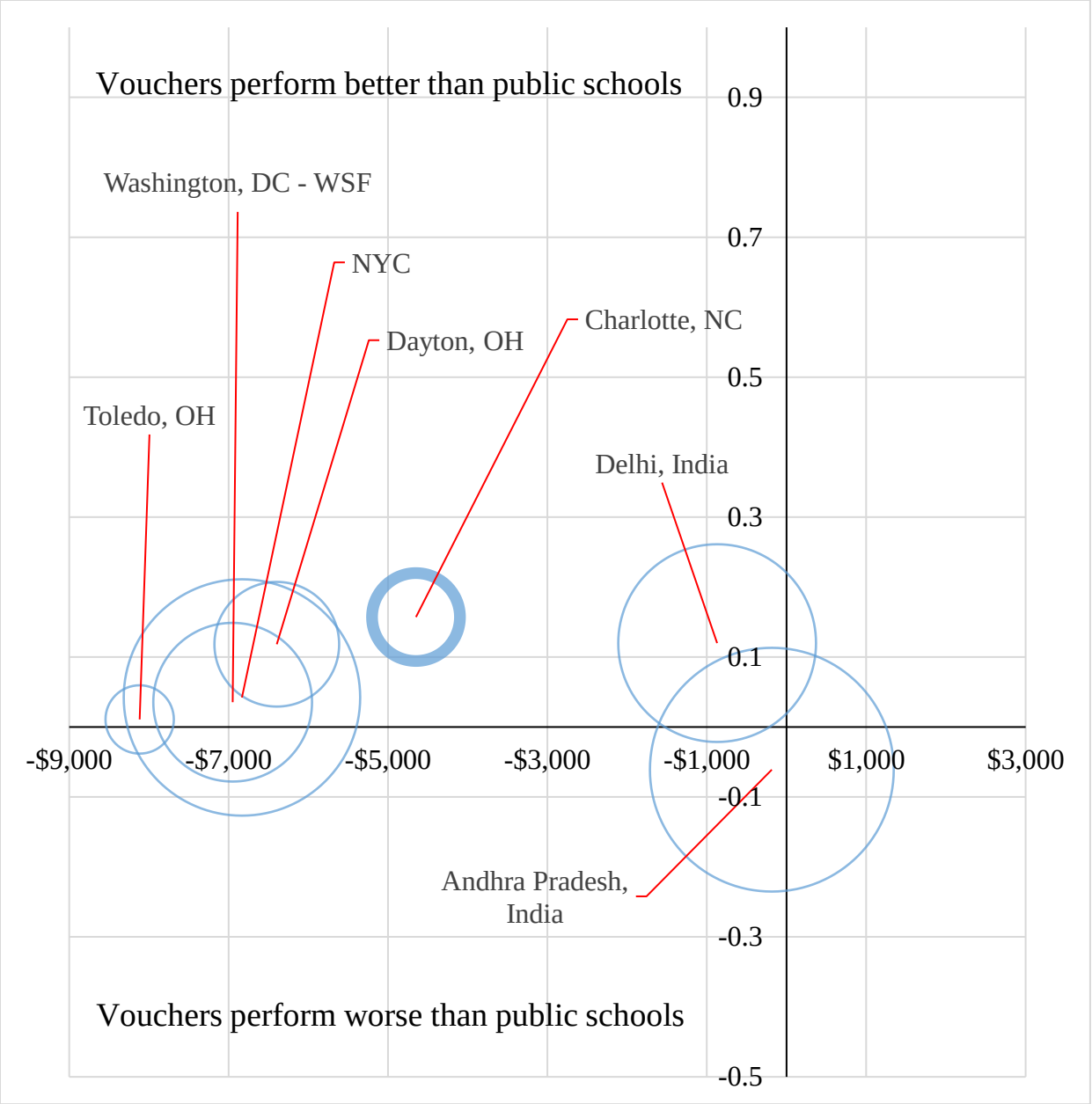

Figure 5: Overall impacts – TOT Math (Privately Funded).

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

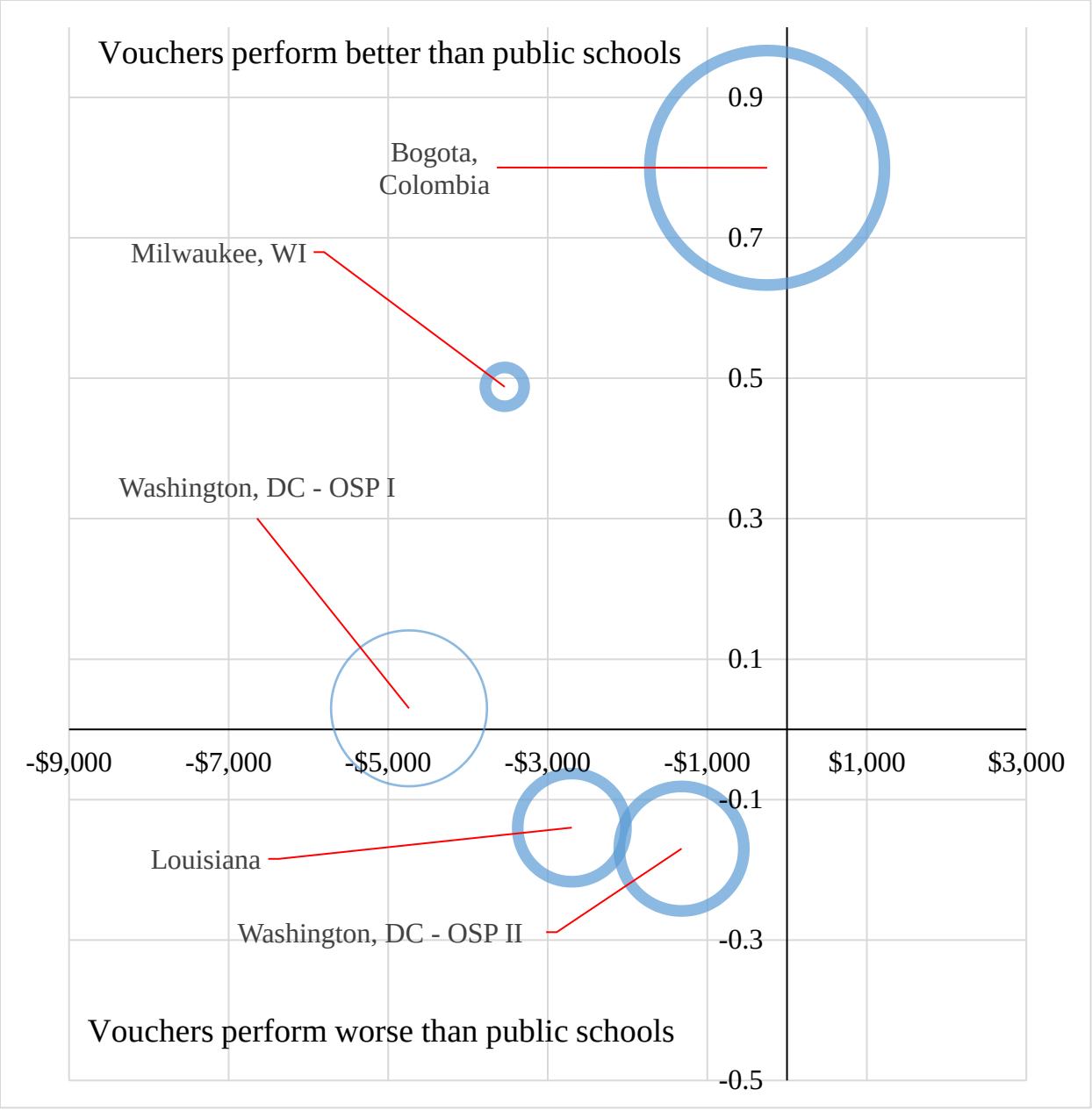

Figure 6: Overall impacts – TOT Math (Publicly Funded).

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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-C. Overall impacts 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, as evidenced by the circles lying to the left of the origin and generally having null to positive achievement impacts. While we cannot determine the exact form of this relationship, generally a positive relation is visible in the graphs.

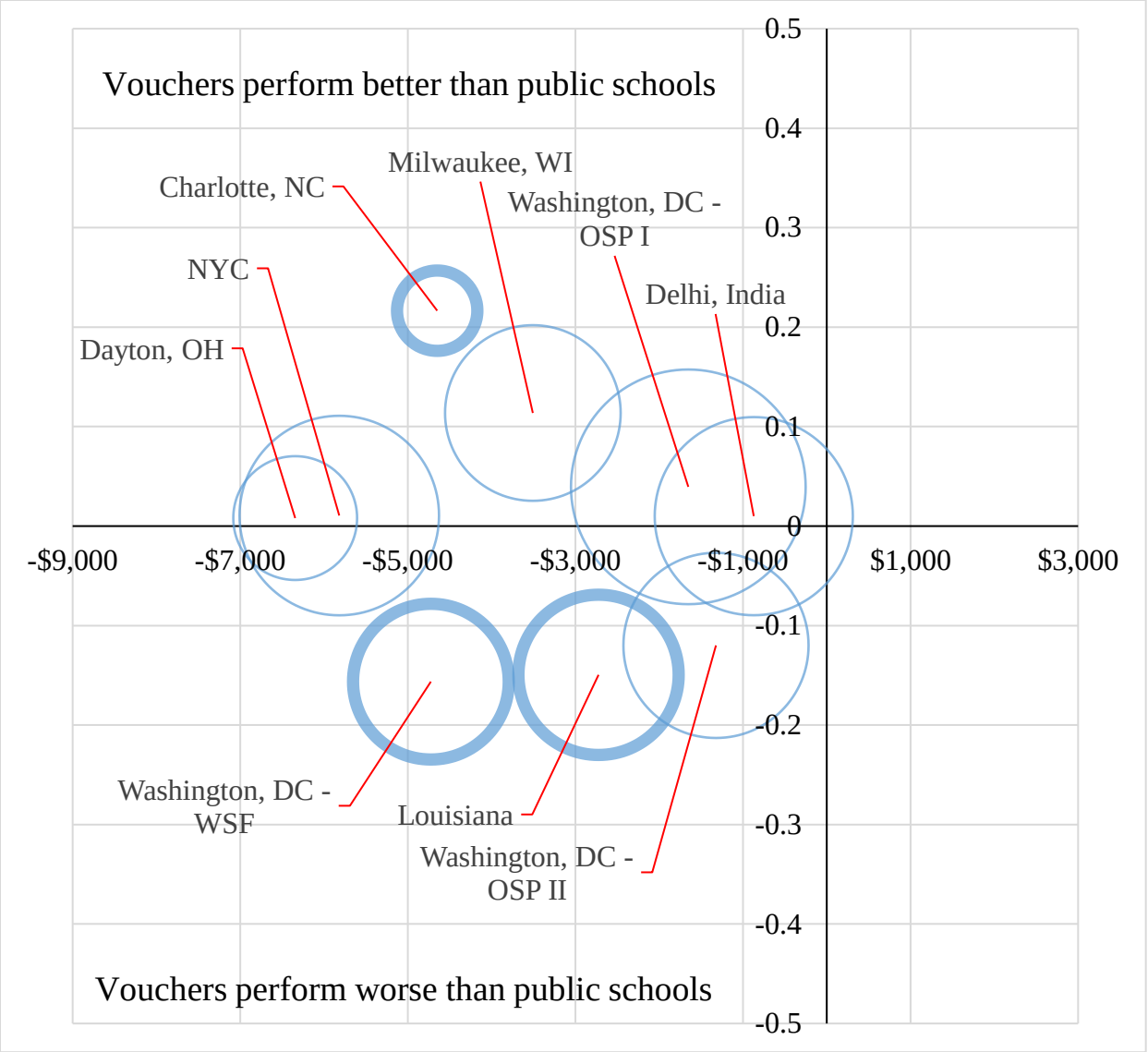

Figure 7: Impacts for Year 1 – TOT Reading.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

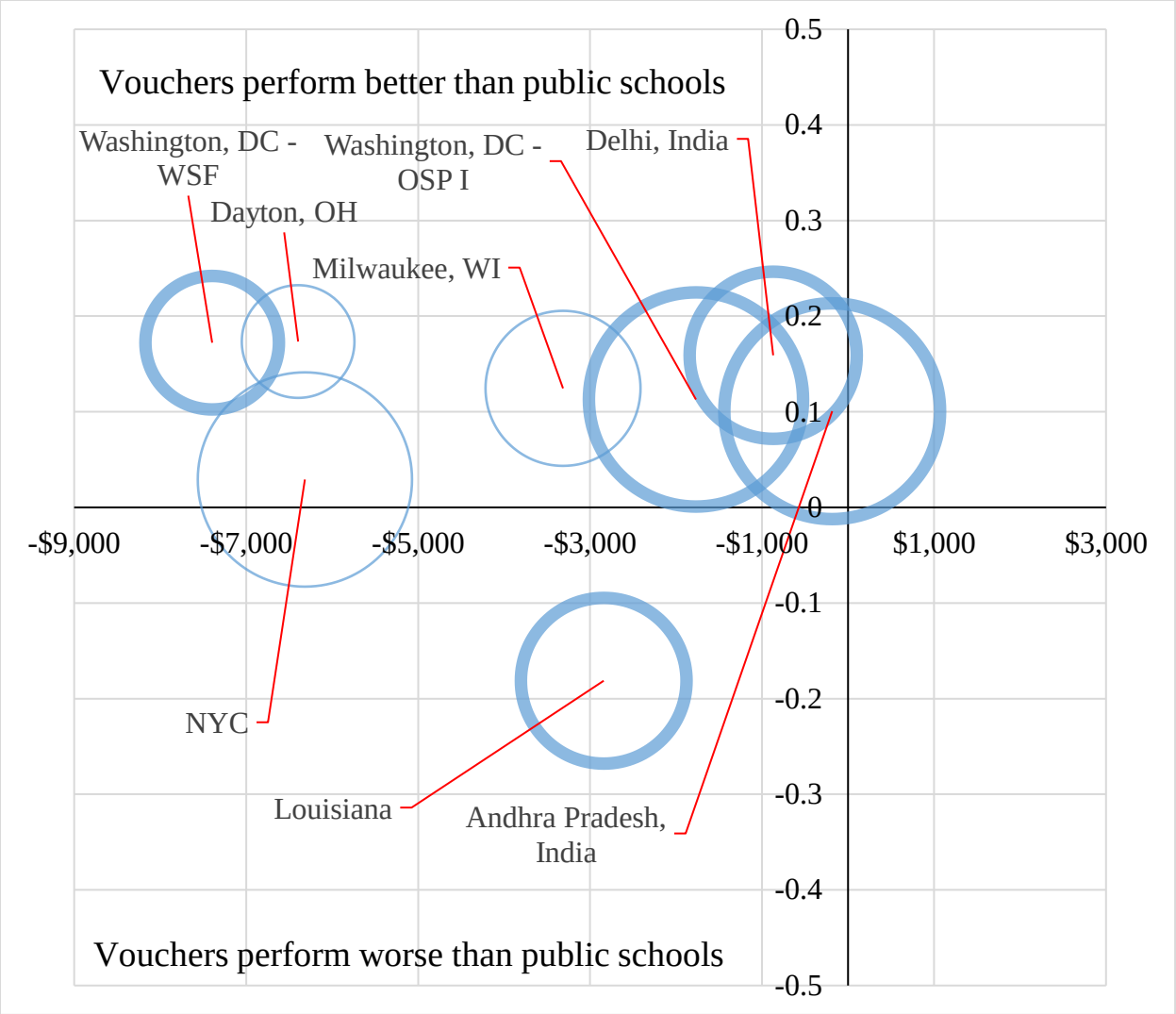

Figure 8: Impacts for Year 2 – TOT Reading.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

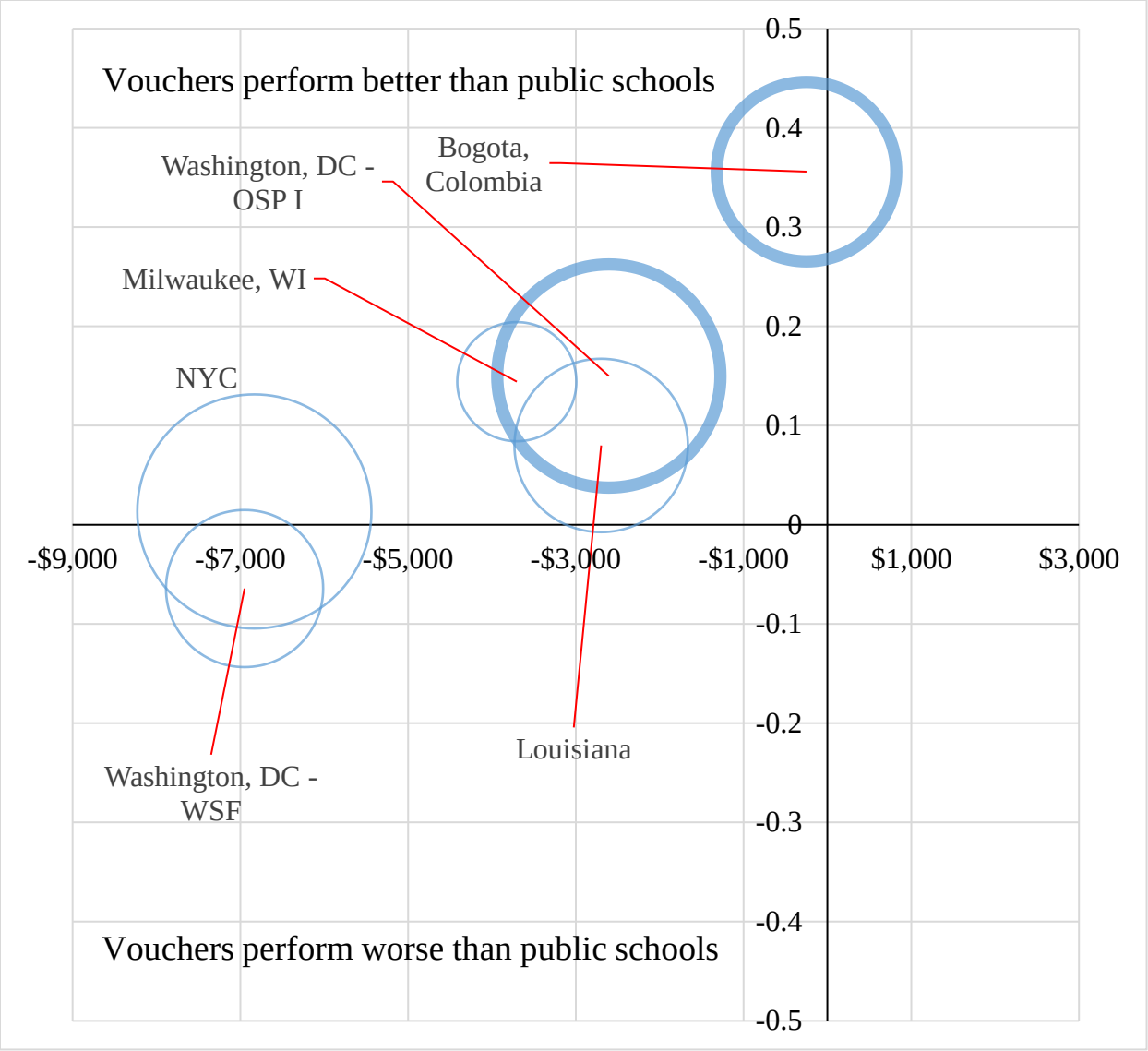

Figure 9: Impacts for Year 3 – TOT Reading.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

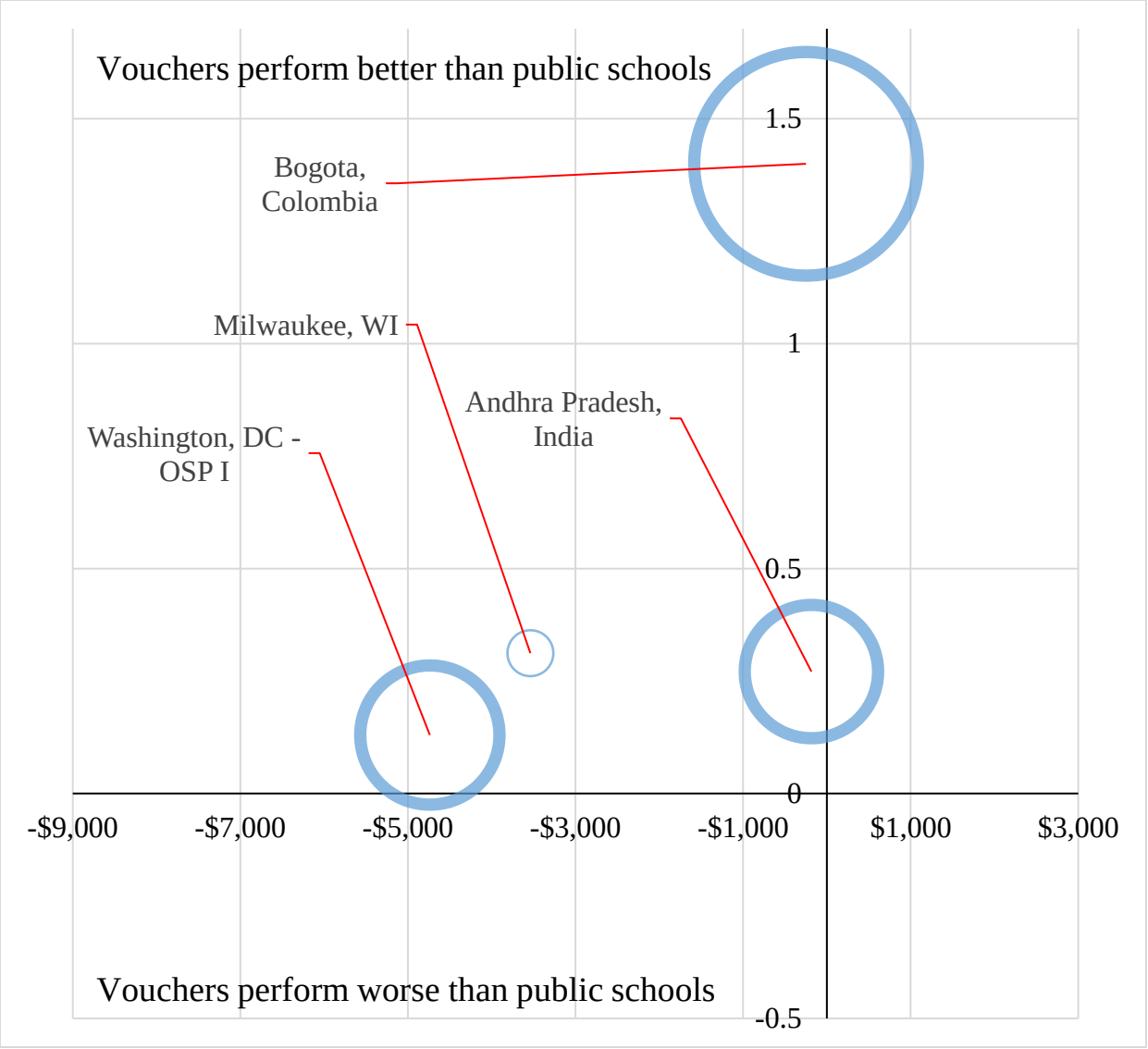

Figure 10: Impacts for Year 4 or more – TOT Reading.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

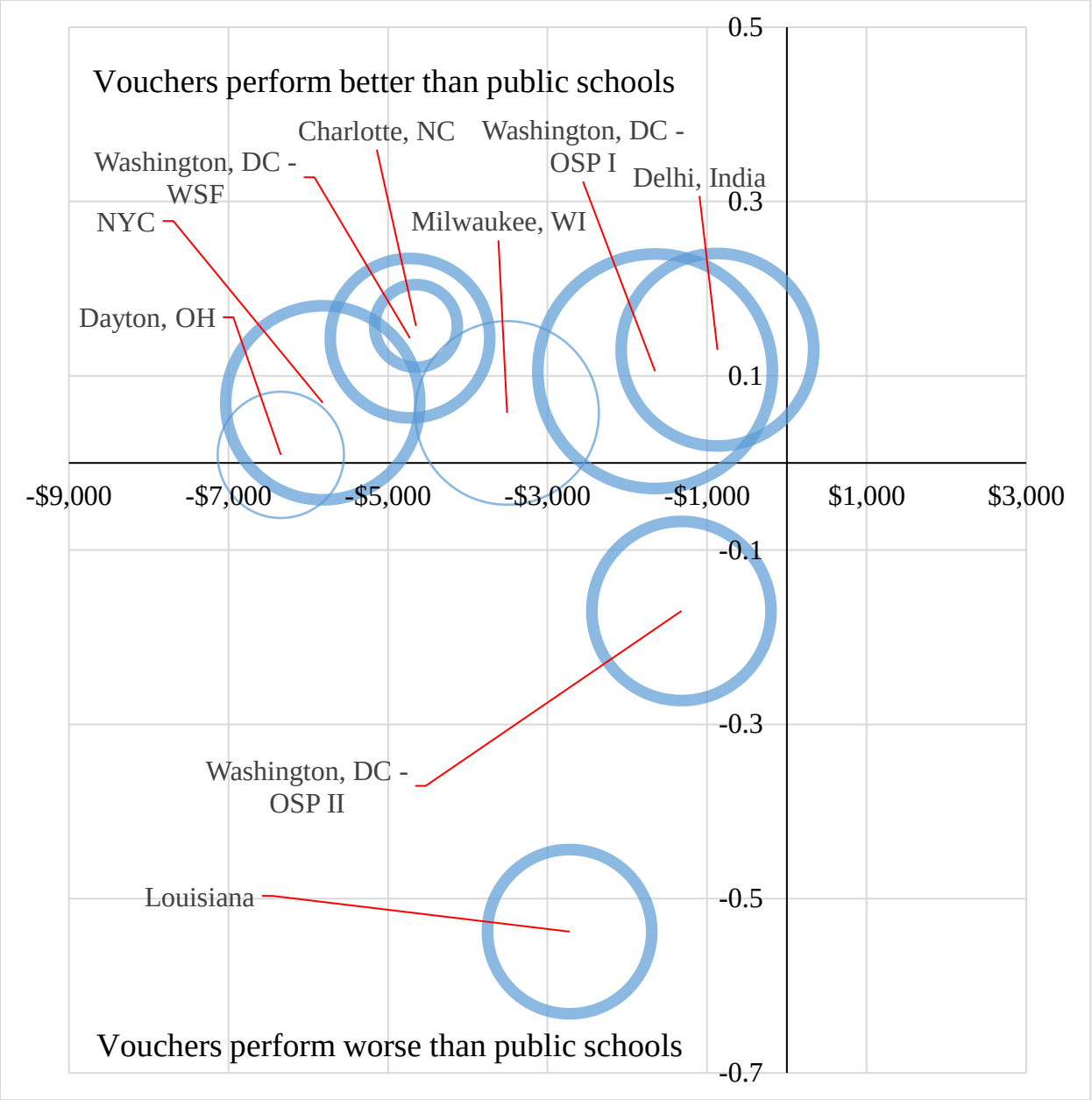

Figure 11: Impacts for Year 1 – TOT Math.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

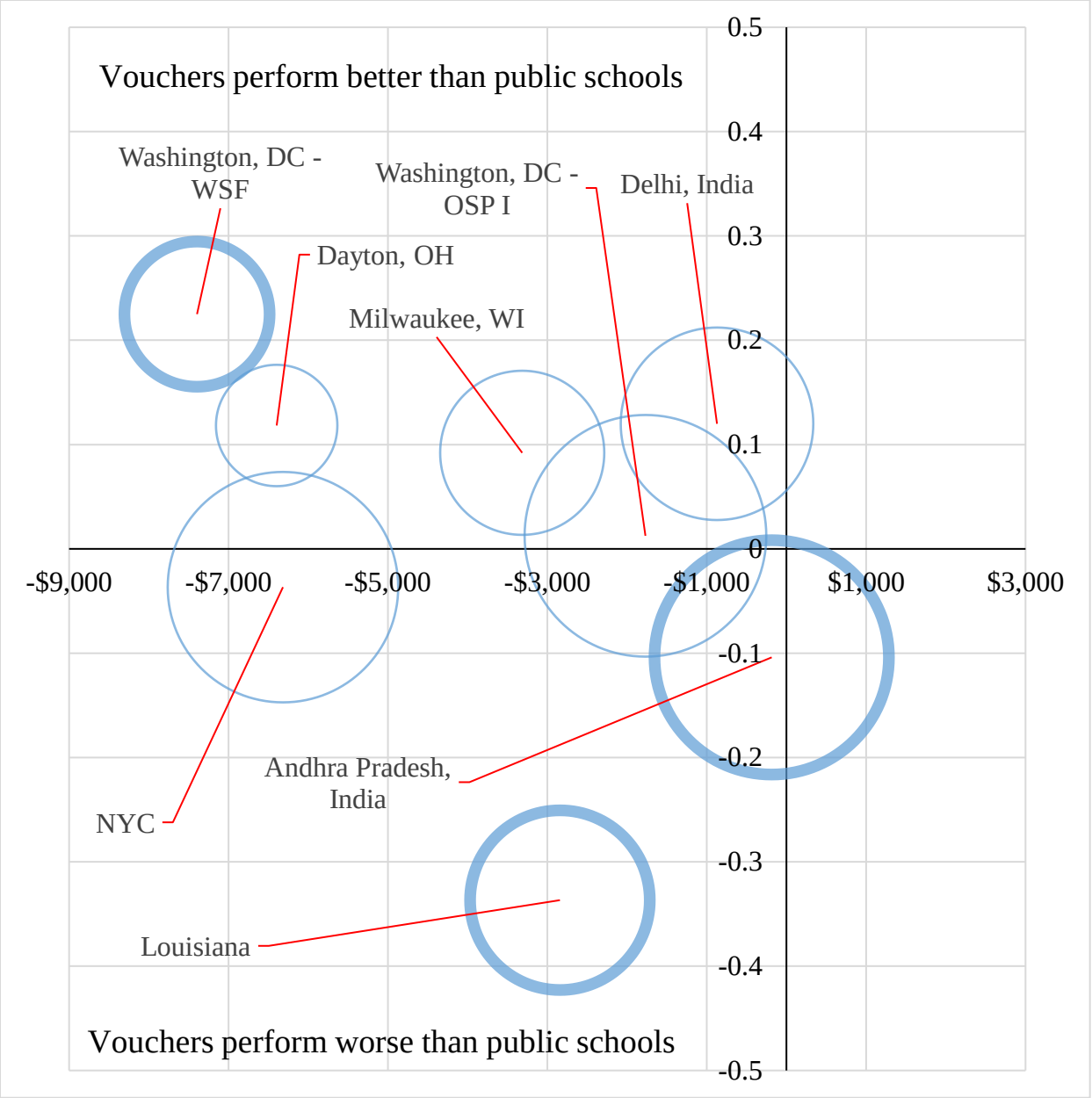

Figure 12: Impacts for Year 2 – TOT Math.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

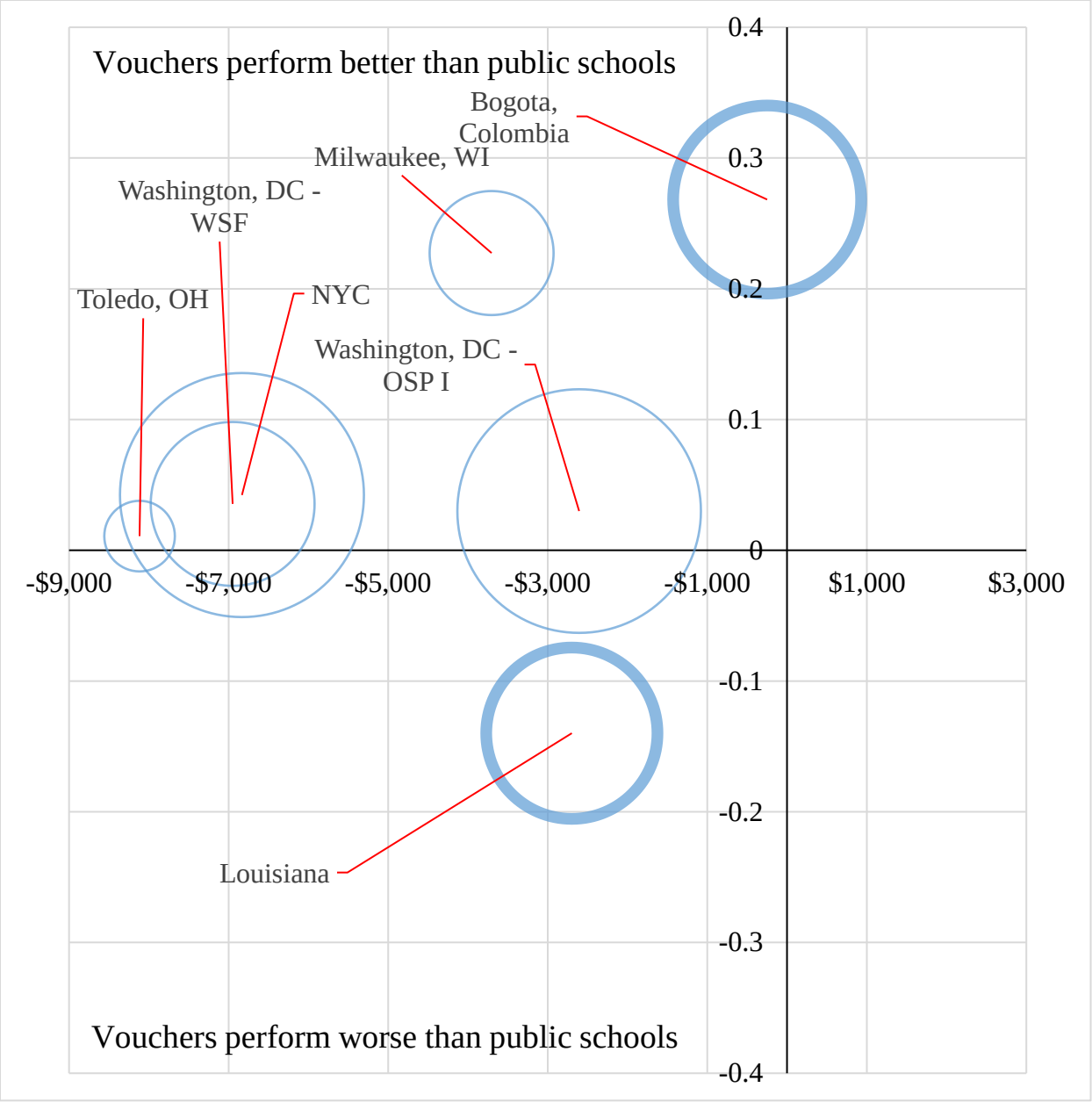

Figure 13: Impacts for Year 3 – TOT Math.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

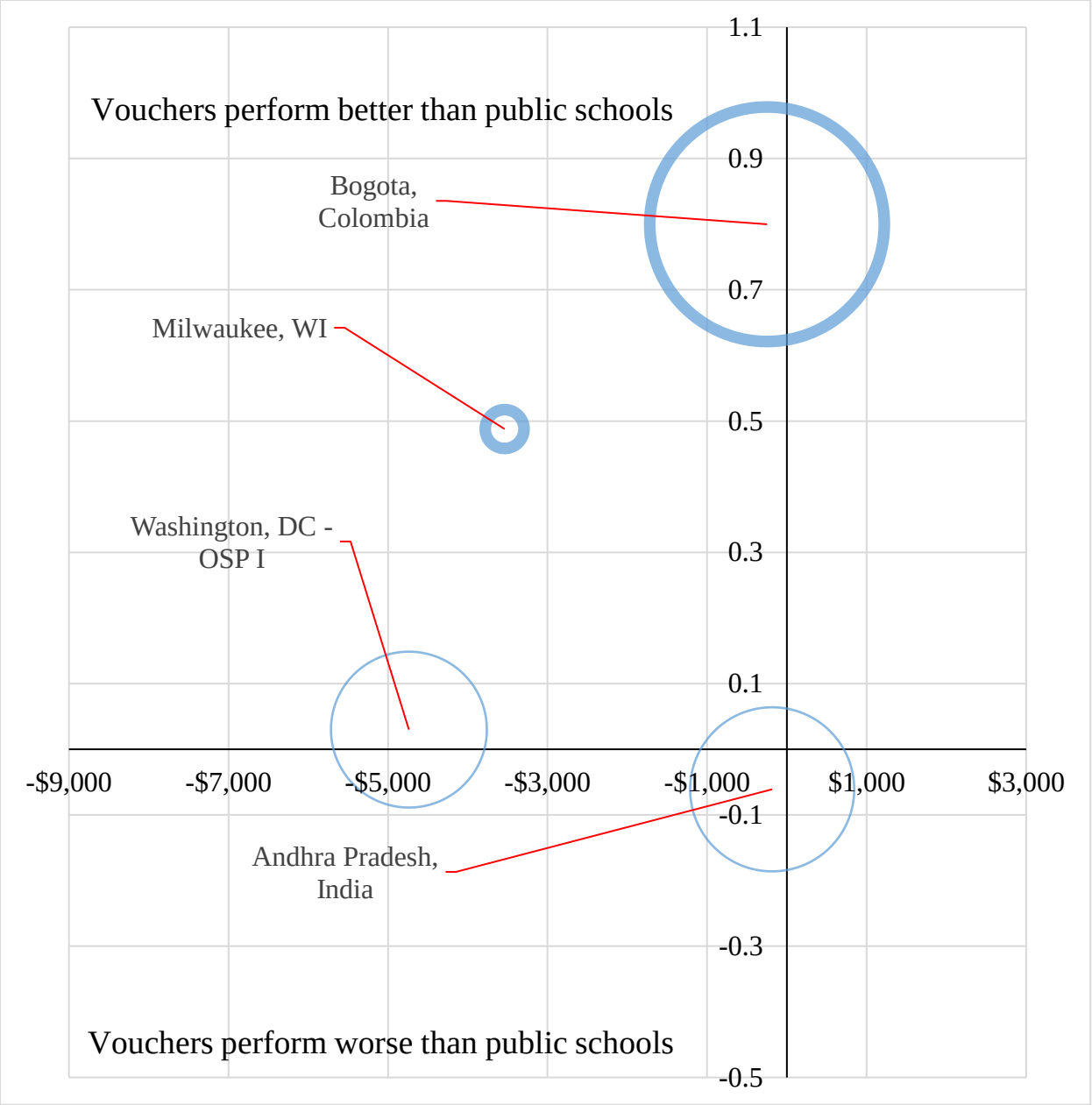

Figure 14: Impacts for Year 4 or more – TOT Math.

Note: On the vertical axis, we plot the effect sizes, Hedges' *g* estimates based on the final year for each study from chapter 2 of this dissertation. 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. TOT impacts for Milwaukee are non-experimental. 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.

There are two main takeaways from the preceding figures. First, school vouchers are consistently more cost-effective than the public schools, even if most of the achievement effects are null. Second, 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 consistently 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 overall 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 that 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.

6. Conclusions and policy implications

This cost-effectiveness analysis contributes to the field of school choice by combining rigorous evidence from RCT studies on school vouchers with actual public school expenditures. 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 in chapter 2 of this dissertation has shown generally higher impact of voucher programs on participants' reading scores in comparison to math scores. In the future, it would be interesting to see how the impacts and cost-effectiveness of LSP and DC OSP II change with later years of evaluation of the program. Although we distinguished between publicly- and privately-funded voucher programs, our study is limited with the difficulty 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 Louisiana 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. 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. 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).

References

Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2015). School vouchers and student achievement: First-year evidence from the Louisiana Scholarship Program (NBER Working Paper No. 21839). Cambridge, MA: National Bureau of Economic Research.

Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review, 92*(5), 1535-1558.

Angrist, J., Bettinger, E., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *The American Economic Review, 96*(3), 847-862.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association, 98*(462), 299-323.

Bettinger, E., & Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics, 90*(8-9), 1625-1648.

Bitler, M., Domina, T., Penner, E. K., & Hoynes, H. (2015). Distributional effects of a school voucher program: Evidence from New York City. *Journal of Research on Education Effectiveness, 8*(3), 419-450.

Bush School of Government and Public Service. (2016). Extending the NCES CWI. Retrieved from: http://bush.tamu.edu/research/faculty/Taylor_CWI/

Chakrabarti, R. (2008). Can increasing private school participation and monetary loss in a voucher program affect public school performance? Evidence from Milwaukee. *Journal of Public Economics, 92*(5-6), 1371–1393.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in action: teacher and health worker absence in developing countries. *The Journal of Economic Perspectives*, *20*(1), 91-116.

Chingos, M. M., & Peterson, P. E. (2015). Experimentally estimated impacts of school vouchers on college enrollment and degree attainment. *Journal of Public Economics, 122*, 1-12.

Cowen, J. M. (2008). School choice as a latent variable: Estimating the "complier average causal effect" of vouchers in Charlotte. *Policy Studies Journal, 36*(2), 301-315.

Cowen, J. M., Fleming, D. J., Witte, J. F., Wolf, P. J., & Kisida, B. (2013). School vouchers and student attainment: Evidence from a state- mandated study of Milwaukee's Parental Choice Program. *Policy Studies Journal, 41*(1), 147-168.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics, 5*(2), 29-57.

Dixon, P. (2013). *International aid and private schools for the poor: Smiles, miracles and markets*. Northampton, MA: Edward Elgar.

Dynarski, M., Rui, N., Webber, A., Gutmann, B., & Bachman, M. (2017). Evaluation of the

DC Opportunity Scholarship Program: Impacts after one year. NCEE 2017-4022. Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance. U. S. Department of Education.

EdChoice (2017). The ABCs of school choice: The comprehensive guide to every private school choice program in America. EdChoice.

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.

Forster, G. (2016). A win-win solution: The empirical evidence on school choice. Indianapolis: Friedman Foundation for Educational Choice.

Glewwe, P., & Muralidharan, K. (2016). *Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications*. Handbook of the Economics of Education, 5, 653-743. Amsterdam, Netherlands: Elsevier.

Greene, J. P. (2000). The effect of school choice: An evaluation of the Charlotte children's scholarship fund program. Manhattan Institute for Policy Research Civic Report, 12.

Greene, J. P., Peterson, P. E., & Du, J. (1999). Effectiveness of school choice: The Milwaukee experiment. *Education and Urban Society, 31*(2), 190-213.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Howell, W. G., & Peterson, P. E. (2006). *The education gap: Vouchers and urban schools*. Washington, DC: Brookings Institution Press.

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School vouchers and academic performance: Results from three randomized field trials. *Journal of Policy Analysis and Management, 21*(2), 191-217.

Jin, H., Barnard, J., & Rubin, D. B. (2010). A modified general location model for noncompliance with missing data revisiting the New York City school choice scholarship program using principal stratification. *Journal of Educational and Behavioral Statistics, 35*(2), 154-173.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist, 47*(5), 658-698.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. NCSER 2013-3000. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Lueken, M. F. (2016). The tax-credit scholarship audit: Do publicly funded private school choice programs save money? EdChoice.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., & Howell, W. G. (2002). *School choice in New York City after three years: An evaluation of the school choice scholarships program.* Washington, DC: Mathematica Policy Research, Inc.

Mills, J. N., & Wolf, P. J. (2016). The effects of the Louisiana scholarship program on student achievement after two years. Available at SSRN: https://ssrn.com/abstract=2738805

Mills, J. N., & Wolf, P. J. (2017). The effects of the Louisiana Scholarship Program on student achievement after three years (Louisiana Scholarship Program Evaluation Report #5). New Orleans, Louisiana: Education Research Alliance for New Orleans & School Choice Demonstration Project, Department of Education Reform, University of Arkansas.

Mosteller, F. & Boruch, F. (Eds.) (2002). *Evidence matters: Randomized trials in education research*. Washington, D.C.: The Brookings Institution.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2016). Disrupting education? Experimental evidence on technology-aided instruction in India (No. w22923). National Bureau of Economic Research.

Muralidharan, K., & Sundararaman. V. (2015). The aggregate effect of school choice evidence from a two-stage experiment in India. *The Quarterly Journal of Economics, 130*(3), 1011-1066.

Muralidharan, K. (2015). Field experiments in education in developing countries. *UC San Diego*.

Peterson, P. E., Howell, W. G., Wolf, P. J., & Campbell, D. E. (2003). School vouchers: Results from randomized experiments. In Hoxby, C. M. (Ed.). *The economics of school choice* (pp. 107-144). University of Chicago Press.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Seventh edition. Sage.

Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics, 113*(2), 553-602.

Scafidi, B. (2012). *The fiscal effects of school choice programs on public school districts.* The Friedman Foundation for Educational Choice.

Spalding, J. (2014). *The school voucher audit: Do publicly funded private school choice programs save money?* The Friedman Foundation for Educational Choice.

Sude, Y., DeAngelis, C. A., & Wolf, P. J. (2018). Supplying choice: An analysis of school participation decisions in voucher programs in Washington, DC, Indiana, and Louisiana. *Journal of School Choice*, *12*(1), 8-33.

Taylor, L. L. & Fowler, W. J. (2006). *A comparable wage approach to geographic cost adjustment.* Education Finance Statistical Center Research and Development Report. U.S. Department of Education: NCES 2006-321.

Tooley, J. (2009). *The beautiful tree*. Washington, DC: Cato Institute.

Wolf, P. J. 2008. Vouchers. In *The International Encyclopedia of Education*, (pp. 635-36). Routledge.

Wolf, P. J., Egalite, A. J., & Dixon, P. (2015). Private school choice in developing countries: Experimental results from Delhi, India. *Handbook of International Development and Education*, 456-471.

Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management*, *32*(2), 246-270.

Wolf, P. J. & McShane, M. (2013). Is the juice worth the squeeze? A benefit/cost analysis of the District of Columbia opportunity scholarship program. *Education Finance and Policy*, *8*(1), 74–99.

Chapter 4

Who is more free? A comparison of the decision-making of private and public school principals

M. Danish Shakeel (mdshakee@uark.edu)

Corey A. DeAngelis (cadeange@uark.edu)

1. Introduction

> *While the public school principal is bound most by red tape, the private school principal is bound most by his or her conscience.*
>
> **—John E. Chubb and Terry M. Moe, 1988, p. 1076**

School choice has emerged as a key intervention in school reform globally. President Donald Trump promised massive expansion of private school choice through a reallocation of $20 billion in federal funding to school choice initiatives. Evidence suggests that private schools slightly outperform public schools on improving student achievement within the US and internationally (Forster, 2016; Greene, 2005; Tooley, 2005; Tooley et al., 2011). Chapter 2 of this dissertation showed that children using voucher programs to attend private schools do slightly better in achievement outcomes than comparison kids in public schools. Of the sixteen experimental studies of private school choice in the United States, only three studies have shown consistent negative findings for test scores in math, reading, or both in the early years of evaluation (Abdulkadiroglu, Pathak, & Walters, 2015; Dynarski et al., 2017; Mills & Wolf, 2016).

Most school choice studies focus on student achievement (West & Woessmann, 2010; Witte, 2001; Witte et al., 2014; Wolf et al., 2013). Other studies on private and charter schools examine impacts on the long-term outcomes such as student attainment (Booker et al., 2008; Cowen et al., 2013; Wolf et al., 2013; Zimmer, 2009) and criminal activity (Deming, 2011; DeAngelis & Wolf, 2016; Dobbie & Fryer, 2015). While this evidence is limited, existing research suggests that access to school choice reduces criminal activity and teen pregnancy while increasing the likelihood of graduating from high school. Additionally, access to private school choice may increase performance in public schools through competitive effects (Egalite, 2013; Egalite, 2016; Figlio & Hart, 2014; Greene & Winters, 2003; Sandström & Bergström, 2005) and

increase civic skills such as voting, volunteering, charitable activity, and tolerance of others (Bettinger & Slonim, 2006; Campbell, 2002; Fleming, 2014; Fleming, Mitchell, & McNally, 2014).

Though many studies have examined whether private schools outperform public schools, few address possible mediating factors. For example, Wolf and Hoople (2006) closely examine the Washington Scholarship Fund, finding that the successful private schools allocated fewer resources to facilities and programs. They also find suggestive evidence that more committed teachers and more challenging homework made a difference. Our study fits into the literature by examining a potential explanation for why school choice seemingly produces (modestly) positive outcomes for students.

We examine the differences in the autonomy of school leaders, since such autonomy may enable leaders to respond to the changing needs of students and staff within their schools. Empowering principals to act on their superior knowledge of and interest in school level outcomes may facilitate the creation of successful education for children (Ouchi, 2009). For example, Grissom, Loeb and Master (2013) find that principals spending time on curricula can positively influence student achievement. Conversely, principals allocating more time on activities such as classroom walkthroughs may have a *negative* impact on student growth. Ouchi (2009) and Hess (2013) point out that student learning cannot be improved unless school leaders have control over important school-level activities such as curriculum, personnel, and the budget.

In schooling, it is axiomatic that leaders that are free to influence important decisions can ameliorate inefficiencies (Tekleselassie & Villarreal III, 2011). In contrast, schools with constrained leadership have less capacity to act in response to school level knowledge. It is possible that principals with a lot of freedom may use it poorly or engage in corrupt practices

such as rewarding their friends. Branch, Hanushek and Rivkin (2013) pointed out that highly effective principals increase student learning by two to seven months within a single school year. Chubb and Moe (1988, p. 1065) found that the public and private schools were "distinctively different in environment and organization." They theorized that greater autonomy would exist in private schools with respect to structure, goals and school operations; though they did not test this specific hypothesis. Private school principals had more teaching experience than public school principals.

We provide an empirical test of the hypothesis that the private schooling sector allows for more leadership autonomy than traditional public schools, using nationally representative survey data of principals in the United States for the 2011-12 school year from the School and Staffing Survey. We compare the reported differences between public and private school principals' influence on decision-making activities within their schools.[34]

2. Theory

Families attending private schools face fewer transaction costs associated with school exit because public schools are typically residentially assigned. This difference makes private school operators more prone to the threat of closure (Friedman, 1955; West, 1981). Shorter term enrollment, and thus monetary loss, can, over the medium and longer term, damage brand name, threaten teachers' jobs, and discourage future clients. Since families are more apt to leave private schools if they are dissatisfied, school leaders must have the power to change schooling to maintain customer satisfaction (West, 1997). If a private school principal is able to make the decisions necessary to respond to the signals transmitted by his or her clients, the families would

[34] Since sector is our key hypothetical independent variable, we do not examine subcategories of private schools and public schools.

find a better match between their needs and schools' quality. Since public schools often have a monopoly on public funding, and their customers are normally assigned residentially, their leaders may face less pressure to respond to dissatisfaction (Hoxby, 2007; Peterson, 1998; Peterson & Hassel, 1998). In other words, the transaction costs for a customer leaving a public school are much higher, typically requiring Tiebout choice (Tiebout, 1956) or paying for a private school out of pocket (Friedman & Friedman, 1990; Merrifield, 2008).

Public school families are less able to just pick up and leave a school. Thus, high levels of principal autonomy can be undesirable in that sector (Neal, 2002). If a malicious, or simply ineffective, principal leads a school, we may not want them making school-level decisions that could negatively affect students (Murphy, 2017). If the ineffective principal is free to make bad decisions, many students may be harmed without an exit option, especially if they come from disadvantaged families unable to seek a wide range of options or push for modifications within the system (Lerner, 1934; Ong-Dean, 2009). Thus, the public sector may limit the principals' power to make decisions. As a result, central office officials may be more likely to control important school-level decisions.

Public school principals operate under increased political constraints as they are accountable to their school boards. Due to fewer political and bureaucratic constraints, the private school principal is likely to have more influence in decision-making and enjoy more autonomy in selection of students and daily administration than his or her public school counterparts (Firestone & Shipps, 2005; Shipps & White, 2009; White, 2006). Since private school principals do not require public funding, they are less likely to feel political pressures than their public school peers. Hence, they may have greater influence over school-level activities.

Additionally, private school principals may face a stronger dismissal threat than their public school counterparts. If private school boards have fewer costs associated with dismissing their principals, they will be more likely to be able to hold them accountable for their actions. Possible dismissal either through boards or other superiors or via school closure due to family exits could make private school principals more motivated to make effective decisions. On the other hand, if a school principal is protected through unionization or civil service rules, they will be less likely to face accountability, and perhaps less apt to perform effectively (Chubb & Moe, 1986). Since it is more difficult to fire a principal in the public sector, we expect that district bureaucrats will reduce their autonomy in order to limit negative outcomes for students (Hess, 2013). Furthermore, public school principals theoretically have an incentive to maximize budgets. This condition suggests central offices will not grant them much autonomy over finance decisions (Niskanen, 1971; for schooling, see Levenson, 2012).

3. Data

The data for the public and private school principals comes from the School and Staffing Survey (SASS) 2011-2012 questionnaire. SASS was developed by the National Center for Education Statistics (NCES) and it has been administered seven times from 1987-88 to 2011-2012. Table 1 lists the question categories and what they measure.[35] The public school principal data file contained 7,510 records while the private school principal data file contained 1,720 records. There were some additional questions for public school principals, but in this study, we compare only the common questions related to decision making. As the sample is nationally

[35] For more information, see http://nces.ed.gov/surveys/sass/pdf/1112/SASS2A.pdf (for public school principals) and http://nces.ed.gov/surveys/sass/pdf/1112/SASS2B.pdf (for private school principals).

representative, systematic differences across public and private schools after controlling for relevant characteristics may indicate benefits of school leadership in one sector over the other.

Our dependent variables come from questions on decision-making in SASS 2011-2012. These variables measure the influence principals perceive to have on setting performance standards, establishing curriculum, determining content for professional development, evaluating teachers, hiring teachers, setting discipline policy, and deciding how the budget will be spent. This section asks the principals to rate their ability to influence seven school related activities on a four-item Likert scale (no influence, minor influence, moderate influence and major influence) and it includes a *not applicable* option for each activity (Table 1). We choose these variables as they are proxies for autonomy and earlier work has shown that bureaucratization in public schools acts as a hindrance for a school principal's ability to influence school-level activities (Chubb, & Moe, 1988; Ouchi, 2009; Hill, Pierce, & Guthrie, 2009; Nadelstern, 2013). Hence, we expect to find increased decision making ability of the private school principal to influence the above mentioned school-level activities in comparison to the public school principal.

Table 1: School-Related Activities over Which the Principal Has Influence

Category	School-related activities
A	Setting performance standards for students of this school
B	Establishing curriculum at this school
C	Determining the content of in-service professional development programs for teachers of this school
D	Evaluating teachers of this school
E	Hiring new full-time teachers of this school
F	Setting discipline policy at this school
G	Deciding how your school budget will be spent

We utilize questions from the survey gauging principals' demographics and academic and professional background for summary statistics. Tables 2A and 2B show the population weighted summary statistics expressed as percentages for the principals in public and private schools. Overall, private school principals report more years of principal experience but lower

education levels in comparison to the public school principals. This result is consistent with the findings of Hill et al. (2016). The proportion of private school principals reporting greater than 10 years of experience as a principal or school head is almost double that of public school principals. The proportion of private school principals involved in teaching in addition to their task as a principal or school head is also about twice that for public school principals.

A higher proportion of public school principals report having previous experience as a department head, assistant principal or program director and participation in a school training or development program in comparison to their private counterparts. The proportion of public school principals holding a school administration license is about twice as large as the proportion of private school principals. Almost all public school principals earned an MA or higher degree while only 76% of the private school principals report doing so. The racial composition of principals is largely white[36] in both the sectors (86% in public schools and 90% in private schools). Lastly, private schools have a larger share of females in their leadership in comparison to the public schools.

[36] This excludes mixed race, so it is a lower bound.

Table 2A: Summary Statistics for Principal Characteristics

Measure	Public	Private
Years principal or school head at this or any school prior to this year		
no experience	8.32	8.78
low experience 1-3	24.55	18.82
medium experience 4-10	43.79	30.97
high experience 10+	23.34	41.43
Years principal or school head at this school prior to this year		
no experience	16.46	14.52
low experience 1-3	38.83	27.62
medium experience 4-10	36.07	32.92
high experience 10+	8.64	24.94
Years of elementary or secondary teaching before becoming principal or school head		
no experience	1.70	18.51
low experience 1-3	2.79	7.99
medium experience 4-10	47.34	32.79
high experience 10+	48.16	40.71
Years of elementary or secondary teaching since becoming principal or school head		
no experience	90.41	49.69
low experience 1-3	5.42	21.87
medium experience 4-10	3.30	15.87
high experience 10+	0.87	12.56
Currently teaching at school	37.37	71.89

Notes: Summary statistics presented using population weighted percentages for each italicized category.

Table 2B: Summary Statistics for Principal Characteristics

Measure	Public	Private
Prior to becoming a principal or school head		
Worked as department head	40.4	35.3
Worked as an assistant principal or program director	73.9	43.8
Participated in school training or development program	55.3	31.4
Previous management experience outside education	40.3	46.4
Currently holding license in school administration	95.9	43.4
Having a bachelor's degree	99.9	88.5
Bachelor degree awarded by a university's department or college of education	81.9	67. 8
Having a master's degree	97.6	76.3
Master's degree awarded by a university's department or college of education	97.4	85.4
Earned a MA and higher degree	97.8	68.9
Participated in any professional development activity related to principal or school head in last 12 months	99.3	89.6
Race (white)	86.4	90.2
Gender (male)	48.4	44.6
N	7,510	1,720

Notes: Summary statistics presented using population weighted percentages for each category.

We also present summary statistics on the percent of private and public school principals to state that they have a major influence on the seven outcome categories in Table 3. More than two-thirds of private school principals state having a major influence on establishing curriculum in their schools whereas less than half of public school principals state the same. A slightly higher proportion of private school principals than public school principals state having a major influence on setting performance standards, professional development and discipline policy in their schools. On the other hand, a greater proportion of public school principals than private school principals state having a major influence on teacher evaluation, hiring teachers and spending budget in their schools.

Table 3: Summary Statistics for Principals' Self-Reported Major Influence on Outcome Variables

Measure	Public	Private
Performance Standards	73.3	80.4
Establishing Curriculum	42.6	69.1
Professional Development	69.5	74.2
Teacher Evaluation	95.3	82.1
Hiring Teachers	84.3	83.7
Discipline Policy	79.4	81.5
Budget Spending	63.8	62.1
N	7,510	1,720

Notes: Summary statistics presented using population weighted percentages for each category.

4. Methods

Since the survey responses related to decision-making are ordinal and have four categories (from "No Influence" to "Major Influence"), the analytic technique we employ is an ordered logistic regression (Borooah, 2001; Cohen, Cohen, West, & Aiken, 2003) of the form:

$$DM_i = \alpha_0 + \alpha_1 Private_i + \mu_i \qquad \text{Equation (1)}$$

The dependent variable of interest *DM*, is the reported decision-making ability of a given principal, i, for the following school-level activities: setting student performance standards, establishing curriculum, determining teacher professional development content, evaluating teachers, hiring new full-time teachers, setting discipline policy and deciding how the budget will be spent. The decentralized nature of private schools is likely to induce better decision making ability for the private school principal in comparison to the public school principal to decide what is best for their schools. The dependent variable *DM* is likely to differ across public and private schools due to the former's increased centralization and bureaucratization (Chubb, & Moe, 1988; Hill, Pierce, & Guthrie, 2009; Nadelstern, 2013; Ouchi, 2009). This variable takes the value 1 for the least influence and the value 4 for the most influence.[37] $Private$ is a dummy variable of value 1 if the principal is in a private school and 0 if the principal is in a public school. The coefficient of interest, α_1, measures the mean difference of the decision-making influence reported by private school principals relative to public school principals. The constant, α_0, measures the average principal decision-making influence reported by public school principals.

Since we want to examine the differences between principals based solely on the type of institution they are in, this initial model does not control for any principal or school-level differences. In order to have a conservative estimate of the association between institution-type and decision-making freedom, we analyze the following model that also includes school and principal characteristics as controls:

$$DM_i = \alpha_0 + \alpha_1 Private_i + \alpha_2 PC_i + \alpha_3 X_i + \mu_i \qquad \text{Equation (2)}$$

[37] Since the dependent variable is ordinal, we use ordered logit regression and report average marginal effects for the likelihood of reporting "major influence."

PC is a vector of controls which includes the following principal characteristics: race, gender, education level, years of experience as a principal or school head, years of experience as a teacher in elementary or secondary school, any experience as a department head, any experience as an assistant principal, participation in professional development or training programs, management experience outside of education, and whether the principle holds a license in school administration. Vector X also includes these school-level characteristics: school size, school level, number of full-time teachers, student/teacher ratio, percent of minority teachers, and percent of minority students. This second model includes school and principal level controls in order to examine if the effects are significant after accounting for differences in the types of schools and principals hired across the two institutions.

The restricted use data provided by the NCES are imputed and adjusted for non-response.[38] Based on the stratified probability proportionate to size sampling strategy used by NCES in the SASS, we use the balance repeated replication (BRR) bootstrap methodology[39] so that the results reflect the true population values and not just the sampled units. This methodology does not change our final estimates, but rather corrects the formula for the calculation of the standard errors.

[38] There are two types of non-response in the SASS survey. The unit-level non-response rates represent a fully or partially incomplete response for the key questionnaire items on the survey. If the key questionnaire items are not filled out, the entire survey is discarded. Additionally at least 10 percent of the remaining items should be completed to meet the threshold for inclusion. The weighted unit-level non-response rate is approximately 73 percent for the public school principal and it is approximately 65 percent for the private school principal. The item-level non-response rates represent missing items in a questionnaire that has passed the threshold to be considered complete. Approximately 96 percent items on the survey for the public school principal and 94 percent items on the survey for the private school principal had a response rate of 85 percent or more. Thus, the item-level non-response rate that may affect the analysis in our study, is not high.

[39] Details can be found in the User's Manual for the 2011–12 Schools and Staffing Survey: https://nces.ed.gov/surveys/sass/methods0708.asp

5. Results

We now present the results for our models with and without controls in Table 4. The first row presents results without any controls, the second includes principal-level controls, and the third includes all school and principal-level controls. The results are robust across models, though the model without controls only finds statistical significance for the first four categories.

The model with all controls indicates that private school principals are more likely to report having a major influence on 6 out of 7 types of school decisions. When controlling for school and principal-level differences across sectors, we find evidence that private school principals exercise significantly more influence over decision-making activities. In particular, private school principals have a higher likelihood of reporting major influence over performance standards, curriculum, professional development, hiring teachers, discipline policy, and budget decisions. However, private school principals have a 3.9 percentage point lower likelihood of reporting to have a major influence on the evaluation of teachers; we will discuss this further below. Since private school principals have a 4.9 percentage point higher likelihood of having a major influence over the hiring of teachers, they may not need to provide as much direct feedback. In addition, since private school principals have a 14.4 percentage point higher likelihood of having a major influence on the content of their teacher professional development programs, they may provide feedback through that channel instead. Notably, private school principals have a 20-percentage point higher likelihood of reporting that they have a major influence on establishing their school's curriculum. Furthermore, private school principals have a 14 percentage point higher likelihood of reporting that they have a major influence on their students' performance standards. Having a major influence on their students' performance standards may be especially important for the ability of the principal to positively impact student

achievement. Based on our results, we expect that the reduced regulatory burden found in private schools (Chubb & Moe, 1988) grants the principals the ability to exercise more influence related to school activities in comparison to public school principals. To explore our analysis further, we examine the coefficients on the control variables for our preferred model, found in Table 5.

Table 4: Results Based on Model Used

	Performance Standards	Establishing Curriculum	Professional Development	Teacher Evaluation	Hiring Teachers	Discipline Policy	Budget Spending
No Controls	**0.072*** **	**0.247*** **	**0.126*** **	**-0.064*** **	**0.019**	**0.018**	**0.001**
	(0.018)	(0.017)	(0.019)	(0.009)	(0.014)	(0.014)	(0.017)
Principal Controls	**0.146*** **	**0.259*** **	**0.141*** **	**-0.034*** **	**0.050*** **	**0.060*** **	**0.049** **
	(0.017)	(0.018)	(0.017)	(0.009)	(0.014)	(0.017)	(0.021)
Principal and School Controls	**0.140*** **	**0.200*** **	**0.144*** **	**-0.039*** **	**0.049*** **	**0.067*** **	**0.071*** **
	(0.018)	(0.019)	(0.016)	(0.009)	(0.014)	(0.017)	(0.021)
Observations	9,230	9,230	9,230	9,230	9,230	9,230	9,230

Notes: Table reports average marginal effects of private on the "major influence" category, estimated after running ordered logit models. Demographic variables, academic training, professional development and educational attainment levels are included as controls. Estimates use balanced repeated replication (BRR) bootstrap population weights. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5: Likelihood of Reporting Major Influence (All Controls)

	Performance Standards	Establishing Curriculum	Professional Development	Teacher Evaluation	Hiring Teachers	Discipline Policy	Budget Spending
Private School Principal	**0.140*** **	**0.200*** **	**0.144*** **	**-0.039*** **	**0.049*** **	**0.067*** **	**0.071*** **
	(0.018)	**(0.019)**	**(0.016)**	**(0.009)**	**(0.014)**	**(0.017)**	**(0.021)**
School Size	0.009*	-0.011*	0.009	0.000	0.004	0.007	-0.006
	(0.005)	(0.006)	(0.007)	(0.003)	(0.004)	(0.006)	(0.008)
School Level	0.018*	0.073***	0.017	-0.003	-0.001	-0.017**	-0.044***
	(0.010)	(0.008)	(0.011)	(0.004)	(0.006)	(0.008)	(0.010)
Number of Full Time Teachers	-0.000	-0.001**	-0.001*	-0.000	0.001	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Student/Teacher Ratio	-0.001	-0.002*	-0.001	-0.000	-0.001	-0.001	-0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
Percent of minority teachers	0.001***	0.001***	0.000	0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent of minority Students	-0.000	-0.001**	-0.000	-0.000	-0.001**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Low principal Experience	-0.060**	-0.053**	0.037	-0.001	0.015	-0.085***	-0.095***
	(0.025)	(0.023)	(0.029)	(0.016)	(0.022)	(0.026)	(0.023)
Low Teaching Experience	-0.042	-0.064	0.069	-0.034**	-0.050*	-0.049*	0.028
	(0.035)	(0.040)	(0.076)	(0.016)	(0.029)	(0.029)	(0.043)
Department Head	0.034***	0.052***	0.039***	0.003	0.034***	0.024**	0.041***
	(0.012)	(0.012)	(0.012)	(0.005)	(0.011)	(0.012)	(0.015)
Assistant Principal/Program Director	-0.027*	-0.046***	0.001	-0.004	0.007	-0.029*	0.032**
	(0.015)	(0.013)	(0.013)	(0.006)	(0.013)	(0.015)	(0.013)
School Training/ Development	0.044***	0.015	0.018*	0.007	-0.006	0.015	0.016
	(0.013)	(0.013)	(0.011)	(0.005)	(0.009)	(0.010)	(0.012)
License in School Administration	0.045**	0.032	0.022	0.031***	0.004	0.037*	0.019
	(0.022)	(0.023)	(0.023)	(0.009)	(0.013)	(0.020)	(0.031)
Management Experience	0.005	0.012	-0.023*	0.002	0.001	0.017	-0.006
	(0.014)	(0.014)	(0.014)	(0.007)	(0.011)	(0.012)	(0.014)
Master's Degree or Higher	0.062**	-0.004	-0.024	0.007	0.035*	0.051*	0.075**
	(0.030)	(0.028)	(0.047)	(0.011)	(0.019)	(0.027)	(0.033)

Table 5 (Cont'd): Likelihood of Reporting Major Influence (All Controls)

	Performance Standards	Establishing Curriculum	Professional Development	Teacher Evaluation	Hiring Teachers	Discipline Policy	Budget Spending
Professional Development	0.146***	0.148***	0.034	0.019	0.059	0.054	0.076
	(0.046)	(0.057)	(0.126)	(0.020)	(0.056)	(0.036)	(0.062)
White	-0.006	-0.041*	0.010	0.002	0.035**	0.051***	0.020
	(0.018)	(0.023)	(0.019)	(0.009)	(0.015)	(0.016)	(0.021)
Female	0.022	0.022	0.052***	0.020***	0.015	0.015	0.034**
	(0.015)	(0.014)	(0.015)	(0.006)	(0.011)	(0.012)	(0.014)
Observations	9,230	9,230	9,230	9,230	9,230	9,230	9,230

Notes: Table reports average marginal effects for the "major influence" category, estimated after running ordered logit models. Estimates use balanced repeated replication (BRR) bootstrap population weights. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Most of our school-level controls are unrelated to the seven outcome measures of interest; however, some statistical significance emerges. Principals within larger schools are more likely to report having a major influence on performance standards, but less likely to report so for establishing curriculum. Principals in secondary schools are more likely to report having a major influence in performance standards and curriculum, but less likely to report having influence over discipline and budget decisions. Being in a school with a more diverse set of teachers is associated with a higher likelihood of reporting a major influence on performance standards and curriculum.

The coefficient on the principal's previous experience as a department head is significant and positive in all cases except for the case of teacher evaluation, where it is not statistically different from zero. Hence, previous leadership experience has a systematic positive relationship with the principal's ability to influence school level activities. Lower levels of previous principal experience and previous teaching experience are associated with a lower likelihood of reporting to have an influence on most categories.

The coefficient on *female* is positive throughout and statistically significant for three of the seven activities. Females seem to have systematic advantages over males in their perception of influence over school-related activities, even after controlling for background and types of school. Since about three-fourths of all elementary and secondary-level teachers are female, female principals may be more able to have a strong connection with their employees (Goldring et al., 2013). Female principals are also more apt to have a background in curriculum than their male counterparts, and to stay longer in the principal post rather than seeing it as a stepping stone to the superintendency (Maranto et al., 2016). Minority principals have a lower likelihood of reporting that they have an influence over hiring teachers and setting discipline policy, but a

higher likelihood of reporting that they have an influence over student performance standards and curriculum. This finding may reflect the tendency of such principals to be overrepresented in more bureaucratized, unionized urban school districts (Moe, 2011; Payne, 2008).

6. Conclusion and policy implications

Private and public school principals differ significantly in their influence on school-level activities. Private school principals may have an advantage over their public school counterparts in affecting student success by having significantly more influence on almost all the school related activities. Principal characteristics, like previous experience as a department head and having a masters or higher degree, play a positive role in their ability to exercise higher influence on school activities. Nevertheless, the private school sector may be able to learn from the public school sector in evaluating teachers. Female principals appear to have a systematic advantage over their male counterparts in reporting more school-level influence.

Regarding policy implications, private school principals report having more autonomy than public school principals on every aspect of decision-making ability except the evaluation of teachers. This finding could reflect the emphasis that the Obama administration's *Race to the Top* program placed on teacher evaluations in public schools, though there are indications that these provisions in fact had little impact on personnel decisions (Maranto, McShane, & Rhinesmith, 2016).

This result could also mean that private school principals have less need for direct teacher observation and evaluation since they have more autonomy in hiring decisions and more involvement in the schools generally, as Chubb and Moe (1988) find. If principal autonomy is associated with enhanced educational experiences for children, and the private sector allows for more decision-making freedom, we should increase access to private school choice. However,

these policy decisions would benefit substantially from additional research linking principal autonomy to student-level outcomes.

Ouchi (2009) has emphasized the importance of principal autonomy, arguing that principals know more and have greater interest in what happens at the school-level than do their central office superiors (see also Nadelstern, 2013). Perhaps the relatively short tenure but greater credentialing of public school principals, as well as larger school size may suggest that they are climbers; that is, they see the principal position as a stepping-stone to the superintendency and focus on pleasing superiors rather than serving kids (Downs, 1967; Maranto et al., 2016). Cheng (2015) finds that schools where principals have more autonomy over personnel have greater mission coherence, though his sample only includes public schools. Principals with more autonomy in schools with greater mission coherence may be able to focus on student success differently than principals in schools with no mission coherence.

There are, however, limitations of this analysis. Since we have relied on self-reported measures in school surveys, the results are prone to social desirability bias as well as reference group bias (Dobbie & Fryer, 2015; West et al., 2016). Although SASS is a nationally representative sample and stable results over time can have good external validity, future studies should utilize other measures like value-added measures related to school's graduation rates and teacher turnover to study principal's leadership qualities.

Prior research theorized that placing decision making power at the school level may work better (Chubb & Moe, 1988; Hill, Pierce, & Guthrie, 2009). Principals are those administrators most aware of the daily issues at the school level: more autonomy may help them address the issues faster. Due to the heavy bureaucratization and centralized functioning of the public school system, principal autonomy is more robust in private schools.

References

Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2015). *School vouchers and student achievement: Evidence from the Louisiana Scholarship Program*. Cambridge, MA: National Bureau of Economic Research.

Bettinger, E., & Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics*, 90(8), 1625-1648.

Booker, K., Sass, T. R., Gill, B., & Zimmer, R. (2008). *Going beyond test scores: Evaluating charter school impact on educational attainment in Chicago and Florida*. Working Paper WR-610-BMG, Rand Education, Santa Monica, CA.

Borooah, V. K. (2001). *Logit and probit: Ordered and multinomial models*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-138. Thousand Oaks, CA: Sage.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter. *Education Next, 13*(1). Retrieved from http://hanushek.stanford.edu/sites/default/files/publications/Branch%2BHanushek%2BRivkin%202013%20EdNext%2013(1).pdf

Campbell, D. E. (2002). The civic side of school reform: How do school vouchers affect civic education?. *Education Next, 1*(3), 55.

Cheng, A. (2015). Do personnel policies matter for organizational mission coherence? A public school test. *Review of Public Personnel Administration.* doi:10.1177/0734371X15587981.

Chubb, J. E., & Moe, T. M. (1986). No school is an island: Politics, markets, and education. *The Brookings Review*, 4(4), 21-28.

Chubb, J. E., & Moe, T. M. (1988). Politics, markets, and the organization of schools. *American Political Science Review, 82*(04), 1065-1087.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cowen, J. M., Fleming, D. J., Witte, J. F., Wolf, P. J., & Kisida, B. (2013). School vouchers and student attainment: Evidence from a state-mandated study of Milwaukee's Parental Choice Program. *Policy Studies Journal, 41*(1), 147-168.

DeAngelis, C., & Wolf, P. J. (2016). The school choice voucher: A get out of jail card? EDRE Working Paper 2016-03. Available at SSRN: https://ssrn.com/abstract=2743541

Deming, D. J. (2011). Better schools, less crime?. *The Quarterly Journal of Economics, 126*, 2063-2115.

Dobbie, W., & Fryer, R.G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy, 123*(5), 985-1037.

Downs, A. (1967). *Inside Bureaucracy.* Boston: Little, Brown.

Dynarski, M., Rui, N., Webber, A., Gutmann, B., & Bachman, M. (2017). *Evaluation of the DC Opportunity Scholarship Program: Impacts after one year* (NCEE 2017-4022). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education.

Egalite, A. J. (2013). Measuring competitive effects from school voucher programs: A systematic review. *Journal of School Choice*, 7(4), 443-464.

Egalite, A. J. (2016). The competitive effects of the Louisiana Scholarship Program on public school performance. Available at SSRN 2739783.

Figlio, D., & Hart, C. (2014). Competitive effects of means-tested school vouchers. *American Economic Journal: Applied Economics*, 6(1), 133-156.

Firestone, W. A., & Shipps, D. (2005). How do leaders interpret conflicting accountabilities to improve student learning. In William Firestone and Carolyn Riehl (Eds.) *A New Agenda for Research in Educational Leadership* (pp. 81-100). New York, NY: Teachers College Press, Columbia University.

Fleming, D. J. (2014). Learning from schools: School choice, political learning, and policy feedback. *Policy Studies Journal*, 42(1), 55-78.

Fleming, D. J., Mitchell, W., & McNally, M. (2014). Can markets make citizens? School vouchers, political tolerance, and civic engagement. *Journal of School Choice*, 8(2), 213-236.

Forster, G. (2016). *A win-win solution: The empirical evidence on school choice*. Indianapolis: Friedman Foundation for Educational Choice.

Friedman, M. (1955). In Robert A. Solo (Ed.) *The role of government in education*. New Brunswick, NJ: Rutgers University Press.

Friedman, M., & Friedman, R. (1990). *Free to choose: A personal statement*. Orlando, FL: Houghton Mifflin Harcourt.

Goldring, R., Gray, L., & Bitterman, A. (2013). *Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2011–12 Schools and Staffing Survey* (NCES 2013-314). Washington, DC: National Center for Education Statistics.

Greene, J. P. (2005). *Education myths: What special interest groups want you to believe about our schools--and why it isn't so*. Oxford, UK: Rowman & Littlefield Publishers.

Greene, J. P., & Winters, M. A. (2003). *When schools compete: The effects of vouchers on Florida Public School achievement*. New York, NY: Center for Civic Innovation, Manhattan Institute.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher, 42*(8), 433-444.

Hess, F. M. (2013). *Cage-busting leadership*. Cambridge, MA: Harvard Education Press.

Hill, J., Ottem, R., DeRoche, J., & Owens, C. (2016). *Trends in public and private school principal demographics and qualifications: 1987-88 to 2011-12*. Washington, DC: National Center for Education Statistics.

Hill, P., Pierce, L. C., & Guthrie, J. W. (2009). *Reinventing public education: How contracting can transform America's schools*. Chicago, IL: University of Chicago Press.

Hoxby, C. M. (Ed.). (2007). *The economics of school choice*. Chicago, IL: University of Chicago Press.

Lerner, A. P. (1934). The concept of monopoly and the measurement of monopoly power. *The Review of Economic Studies, 1*(3), 157-175.

Levenson, N. (2012). *Smarter budgets, smarter schools.* Cambridge, MA: Harvard Education Press.

Maranto, R., Teodoro, M., Cheng, A., & Carroll, K. (2016, September 3). *Gender, career paths & bureaucratic ambition in public schools.* Paper presented at annual American Political Science Association meeting, Philadelphia, PA.

Maranto, R., McShane, M. Q. & Rhinesmith, E. (2016). *Education reform in the Obama era: The second term and the 2016 election***.** New York: Palgrave/Macmillan Pivot.

Merrifield, J. (2008). The twelve policy approaches to increased school choice. *Journal of School Choice*, 2(1), 4-19.

Mills, J. N., & Wolf, P. J. (2016). The effects of the Louisiana Scholarship Program on student achievement after two years. *Available at SSRN 2738805*.

Moe, T. M. (2011). *Special interest: Teachers unions and America's public schools. Washington, DC, United States: Brookings Institution Press.*

Murphy, B.G. (2017). *Inside Our Schools: Teachers on the failure and future of education reform*. Cambridge, MA: Harvard Education Press.

Nadelstern, Eric. (2013). *Ten lessons from New York City schools.* New York: Teachers College Press.

Neal, D. (2002). How vouchers could change the market for education. *The Journal of Economic Perspectives*, 16(4), 25-44.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Piscataway, NJ: Transaction Publishers.

Ong-Dean, C. (2009). *Distinguishing disability: Parents, privilege, and special education*. Chicago, IL: University of Chicago Press.

Ouchi, W. G. (2009). *The secret of TSL: The revolutionary discovery that raises school performance*. New York: Simon and Schuster.

Payne, C. M. (2008). *So much reform, so little change: The persistence of failure in urban schools*. Cambridge, MA: Harvard Education Press.

Peterson, P. E. (1998). School choice: A report card. *Virginia Journal of Social Policy & the Law*, 6, 47.

Peterson, P. E., & Hassel, B. C. (Eds.). (1998). *Learning from school choice*. Washington, DC: Brookings Institution Press.

Sandström, F. M., & Bergström, F. (2005). School vouchers in practice: Competition will not hurt you. *Journal of Public Economics*, 89(2), 351-380.

Shipps, D., & White, M. (2009). A new politics of the principalship? Accountability-driven change in New York City. *Peabody Journal of Education*, 84(3), 350-373.

Tekleselassie, A. A., & Villarreal III, P. (2011). Career mobility and departure intentions among school principals in the United States: Incentives and disincentives. *Leadership and Policy in Schools*, 10(3), 251-293.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, *64*(5), 416-424.

Tooley, J. (2005). *Reclaiming education*. London, UK: Bloomsbury Publishing.

Tooley, J., Bao, Y., Dixon, P., & Merrifield, J. (2011). School choice and academic performance: Some evidence from developing countries. *Journal of School Choice*, 5(1), 1-39.

West, E. G. (1981). Choice or monopoly in education. *Policy Review*, 15(Winter), 103.

West, E. G. (1997). Education vouchers in principle and practice: a Survey. *The World Bank Research Observer*, 12(1), 83-103.

West, M. R., & Woessmann, L. (2010). Every Catholic child in a Catholic school: Historical resistance to state schooling, contemporary private competition and student achievement across countries. *The Economic Journal*, 120(546), F229-F255.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148-170.

White, M. A. (2006). A case of New York City high school principals' experiences with multiple accountability pressures and accompanying models for decision-making. *Columbia University Teacher's College*.

Witte, J. F. (2001). *The market approach to education: An analysis of america's first voucher program*. Princeton: Princeton University Press.

Witte, J. F., Wolf, P. J., Cowen, J. M., Carlson, D., & Fleming, D. F. (2014). High stakes choice: Achievement and accountability in the nation's oldest urban voucher program. *Education Evaluation and Policy Analysis*, *36*(4), 437-456.

Wolf, P. J., & Hoople, D. S. (2006). Looking inside the black box: What school factors explain voucher gains in Washington, DC?. *Peabody Journal of Education*, 81(1), 7-26.

Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management*, *32*(2), 246-270.

Zimmer, R. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition*. Santa Monica, CA: RAND Corporation.

Chapter 5

Who stays, who leaves? Determinants of principal attrition across school sectors.

M. Danish Shakeel (mdshakee@uark.edu)

An earlier version of this chapter was presented at the conference of the Association for Education Finance & Policy in spring 2018.

1. Introduction

Principal attrition poses a challenge to the field of educational leadership. School principals play an important role in a schools' performance and, presumably, they are most informed about the issues that affect a school's environment (Hess, 2013; Ouchi, 2009). Principal turnover likely generates short-term shocks into the school system, which in turn may affect school environment and student learning (Branch, Hanushek, & Rivkin, 2013; Burkhauser et al., 2012; Mascall & Leithwood, 2010; Miller, 2013). High turnover may make it difficult for schools to effectively implement policies (Miller, 2013), improve school environment and lay out a plan (Day, Gu, & Sammons, 2016; Hitt & Tucker, 2016; Robinson, Lloyd, & Rowe, 2008) and it may also affect hiring and retaining effective teachers (Baker & Cooper, 2005; Burkhauser et al., 2012; Ronfeldt, Loeb, & Wyckoff, 2013). The School Leaders Network (2014) estimates that it costs $75,000 for a district to hire and nurture a principal. Expenses on principals' nurturing extend beyond salary, and include costs related to participation in professional development activities. The cost of turnover is likely to be higher for districts as the loss incurred is on a school's human resources and student learning. A reduction in turnover of effective principals is likely to benefit high poverty districts, where the turnover rates are high. Alternately, in some cases where student learning is on the decline and teachers are dissatisfied with the school environment, the turnover of ineffective principals may yield benefits. Nevertheless, principal turnover appears to be very important for educational leadership.

Earlier studies found heterogeneity in the types of principals who exit their schools, such as the satisfied and the disaffected principals. For example, the majority of the satisfied principals are pulled out of their job as they move on to better jobs with increased salary and professional benefits, and the low performing disaffected principals are pushed out of their

positions (Boyce & Bowers, 2016; Farley-Ripple, Raffel, & Welch, 2012; Farley-Ripple, Solano, & McDuffie, 2012). Principal satisfaction may be related to principals' intention to stay in the profession as disaffected principals are more likely to return to teaching (Boyce & Bowers, 2016). Finally, their actual decisions to stay in the profession may vary based on their experience as school leaders.

Estimates using state-level data find average principal turnover for public schools as 30 percent for Texas between 1995 and 2001 (Branch, Hanushek, & Rivkin, 2009) and 21 percent for Illinois between 2001 and 2008 (DeAngelis & White, 2011). Using Texas state data between 1995 and 2008, Fuller and Young (2009) find that almost half of newly hired principals leave within three years and more than two-thirds leave within five years. The nationally representative 2012-13 Principal Follow-up Survey (PFS) shows that 78 percent of school principals (public and private) continued to be in the same school, while 12 percent left the principalship and 6 percent moved to a different school in the following year (Goldring, Taie, & Owens, 2014). Similar statistics for the 2008-09 Principal Follow-up Survey (PFS) that also included Bureau of Indian Education schools are 80 percent, 12 percent and 6 percent (Battle, 2010). Studying principal turnover is important as in the 2012-2013 PFS among the principals who left the principalship, more than 60 percent of public school principals and 70 percent of private school principals did not leave due to retirement.

Although some research has explored the determinants of principal turnover in traditional public schools[40] using nationally representative samples (Boyce, & Bowers, 2016; Mitani, 2017; Sun & Ni, 2016; Tekleselassie & Villarreal, 2010), a comparative analysis of the attrition patterns of public and private school principals using nationally representative samples does not

[40] Sun and Ni (2016) studied turnover behavior between charter and traditional public schools.

exist (Rangel, 2018). A comparative analysis of attrition patterns of school principals may be informative of the underlying school-level mechanisms that differentiate public schools from private schools, which in turn may offer explanations for intersectoral differences in student learning.

Private school choice has gained attention with Betsy Devos as the Secretary of Education. The effectiveness of private schools in general and private school choice in particular is debated (Chubb & Moe, 1990; Coleman, Hoffer, & Kilgore, 1982; Doolittle & Connors, 2001; Lubienski & Lubienski, 2014; Ravitch, 2013). Chapter 2 of this dissertation showed that experimental evidence indicates moderate positive effects of private school vouchers over time on student achievement in the U.S., especially for students who remain in the program for more than two years. However, much less is known about the mechanism through which private schools may produce different outcomes than public schools. Differences in principals across sectors may explain some of the observed differences between public and private schools. Research shows that public school principal stability is positively associated with work-related autonomy (Farley-Ripple, Raffel, & Welch, 2012; Tekleselassie & Villarreal, 2010). Some studies argue that there is increased principal autonomy in the private school sector (Chubb & Moe, 1988; chapter 4 of this dissertation). Yet, prior work has not explored whether more autonomy for private school principals explains achievement or attainment differences across school sectors.

This study utilizes nationally representative datasets to study the determinants of attrition for public and private school principals. For this study attrition is defined as a principal leaving the profession.[41] The hypothesis tested in the study is H0: Private school principals are more

[41] The questions asked to principals at baseline year of survey (to be discussed in section 3 of this study) ask about their intentions to stay in or leave the profession. One could also study attrition as attrition from schools. The policy

likely to remain in the profession. Private school principals may be more likely to remain in the profession as they enjoy increased autonomy over managing a school (Chubb & Moe, 1988; chapter 4 of this dissertation) and self-selection of private school principals based on school mission and community orientation (Brinig & Garnett, 2014; Bryk, Lee, & Holland, 1993). On the other hand, the existence of tenure related job security, higher average salaries for principals in public schools than private schools, availability of professional development opportunities (that may strengthen their resume and facilitate principals' promotion to superintendent), may make public school principals more likely to stay in the profession than private school principals, posing a challenge to the hypothesis. The availability of more resources in the public school districts may incentivize public school principals to stay in the profession. Thus, the stay of private school principals in the profession may be socially conditioned whereas the stay of public school principals in the profession may be vocationally conditioned. Presumably, socially conditioned motivations are ideological in nature and the principals may be actively making informed selections for long-term consequences, whereas vocationally conditioned motivations are based on an expectation of availability of better professional opportunities that may not actually be available. Thus, socially conditioned motivations may translate into smaller differences in stated versus revealed preferences than vocationally conditioned responses. Hence, the study anticipates a larger gap between stated and revealed preferences of principals in public school than of principals in private schools. Due to distinct environment and organization, daily internal administrative and paperwork related requirements concerning principalship may be less

implications of studying attrition from principal profession versus studying attrition from schools would be more conservative in relation to effects on student outcomes at the school-level. However, in the current study, it is less concerning as the rates of attrition between schools are similar across sectors. Merely seven and two percent of principals within the public and private school sectors moved from their original school to another school in the following year. Seventy-seven percent and 80 percent of principals in the public and private sector schools stayed in the same school in the following year. Hence, attrition from the principal profession is not vastly different in magnitude than attrition from original schools.

demanding in private schools than public schools (Chubb & Moe, 1988). The study expects to find internal administrative pressures as a destabilizing force for public school principals.

There could be heterogeneity in principal turnover even within the public and private sectors. Principal turnover in public charter schools, due to their mission orientation and relative decentralized models in comparison to traditional public schools, may resemble differences in turnover between public and private schools. It is also possible that principal turnover rates may differ within the private sector. For example, a majority of private schools have a religious mission orientation whereas the non-sectarian private schools have a money or status driven orientation. These differences may be associated with principal satisfaction and their desire to stay in/exit from the profession. In a Catholic school network, a priest or nun leading a school may be transferred by the Archdioceses to a school (or a role other than principal) where they are needed more, particularly if they have trained a successor. In this way Catholic private schools, at least those connected to an archdiocese rather than independent, might resemble well run school districts. Thus, the study briefly tests the within school sector hypotheses: a) public charter school principals are more likely to remain in the profession that traditional public school principals and b) religious school principals are more likely to remain in the profession that non-sectarian school principals.

Using data from the School and Staffing Survey 2011-2012 and the Principal Follow-up Survey 2012, the study analyzes principal attrition between public and private schools. Principals' stated preferences in 2011-2012 are compared to their revealed status a year later. The study is relevant to the fields of educational leadership, school choice and education policy. As a good proportion of teachers take the route to become school principals, the study

contributes to our understanding of the journey of those who choose to become a principal by determining who chooses to stay and who decides to leave.

2. Literature review

Research on principal turnover has gained prominence in the area of educational leadership, primarily due to the high rates of turnover and the impact generated by turnover on schooling and principal labor markets. Rangel (2018) conducts a systematic review of 36 empirical studies on principal turnover. Studies measure turnover in many ways, ranging from amount of time principals stay or have tenure in a given school (Baker, Punswick, & Belt, 2010; Fuller & Young, 2009) to leaving a school (Li, 2015; Loeb, Kalogrides, & Horng, 2010). Other studies on principal turnover analyze career changes (Farley-Ripple, Raffel, & Welch, 2012; Li, 2015), effectiveness of the principal in improving student achievement (Branch, Hanushek, & Rivkin, 2009) and job satisfaction (Boyce & Bowers, 2016). Turnover can be based on a principal's stated decision (Tekleselassie & Villarreal, 2010) to leave a school/principalship or their revealed status later on (Goldring, Taie, & Owens, 2014; Li, 2015).

The definition of turnover depends on the policy question being answered. For example, if a principal leaves a school but remains within a school district, the policy implications of turnover would be different from a principal leaving the district. Similarly, the policy implications for retirement related turnover would be different from job dissatisfaction related turnover. Policymakers may also be interested in turnover that affects principals after their initial hiring year (Burkhauser et al., 2012).

Due to a lack of experimental studies and variation in the definition of principal turnover across studies, it is difficult to comment on the causal determinants of principal turnover. Rangel (2018) notes that the different factors vary in their importance for determining principal turnover,

and the determinants can be classified as relatively weak and relatively strong. A caveat with this classification as weak and strong is the low number of studies on principal turnover and differences in contexts, definitions and methodologies employed across studies. Relatively weak and relatively strong determinants of principal turnover are discussed below for potential inclusion in the analytical models employed in this study. For example, relatively weak determinants of turnover include principal characteristics such as gender (Fuller, Young, & Orr, 2007, Ni, Sun, & Rorrer, 2015; Sun & Ni, 2016), race (Baker, Punswick, & Belt, 2010; Fuller, Young, & Orr, 2007), age (Fuller & Young, 2009; Tekleselassie & Villarreal, 2010), principal experience (DeAngelis & White, 2011; Podgursky et al., 2016; Tran & Buckman, 2017) and principal education (Baker, Punswick, & Bel, 2010; Gates et al., 2006; Tekleselassie & Villarreal, 2010).

Analysis of a nationally representative sample shows a higher likelihood of minority principals intending to leave the principalship in comparison to their white counterparts (Carroll et al., 2018; Tekleselassie & Villarreal, 2010). Some African American principals might use principalship as a stepping stone to get better jobs elsewhere. They may see other employment markets as racist and hence may be more likely to go into public education (Carroll et al., 2018; Maranto et al., 2017). Evidence also suggests a non-linear relationship between principal age and turnover (Rangel, 2018). Younger and older principals are more likely to move or leave in comparison to middle-aged principals. The opportunity cost of changing profession may be lower for young principals than middle age principals. Old age principals may leave the profession due to accrual of adequate social security and retirement related benefits.

Work related conditions such as principal satisfaction (Boyce & Bowers, 2016; Tekleselassie & Villarreal, 2010), principal autonomy (Farley-Ripple, Raffel, & Welch, 2012;

Tekleselassie & Villarreal, 2010), relationships with superiors, peers and subordinates (Farley-Ripple, Raffel, & Welch, 2012) and the changing nature of principalship (Oberman, 1996) are weakly related to turnover. Participation in professional development activities has become an important component of the principal profession. Such changes in the nature of principalship have been building for a long time (Rousmaniere, 2013). One experimental study found that participation in professional development leads to a decline in principal turnover (Jacob et al., 2015). As participation in professional development seems to be the norm for educational leadership, the quality and type of professional development may be more relevant for principal satisfaction than mere availability of professional development activities.

School characteristics such as school level (Baker, Punswick, & Bel, 2010; Fuller, Young, & Orr, 2007; Ni, Sun, & Rorrer, 2015), percent of students at the school qualifying for special education services (Solano et al., 2010; Ni, Sun, & Rorrer, 2015) and urbanicity (Gates et al., 2006; DeAngelis & White, 2011; Tekleselassie & Villarreal, 2010) are relatively weak determinants while the school size (Baker, Punswick, & Bel, 2010; Podgursky et al., 2016), school's performance (Cullen & Mazzeo, 2007; Béteille, Kalogrides, & Loeb, 2012; Podgursky et al., 2016; Tran & Buckman, 2017), conditions (Béteille, Kalogrides, & Loeb, 2012; Sun & Ni, 2016; Tekleselassie & Villarreal, 2010), and some student demographics such as race/ethnicity and level of poverty (Gates et al., 2006; Loeb, Kalogrides, & Hong, 2010; Podgursky et al., 2016) are relatively strong determinants of turnover. Teacher characteristics such as certification (DeAngelis & White, 2011; Sun & Ni, 2016) are weakly related to principal turnover. Large districts are their own job market. Hence, within district turnover may not be a first order issue in principal turnover for larger districts, unless it leads to geographical stratification of lower quality schools. It is interesting to note that across the studies, increased school size seems to add

to principal stability. Perhaps large school size permits more division of labor. The findings may also hold for principals in charter school networks as opposed to individual schools (Foreman & Maranto, 2017).

The policies related to flexibility from the district central office over human resources, such as hiring and firing of teachers (Oberman, 1996; Sun & Ni, 2016), and state's accountability policy (DeAngelis & White, 2011; Li, 2015; Mitani, 2017) are relatively strong determinants of principal turnover. The relation between principal salary and turnover is the most researched and has generally showed that increased salary adds to principal stability (Baker, Punswick, & Bel, 2010; Cullen & Mazzeo, 2007; Ni, Sun, & Rorrer, 2015; Solano et al., 2010; Tekleselassie & Villarreal, 2010). Oberman (1996) found principal attrition positively related to early retirement options in Chicago Public Schools. Finally, district expenditures are weakly related to principal turnover (Solano et al., 2010). Boyce and Bowers (2016) show that among principals satisfied with jobs, salary is not related to turnover. Hence, beyond a mere increment in salary, factors such as school quality, school climate, working conditions and professional development of principals should be analyzed for policy relevant conclusions.

Overall effect sizes across studies show that school performance, accountability policy and professional development are consistently strong determinants of principal turnover (Rangel, 2018). Differences in school performance are related to principal turnover with as low as 3% to as high as 350% change in the likelihood of turnover. Two studies found significant differences in principal turnover in traditional public and charter schools (Ni, Sun, & Rorrer, 2015; Sun & Ni, 2016); however, no study has yet examined the differences between public and private schools in a nationally representative sample. Findings from two studies caution us to account for principal effectiveness at improving student achievement (Branch, Hanushek, & Rivkin,

2009) and satisfaction based on work-related conditions (Boyce & Bowers, 2016) when comparing principals who either stay or leave.

3. Data

The data come from the Public School Principal Status and Private School Principal Status Data Files of the 2012–13 Principal Follow-up Survey (PFS). The survey is developed by the National Center for Education Statistics (NCES) and it assesses principal attrition and mobility after a survey conducted in 2011-2012. The survey is informative of 1,720 records (Catholic, other religious and non-sectarian schools) for private school principals and 7,510 records (traditional public schools (TPS) and public charter schools) for public school principals. Principal switching between either school type does not seem to be a first order issue in principal turnover as approximately 1 percent of principals switched from public to private schools and approximately 5 percent moved vice versa. Two definitions of principal turnover are central to this study. A principal is a: a) *stayer* if they stayed as a principal and b) *leaver* if they left principalship altogether.

Tables 1A and 1B present information on principals' experience, training and professional development. A greater proportion of private than public school principals report ten or more years of experience as a principal at their current school or any school. Just nine percent of public school principals report being principal at their current school beyond ten or more years, while 23 percent report so for any school. Comparable statistics for private school principals are 25 and 41 percent. The findings may be associated with a socially versus vocationally conditioned response of leaders between private and public school sectors, wherein public school leaders may lose the vocational incentive to continue in the same school and generally in the principal profession for more than 10 years. Thus principal experience may be

weakly related to turnover (DeAngelis & White, 2011; Podgursky et al., 2016; Tran & Buckman, 2017) especially after 10 years in the profession. Public schools are more hierarchical (Gerth & Mills, 1946) than private schools. There are very few superintendents of private school systems but many more superintendents of public school systems. For most private school administrators, school principal is the highest ranking they might attain, but that is not true for most public school principals. Public school administrators may view the principalship as just another temporary step towards the ultimate prize: superintendent.

Among principals having four or more years of teaching experience, a higher percentage of public school principals have teaching experience before becoming a principal but the relation is the opposite for principals having teaching experience since their principalship. This finding could be due to the distinct nature of private schooling where a higher proportion of principals (72 percent) currently teach at school in comparison to principals at public schools (37 percent). A higher percentage of public school principals have higher education, a license/certificate in school administration and previous experience and training in school administration in comparison to their public school counterparts (Hill et al., 2016). A higher proportion of public school principals participated in professional development activities in the past 12 months. This finding could be due to the availability of extra resources and administrative requirements for professional development laid down by public authorities.

As professional development is a strong determinant of principal stability (Jacob et al., 2015; Rangel, 2018), increased opportunities for professional development in the public schools may add to higher satisfaction for a principal and consequently they may be more likely to *stay* in comparison to their private counterparts. If professional development is a source of satisfaction to the principal profession, more public school principals are likely to be *stayers* than

are private school principals. A majority of the public school principals currently hold a license/certificate in school administration. As licensure/certification may be relevant to the principal profession in public schools, it may incentivize principals to stay in the profession (DeAngelis & White, 2011; Sun & Ni, 2016). Finally, most principals in either school sector have white ethnicity and female gender; the proportions are higher for private school principals.

Table 1A: Summary statistics for principal experience and training

Measure	Public	Private
Years principal or school head at this or any school prior to this year		
no experience	8.3	8. 8
low experience 1-3	24.6	18.8
medium experience 4-10	43.8	30.9
high experience 10+	23.3	41.4
Years principal or school head at this school prior to this year		
no experience	16.5	14.5
low experience 1-3	38.8	27.6
medium experience 4-10	36.1	32.9
high experience 10+	8.6	24.9
Years of elementary or secondary teaching before becoming principal or school head		
no experience	1.7	18.5
low experience 1-3	2.8	7.9
medium experience 4-10	47.3	32.8
high experience 10+	48.2	40.7
Years of elementary or secondary teaching since becoming principal or school head		
no experience	90.4	49.7
low experience 1-3	5.4	21.9
medium experience 4-10	3.3	15.9
high experience 10+	0.8	12.6
Years of total elementary or secondary teaching		
no experience	1.4	11.1
low experience 1-3	2.5	7.1
medium experience 4-10	45.5	25.4
high experience 10+	50.6	56.4
Currently teaching at school	37.4	71.9
N	7,510	1,720

Notes: Summary statistics presented using population weighted percentages for each category.

Table 1B: Summary statistics for principal experience and training

Measure	Public	Private
Prior to becoming a principal or school head		
Worked as department head	40.4	35.3
Worked as an assistant principal or program director	73.9	43.8
Participated in school training or development program	55.3	31.4
Previous management experience outside education	40.3	46.4
Currently holding license/certificate in school administration	95.9	43.4
Having a bachelor's degree	99.9	88.5
Bachelor degree awarded by a university's department or college of education	81.9	67.8
Having a master's degree	97.6	76.3
Master's degree awarded by a university's department or college of education	97.4	85.4
Earned a MA and higher degree	97.8	68.9
Participated in any professional development activity related to principal or school head in last 12 months	99.3	89.6
Race (white)	86.4	90.2
Gender (male)	48.4	44.6
N	7,510	1,720

Notes: Summary statistics presented using population weighted percentages for each category.

Table 2 shows that, in general, a greater proportion of private school principals report a specialized status for their schools. This difference could be due to an often alluded to consumer oriented nature of private schools. Specialized schools may attract leaders wanting to work in specialized environments. Conversely, specialized schools may pose challenges to leaders, and require leadership qualities that are different than the ones required for regular schools.

Table 2: Summary statistics for school's specialization

Measure	Public	Private
Program type of school		
Regular	88.4	81.1
Montessori		5.2
Special program emphasis	3.1	1.8
Special Education	1.3	6.6
Career/Technical/Vocational Education	1.1	
Alternative	6.3	4.8
Early Childhood Program/Daycare Center		0.5

Notes: Summary statistics presented using population weighted percentages for each category.

Table 3 presents the average statistics for principals' working conditions. The average statistics do not show a meaningful difference in the principals' working conditions related to hours spent on school related activities and number of days per year required to work under their current contract. Thus, internal administrative pressures may be generally similar for the principal profession, regardless of sector. The average salary for a public school principal is one and a half times that of a private school principal. Also, on average, the public school principal is slightly younger than his/her private school counterpart. If salary is a source of satisfaction for the principalship (Baker, Punswick, & Bel, 2010; Cullen & Mazzeo, 2007; Ni, Sun, & Rorrer, 2015; Solano et al., 2010; Tekleselassie & Villarreal, 2010), the public school principal is more likely to be a *stayer*. On the other hand, the religious, mission and community orientation of

private schools (Brinig & Garnett, 2014; Bryk, Lee, & Holland, 1993; Chubb & Moe, 1988) may incentivize private school leaders to *stay* for reasons unrelated to salary.

Table 3: Summary statistics (mean) for principals' working conditions

Measure	Public	Private
Hours spent on all school-related activities during a typical full week	58.1	53.2
Hours spent interacting with students during a typical full week	22.5	21.2
Number of days per year required to work under current contract	230.8	230.9
Current annual salary (without tax and deductions)	90,510	57,560
Age	48.0	51.7

Notes: Summary statistics presented using population weighted percentages for each category. Salaries are rounded to nearest dollars.

The SASS data allow for looking at principal turnover from principals' stated preferences in 2011-2012 and revealed status at follow-up in 2012-2013 (table 4A). For the SASS 2011-2012 baseline year, the principals were asked to select from among eight categories as the projected duration for them to remain as principal, such as: *as long as able to, until eligibility for retirement from current/previous job, until eligible for social security benefits or occurrence of a special life event, until more desirable job opportunity comes along, plan to leave as soon as they can* or *undecided*. For this study, the categories are recoded as a *stayer* for *as long as able to, leaver* for intending to leave for any reason and the last category remains *undecided*. As the questions at the baseline year do not ask the principals if they want to stay in the same school, the *stayer* category represents principals' intention to stay in the profession.

Principals' revealed status at the follow-up year in 2012-2013 PFS was based on a four category status for schools in which the 2011-2012 SASS survey was carried out. The four categories were: *stayer* (principals who stayed in the same schools), *mover* (principals who were principals in a different school), *leaver* (principals who left the profession) and *other* (principals for whom it was not possible to determine a mover or leaver status). For this study, the PFS 2012-2013 definitions of *stayer* and *mover* at the follow-up year are recoded as *stayer* in the

principal profession and the *other* category is recoded as missing. Thus, comparisons can be made among the intended stayers at the base year and actual stayers at the follow-up year. Table 4B shows that among the stated baseline category of *leaver*, most principals (more among the public schools) continued to be stayers at the follow-up year.

Table 4A: Summary statistics for principals' turnover from SASS and PFS

Measure	Public	Private
Length planned to remain principal (stated at base year)		
Intended Stayer	42.7	56
Intended Leaver	32.2	15.7
Undecided	25.1	28.3
Four category principal status (revealed at follow-up)		
Actual Stayer	84.4	82
Actual Leaver	11.5	11.6
Other	4.2	6.4

Notes: Summary statistics presented using population weighted percentages for each category.

Table 4B: Summary statistics for principals' stated vs. revealed preferences

	Actual Stayer		
Length planned to remain principal	All schools	Public	Private
Intended Stayer	91.5	91.4	91.9
Intended Leaver	85.7	86.5	79.4
Undecided	83.9	84.3	82.6

Notes: Summary statistics presented using population weighted percentages for each category.

Over half of private school principals and less than half of public school principals stated they expected to remain principals the following year unconditionally. The revealed status at the follow-up year shows that more than two-thirds of principals continued to be *stayers*. The percent of actual *stayers* at the follow-up year doubled in comparison to the percent of intended *stayers* at the base year (84% versus 42%) for public schools. A smaller gap (82% versus 56%) is observed for intended *stayers* and actual *stayers* for private schools. There is a disconnect between principals' stated preferences at the base year and their revealed status at follow-up. It could be that principals wanted to leave but they ended up staying because of the high

opportunity cost of selecting alternatives. The larger gap between principals' stated versus revealed preferences for public schools in comparison to private schools may be related to the former's vocationally conditioned response versus the latter's socially conditioned response.

Chubb and Moe (1988) and chapter 4 of this dissertation argue that there is increased principal autonomy in the private schools. Principal autonomy is weakly associated with turnover in public schools (Farley-Ripple, Raffel, & Welch, 2012; Tekleselassie & Villareal, 2010). Work related barriers; especially barriers related to poor human resources that affect student learning, are likely to be a source of dissatisfaction for principals (Oberman, 1996; Sun & Ni, 2016). Table 5 shows a greater proportion of public school principals report barriers related to human resources and bureaucracy. Private school principals report more stress and a higher resistance from parents in the firing and replacement of teachers. This pattern could be due to a higher degree of parental involvement in private schools (Hiatt-Michael, 2017) and presumably high opportunity cost of replacing teachers that may be self-selecting into the private school sector.

Table 5: Summary statistics for principals' working barriers

Measure	Public	Private
Barriers to dismissal of poor-performing teachers		
Personnel policies	52.6	19.1
Termination decisions not upheld	20.3	7.9
Length of time required for termination process	61.8	14.2
Effort required for documentation	65.7	27.7
Tight deadlines for completing documentation	32.2	8.7
Tenure	69.4	13.0
Teacher associations or unions	60.2	5.9
Dismissal is too stressful and/or uncomfortable	11.5	13.2
Difficulty in obtaining suitable replacements	11.4	23.4
Resistance from parents	3.6	11.3

Notes: Summary statistics presented using population weighted percentages for each category.

Principal turnover is likely to be affected by school climate. Table 6 shows that principals report less crime in private schools (Andrade, 2013; Brinig & Garnett, 2012; Waasdorp et al., forthcoming, 2017; Zhang, Musu-Gillette, & Oudekerk, 2016). This difference could be due to possible student selection, smaller school size, mission orientation and the higher degree of parental involvement in private schools.

Table 6: Summary statistics for frequency of school problems (never occurring)

Measure	Public	Private
Physical conflicts among students	4.7	27.3
Robbery or theft	16.2	52.5
Vandalism	20.4	54.4
Student use of alcohol	69.4	85.3
Student use of illegal drugs	63.4	86.1
Student possession of weapons	58.2	93.5
Physical abuse of teachers	79.6	93.7
Student racial tensions	54.7	78.8
Student bullying	3.7	19.0
Student verbal abuse of teachers	31.2	71.0
Widespread disorder in classrooms	72.2	85.3
Student acts of disrespect for teachers	12.4	36.0
Gang activities	80.4	97.7

Notes: Summary statistics presented using population weighted percentages for each category.

Lower salaries and lower availability of professional development opportunities may put the private school principal at a comparative disadvantage. Conversely, increased autonomy, lower barriers related to human resources that affect student learning and a perceived safer school environment may put the private school principal at a comparative advantage. A comparative disadvantage/advantage would mean the private school principal is less/more likely to be a *stayer* in comparison to the public school principal. It remains to be seen if self-perceived comparative advantages/disadvantages by school principals in either sector actually translate to similar revealed status at follow-up year after controlling for various covariates.

4. Research design

The study utilizes multinomial logistics regressions to analyze the determinants for the dependent variable $Turnover$: a) for the 2011-2012 baseline year as the length planned to remain principal expressed: *stayer*, *leaver* or *undecided* and b) for the 2012-2013 follow-up year as a *stayer* or *leaver*. The baseline category *stayer*/*leaver* is comparable to the *stayer*/*leaver* category at the follow-up year. The *undecided* category at baseline year is not comparable with any category at the follow-up year. Hence, results for the *undecided* category are not reported.

$$Turnover_i = \alpha + \beta_1 X_i + \mu_i \qquad \text{Equation (1)}$$

$$Turnover_i = \alpha + \beta_1 Private_i + \beta_2 X_i + \mu_i \qquad \text{Equation (2)}$$

Equation (1) is run separately for the public and private school samples to examine the within sector determinants of turnover. Additionally equation (2) includes a school type dummy $Private$ to study if the turnover rates are consistently higher in the private sector than in the public sector. For within public school (public charter and traditional public schools) and private school (religious, Catholic and non-sectarian) sector analyses, the dummy *Private* is replaced by dummies *Public charter*, *Private religious* and *Private Catholic* in equation (2).

In equation (1), coefficient β_1 measures the likelihood of principal turnover for a unit increase in the control vector *X*. μ contains the unobserved error term. In equation (2), β_1 measures the likelihood of difference in principal turnover between private and public schools after controlling for vector *X*. μ contains the unobserved error term. The determinants of principal staying in or leaving the profession are contained in the control vector *X*. *X* comprises several measures that include principal-level characteristics such as principals' total experience, race, gender, salary and age and school-level characteristics such as school size, school level, percent of minority students and percent of minority teachers. The variables are shown in detail

in table 7. When multiple survey items were listed under a broad theme of questions in the SASS survey, the items were combined into factors using a factor analysis. Average marginal effects for equations (1) and (2) are reported in the regression coefficients. All regressions are weighted using the survey weights provided in the principal status data files (the weights are the same for baseline year and the follow-up survey). As principal labor markets are likely to differ by state, state dummies are included in all models.

Table 7: Measures used as controls in multinomial regression models

Measure	Variables
Total experience	Total experience (dummies as per table 1) as a principal, teacher, and previous experience as a department head, assistant principal or program director, participated in school training program, holding a license/certificate and management experience
Race and gender	Race and gender
Other variables	Salary in log, age (dummies created for young age ≤40, 41≤medium age≤60 and old age≥61)
School controls	School size, school level (dummies for elementary, secondary and combined), total school enrollment, number of full-time teachers, student-teacher ratio, percent of minority teachers, percent of minority students, urbanicity (dummies for city, suburban, town and rural) and percentage of enrolled students approved for the National School Lunch Program (NSLP) at school, dummies for states.
School specialization	Dummy variable for specialized schools (for equation 2, the dummy is recoded for only comparable categories in table 3)
School problems	Factor variable of school problems (loads onto two factors with Cronbach's alpha = 0.8526). Frequency of physical conflict among students, physical abuse of teachers, student bullying, student verbal abuse of teachers, widespread disorder in classroom and student acts of disrespect for teachers load on to factor 1 whereas frequency of robbery or theft, vandalism, student use of alcohol, illegal drugs, possession of weapons, racial tensions and gang activities load on to factor 2.
Principal autonomy	Factor variable of principal's decision making (loads onto two factors with Cronbach's alpha = 0.6095). Determining content of in-service professional development programs for teachers, evaluating and hiring teachers, setting discipline policy and deciding how to spend school budget load on to factor 1 whereas setting performance standards for students and establishing curriculum at school load on to factor 2.
Work conditions (bureaucratic)	Weekly hours spent on all school-related activities, avg. percent time spent on internal administrative tasks

Table 7 (Cont'd): Measures used as controls in multinomial regression models

Measure	Variables
Barriers (teacher dismissal)	Factor of variables related to barriers faced by principals in incompetent teacher dismissal (loads onto two factors with Cronbach's alpha = 0.7352). Personnel policies, termination decisions not upheld, length of time required for termination process, effort required for documentation, tight deadlines for completing documentation, tenure and teacher associations or unions load on to factor 1 whereas dismissal is too stressful and/or uncomfortable, difficulty in obtaining suitable replacements and resistance from parents load on to factor 2.

5. Results

For the 2011-2012 baseline year, the dependent variable has three categories: *stayer*, *leaver* and *undecided*. The results for *stayers* generally mirror *leavers*. For findings where results for *stayers* and *leavers* mirror each other, the discussion focuses only on *stayers*. For the 2012-2013 follow-up year, the dependent variable is dichotomous: *stayers* and *leavers*. Hence, the discussion focuses only on *stayers*.

5.1 Determinants of principals' stay in the profession and attrition based on principals' stated preferences

Results for different models are shown in table 8A. At the 2011-2012 baseline year, the significantly positive determinants of a private school principal to intend to remain as principal are principal autonomy, low principal experience (reference category experience>10 years), no teaching experience (reference category experience>10 years) and previous management experience outside education. Results for autonomy being perceived as a stabilizer by principals accord with Chubb and Moe (1988) and chapter 4 of this dissertation. It could be that private school principals with no teaching experience may be using the principalship as a route to teaching. Conversely, the significantly negative determinants for private school principals' intention to stay as principal are average percent of time spent on internal administrative tasks, medium principal experience (reference category experience>10 years), having worked as a

department head, being white, principal salary, principals' young age≤40 (reference category age>60) and elementary school level (reference category combined school level). Furthermore, school's location in a city (reference category rural) and increase in the frequency of school problems is associated with private school principals' intention to leave. As urbanicity is a weak determinant of principal turnover (Gates et al., 2006; DeAngelis & White, 2011; Tekleselassie & Villarreal, 2010), the results may be policy relevant if they remain statistically significant at the follow-up year.

The results for the average percent of time spent on internal administrative tasks suggest that private sector institutions face administrative pressures similar to public institutions. It is surprising to note that a percent increase in principal salary is associated with a 10 percentage point increase in the likelihood of a private principal to state their intention to leave the principalship. The negative coefficient on salary suggests that private schools attract principals for reasons other than salary. The change in sign of principal experience from positive to negative with an increase in experience (low: 1-3 years to medium: 4-10 years) indicates lesser opportunities for professional development or incentives for private school principals.

The significantly positive determinants of a public school principal to intend to remain as principal are principal autonomy, working in a specialized school, no and low (1-3 years) principal experience (reference category experience>10 years), no and medium (4-10 years) teaching experience (reference category experience>10 years), participation in a school training or development program and currently holding a license/certificate in school administration. Seemingly public school principals have slightly lower autonomy (Chubb & Moe, 1988; chapter 4 of this dissertation) than their private counterparts, hence, they perceive increased autonomy positively (Farley-Ripple, Raffel, & Welch, 2012; Tekleselassie & Villarreal, 2010). The

significantly negative determinants of public school principals' intention to stay as principal are barriers to dismissal of poor-performing teachers (Oberman, 1996; Sun & Ni, 2016), average percent of time spent on internal administrative tasks and principals' medium (41≤medium≤60) age (reference category age>60). It seems that the bureaucratic environment of principals' office concerning paperwork, human resource and job related constraints adds to public school principals' negative perception to remain as principals. Furthermore, the frequency of school problems, principals' white race, principal salary, previous management experience outside education and percent of minority students at the school are associated with public school principals' intention to leave. Results for principal race stand in contrast to earlier research (Rangel, 2018).

When we compare the results for public and private schools, principal autonomy is a positive determinant for principals' intention to stay as principal for both public and private school principals. However, principals in either sector value different factor loadings of principal autonomy. The private principals value autonomy concerning the setting of performance standards for students and establishing curriculum whereas the public principals value autonomy related to determining the content of professional development programs for teachers, evaluating and hiring teachers, setting the discipline policy and deciding how to spent the school budget. In both school sectors, the average percent of time spent on internal administrative tasks is a significantly negative determinant of principals' intention to stay as a principal. Low (1-3 years) principal experience (reference category experience>10 years) and no teaching experience (reference category experience>10 years) are significantly positive determinants of public and private school principals to state the desire to remain as principal. The combined model in column 3 yields a statistically significant and positive coefficient on the dummy *Private*. On

average, the private school principal is 12 percentage points more likely to state the intention to stay as principal after controlling for principal background, school and work related characteristics.

Table 8A: Results for principals' stated status at baseline year

	Intended Stayer			Intended Leaver		
	Private	Public	Pri / Pub	Private	Public	Pri / Pub
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Private			0.118***			-0.075*
			(0.039)			(0.041)
Days/year required to work	-0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Teacher dismissal barrier 1	-0.024	-0.032***	-0.031***	0.007	0.038***	0.037***
	(0.032)	(0.010)	(0.009)	(0.027)	(0.009)	(0.008)
Teacher dismissal barrier 2	-0.010	0.006	0.002	-0.011	0.005	0.006
	(0.023)	(0.009)	(0.009)	(0.015)	(0.009)	(0.008)
Weekly hrs. spent on all school	-0.001	0.001	0.001	0.001	-0.001*	-0.001
related activities	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Avg. percent time spent on	-0.004***	-0.003***	-0.003***	0.001	0.002***	0.002***
internal administrative tasks	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Principal autonomy 1	0.130***	0.016	0.020**	-0.070***	0.011	0.007
	(0.028)	(0.010)	(0.010)	(0.015)	(0.009)	(0.008)
Principal autonomy 2	0.064	0.033***	0.035***	-0.065***	-0.024***	-0.025***
	(0.040)	(0.009)	(0.009)	(0.023)	(0.008)	(0.007)
School problems 1	-0.001	-0.012	-0.012	0.035*	0.025***	0.026***
	(0.025)	(0.009)	(0.009)	(0.019)	(0.009)	(0.008)
School problems 2	-0.037	-0.017	-0.019*	-0.011	0.015	0.014
	(0.046)	(0.011)	(0.011)	(0.035)	(0.011)	(0.010)
Specialized school	-0.102	0.061*	0.053	-0.124	-0.007	-0.010
	(0.111)	(0.036)	(0.034)	(0.086)	(0.034)	(0.032)
Principal experience (no)	0.072	0.132***	0.123***	-0.133*	-0.123***	-0.118***
	(0.090)	(0.036)	(0.034)	(0.069)	(0.037)	(0.035)
Principal experience (low: 1-3 years)	0.219***	0.068**	0.068***	-0.133*	-0.067**	-0.062**
	(0.078)	(0.028)	(0.026)	(0.076)	(0.027)	(0.026)
Principal experience (medium: 4-10 years)	-0.167***	0.007	-0.006	0.120***	-0.020	-0.009
	(0.055)	(0.023)	(0.022)	(0.039)	(0.021)	(0.020)

Table 8A (Cont'd): Results for principals' stated status at baseline year

	Intended Stayer			Intended Leaver		
	Private	Public	Pri / Pub	Private	Public	Pri / Pub
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Total teaching experience (no)	1.009***	0.139**	0.121*	-1.996***	-0.012	-0.034
	(0.199)	(0.068)	(0.064)	(0.269)	(0.066)	(0.062)
Total teaching experience (low: 1-3 years)	0.128	-0.057	-0.038	0.118	-0.006	0.005
	(0.102)	(0.049)	(0.047)	(0.073)	(0.048)	(0.046)
Total teaching experience (medium: 4-10 years)	-0.038	0.056***	0.049***	-0.045	-0.020	-0.021
	(0.058)	(0.019)	(0.018)	(0.040)	(0.017)	(0.017)
Worked as department head	-0.096*	0.020	0.013	0.108***	-0.017	-0.011
	(0.049)	(0.017)	(0.017)	(0.036)	(0.016)	(0.016)
Worked as an assistant principal	0.026	0.003	0.001	-0.039	0.032	0.027
or program director	(0.048)	(0.021)	(0.020)	(0.038)	(0.020)	(0.019)
Participated in school training	-0.045	0.068***	0.060***	-0.021	-0.020	-0.019
or development program	(0.048)	(0.017)	(0.017)	(0.041)	(0.016)	(0.016)
Currently holding license/certificate	0.060	0.118**	0.087**	0.006	-0.031	-0.026
in school administration	(0.054)	(0.046)	(0.036)	(0.044)	(0.046)	(0.037)
Previous management experience	0.104**	0.018	0.025	-0.062	-0.035**	-0.036**
outside education	(0.048)	(0.017)	(0.016)	(0.039)	(0.016)	(0.015)
Race (white)	-0.157*	-0.027	-0.023	0.046	0.050*	0.047*
	(0.093)	(0.029)	(0.028)	(0.071)	(0.028)	(0.027)
Gender (male)	0.011	-0.014	-0.011	-0.052	-0.005	-0.004
	(0.053)	(0.018)	(0.017)	(0.042)	(0.016)	(0.016)
Principal salary (log)	-0.123**	-0.092	-0.094**	0.102***	0.097*	0.102**
	(0.052)	(0.057)	(0.042)	(0.040)	(0.056)	(0.043)
Principal age (young≤40)	-0.248***	-0.031	-0.033	0.049	0.050	0.042
	(0.087)	(0.038)	(0.035)	(0.059)	(0.038)	(0.035)
Principal age (41≤medium≤60)	-0.039	-0.056*	-0.050*	0.015	0.104***	0.088***
	(0.054)	(0.030)	(0.028)	(0.043)	(0.030)	(0.028)
School size	-0.021	0.002	0.007	-0.021	0.004	-0.000
	(0.035)	(0.011)	(0.010)	(0.026)	(0.011)	(0.010)

Table 8A (Cont'd): Results for principals' stated status at baseline year

	Intended Stayer			Intended Leaver		
	Private	Public	Pri / Pub	Private	Public	Pri / Pub
Variables	(1)	(2)	(3)	(4)	(5)	(6)
School level (elementary)	-0.309***	0.020	-0.010	0.133**	-0.030	-0.018
	(0.072)	(0.030)	(0.028)	(0.058)	(0.028)	(0.026)
School level (secondary)	-0.126	0.023	0.000	0.057	-0.022	-0.013
	(0.078)	(0.029)	(0.027)	(0.061)	(0.027)	(0.026)
Total school enrollment	0.000	-0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of full time teachers	-0.000	0.002	0.002*	-0.000	-0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Student/teacher ratio	-0.003	0.004	0.004	0.006	-0.003	-0.002
	(0.008)	(0.003)	(0.003)	(0.006)	(0.004)	(0.003)
Percent of minority teachers	-0.001	-0.000	-0.000	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Percent of minority students	-0.000	0.000	0.000	0.000	-0.001**	-0.001**
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
City	0.036	0.016	0.019	-0.124*	-0.035	-0.036
	(0.091)	(0.028)	(0.026)	(0.070)	(0.026)	(0.025)
Suburban	-0.004	0.018	0.014	-0.053	-0.033	-0.030
	(0.093)	(0.025)	(0.024)	(0.067)	(0.023)	(0.022)
Town	-0.050	-0.002	-0.007	-0.036	-0.002	0.002
	(0.088)	(0.024)	(0.023)	(0.063)	(0.023)	(0.022)
Percentage of enrolled students approved for the	-0.000	-0.000	-0.000	0.001	0.000	0.000
National School Lunch Program (NSLP) at school	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	420	6,970	7,380	420	6,970	7,380

All models include controls for state dummies. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

5.2 Determinants of principals' stay in the profession and attrition based on principals' revealed preferences

For the revealed status at follow-up year in table 8B, principals' young and medium age (reference category age>60) are the only significantly positive determinants of a private school principal to stay as a principal. Average percent of time spent on internal administrative tasks, frequency of school problems and low principal experience (reference category experience>10 years) are negative determinants of a private school principal to stay as a principal. Principals aged 40 or lower are associated with a 22 percentage point increased likelihood for the private school principal to remain as a principal at the follow-up year when compared to principals aged 60 or more. The coefficient for medium age principal diminishes in magnitude as compared to the coefficient for young age principal indicating that with an increase in age, the private school principal may be less likely to stay in the profession.

Comparing the private principals' responses at the baseline and the follow-up year shows a contradiction between their stated versus revealed preferences. The sign on the coefficient of low principal experience (reference category experience>10 years) and principal age change at the baseline versus follow-up year. The only consistent determinant for baseline and follow-up years is average percent time spent on internal administrative tasks. Principal autonomy, low principal experience (reference category experience>10 years), participation in school training or professional development, currently holding license/certificate in school administration and principals' young and medium age (reference category age>60) are strong positive determinants while average percent of time spent on internal administrative tasks is a negative determinant of a public school principal to stay as a principal. Comparing results for public principals at baseline and follow-up years shows that average percent of time spent on internal administrative

tasks, principal autonomy, low principal experience (reference category experience>10 years) and principals' young age (reference category age>60) remain consistent determinants of the public school principal to remain a *stayer* at both stages.

Table 8B: Results for principals' revealed status (Actual stayer) at follow-up year

	Private	Public	Private / Public
Variables	(1)	(2)	(3)
Private			0.029
			(0.028)
Days/year required to work	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Teacher dismissal barrier 1	0.017	-0.001	0.001
	(0.024)	(0.007)	(0.007)
Teacher dismissal barrier 2	0.005	-0.005	-0.005
	(0.015)	(0.005)	(0.005)
Weekly hrs. spent on all school	0.001	-0.000	-0.000
related activities	(0.002)	(0.000)	(0.000)
Avg. percent time spent on internal administrative tasks	-0.002*	-0.001*	-0.001**
	(0.001)	(0.000)	(0.000)
Principal autonomy 1	0.016	0.005	0.006
	(0.021)	(0.006)	(0.006)
Principal autonomy 2	-0.008	0.009*	0.010*
	(0.033)	(0.005)	(0.005)
School problems 1	-0.034*	0.001	-0.001
	(0.019)	(0.006)	(0.006)
School problems 2	-0.054**	-0.000	-0.003
	(0.024)	(0.007)	(0.007)
Specialized school	0.038	-0.019	-0.024
	(0.079)	(0.027)	(0.025)
Principal experience (no)	-0.096	0.039	0.022
	(0.072)	(0.029)	(0.027)
Principal experience (low: 1-3 years)	-0.120**	0.054***	0.041**
	(0.061)	(0.019)	(0.019)
Principal experience (medium: 4-10 years)	-0.072	0.022	0.016
	(0.048)	(0.014)	(0.013)

Table 8B (Cont'd): Results for principals' revealed status (Actual stayer) at follow-up year

	Private	Public	Private / Public
Variables	(1)	(2)	(3)
Total teaching experience (no)	0.026	0.039	0.041
	(0.111)	(0.050)	(0.046)
Total teaching experience (low: 1-3 years)	-0.026	-0.030	-0.023
	(0.091)	(0.038)	(0.037)
Total teaching experience (medium: 4-10 years)	-0.013	0.019	0.018
	(0.052)	(0.012)	(0.012)
Worked as department head	-0.057	0.015	0.010
	(0.043)	(0.012)	(0.012)
Worked as an assistant principal or program director	-0.036	0.003	0.001
	(0.045)	(0.013)	(0.013)
Participated in school training or development program	0.032	0.019*	0.017
	(0.040)	(0.012)	(0.011)
Currently holding license/certificate in school administration	0.018	0.063**	0.042
	(0.043)	(0.032)	(0.026)
Previous management experience outside education	0.049	0.001	0.005
	(0.039)	(0.012)	(0.012)
Race (white)	0.026	-0.015	-0.013
	(0.098)	(0.019)	(0.018)
Gender (male)	-0.004	-0.002	-0.001
	(0.046)	(0.012)	(0.012)
Principal salary (log)	-0.005	-0.016	-0.014
	(0.043)	(0.039)	(0.026)
Principal age (young≤40)	0.223***	0.120***	0.122***
	(0.081)	(0.023)	(0.022)
Principal age (41≤medium≤60)	0.141***	0.111***	0.107***
	(0.039)	(0.016)	(0.015)
School size	-0.029	-0.003	-0.002
	(0.040)	(0.008)	(0.007)

Table 8B (Cont'd): Results for principals' revealed status (Actual stayer) at follow-up year

	Private	Public	Private / Public
Variables	(1)	(2)	(3)
School level (elementary)	-0.023	0.014	0.011
	(0.063)	(0.023)	(0.021)
School level (secondary)	0.079	-0.006	-0.002
	(0.053)	(0.020)	(0.019)
Total school enrollment	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Number of full time teachers	0.001	-0.000	-0.000
	(0.002)	(0.001)	(0.001)
Student/teacher ratio	0.007	-0.001	-0.001
	(0.010)	(0.002)	(0.001)
Percent of minority teachers	0.002	0.000	0.000
	(0.001)	(0.000)	(0.000)
Percent of minority students	-0.001	-0.000	-0.000
	(0.001)	(0.000)	(0.000)
City	-0.106	0.012	0.005
	(0.067)	(0.018)	(0.017)
Suburban	-0.082	0.017	0.014
	(0.064)	(0.015)	(0.015)
Town	-0.068	0.019	0.015
	(0.067)	(0.017)	(0.016)
Percentage of enrolled students approved	0.001	0.000	0.000
for the National School Lunch Program (NSLP) at school	(0.001)	(0.000)	(0.000)
Observations	400	6,730	7,120

All models include controls for state dummies. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Young and medium age of principals are significantly positive determinants whereas average percent time spent on internal administrative tasks is a significantly negative determinant for both public and private school principals to stay as principal at the follow-up year. Low principal experience is a negative determinant for the private school principal but a positive determinant for the public school principal to remain a *stayer* at the follow-up year. The overall model with the dummy *Private* in column 3 has a positive coefficient but is not statistically significant. Although the private school principals were more likely to express their intention to stay as principal at the baseline year in comparison to public school principals, their stated intentions do not translate to the follow-up year in a statistically significant way. The results hint towards context dependent complexity in the nature of attrition and principal satisfaction (Boyce & Bowers, 2016; Farley-Ripple, Raffel, & Welch, 2012; Farley-Ripple, Solano, & McDuffie, 2012). In general, there is a disconnect between principals' stated versus revealed preferences. Results indicate that public school principals might be climbers (Downs, 1967) who may stay in the profession for a few years and then get promoted into higher administration. Some principals may stay in the profession to gain teaching experience and later on go to teaching (Boyce & Bowers, 2016).

Comparing principals' responses at the baseline and the follow-up year shows that the average percent of time spent on internal administrative tasks, low principal experience (reference category experience>10 years) and principals' medium age (reference category age>60) are the only significantly consistent determinants of private principals leaving the profession. Whereas average percent of time spent on internal administrative tasks, principal autonomy, low principal experience (reference category experience>10 years) and principals' medium age (reference category age>60) are the only significantly consistent determinants of

public principals leaving the profession. In the overall model, the dummy *Private* loses statistical significance at follow-up as compared to the baseline year.

5.3 Within school sector comparisons

Tables 9A and 9B show results for within school sector comparisons between public charter and traditional public school (TPS) (dummy *Public charter*), religious and non-sectarian (dummy *Private religious*) and Catholic and other private schools (dummy *Private Catholic*). For within public school sector comparisons, the stated preferences at the baseline year show that public charter school principals are significantly more likely to intend to remain principals and much less likely to leave the profession in comparison to TPS principals. However, their revealed preferences at the follow-up year show a loss of statistical significance for the *Public charter* dummy. Results for within private sector comparisons yield statistically null coefficients on the dummy *Private religious* and *Private Catholic*. Principals' stated preferences at the baseline year suggest that religious school principals are more likely to leave the profession in comparison to non-sectarian school principals, however the revealed status at the follow-up year indicates the contrary. Catholic school principals are less likely to intend to stay and also less likely to intend to leave the profession at the baseline year in comparison to other private school principals. At the follow-up year Catholic school principals are less likely to stay and more likely to leave the profession. The results reflect the high rate of closing of Catholic schools across the country (Brinig & Garnett, 2014). Thus, Catholic school principals would be more likely to expect to leave the profession and also be more likely actually to leave it, as for some of them their school closed in the meantime.

Table 9 A: Results for principals' stated status (Intended stayer) at baseline year

	Intended Stayer			Intended Leaver		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Public charter	0.075*			-0.081*		
	(0.040)			(0.042)		
Private religious		-0.040			0.053	
		(0.093)			(0.080)	
Private Catholic			-0.072			-0.021
			(0.061)			(0.048)
Observations	6,970	420	420	6,970	420	420

All models include controls for covariates in table 7 and also for state dummies.
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 9 B: Results for principals' revealed status (Actual stayer) at follow-up year

VARIABLES	(1)	(2)	(3)
Public charter	0.001		
	(0.028)		
Private religious		0.024	
		(0.086)	
Private Catholic			-0.023
			(0.049)
Observations	6,730	400	400

All models include controls for covariates in table 7 and also for state dummies.
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

At the overall level, the dummy variables indicate that within school sector differences are similar to between school sector differences at the follow-up year. However, at the baseline year only the within public school sector differences are similar to differences between the public and private school sectors. The overall results do not dismiss the possible heterogeneity that could exist within school sectors. Future research should consider the role of sector variables interacted with control variables in drawing policy relevant conclusions for across school sector differences.

6. Conclusions

This study contributes to fields of educational leadership, school choice and education policy by presenting a comparative analysis of the determinants of principal attrition in public

and private schools in a nationally representative sample. There is a mismatch between principals' stated preferences and revealed status for both public and private school principals. About one-third of control variables lose statistical significance at the follow-up year in comparison to the baseline year. Principals' stated intentions to leave the principalship or not are more predictable, based on descriptive characteristics, than their subsequent revealed behaviors. Thus, there is a need to understand the mechanism that contributes to the disconnect between principals' stated preferences versus their revealed status. The findings call for more research on principal satisfaction as a key determinant of principal turnover (Boyce & Bowers, 2016; Farley-Ripple, Raffel, & Welch, 2012; Farley-Ripple, Solano, & McDuffie, 2012).

The study finds moderate evidence that private school principals are more likely to remain in the profession in comparison to public school principals and charter school principals are more likely to remain in the profession in comparison to TPS principals. Charter school principals are often the founder of their school so they may feel a personal attachment to it. The findings for a greater likelihood of religious school principals to remain in the profession in comparison to non-sectarian school principals are seen at the follow-up year but they contradict the results for principals' stated preferences. The larger gap between public school principals' stated preferences versus actual status suggest that private school principals provide a socially conditioned response whereas their public school counterparts provide a vocationally conditioned response. As a greater proportion of principals in private schools currently teach at their school, they are more likely to switch to teaching at their schools in the future (Chubb & Moe, 1988). A move from principalship to teacher is relatively easier for the private school principal than public school principal due to the existence of bureaucratic rules and

specialization (Gerth & Mills, 1946) that limit such a transition and greater social distance between principals and teachers in public schools (Ingersoll, 2003).

Participation in professional development activities, currently holding a license/certificate in school administration and principal autonomy are found to be significant determinants only for a public school principal to remain in the profession. Similarities in results at the follow-up year indicate that principals in both school sectors perceive the average percent of time they spend on internal administrative tasks as negative for their stay in the profession. Thus, internal administrative pressures are innate to the principal profession. Both young and medium age principals are more stable in the profession in comparison to old age principals. The results contrast with those of Rangel (2018) where the relation between age and turnover hints that younger and older principals are more prone to turnover in comparison to middle aged principals. Low principalship experience adds to the likelihood of public school principals to remain in the profession whereas it weakens the likelihood of private school principals to remain in the profession. This pattern could be due to presumably better opportunities for professional growth and job related stability for leaders in public schools.

The current study has some limitations. For example, it did not differentiate between various types of attrition. Inclusion of variables related to school performance, accountability, parental involvement, teacher turnover and the role of sector variables interacted with control variables may offer opportunities for better understanding the relative importance of underlying mechanisms that affect turnover in each school sector. Researchers may investigate if principal training programs could address their common concerns related to average percent of time spent on internal administrative tasks. Replication of a similar analysis using different datasets may help policymakers and educators address issues of principal attrition in a better way.

References

Andrade, F. H. (2013). Can the school context moderate the protective effect of parental support on adolescents' alcohol trajectories in urban Chicago? *Drug and alcohol dependence, 133*(2), 330-337.

Baker, B. D., & Cooper, B. S. (2005). Do principals with stronger academic backgrounds hire better teachers? Policy implications for improving high-poverty schools. *Educational Administration Quarterly, 41*(3): 449-479.

Baker, B. D., Punswick, E., & Belt, C. (2010). School leadership stability, principal moves, and departures: Evidence from Missouri. *Educational Administration Quarterly, 46*(6), 523–557.

Battle, D. (2010). *Principal attrition and mobility: Results from the 2008-09 Principal Follow-up Survey* (NCES 2010-337). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research, 41*(4), 904-919.

Bettinger, E., & Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics, 90*(8), 1625-1648.

Boyce, J., & Bowers, A. J. (2016). Principal turnover: Are there different types of principals who move from or leave their schools? A latent class analysis of the 2007–2008 Schools and Staffing Survey and the 2008–2009 Principal Follow-up Survey. *Leadership and Policy in Schools, 15*(3), 237-272.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2009). Principal turnover and effectiveness. Presented at the American Economics Association, San Francisco.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter. *Education Next, 13*(1).

Brinig, M. F., & Garnett, N. S. (2012). Catholic schools, charter schools, and urban neighborhoods. *The University of Chicago Law Review, 79*(1), 31-57.

Brinig, M. F., & Garnett, N. S. (2014). *Lost classroom, lost community: Catholic schools' importance in urban America*. University of Chicago Press.

Bryk, A. S., Lee, V. E., & Peter, B. Holland. 1993. *Catholic schools and the common good*. Harvard University Press.

Burkhauser, S., Gates, S. M., Hamilton, L. S., & Ikemoto, G. S. (2012). *First-Year principals in urban school districts: How actions and working conditions relate to outcomes* (Technical Report). Santa Monica, CA: RAND Corporation.

Carroll, K., Cheng, A., Maranto, R. & Teodoro, M. P. (2018, in press). Race, career paths, and bureaucratic ambition: Evidence from public school principal posts.

Campbell, D. E. (2008). The civic side of school choice: An empirical analysis of civic education in public and private schools. *BYU L. Rev.*, 487.

Chubb, J. E., & Moe, T. M. (1988). Politics, markets, and the organization of schools. *American Political Science Review, 82*(04), 1065–1087.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.

Cullen, J. B., & Mazzeo, M. J. (2007). *Implicit performance awards: An empirical analysis of the labor market for public school administrators*. University of California, San Diego.

Day, C., Gu, Q., & Sammons, P. (2016). The impact of leadership on student outcomes: How successful school leaders use transformational and instructional strategies to make a difference. *Educational Administration Quarterly*, *52*(2), 1-38.

DeAngelis, K. J., & White, B. R. (2011). *Principal turnover in Illinois Public Schools, 2001-2008* (Policy Research No. IERC 2011-1). Illinois: Illinois Education Research Council.

Downs, A. (1967). *Inside Bureaucracy.* Boston: Little, Brown.

Farley-Ripple, E. N., Raffel, J. A., & Welch, J. C. (2012). Administrator career paths and decision processes. *Journal of Educational Administration, 50*(6), 788–816.

Farley-Ripple, E. N., Solano, P. L., & McDuffie, M. J. (2012). Conceptual and methodological issues in research on school administrator career behavior. *Educational Researcher, 41*(6), 220-229.

Foreman, L. M., & Maranto, R. (2017). Why mission matters more for (some) charter principals. *School Leadership & Management, 37*(3).

Fuller, E., Young, M. D., & Orr, M. T. (2007). Career pathways of principals in Texas. Presented at the American Educational Research Association, Chicago, IL.

Fuller, E. J., & Young, M. D. (2009). *Tenure and retention of newly hired principals in Texas*. University Council for Educational Administration, Department of Educational Administration, University of Texas at Austin.

Gates, S. M., Ringel, J. S., Santibanez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review, 25*, 289–302.

Gerth, H. H., & Mills, C. W. (1946). *From Max Weber: Essays in sociology.* New York, NY: Oxford University Press.

Goldring, R., Taie, S., & Owens, C. (2014). *Principal attrition and mobility: Results from the 2012-13 Principal Follow-up Survey. First Look* (No. NCES 2014-064). Washington, DC: National Center for Education Statistics.

Hess, F. M. (2013). *Cage-busting leadership*. Cambridge, MA: Harvard Education Press.

Hiatt-Michael, D. B. (Ed.). (2017). *Family involvement in faith-based schools*. Charlotte, NC: IAP.

Hitt, D. H., & Tucker, P. D. (2016). Systematic review of key leader practices found to influence student achievement: A unified framework. *Review of Educational Research, 86*(2), 531-569.

Hill, J., Ottem, R., DeRoche, J., & Owens, C. (2016). *Trends in public and private school principal demographics and qualifications*: 1987-88 to 2011-12. Washington, DC: National Center for Education Statistics.

Ingersoll, R. M. (2003). *Who controls teachers' work?: Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), 314-332.

Li, D. (2015). School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders. Harvard Business School Working Paper 16-052.

Loeb, S., Kalogrides, D., & Horng, E. L. (2010). Principal preferences and the uneven distribution of principals across schools. *Educational Evaluation and Policy Analysis, 32*(2), 205–229.

Lubienski, C. A., & Lubienski, S. T. (2014). *The public school advantage: Why public schools outperform private schools*. Chicago, IL: University of Chicago Press.

Maranto, R., Teodoro, M. P., Carroll, K., & Cheng, A. (2017). Gendered ambition: Career advancement in public schools. EDRE Working Paper No. 2017-2018. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3045990

Mascall, B. & Leithwood, K. (2010). Investing in leadership: The district's role in managingprincipal turnover. *Leadership and Policy in Schools, 9*(4), 367-383.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review, 36*, 60-72.

Mitani, H. (2017). Principals' Working Conditions, Job Stress, and Turnover Behaviors under NCLB Accountability Pressure. Paper presented at the annual meeting of the Association for Education Finance and Policy.

Ni, Y., Sun, M., & Rorrer, A. (2015). Principal turnover: Upheaval and uncertainty in charter schools? *Educational Administration Quarterly*, *51*(3), 409–437.

Oberman, G. L. (1996). *A report on principal turnover in the Chicago Public Schools* (Research Report). Chicago, IL: Chicago Public Schools Department of Research, Evaluation, and Planning.

Ouchi, W. G. (2009). *The secret of TSL: The revolutionary discovery that raises school performance*. New York, NY: Simon and Schuster.

Podgursky, M., Ehlert, M., Lindsay, J., & Wan, Y. (2016). *An examination of the movement of educators across and within three Midwest region states*. Regional Educational Laboratory Midwest.

Rangel, V. S. (2018). A review of the literature on principal turnover. *Review of Educational Research*, *88*(1), 87-124.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635–674.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Rousmaniere, K. (2013). *Principal's Office, The: A Social History of the American School Principal*. SUNY Press.

School Leaders Network. (2014). *CHURN: The high cost of principal turnover*. School Leaders Network.

Solano, P. L., McDuffie, M. J., Farley-Ripple, E. N., & Bruton, J. (2010). *Principal retention in the state of Delaware*. Newark, DE: University of Delaware, Center for Community Research and Service.

Sun, M., & Ni, Y. (2016). Work Environments and Labor Markets: Explaining Principal Turnover Gap Between Charter Schools and Traditional Public Schools. *Educational Administration Quarterly*, *52*(1), 144–183.

Tekleselassie, A. A., & Villarreal, III, P. (2010). Career mobility and departure intentions among school principals in the United States: Incentives and disincentives. *Leadership and Policy in Schools*, *10*(3), 251–293.

Tran, H. & Buckman, D. G. (2017). The impact of principal movement and school achievement on principal salaries. *Leadership and Policy in Schools*, *16*(1), 106-129.

Waasdorp, T. E., Berg, J., Debnam, K. J., Stuart, E. A., & Bradshaw, C. P. (forthcoming, 2017). Comparing social, emotional, and behavioral health risks among youth attending public versus parochial schools. *Journal of School Violence*.

Zhang, A., Musu-Gillette, L., & Oudekerk, B. A. (2016). Indicators of school crime and safety: 2015. NCES 2016-079. NCJ 249758. *National Center for Education Statistics*.

Chapter 6

Conclusion

This dissertation presented evidence from four studies of differences between the public and private school sectors. Two studies combined causal evidence from studies on school vouchers globally to test the competing claims concerning the efficiency of public versus private schools at generating test scores. The overall evidence on student achievement in math and reading scores from all available experimental studies on school vouchers was analyzed. The findings were also explored from a cost-effectiveness perspective. Beyond voucher interventions, nationally representative datasets for school principal surveys were analyzed. Two studies separately examined principal autonomy and principal attrition between the public and private school sectors. This concluding chapter of the dissertation summarizes the findings from chapters 2, 3, 4 and 5. The study limitations and policy implications are discussed.

Summary of findings

Evidence is presented from experimental studies on private school vouchers and educational leadership. Four key issues analyzed in the dissertation relate to: a) achievement effects of vouchers, b) cost-effectiveness of vouchers, c) comparative autonomy and d) comparative attrition of school leaders in private versus public schools.

Chapter 2 is a global meta-analysis that combined evidence on student math and reading test scores from 20 randomized control trial (RCTs) on school vouchers. Sixteen studies took place in the U.S. whereas two studies each took place in India and Colombia. The 20 studies represent 11 distinct voucher programs. For each study, the level of randomization was at the child/family level. The voucher programs targeted low-income students through either income

limits and/or program location. The meta-analytic effects represent the achievement effects of low-cost private school vouchers on low-income, primarily urban minority children.

The findings from the meta-analysis show a statistically null impact of the use of school vouchers on student achievement in math and reading inside the U.S. Overall impacts inside the U.S. for reading are 0.03 standard deviation (SD) whereas for math it is 0.01 SD. However, outside the U.S., the impacts are significantly positive: the reading impact is 0.51 SD and the math impact is 0.35 SD. The large impacts outside the U.S. are driven by the voucher program in Colombia that represents a combination of student incentives, additional education spending and private school productivity (Glewwe & Muralidharan, 2016). When the findings are analyzed by years of treatment, initially a decline in test scores impacts is seen but, by year 3 and beyond, the test scores catch up and improve, more so for reading scores.

Chapter 3 analyzed the voucher studies from the perspective of cost-effectiveness. The programs were categorized as either publicly or privately funded, with the former defined as a program that received any public funds. Furthermore, some programs required participating private schools to accept the cost of the voucher as the full amount for educating students in their schools. Such programs were labelled 'fully funded.' Other programs covered only partial tuition costs and allowed the parents to top-up with extra amounts to cover the remaining cost of educating their children.

The results of the cost-effectiveness analysis do not establish whether publicly versus privately funded programs are more efficient, as 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. However, in accord with prior findings (Muralidharan & Sundararaman, 2015; Wolf & McShane, 2013) the cost-effectiveness analysis

found that generally, private school vouchers save money, even in cases where the voucher impacts are statistically null for student achievement in math and reading. The lower bound net savings for US programs was approximately $55 million. The lower bound net savings for the program in Colombia was $4 million and for the programs in India was $2 million. Globally the voucher interventions resulted in a lower bound net savings of approximately $60 million.

Assuming the results for a modest private school advantage in student achievement with time hold across contexts, we want to know why they do so. Chapter 4 analyzed differences in private and public sector principals' responses to having control over seven school level activities. In six out of seven activities such as hiring and firing teachers and having autonomy over spending the budget, the private school principals reported having significantly more autonomy than their public counterparts. The results comport with Chubb and Moe (1988) who theorized there was increased autonomy for leaders in the private school sector in comparison to leaders in the public school sector.

Chapter 5 investigated whether private school principals are more likely to stay in the profession than public school principals. Principals' stated preferences at the baseline year showed that principals in private schools are 12 percentage points more likely to intend to remain in the profession in comparison to principals in public schools. A year later, the differences in principals' revealed status loses statistical significance. The determinants of principal attrition at the baseline year stand in contrast to determinants of attrition at the follow-up year.

Study limitations

The first two studies presented in this dissertation suffer from some limitations. For example, the meta-analysis examined 11 voucher programs, three of which are outside the U.S., thereby limiting its global scope. There is a dearth of causal studies on school vouchers,

especially outside the U.S., in developing countries having a larger gap between public and private institutions. A second issue is that most voucher programs have not been evaluated for three years or more, thus, severely restricting the ability of researchers to test dosage effects of voucher use. Thirdly, limited data does not establish differences between publicly and privately funded programs. More granular data related to funding is necessary for a more precise benefit/cost analysis of vouchers. Fourthly, most voucher programs included in the meta-analysis are small scale interventions. Programs with more participants may allow a better understanding of issues that relate to the scaling up of vouchers. Lastly, the data do not allow us to examine how voucher impacts differ between lower, middle and high grade levels.

Some limitations also concern the last two studies in this dissertation. Private institutions tend to have less public accountability than public institutions. Hence, increased autonomy for private school principals needs to be connected to educational outcomes for establishing its role in producing educational benefits. Furthermore, the dissertation did not test if principal autonomy is a moderator of student achievement, attainment or civic outcomes. Whether increased autonomy for school principals is an educational good or if principals' self-perceived increased autonomy actually translates into revealed practices of education innovation needs to be studied. Similarly, the dissertation did not test if principals' longer stay in the profession is good or bad. The results show no statistically significant difference between principals' revealed attrition status in the public and private sectors. Lastly, the findings from leadership studies are not causal. More studies are needed in different contexts to establish differences in leadership between public and private institutions.

Important lessons for policy

The general pattern of the results indicates that vouchers tend to increase reading scores more than math scores. The initial decline in test scores could be due to the school switch which most voucher participants experience at the start of their participation in the program. Students may require time to adjust to their new private schools. Conversely, the private schools that participate in voucher programs may require time to adjust to an influx of disadvantaged students (Wolf, 2018). Thus, longer term achievement effects are more relevant for drawing policy conclusions (Das et al., 2013; Glewwe & Muralidharan, 2016).

Generally, school vouchers are cost-effective as they tend to generate student achievement outcomes either as good as or somewhat better than traditional public schools, at a fraction of the cost (Muralidharan & Sundararaman, 2015; Wolf & McShane, 2013). Thus null achievement impacts in voucher programs, if obtained at lower per-student expense than public schools, may have a net benefit. For policy, results establish that vouchers are generally cost-effective and produce moderate effects over time.

Results also establish a private school leadership advantage in autonomy over school-level activities such as establishing curriculum and setting the discipline policy of the school. At the baseline year, private school principals are much more likely to perceive that they will stay in the principalship than are public school principals. The study on principal attrition yields a private sector advantage for principals' perceptions but not for their revealed status. Overall, the leadership research shows a private school advantage for school principals' autonomy and modest advantage for principals to remain in the profession. Results indicate that the response of private school principals to questions about their continuance in the job is socially conditioned whereas the response of public school principals is vocationally conditioned.

This dissertation presents research showing a clear and consistent but moderate benefit of school vouchers on student test scores that is also cost-effective. Second, research shows a more free private school principal, who is not bound by as much red tape of bureaucracy. The impacts may vary across contexts so more research is needed. Nevertheless, private schools offer a hope for educational improvement. Use of empirical data in educational research is both wonderful and helpful for a graduate student. However, Hess (2017) informs is in his book *Letters to a Young Education Reformer* that even educated and well-trained researchers may look and interpret the data in opposing ways.

References

Chubb, J. E., & Moe, T. M. (1988). Politics, markets, and the organization of schools. *American Political Science Review, 82*(04), 1065–1087.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics, 5*, 29-57.

Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. *Handbook of the Economics of Education, 5*, 653-743. Amsterdam, Netherlands: Elsevier.

Hess, F. M. (2017). *Letters to a Young Education Reformer*. Harvard Education Press. Cambridge, MA.

Muralidharan, K., & Sundararaman. V. (2015). The aggregate effect of school choice evidence from a two-stage experiment in India. *The Quarterly Journal of Economics, 130*, 1011-1066.

Wolf, P. J. & McShane, M. (2013). Is the juice worth the squeeze? A benefit/cost analysis of the District of Columbia opportunity scholarship program. *Education Finance and Policy, 8*, 74–99.

Wolf, P. J. (2018). Programs benefit disadvantaged students. In forum: Taking stock of private-school choice. *Education Next, 18*(2).

Appendix F: Institutional Review Board Approval Letter

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

August 31, 2017

MEMORANDUM

TO:	M. Danish Shakeel Patrick Wolf
FROM:	Ro Windwalker IRB Coordinator
RE:	EXEMPT PROJECT CONTINUATION
IRB Protocol #:	16-10-137
Protocol Title:	*Who is More Free? A Comparison of the Decision-Making of Private and Public School Principals*
Review Type:	☒ EXEMPT
New Approval Date:	08/31/2017

Your request to extend the referenced protocol has been approved by the IRB. We will no longer be requiring continuing reviews for exempt protocols.

If you wish to make any modifications in the approved protocol that may affect the level of risk to your participants, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.