7-2015

The Effectiveness of Cash Transfers as a Policy Instrument in K-16 Education

Jonathan Norman Mills
University of Arkansas, Fayetteville

The Effectiveness of Cash Transfers as a Policy Instrument in K-16 Education

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

by

Jonathan N. Mills
University of Missouri
Bachelor of Science in Economics, 2007
University of Missouri
Master of Arts in Economics, 2009

July 2015
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dr. Gema Zamarro Rodriguez
Dissertation Director

Dr. Jay P. Greene
Committee Member

Dr. Patrick J. Wolf
Committee Member

Abstract

Cash transfers, defined as direct transfers of purchasing power from an institution or individual to another individual, are an increasingly popular policy instrument both abroad and in the United States. This dissertation investigates how two educational interventions utilizing cash transfers affect participating students. The first, the Louisiana Scholarship Program, is a statewide program offering publicly financed scholarships for low income students attending poorly performing public schools to attend K-12 private schools. The second program, the Arkansas Academic Challenge Scholarship, is a state financed broad-based merit-aid scholarship for students in college within the state of Arkansas. In general, the results presented in this dissertation suggest that students using LSP scholarships performed significantly behind their counterparts in math and English Language Arts after one year, but did not differ substantially from student who did not receive a scholarship on measures of non-cognitive skills or political tolerance. In addition, the findings presented here suggest that currently enrolled students who barely qualified for an Academic Challenge Scholarship performed no differently after one year than students who barely missed the academic requirements, but earned significantly lower final GPAs and were less likely to graduate within four years. These differences disappeared after five years, suggesting that scholarship recipients may have delayed graduation in response to the program. These studies contribute to the literature on school vouchers and college merit-aid scholarships by providing the first experimental evaluation of a statewide voucher program on student achievement, the first descriptive evaluation of a voucher program on student non-cognitive skills and political tolerance, and the first examination of the effects of a merit-aid scholarship on students who were currently enrolled in college at the time of scholarship receipt.

Acknowledgements

First and foremost, I would like to thank my committee members Gema Zamarro Rodriguez, Jay P. Greene, and Patrick J. Wolf, without whose guidance and support this work could not exist. In addition, I would like to thank my colleagues who have contributed to the research comprising this dissertation--Albert Cheng, Collin Hitt, and Gary W. Ritter—as well as the panel participants at the annual meetings of the Association for Education Finance and Policy and Association for Public Policy Analysis and Management who provided valuable advice in the research process. I am also particularly indebted to Anna Egalite and Brian Kisida, whose advice and support played integral roles in this work as well as my development as a researcher. In addition, I would like to thank the remaining faculty, staff, and students of the Department of Education Reform at the University of Arkansas for contributing to my growth as a researcher over the last few years.

I would also like to acknowledge my good friends, especially Justin Adams, Karl Doege, and April Johnson, who have made Fayetteville feel like home for the last six years. Finally, I would like to thank my family—Reenie and Tom Thornton, Ellie Mills, Zach Thornton, and Laura Thornton—for their continued support during my time in Fayetteville.

Dedication

This edition of *The Effectiveness of Cash Transfers as a Policy Instrument in K-16 Education* is dedicated to my father, Frank Mills, whose memory I attempt to serve every day.

Table of Contents

Introduction

Education is typically viewed in the United States as one of the most important means to improve individual economic outcomes (Heckman, 2008) while also developing the nation's citizenry (Dewey, 1916; Gutmann, 2003). Unfortunately, despite numerous proposals to improve the quality of education in the US, student performance has largely stagnated over the last 40 years. While 9 and 13 year olds have made significant gains on the long-term National Assessment of Educational Progress (NAEP), math and reading achievement has remained stagnate for 17 year olds over the past 4 decades. Similarly, while Black and Hispanic students have made slight gains against White students on the NAEP since 1970, achievement gaps persist today. At the same time, real current dollar expenditure per pupil has more than doubled since the 1970s (Mills, 2013). In addition, while America's higher education system is generally viewed as having played an important role in the nation's economic success (Goldin & Katz, 2008), college graduation rates have risen only slightly since 2004 (National Center for Education Statistics, 2014). This dissertation presents an investigation of two educational interventions—one focused on K-12 education and the other on higher education—utilizing cash transfers, defined as a direct transfer of purchasing power from an institution or individual to another individual, to address these issues in the K-16 education system.

Cash transfer programs have become a popular education policy instrument in developing nations (Baird, Ferreira, Ozler, & Woolcock, 2013; Baird, McIntosh, & Ozler, 2011; Handa & Davis, 2006; World Bank, 2009). Cash transfers can be implemented in two types: unconditional and conditional cash transfers. While conditional cash transfer (CCT) programs build in rules for program eligibility as well as continual receipt of funds, unconditional cash transfer (UCT) programs typically do not have such on-going requirements (Baird et al., 2013). In general, studies of CCT and UCT interventions in developing nations using the most rigorous empirical methods available generally find positive effects on schooling outcomes. Baird et al. (2013) provide a meta-analysis of the effects of CCTs and UCTs on schooling and later life outcomes in developing nations. Their work reviews evidence from 35 experimental and quasi-experimental evaluations of cash transfer programs, with five focused exclusively on UCTs, 26 on CCTs, and four studies comparing the two types of interventions to one another. In general, Baird et al. report that UCTs and CCTs significantly improve the likelihood of enrolling in and attending school. Moreover, the authors report evidence suggesting that stronger conditions are generally correlated with stronger effects among CCT programs.

Nevertheless, there is evidence suggestive of some tradeoffs between CCTs and UCTs. Baird, McIntosh, and Ozler (2011) note that CCT proponents argue that conditions are needed in cash transfer programs to address existing market failures that led to the under investment in education in the first place. This argument is supported by empirical evidence suggesting relatively stronger effects associated with CCTs over UCTs (Baird, McIntosh, & Ozler, 2011; Baird et al., 2013; Todd & Wolpin, 2006). In addition, CCT proponents suggest that conditions are necessary in order to gain political support from the middle- and upper-classes (Baird, McIntosh, & Ozler, 2011). Proponents, on the other hand, argue that UCTs should serve as the default because policy designers actually have a poor understanding of how incentives operate in CCTs (Baird, McIntosh, & Ozler, 2011). For example, Baird, McIntosh, and Ozler (2011) find that teen pregnancy and marriage rates were significantly lower among girls randomly assigned to a UCT-style intervention in Malawi compared to girls assigned to a CCT intervention. In addition, Benhassine, Devoto, Duflo, Dupas, and Pouliquen (2013) find that simple informational nudges may be all that is required, as individuals in Morocco randomly assigned to an UCT intervention that was explicitly identified as an educational intervention were no less likely enroll in school as students assigned to a CCT program. Thus, while evidence from developing nations indicates that cash transfers generally can positively affect schooling outcomes, there is evidence that program design can have important—and potentially unintended—consequences for program effectiveness.

Cash transfers are also becoming an increasingly popular policy instrument in K-12 and higher education in the United States. Cash transfers have long played a prominent role in higher education policy in the U.S. (Dynarski, 2004). Examples include Pell Grants and the G.I. Bill at the federal level, state financed merit-aid scholarships such as the Georgia HOPE scholarship and the West Virginia PROMISE scholarship at the state level, and privately financed scholarships offered directly by institutions. In addition, school vouchers, or publicly financed scholarships for K-12 students to attend private schools (Wolf, 2008), have recently begun to take on a more prominent role in K-12 education policy. Motivated by low achievement, especially among the relatively poor, several cities and states have implemented voucher programs in the hopes of improving student achievement by allowing for better matches of students to schools as well as improving the general quality of K-12 education systems by introducing competitive forces into an otherwise monopolistic system. As of 2013, 12 states and the District of Columbia have implemented school voucher programs (The Friedman Foundation for Educational Choice, 2013). In all cases, the vouchers are conditional cash transfers, with program eligibility determined by a number of requirements such as low family income or evidence of a disability (The Friedman Foundation for Educational Choice, 2013).

This dissertation provides an investigation into the effects of two statewide conditional cash-transfer based educational interventions on participating students. The first program, the

Louisiana Scholarship Program (LSP), is a school voucher program offering publicly financed scholarships to low-income children in poorly performing public schools to attend K-12 private schools. The second program, the Arkansas Academic Challenge Scholarship (ACS), offers sizable college scholarships to incoming freshman and currently enrolled college students who have met a series of academic-based eligibility requirements. Both interventions fall under the broad category of conditional cash transfer programs because of their incoming and continuing eligibility requirements.[1]

The following pages provide rigorous investigations into the effects of these programs on student outcomes. Chapters 1 and 2 first investigate the experiences of eligible students applying to the Louisiana Scholarship Program in the first year of the statewide expansion of the program. Chapter 1 employs an experimental design to estimate the impact of LSP usage on student achievement on Louisiana's state assessments after one year. In contrast with prior experimental studies of K-12 school voucher programs in the U.S. which have largely found null or positive effects, the results presented in Chapter 1 suggest significant negative impacts for students induced to use the scholarship to attend private schools. Chapter 2 compliments Chapter 1 by providing a descriptive comparison of the development of non-cognitive skills[2] and levels of political tolerance among the first cohort of students in the statewide expansion of the LSP. Utilizing results from phone surveys administered to a subsample of eligible LSP applicants, we

[1] LSP is a conditional cash transfer program due to its eligibility requirements and because continued receipt is made conditional on enrollment in the private school. Currently enrolled college students were eligible for the ACS if they had been continually enrolled in up to 12 credit hours per semester since their initial matriculation and had a cumulative GPA of 2.5. ACS recipients would continue to receive an ACS scholarship if they successfully completed 15 credit hours a semester and maintained a cumulative GPA of 2.5.

[2] Non-cognitive skills has become a catch all phrase commonly used among economists to describe non-academic outcomes (West et al., 2014). Chapter 2 focuses on the non-cognitive skills of Grit (Duckworth, Peterson, Matthews, & Kelly, 2007), Locus of Control (Rotter, 1966), and Self-esteem (Rosenburg, 1965).

find no evidence of significant differences between students who received an LSP scholarship and those who did not receive a scholarship on any of the scales examined.

Chapter 3 provides a quasi-experimental analysis of the effects of the ACS on short- and long-run college outcomes among students who were enrolled at a Large Arkansas University (LAU) during a dramatic expansion of the program in the fall of 2010. In particular, we take advantage of detailed panel data on students along with knowledge of the ACS eligibility requirements to compare the outcomes of students who barely qualified for the scholarship with those of students who just missed the eligibility requirements in a fuzzy regression discontinuity design. In general, while the analyses presented in Chapter 3 admittedly suffer from concerns regarding statistical power, our models indicate that ACS receipt did not affect student outcomes in the short-run but did have significant negative effects on final GPA as well as the likelihood of graduating within four years. Nevertheless, our results also suggest that ACS recipients eventually caught up to their counterparts, as we find no evidence of significant differences in the likelihood of graduation after five years.

The evidence presented in this dissertation benefits the literatures on school vouchers, state financed higher education scholarships, and cash transfers more generally in several ways. First, Chapter 1 provides the first experimental analysis of a statewide K-12 voucher program in the U.S. In doing so, this analysis offers insight into how school voucher programs may affect student achievement when brought to scale. With an increasing number of states implementing such large-scale programs, this research is particularly timely.

Chapter 2 additionally benefits the school voucher literature by providing the first examination of the distribution of non-cognitive skills associated with a school voucher program. While a growing body of research demonstrates the importance of such skills for later life success (Almlund et al. 2011; Heckman & Kautz, 2012; Tangney, Baumeister, & Boone, 2004), evaluations of school voucher programs have generally restricted themselves to measures of student achievement, attainment, and satisfaction. While our results indicate no evidence of significant differences between the students receiving and not receiving an LSP scholarship, this work nevertheless expands the school voucher literature by providing a first attempt at examining the distribution of such skills among participating students.

Finally, Chapter 3 benefits the literature on state financed scholarships in two ways. First, Chapter 3 compliments exiting research by focusing on a program with relatively low eligibility requirements. Second, by focusing on students who were currently enrolled in college at the time of the program's expansion, as opposed to incoming Freshman, our study represents the first attempt to estimate the motivational effects of a merit-aid scholarship on student outcomes that is unaffected by any motivational effects occurring prior to enrollment in college.[3]

Taken together, the research presented in the following chapters represents important contributions to the understanding regarding the effectiveness of cash transfers as policy instruments in K-12 education.

[3] For example, the mere existence of merit-aid scholarships may incentivize students to work harder in high school to meet the program eligibility requirements (Dynarski, 2004).

References

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. (2011). Personality psychology and economics (NBER Working Paper No. 16822). Cambridge, MA: National Bureau of Economic Research.

Baird, S., Ferreira, F. H. G., Ozler, B., & Woolcock, M. (2013). Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review. Campbell Systematic Reviews, 8.

Baird, S., McIntosh, C., & Ozler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. The Quarterly Journal of Economics, 126, 1709–1753.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., & Pouliquen, V. (2013). Turning a shove into a nudge? A "Labeled Cash Transfer" for education (NBER Working Paper No. 19227). Cambridge, MA: National Bureau of Economic Research.

Dewey, J. (1916). *Democracy and education*. New York, NY: The Macmillian Company.

Dynarski, S. (2004). The new merit aid. In Hoxby, C. M. (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*. Chicago, IL: University of Chicago Press.

Goldin, C., & Katz, L. F. (2008). The race between education and technology. Cambridge, MA: Belknap Press of Harvard University Press.

Gutmann, A. (2003). Assessing arguments for school choice: Pluralism, parental rights, or educational results? In A. Wolfe (Ed.), *School Choice: The Moral Debate* (pp. 126–148). Princeton, NJ: Princeton University Press.

Handa, S., & Davis, B. (2006). The experience of conditional cash transfers in Latin America and the Caribbean. Development Policy Review, 24(5), 513–536.

Heckman, J. J. (2008). Schools, skills, and synapses. Economic Inquiry, 46(3).

Mills, J. N. (2013b). The achievement impacts of Arkansas open-enrollment charter schools. Journal of Education Finance, 38(4), 320–342.

National Center for Education Statistics. (2014). *Digest of education statistics* (Table 326.10.). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_326.10.asp

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Rotter, J. R. (1966). Generalized expectations for internal versus external control of reinforcement. Psychological Monographs: General and Applied, 80(1), 1–28.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72(2), 271–322.

The Friedman Foundation for Educational Choice (2013). The ABCs of school choice, 2013 edition. Indianapolis, IN: The Friedman Foundation for Educational Choice.

Todd, P. E., & Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in Mexico: using a social experiment to validate a dynamic behavioral model of child schooling and fertility. American Economic Review, 96, 1384–1417.

Turner, S. E. (2004). Going to college and finishing college: Explaining different educational outcomes. In College Choices: The Economics of Where to Go, When to Go, and How to Pay for It (pp. 13–62). Chicago, IL: The University of Chicago Press.

West, M. R., Gabrieli, C. F. O., Finn, A. S., Kraft, M. K., & Gabrieli, J. D. E. (2014). What effective schools do. Education Next, 14(4).

Wolf, P. J. (2008). Vouchers. The International Encyclopedia of Education (McCulloch, G., & Crook, D., eds.). London: Routledge, pp. 635-636.

World Bank. (2009). Conditional cash transfers: Reducing present and future poverty. Washington, DC: World Bank.

Chapter 1[4]

First Year Participant Effects of the Louisiana Scholarship Program

Introduction

The Louisiana Scholarship Program (LSP) is a statewide school voucher program, providing public funds for low-income students in underperforming public schools to attend participating private schools.[5] Originally piloted in New Orleans in 2008, the statewide expansion of the LSP program in 2012-13 allowed almost 5,000 low- to moderate-income students across the state of Louisiana to transfer out of their traditional public schools and into private schools at state expense. The empirical evidence presented here examines how the LSP has impacted student achievement in the first year of the statewide expansion.

Our analysis uses the results of the oversubscription lotteries for nearly 10,000 eligible applicants to analyze the achievement impacts of LSP as a randomized control trial (RCT). In particular, we estimate the effect of using an LSP scholarship to enroll in a private school for applicants to oversubscribed lotteries who were induced to attend a private school as a result of winning the lottery. Our analysis uses student-level data obtained via a data-sharing agreement with the state of Louisiana.

In general the results presented in this chapter indicate that the use of an LSP scholarship to enroll in private schools is associated with statistically significant—and substantively large—negative effects on student achievement. Specifically, LSP users are found to be nearly a quarter of a standard deviation behind their control group counterparts in English Language Arts and

[4] This paper was co-authored with Patrick J. Wolf and Jay P. Greene.
[5] The program was initially called the Student Scholarships for Educational Excellence Program but is now referred to as the Louisiana Scholarship Program.

two-thirds of a standard deviation behind in math. The magnitude of these negative estimates is unprecedented in the literature of random assignment evaluations of school voucher programs.

The remainder of this chapter proceeds as follows. In the next section, we define terms that are key to our analysis. Then we summarize the existing literature on random assignment evaluations of the participant effects of school voucher programs. After that, we provide a brief description of the LSP and the lottery process that enabled the experimental analysis. Next we describe the data and analytical strategy used to estimate the participant effects of the first year of the statewide expansion of the LSP. We then describe the results of our analyses and conclude with a discussion of our findings.

School Vouchers and K-12 Education

School vouchers are a mechanism by which government resources are provided to families that enable them to attend a private school of their choosing (Wolf, 2008). Strictly speaking, a private school choice program is only a "voucher" program if the government funds the program directly out of an appropriation. Other private school choice programs 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 scholarship programs. Since tax-credit scholarship programs accomplish the same general purpose as voucher programs we will treat both types of private school choice programs as functionally equivalent for purposes of this study, although we will specify whether individual initiatives are voucher or tax-credit scholarship programs when discussing them.

Although the origin of the voucher idea generally is linked to economist Milton Friedman (1955), political philosophers Thomas Paine (1791) and John Stuart Mill (1962 [1869]) supported the theoretical debate about their desirability. It would seem that high-quality research on the question is an imperative (Doolittle & Connors, 2001). For example, Richard Murnane (2005) 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. (p. 181)

Experimental design is critical in the case of evaluating school voucher programs because of concerns about selection bias due to more motivated and able families self-sorting into private schools on their own or through access to a voucher. Fortunately, much of the research on school vouchers in the U.S. has taken the form of random assignment experiments.

Prior Random Assignment Evaluations of School Voucher Programs

Prior rigorous empirical studies of the effects of school vouchers on participants' achievement, though generally modestly positive, have been inconsistent in their pattern of results and have yet to produce a scholarly consensus about the impacts of vouchers on students' academic outcomes (Wolf, 2008; Barrow & Rouse, 2008).

A total of 13 analyses have applied experimental or regression discontinuity design (RDD) methods to data from voucher and voucher-type scholarship programs in Charlotte, Dayton, the District of Columbia, Florida, Milwaukee, and New York to determine their impacts on student achievement. Both analyses of the Charlotte data reported that the scholarship program produced positive and statistically significant achievement impacts (Greene, 2001; Cowen, 2008). The experimental evaluation of the Dayton scholarship program concluded that it produced achievement gains, but only for the African American subgroup of participants (Howell et al., 2002). A single analysis of experimental data from an early scholarship program in the District of Columbia concluded that achievement gains from the program that were evident after two years disappeared in the third and final year of the evaluation (Howell & Peterson, 2006). The congressionally mandated evaluation of the District of Columbia Opportunity Scholarship (voucher) Program, established in 2004, reported achievement impacts, but only in reading, that were statistically significant at a 99 percent level of significance after three years (Wolf et al. 2009, p. 36) but only at a 94 percent level of significance in the fourth and final year of the study (Wolf et al. 2013; Wolf et al. 2010, p. 35). An RDD analysis of the tax-credit scholarship program in Florida concluded that students near the income eligibility cutoff experienced clear achievement gains in reading, but not necessarily in math, if they had access to the program (Figlio, 2011).

Two different analyses of experimental data from the early years of the Milwaukee voucher program reached slightly different conclusions, with one reporting that voucher students realized statistically significant achievement gains in both reading and math (Greene, Peterson, & Du, 1999) and the other stating that the voucher achievement gains were limited to just math (Rouse, 1998). Five different analyses of data from the New York scholarship experiment also reached somewhat divergent conclusions. One study reported no significant achievement gains from the scholarship program, overall or for any subgroup of participants (Krueger & Zhu, 2004). Two other analyses employing alternative methods for addressing missing data found program-induced gains, but only for African Americans in math (Barnard, Frangakis, Hill, & Rubin, 2003; Jin, Barnard, & Rubin, 2010). The original experimental analysis concluded that African American scholarship students outperformed the control group students on a combined measure of math and reading scores (Mayer et al., 2002).[6] Finally, Bitler, Domina, Penner, and

[6] Nevertheless, as Peterson and Howell (2004) note, Krueger and Zhu's insignificant subgroup findings appear to be driven in part by the particularly unique way in which they chose to classify students as African American.

Hoynes (2013) examine using quantile analysis if effects differed across the achievement distribution, finding little evidence of heterogeneous effects.

Since the pattern of results from previous experimental and RDD evaluations of voucher programs has ranged from neutral to positive, with no statistically significant negative impacts of vouchers on student achievement having been reported to date, our operating hypothesis is that the LSP will have a positive impact on student achievement.

Description of the Intervention

The Louisiana Scholarship Program (LSP) is a statewide school voucher program available to moderate- to low-income students in low-performing public schools across the state of Louisiana. Student eligibility for the scholarship program is determined by family income—which must not exceed 250 percent of the federal poverty line—and where the student previously attended public school. Income-eligible students must have attended a public school that was graded C, D, or F for the prior school year; be entering kindergarten; or have been previously enrolled in the Recovery School District in order to be fully eligible for the program. In the program's first year, 9,809 students were deemed to be fully eligible applicants, with a majority of them located outside of Orleans parish.

The LSP was created by Act 2 of the 2012 Regular Session of the Louisiana Legislature and Senate. Act 2 requires the state board to allocate the funds for the program annually from the minimum foundation program. The voucher size is the lesser of the amount allocated to the local school system in which the student resides or the tuition charged by the participating private school that the student attends. Average tuition at participating private schools ranges from $2,966 to $8,999, with a median cost of $4,925, compared to an average total minimum foundation program per pupil amount of $8,500 for public schools.

Private schools that wish to participate in the program must go through a screening process and schools that do not meet the required participation criteria are declared ineligible to participate in the program. There are four areas that are evaluated for determining school eligibility: (1) enrollment; (2) financial practice; (3) student mobility; and (4) health, safety and welfare of students. In the 2012-13 school year, only a third of eligible private schools participated in the program (Kisida, Wolf, & Rhinesmith, 2013). A recent survey of participating and non-participating private schools in Louisiana suggests that the program's regulatory barriers have strongly influenced schools' choices to participate (Kisida, Wolf, & Rhinesmith, 2013).

Research Methodology

Experimental Design

When the LSP was expanded to a statewide program in 2012, the Louisiana Department of Education also changed the lottery process determining scholarship awards. While the original application process in the New Orleans pilot version of the LSP allowed families to submit the name of only one private school for admission, the revised application process allowed individuals to offer up to five private school preferences. This difference was the result of switching to a lottery process similar to the deferred acceptance lotteries used in New York City to assign students to schools through the city's public school choice program (see Abdulkadiroglu, Pathak, & Roth, 2005). The deferred acceptance algorithm is designed to encourage families to reveal their true school preference rankings and thereby reduce the likelihood of gaming.

While it is not the case that all eligible LSP applicants were awarded scholarships through a lottery process in the 2012-13 school year, we can isolate cases in which lotteries occurred in order to perform an experimental evaluation of the program.

Specifically, eligible LSP applicants were allowed to submit up to five private school preferences and the LSP lottery algorithm attempted to place students into schools while taking into account several lottery priorities. First, students with disabilities and "multiple birth siblings"[7] are manually awarded LSP scholarships if there is available space at their given school preference. Remaining students are grouped into one of six priority categories:

- **Priority 1** - Students who received LSP scholarships in the prior school year who are applying to the same school
- **Priority 2** - Siblings of Priority 1 awardees in the current round
- **Priority 3** - Students who received LSP scholarships in the prior school year who are applying to a different school
- **Priority 4** – New applicants who attended public schools that received a "D" or "F" grade in Louisiana's school accountability system at baseline
- **Priority 5** – New applicants who attended public schools that received a "C" grade in Louisiana's school accountability system at baseline
- **Priority 6** – New applicants who are applying for kindergarten placements

The LSP award process is summarized in Figure 1. The process begins by attempting to place all Priority 1 category students into their first choice school. The algorithm first groups Priority 1 students applying to the same school and grade combination and then checks the number of available seats for that grouping. If there are more seats than applicants, all students receive an LSP scholarship. If there are no seats available, no students in the given group receive a scholarship. Finally, if there are more applicants than seats, students are awarded LSP

[7] "Multiple birth siblings" are twins, triplets, etc.

scholarships through a lottery. Once the process is complete for all Priority 1 students, the algorithm attempts to place Priority 2 students into their first choice school. After cycling through all remaining priority categories, the LSP algorithm attempts to place students who have yet to receive a scholarship in their second choice schools. The LSP algorithm continues until all eligible applicants have either been awarded or not awarded an LSP scholarship.

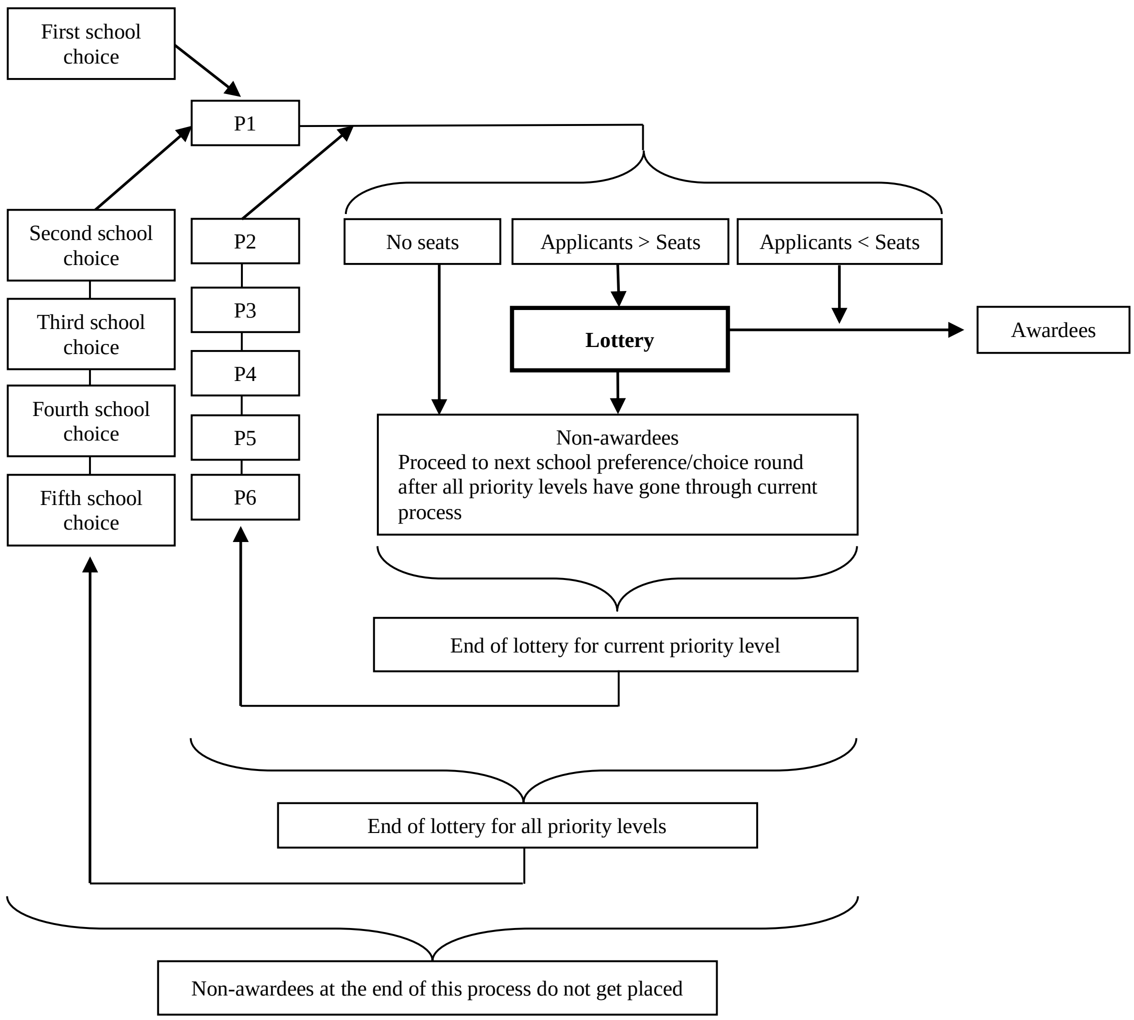

Figure 1. The Louisiana Scholarship Program award allocation process for the 2012-13 school year. This figure illustrates the iterative process used to allocate LSP scholarships to students. In addition, this figure highlights the fact that only a subset of students was awarded LSP scholarships via lotteries. Our analysis focuses on isolating lotteries for one's first choice school.

It is important to note that only a subset of eligible applicants were awarded or not awarded an LSP scholarship via a lottery process. Specifically, only those students in priority categories one through six whose school-grade combination had more applicants than available seats participated in a lottery. Fortunately, using data on student characteristics and school preferences, we can identify the subset of eligible applicants who experienced a lottery process.[8] We will focus on this subset of lottery participants to estimate the effects of the LSP on student achievement after one year of program participation because these are the only applicants for whom LSP scholarship award was randomly determined.[9]

Data Description

Most of the data for this study come from student-level datasets provided by the Louisiana Department of Education (LDE) in compliance with our data agreement with the state. The LDE had provided us with their:

- Student Information Systems (SIS) files for 2011-12 ("Baseline") and 2012-13 ("Year 1 Outcome") which includes data on student enrollment and demographic background;
- LSP eligible applicant file, which includes information on the school choice sets of all eligible applicants as well as the results of the 2011-12 placement lottery;
- State assessment files for the 2011-12 and 2012-13 school years, which include data on each student's participation in the annual accountability assessments and their scores.

The LDE also provided information on participating public and private schools, and this information is supplemented with publicly available data from the National Center for Education

[8] We infer that a lottery has occurred when the LSP award percentage is between 0 and 100 percent for a given school preference by grade by priority category combination. For example, if 60 percent of Priority 1 category students applying to third grade at school "A" as their first choice school actually received scholarships, we assume that all students in that combination were subject to a lottery.

[9] After accounting for student testing, demographic, and school data while also limiting our analysis to students in binding lotteries, approximately 7 percent of the analytical sample were in Priority Category 1, less than 1 percent were in Priority Category 3, 74 percent were in Priority Category 4, and 18 percent were in Priority Category 5. Priority Category 2 is not represented in our analysis.

Statistics (NCES) Common Core of Data (CCD) and Private School Universe Survey (PSS) when necessary.

The Louisiana state accountability system places a strong emphasis on test-based accountability, with standardized assessments offered in most grades—including alternative assessments, end-of-course exams, and exams measuring college-readiness. The participant effects portion of this study uses student performance on the Louisiana state assessments in grades three through eight as our primary outcome measure of interest.[10] All students participating in the LSP are required to be tested by their private schools, using the state accountability assessments, for any grade in which the public school system also tests its students.

Students in Louisiana who are not classified as having a special need that qualifies them for alternative programs take one of two state assessments in grades three through eight. In grades four and eight, students take the Louisiana Educational Assessment Program (LEAP) exams, a series of criterion-referenced tests aligned with Louisiana's state standards for the subjects of math, English language arts, science, and social studies. In the remaining grades, students take the Integrated Louisiana Educational Assessment Program (iLEAP) exams, a series of hybrid exams including both criterion-referenced and norm-referenced test items in the same subjects as the LEAP exams. Performance on both sets of exams ranges between a minimum possible score of 100 and a maximum possible score of 500. All exams are scaled with means of

[10] Our initial investigations of the test databases revealed 391 eligible LSP applicants in tested grades with missing testing data at baseline and 516 observations in year 1. These observations represent approximately 10 and 15 percent of the eligible LSP applicants in relevant grade ranges for these years. Further investigation revealed that 82 of the missing observations took the Louisiana alternative assessments at baseline and 115 took them in year 1. Given our requirements for baseline equivalence, all records with missing baseline testing data are excluded from our analysis.

300 and standard deviations of 50 (Louisiana Department of Education, 2013a; 2013b). Rather than rely on these scale score values, which differ across grades by design, our analysis is primarily based on standardized values of individual LEAP and iLEAP performance.[11] While the LEAP and iLEAP item differences introduce noise into our model, the fact that both treatment and control students in a particular grade take the same exam (either LEAP or iLEAP), and our inclusion of prior achievement on the right-hand side of the model, should reduce the likelihood of bias due to these test differences.

There are a small number of eligible LSP applicants with duplicate records in the baseline (12 duplicate pairs) and year 1 testing data (42 duplicate pairs). When possible, we have resolved duplicates by keeping records with the most complete data on LSP participants. For the remaining observations, we have randomly kept one record and dropped the other. These records represent less than 1 percent of the LSP applicants in both years.

In addition to individual performance outcomes, the state-provided assessment data files include information on student demographics as well as participation in school programs such as the free- and reduced-price lunch (FRL) program and special education program enrollments. Our analysis includes these baseline covariates in order to improve effect estimate precision.[12]

Sample Selection Process

The student-level data provided by the LDE indicate an initial sample of 9,809 eligible LSP applicants in the first year of the statewide expansion of the program. Of these, 5,777

[11] Values have been standardized within grade and year based on the observed means and standard deviations for control group members on the ELA and math portions of the iLEAP and LEAP.

[12] A single individual in our final analysis sample has missing data for their gender status as baseline (2011-12). We have updated this individual's gender status using their reported gender in the 2012-13 assessment data. After making this substitution, all records in our final analysis sample have complete information on baseline covariates.

students received LSP scholarship placements in a specific private school and 4,038 did not receive a voucher-supported placement. Our analysis relies on a non-random sample of this original population comprised of eligible applicants with baseline testing data in grades three through seven who did not list a special education exclusion on their application and who were not multiple birth siblings. Of the 2,897 observations meeting these criteria, we identify 1,908 individuals as participating in LSP scholarship lotteries. Of these, 727—or 38 percent—won LSP scholarships.

Analytical Strategy

This section outlines the empirical strategy we use to estimate the impacts of LSP scholarship usage on student achievement after two years. We begin with a description of our primary analyses, which use the results of eligible applicants' first school choice lotteries to estimate the impact of LSP scholarship usage on student achievement in a two-stage least squares (2SLS) framework. We then outline a series of subgroup analyses conducted to examine possible effect heterogeneity of the LSP scholarship program.

Local Average Treatment Effect estimation. As Bloom and Unterman (2014) note, because students can participate in multiple lotteries in a deferred-acceptance award process, the traditional intent-to-treat estimator has limited policy relevance.[13] Instead, we intend to estimate the impact of LSP scholarship usage on student achievement—also known as the Local Average Treatment Effect (LATE) (Angrist & Pischke, 2009, Cowen, 2008)—by using the result of one's first choice school lottery as an instrumental variable to predict scholarship usage in a 2SLS framework. The lottery is the ideal instrumental variable as the high placement take-up rate for this program ensures that it is will be a strong predictor of private schooling while the random

[13] For example, a student who loses her first lottery can still win an LSP scholarship to her second choice school via lottery.

nature of the lottery process assures that scholarship receipt is uncorrelated with the error term. In addition, because students could only receive a chance to use an LSP scholarship to enroll in a private school through the lottery, we can be confident that the variable only influences student outcomes through the private schooling that it enables.

Specifically, we use the following 2SLS model to estimate the effects of LSP scholarship usage on student achievement after two years:

1. $E_i = \sum \pi_j R_{ji} + \delta T_i + \boldsymbol{X\beta} + u_i$

2. $Y_i = \sum \alpha_j R_{ji} + \tau \widehat{E_i} + \boldsymbol{X\_i\gamma} + \epsilon_i$

Where i denotes student and j denotes lottery:

- *E* is a variable indicating the number of years a student used an LSP to enroll in a private school[14]
- *R_i* is a fixed effect for a student's first choice school lottery[15]
- *T_i* is a variable indicating if a student received an LSP scholarship to their first choice school
- *Y_i* is student standardized math or English Language Arts achievement in year 1 of the program (2012-13)[16]

[14] Although some prior studies have used "ever attended private school" (Wolf et al., 2010, Appendix D) or "consistently enrolled in private school" (Mayer et al., 2002) as the endogenous treatment "dosage" variable instrumented for, we follow the approach used by Rouse (1998) and employ the lottery to instrument for total years of private schooling as a more global and precise treatment dosage measure.

[15] We include a fixed effect for first school choice lottery to account for differing probabilities of success across lotteries (Gerber & Green, 2012). By using fixed effects, we are essentially comparing lottery winners and losers within the same strata to calculate unbiased estimates of the effect of being randomly offered an LSP scholarship. The approach is comparable to analyzing the impact of hundreds of "mini-experiments" and aggregating the results across them.

[16] Student achievement scores are standardized using distributional parameters of outcomes from the control group.

- *X_i* is a vector of student characteristics—including achievement—collected at baseline (2011-12)

The 2SLS procedure first uses one's treatment status to predict the number of years they will actually use a scholarship; and then uses this predicted value to provide an unbiased LATE effect estimate ($\hat{\tau}$) for the program. It is important to note that the proposed method instruments for LSP usage using the result of one's first choice school lottery outcome (T). This implies that the 2SLS procedure will effectively treat students who lose their first choice lottery but go on to win an LSP to a lower school preference as control-group cross overs (Bloom & Unterman, 2014).

In addition, it is important to note that there are at least two types of nesting in the LSP data that can lead to biased inference (Angrist & Pischke, 2009). First, members of both the treatment and control group are nested within schools in the first year of the program analysis. Second, observations can be nested within family units, with the potential for several children participating from the same family.[17] This was also the case in the evaluation of the Opportunity Scholarship Program (OSP) in Washington, DC, in which researchers used standard errors clustered at the family level to account for error-covariance (Wolf et al., 2013). Unfortunately, the results presented here do not account for these types of nesting due to the complex nature of multi-level clustering. Instead, we currently only account for nesting of observations within risk set.[18] We intend to address this issue in future versions of this research.

Given our dataset, there are two broad methods available to account for non-response bias: employing non-response weights or making assumptions about the nature of non-response

[17] Approximately 23 percent of individuals in our final analytical sample have siblings that also appear in the sample.
[18] Clustering on risk set should capture a large amount of the nesting of individuals within current school as risk set includes school of application.

to estimate bounds around the program's true effect (Gerber & Green, 2012). Nonresponse weights effectively reweight the data to allow respondent values to account for the values of non-respondents (Kreuter & Valliant, 2007). Nevertheless, it is important to note that nonresponse weights do not account for potential unobservable forces that may be driving patterns of nonresponse. If, for example, those in the control group with higher expected outcomes both in public and private school leave the sample with higher probability, our LATE estimates will be positively biased. Given the likelihood that the observed control non-response reflects potential selection effects, we prefer instead to estimate the degree to which attrition affects our estimates via a bounding exercise (Angrist, Bettinger, Kremer, 2006; Lee, 2009). Specifically, if we assume that the causes of missing data are monotonic[19], we can estimate an upper and lower bound for the LSP effect by omitting a portion of the control group from the data in order to balance non-response probabilities among treated and controls. While our primary estimates of the effects of LSP scholarship usage on student achievement after one year do not account for differential attrition, we include as a robustness check estimates based on the bounding exercises. In general, the results from our bounding analyses do not suggest that differential attrition has strongly influenced our primary LATE estimates.

Subgroup analysis. In addition to examining overall program impacts, we examine the extent to which LSP program effects differ across different subgroups. In particular, we examine if there are differential impacts experienced for three subgroups: (1) males relative to females, (2) African American participants compared to all other program participants, and (3) New

[19] The monotonicity assumption requires that treatment assignment influences data missingness in one unique direction (Gerber & Green, 2012). This assumption seems plausible in our case, as we are most concerned with sample attrition resulting from control group students attending private schools on their own accord and it is highly unlikely that a student who won a scholarship to their first choice school chose to attend an alternative private school.

Orleans participants compared to other participants. The first two subgroup analyses are motivated by prior evaluations of school choice programs, which have found differential effects by gender and ethnicity groups.[20] The final subgroup comparison is motivated by the strong existing market for school choice in New Orleans in comparison with the rest of the state. In addition to having a pilot version of the LSP in place since 2008, New Orleans has a thriving charter school market and a history of public school choice (Cowen Institute, 2013). As such, it is highly likely the experiences of LSP participants in New Orleans are substantially different than other participants; and we should therefore examine the extent to which these differences translate into our estimated program effects.

Finally, in addition to the subgroup analyses described, we examine if there are heterogeneous treatment effects by baseline achievement category. Unfortunately, we are unable to use the same analytical strategy to estimate performance subgroup effects using the interaction models outlined above while also controlling for baseline achievement. Instead, we run our primary regression models on three separate subgroups of students: those in the bottom third of the baseline achievement distribution, those in the middle third, and those in the top third. While this method does not allow for a statistical test of the differences between the estimated treatment effects across the three models, it does provide descriptive insight into the extent to which students in different parts of the achievement distribution experienced the program differently.[21]

[20] Analyses of the New York Scholarship Program have found significant effects for African Americans, but insignificant effect estimates overall (Mayer et al., 2002; Barnard et al., 2003; Chingos & Peterson, 2013). Similarly, Wolf and colleagues (2013) report significant improvement in reading for female participants in the DC OSP evaluation, but no significant differences for males.

[21] Bitler and colleagues (2013) focus on differences in treatment effects across the achievement distribution associated with the New York City School Choice Scholarships Program. The results of their quantile treatment effects analysis provide limited evidence of heterogeneous effects.

Treatment-Control Contrast

Before moving on to our primary analyses, we examine two features of the data that have implications for our analysis. First, we examine the extent to which treatment assignment is correlated with school enrollment by looking at school enrollments for lottery winners and losers. Next, we examine the extent to which the lottery process resulted in covariate balance at baseline for our analysis sample.

Scholarship Usage

While eligible applicants were randomly assigned to receive or not receive an LSP scholarship with private school placement, participating families are not required to use the scholarship to attend their designated school. Lottery winners, for example, could choose to attend traditional public schools or charter schools rather than use their scholarship offer to attend the private school in which they were placed by the lottery. Lottery losers, on the other hand, could choose to attend charter schools or elect to enroll in private schools without a scholarship in addition to attending a traditional public school.

Table 1 describes the patterns of enrollment for students who received and did not receive LSP scholarships to their first choice schools in the first year of the analysis. It should be noted that, because our LATE analysis focuses on the results of first choice school lotteries, the control group includes both students who were never awarded a scholarship and students who received a scholarship to one of their remaining school preferences. The latter group, who account for the 60 control group members appearing in private schools in 2012-13, are effectively treated as control-group crossovers in our LATE analysis.

The results in Table 1 indicate that the majority of lottery winners used their scholarships to attend private schools, while 85 percent of students who did not receive scholarships attended public sector schools. In addition, it appears that control group contamination should not strongly impact our analysis, as only 5 percent of students who did not receive a scholarship to their first choice school are identified as enrolled in private schools in 2012-13. Nevertheless, it is also important to note that over 10 percent of lottery losers do not appear in our testing data in year 1. In contrast, only 5 percent of lottery winners are missing year 1 outcome data. These missing outcome observations could represent control-group crossovers attending private schools or students that moved out of Louisiana in the 2012-13 school year. Unfortunately, our reliance on the state testing data does not allow us to distinguish the causes behind these missing data.

Table 1

Year 1 Enrollment Patterns by Scholarship Award

	Treatment Group		Control Group		
	Received Scholarship to First Choice School		Did Not Receive Scholarship to First Choice School		Total
	N	%	N	%	
Private School	550	76%	60	5%	610
Public School	143	20%	1,005	85%	1,148
Unknown School	34	5%	116	10%	150
Total	727	100%	1,181	100%	1,908

Notes. Sample represents all students with baseline testing data in grades three through seven who did not list a special education exclusion on the LSP application and who were not identified as multiple birth siblings. For students in the treatment group attending public schools, 72 percent attended a traditional public school, 25 percent attended a charter school, and 3 percent attended a magnet school. The corresponding percentages for the control group members attending a public school are: 74 percent attended a traditional public school, 16 percent attended a charter school, and 9 percent attended a magnet school.
Source. Authors' calculations.

Finally, it should be noted that our outcome data response rate of nearly 90 percent overall for the analytic sample is much higher than the outcome response rate for any previous experimental evaluation of a voucher program, giving us some confidence that any nonresponse bias is trivial. Nevertheless, the differential rates of non-response between the treatment (5 percent) and control (10 percent) groups are concerning (What Works Clearinghouse, 2014). While our primary results will not account for these differences, we examine the extent to which differential attrition biases our results using two separate bounding procedures. In general, the results of these analyses do not suggest the estimated results are substantially influenced by differential attrition.

Baseline Equivalence

The final step required before moving on to our empirical analysis of the participant effects of the statewide expansion of the LSP is to analyze the extent to which the LSP lottery process actually ensured independence of assignment to the treatment and control groups. While we cannot know the extent to which members of the treatment and control group differ on unobservable characteristics, we can get a good idea of the success of the lottery process by examining if there is baseline equivalence in observable characteristics between lottery winners and losers. The results of this analysis are presented in Table 2, which provides the results of t-tests for differences in means on key baseline covariates between members of the treatment and control groups included in our analysis sample, with $p < .10$ as the lowest threshold of statistical significance.[22]

The results presented in Table 2 are favorable for our analysis, as nearly all of the estimated differences between lottery winners and losers are statistically insignificant, suggesting that we have adequately identified lotteries in our analytical sample. The sole exception is that lottery winners provided significantly fewer school preferences on average than lottery losers.[23]

[22] All analyses include fixed effects for one's first school choice lottery to account for different probabilities of selection.

[23] While our primary models do not explicitly account for these differences, we perform a series of robustness checks examining the extent to which our findings are sensitive to the number of school choices offered. In general, the estimated effects in models controlling for the number of school preferences offered are not distinguishably different from our primary model findings.

Table 2

Baseline Equivalence of Treatment & Control Groups on Covariates

	Lottery Winners (N=693)	Lottery Losers (N=1,073)	Diff.	p
Female	0.51	0.51	0.00	0.99
Race/Ethnicity				
African American	0.90	0.91	-0.01	0.57
Hispanic	0.02	0.01	0.01	0.40
White	0.07	0.06	0.00	0.97
Other	0.02	0.01	0.00	0.70
Free-or-Reduced Price Lunch	0.01	0.01	0.00	0.93
Limited English Proficiency	0.86	0.87	-0.01	0.47
Number of School Preferences Listed	2.14	2.32	-0.18***	<0.01
Standardized Performance†				
ELA Scale Score	-0.38	-0.37	-0.02	0.71
Math Scale Score	-0.41	-0.46	0.05	0.30
Science Scale Score	-0.50	-0.51	0.00	0.94
Social Studies Scale Score	-0.42	-0.42	0.00	0.98

*** - p<.01, ** - p<.05, * - p<0.10

† Scores are standardized within grade based on the observed distributions of scale scores across Louisiana.

Notes. The analysis sample excludes students with disabilities, multiple birth siblings, and individuals without baseline testing data in grades three through seven. All analyses include fixed effects for one's first school choice lottery.

Source. Authors' calculations.

Results

This section presents the results of our preliminary analyses of the first year impacts of the statewide expansion of the LSP on student achievement.

Primary Estimates of the Impact of Using an LSP on Student Achievement

The primary results of our preliminary LATE analyses are presented in Table 3. Columns (1) through (3) present the results based on ELA performance and columns (4) through (6) present the results for math. In addition, it should be noted that the results from the first stage regressions for the models presented in Table 3 (see Table A1 in the Appendix) all indicate that

LSP scholarship receipt is highly predictive of usage.[24] These findings support our claim that our results represent unbiased estimates of the effect of LSP scholarship usage on student achievement after one year.

In general, LSP scholarship users are found to score significantly worse than their control group counterparts on the state's ELA and math exams. In particular, students who used an LSP scholarship to enroll in private schools are found to score about a quarter of a standard deviation behind their control group counterparts in ELA achievement and over a two-thirds of a standard deviation behind in math per year of private school attendance.[25] These estimates are statistically significant, substantial, and unprecedented in the literature.

The performance of the baseline covariates in the regressions provides us with some confidence that the estimation of the voucher impact is valid and reliable. Students' baseline achievement scores are highly predictive of their outcome scores, at a level comparable to or even better than those observed in previous analyses (e.g. Howell et al., 2002). Female students outperformed males, in both ELA and math, all else equal. White and Hispanic students tend to score higher than African American students in ELA; and FRL eligible students tend to perform worse than their counterparts in math. While the majority of these estimates are not statistically significant, it is important to see that they are largely in the expected direction.

[24] In particular, LSP scholarship receipt is associated with a 70 percent increase in the likelihood of using a scholarship to enroll in a private school across all models. In addition, the associated joint F-statistics are substantially greater than Staiger and Stock's (1997) recommended threshold for instrumental variable relevance.

[25] The results presented in Table 3 do not appear to be driven by a dramatic achievement gains in the control group relative to the treatment group (see figures A1 and A2 in the Appendix).

Table 3

Estimated Effects of LSP Enrollment Exposure on Student Achievement

	ELA			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Enrollment Exposure	-0.21***	-0.22***	-0.22***	-0.65***	-0.65***	-0.65***
	(0.07)	(0.07)	(0.08)	(0.09)	(0.09)	(0.09)
Baseline Achievement	0.66***	0.65***	0.65***	0.63***	0.62***	0.62***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Female		0.15***	0.15***		0.10**	0.10**
		(0.04)	(0.04)		(0.04)	(0.05)
FRL		-0.18**	-0.18***		0.03	0.03
		(0.08)	(0.07)		(0.07)	(0.07)
Race/Ethnicity†						
Hispanic		0.13	0.13		0.25**	0.25*
		(0.18)	(0.17)		(0.11)	(0.14)
White		0.11	0.11		0.24***	0.24***
		(0.09)	(0.09)		(0.07)	(0.06)
Other		-0.04	-0.04		0.04	0.04
		(0.15)	(0.11)		(0.21)	(0.17)
In New Orleans School			0.16*			0.04
			(0.09)			(0.10)
N	1,766	1,766	1,766	1,766	1,766	1,766
Risk Sets	194	194	194	194	194	194
Within R-squared	0.42	0.43	0.43	0.39	0.39	0.39
Between R-Squared	0.44	0.42	0.41	0.30	0.30	0.31
Overall R-Squared	0.42	0.42	0.42	0.38	0.38	0.38

*** - p<.01, ** - p<.05, * - p<0.10

† Excluded comparison group is African American students

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

At the same time, it is important to note that these results are based on a subset of students (approximately twenty percent of all eligible LSP applicants). Furthermore, given that the observed grades are not typical entry grades for schools, these effects may not be representative of the experiences of participants to the all LSP participants. Nevertheless, these caveats aside, the results presented in Table 3 indicate sizeable negative one-year achievement impacts, especially in math, associated with winning an LSP scholarship.[26]

Subgroup Analysis

In addition to estimating the general impacts of participation in the LSP on student achievement, we are interested in how various student subgroups respond to the treatment. Table 4 presents LATE estimates for three subgroup comparisons: females versus males, African Americans compared to other students, and LSP scholarship users enrolling in New Orleans schools compared to students enrolling in private schools in other locations.[27] Joint F statistics from first stage regressions predicting LSP usage and an interaction of LSP usage and a subgroup identifier are presented along with the overall results. Each of the reported F-statistics suggests that the LSP scholarship receipt is a relevant predictor of usage.

In general, the results presented in Table 4 do not reflect substantial heterogeneous effects among the examined subgroups of students. In general, while the point estimates are slightly different for each subgroup, the results suggest that LSP scholarship users performed significantly worse than their counterparts in both ELA and math, irrespective of their backgrounds.[28] A single exception is observed in the New Orleans subgroup analysis, which suggests that students using the scholarship in New Orleans experienced significantly more negative math impacts than other students. Nevertheless, it is important to note that these results

[26] While our analysis focuses on differences in ELA and math achievement, we have confirmed that these large negative effects are equally present in both science and social studies achievement. In particular, the results presented in Appendix Table A2 indicate that eligible applicants offered an LSP scholarship were about a third of a standard deviation behind their control group counterparts in science and over forty percent of a standard deviation behind in social studies.

[27] The results are based on the models that include terms interacting predicted LSP exposure with the particular subgroup of interest.

[28] It is interesting to note that our results suggest that African American LSP users experience worse performance impacts than other LSP users when one considers prior RCT evaluations of school voucher programs (see, for example, Howell, Wolf, Campbell, & Peterson, 2002).

are likely reflective of a collinearity issue because risk set fixed effects include one's application school. By including risk set fixed effects, we are effectively limiting our comparisons to students in the same priority category applying to the same school and grade. Our regression results are therefore based on within-risk set variation, or cases in which there is sufficient variation within the risk set group in the given characteristics. The New Orleans subgroup analysis, for example, is based solely on those risk sets that include both students living in New Orleans and those not living in New Orleans in 2011-12. Given that individuals are highly likely to choose a private school in their parish, it is unlikely that many risk sets are contributing to the subgroup estimates for New Orleans. We therefore strongly recommend against generalizing them to New Orleans as a whole.

Finally, we examine the extent to which the estimated effects vary across the LSP applicant baseline achievement distribution. It would be particularly concerning, for example, to find that the negative effects are largely captured by students who were already among the worst performers at entry. The results presented in Table 5 provide a descriptive examination of the different ways students in the bottom, middle, and top third of the baseline achievement distribution experienced the LSP after one year. Specifically, Table 5 presents results of regressions run separately on these three performance subgroups. Column 1 presents the subgroup sample size; column 2 provides descriptive information for the control group's year 1 performance distribution for the selected sample; column 3 presents results for regressions that only control for baseline achievement and student risk set; and column 4 presents results for fully specified models additionally controlling for student demographics and residence.

Table 4

Estimated Effects of LSP Usage by Subgroup

	ELA			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Gender Subgroup Analysis						
Female	-0.22***	-0.22**	-0.23***	-0.65***	-0.65***	-0.65***
	(0.08)	(0.09)	(0.08)	(0.11)	(0.11)	(0.10)
Male	-0.20**	-0.21*	-0.21**	-0.65***	-0.66***	-0.66***
	(0.08)	(0.11)	(0.08)	(0.11)	(0.12)	(0.11)
Difference	-0.01	-0.02	-0.02	0.01	0.01	0.01
	(0.10)	(0.12)	(0.09)	(0.11)	(0.12)	(0.10)
First stage F						
Exposure	507.3	255.6	227.9	505.8	254.7	227.0
Interaction	881.1	443.9	395.5	881.7	444.1	395.6
Race/Ethnicity Subgroup Analysis						
African American	-0.22***	-0.22***	-0.23***	-0.68***	-0.68***	-0.68***
	(0.07)	(0.08)	(0.08)	(0.10)	(0.09)	(0.11)
Not African American	-0.15	-0.14	-0.15	-0.44***	-0.44***	-0.44***
	(0.19)	(0.16)	(0.17)	(0.12)	(0.16)	(0.16)
Difference	-0.07	-0.08	-0.08	-0.23	-0.24	-0.24
	(0.19)	(0.15)	(0.16)	(0.14)	(0.15)	(0.16)
First stage F						
Exposure	507.0	340.8	292.9	505.5	339.5	291.7
Interaction	541.2	362.0	310.3	539.8	360.9	309.3
New Orleans Subgroup Analysis						
New Orleans Student	-0.28	-0.29		-1.00***	-1.00***	
	(0.23)	(0.22)		(0.32)	(0.22)	
Other Student	-0.21***	-0.21**		-0.61***	-0.61***	
	(0.07)	(0.09)		(0.09)	(0.11)	
Difference	-0.08	-0.08		-0.40*	-0.39*	
	(0.21)	(0.25)		(0.23)	(0.24)	
First stage F						
Exposure	522.6	234.1		521.1	233.2	
Interaction	364.9	161.8		360.9	159.9	
Model Summary						
Demographic Controls		X		X		X	X

New Orleans Control			X		X	X
N	1766	1766	1766	1766	1766	1766
Risk Sets	194	194	194	194	194	194

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

The results presented in Table 5 suggest that students in the bottom third of the ELA achievement distribution at baseline performed no differently than control group students after one year. On the other hand, the LATE estimates for the middle and top performers largely mirror the overall significant negative finding presented in Table 3.

Table 5

Estimated Effects of LSP Usage by Baseline Performance Group

Baseline achievement group	N	Control group mean/sd	LSP effect without covariates	LSP effect with covariates
	(1)	(2)	(3)	(4)
ELA achievement				
Bottom third	591	-0.788	0.11	0.09
		(0.97)	(0.13)	(0.15)
Middle third	596	0.021	-0.27**	-0.30**
		(0.66)	(0.13)	(0.14)
Top third	579	0.721	-0.28***	-0.26**
		(0.73)	(0.10)	(0.12)
Math achievement				
Bottom third	608	-0.651	-0.71***	-0.73***
		(0.68)	(0.21)	(0.19)
Middle third	582	-0.062	-0.60***	-0.61***
		(0.23)	(0.15)	(0.15)
Top third	576	-0.765	-0.68***	-0.67***
		(0.73)	(0.14)	(0.13)

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Results presented in column 3 are based on models that only control for student baseline achievement. Results presented in column 4 are based on models additionally including controls for demographics and being a New Orleans resident. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

In contrast, the findings for math suggest that all groups experienced significant declines in achievement when using an LSP scholarship. While the coefficient estimates indicate that students in the bottom and top thirds of the math achievement distribution at baseline performed relatively worse than those in the middle, it is important to note that these estimates are particularly noisy—which is to be expected given the smaller sample sizes—and the 95 percent confidence intervals for each of the regressions largely overlap.

Thus, in general, the results presented in Table 5 suggest that students performing poorly in ELA at baseline did not experience large achievement declines after using an LSP scholarship and the negative effects of the program were largely experienced uniformly among remaining students.

Robustness Checks

In general, our analyses indicate that participation in the first year of the statewide expansion of the LSP negatively impacted student achievement on Louisiana's state assessments. These negative findings are unique among random assignment evaluations of school voucher programs, all of which have found insignificant or positive outcomes. Given the uniqueness and magnitude of these estimated effects, especially in math, it is important to test if the results are sensitive to our chosen analytic strategy. This section presents the results from sensitivity analyses designed to test the robustness of our findings.

Sensitivity of results to differential attrition. Our first robustness check examines the extent to which our estimated effects are sensitive to the different rates of attrition observed between treatment and control group members in our sample. Specifically, we find that 108—or 9.1 percent—of students who did not win an LSP to their first choice school do not appear in the state's assessment data in 2012-13; whereas only 4.4 percent of LSP winners are missing. While this difference is not cause for great concern (What Works Clearinghouse, 2014), it is important to consider if differential attrition is driving our primary findings.

If we can assume the observed differences in attrition are due to random factors, our LATE estimates are generally less precise but are not biased by differential attrition (Gerber &

Green, 2012). On the other hand, if the observed differences are due to systemic, yet unobservable, sample selection effects, our primary estimates of the effect of using an LSP scholarship on student achievement are biased (Gerber & Green, 2012; Lee, 2009). If, for example, those in the control group with higher expected outcomes both in public and private school leave the sample with higher probability, our LATE estimates will be positively biased.

In this section, we examine the extent to which differential attrition may be biasing our results by using two bounding strategies. The first, developed by Lee (2009) and hereafter referred to as "Lee Bounds", involves removing a subset of applicants from the treatment group in an attempt to parse out marginal individuals who have selected into the sample only because they received an LSP scholarship.[29] In particular, Lee shows that if you can assume that problematic attrition is only present in either the treatment or control group, then you can effectively bound the average treatment effect for individuals whose treatment status does not influence their sample selection likelihood by trimming away from that group a percentage of applicants equal to the attrition difference from the bottom and top performers. These trimming procedures produce upper and lower bounds of the effect, respectively.

One of the primary benefits of Lee's bounding method is that it does not require strong assumptions on the selection mechanism producing the attrition problems beyond the assumption that the effect is only present in either the treatment or control groups. For example, one need not assume that control group attritors are either more- or less-academically able than students who

[29] Lee's (2009) bounding method is built on two assumptions: that the assignment mechanism is random and that sample selection is a monotonic function of treatment status. The first assumption is easily satisfied by the LSP lottery process. The second assumption essentially requires that there are no LSP applicants who were assigned an LSP scholarship but decided to forgo their scholarship and instead enroll in a private school at their own expense. While we cannot validate this assumption empirically, it seems highly unlikely that such "defiers" exist in our data—especially given the program's income threshold.

actively choose to remain in the sample. Nevertheless, the simplicity of Lee's method comes at a cost: Lee bounds can be quite large—especially in the presence of large differences in nonresponse rates. In contrast, Angrist, Bettinger, and Kremer (2006)—hereafter "ABK"—note that one can estimate tighter bounds of a program's effect on individuals who would always select into the sample by assuming that attrition is more likely to come from one of the tails of the outcome variable. For example, if we assume that the differential attrition rate we observe in our control group is largely explained by more advantaged families opting to enroll in private schools at their own expense, we can use the ABK method by focusing on estimating effects at the lower end of the observed outcome distribution.[30] Effectively, the ABK method assumes that individuals whose sample selection decision is conditional on their treatment status are likely to be present around a particular margin in the outcome distribution; and we should therefore focus on the other side of the outcome distribution to bound the estimated effect for the subgroup of students who would always select into the sample.

Table 6 presents both the original LATE estimates produced in Table 3–included as a reference—as well as results from the two bounding exercises described. Columns 2 through 4 present models controlling only for baseline achievement and risk set while columns 5 through 7 present models that additionally include controls for demographics and residence.

As expected, the Lee bounds presented in Table 6 are quite large, with gaps of over 30 percent of a standard deviation in achievement. Despite the magnitude of these gaps, the results for math are consistent with LSP scholarship usage having a negative effect on achievement. In

[30] Simple comparisons of baseline characteristics indicate that control-group attriters performed slightly worse in both math and ELA on average relative to the control group average. While this suggests that we should instead focus on the top end of the performance distribution, we cannot rule out the possibility that these applicants had stronger growth potential than other control group students.

contrast, the results for ELA suggest that—in a best case scenario—LSP scholarship usage may have had an insignificant effect on student achievement after one year. Given the magnitude of the estimated effect, along with the knowledge that these estimates are based on removing the lowest performers from the treatment group, we are hesitant to conclude that the overall effect of the LSP on ELA achievement was null in the first year of the statewide expansion. Nevertheless, if this assumption is met, the results presented in Table 6 suggest the possibility that the two groups did not differ in ELA achievement after one year.

The ABK bounding procedure yields similar results for math: for the most part all lower and upper ABK bounds indicate a large negative effect of LSP scholarship usage on math achievement after one year. In contrast, the ABK results for ELA are insignificant in all cases. While this is likely due to the substantially smaller sample sizes used to construct ABK bounds, this finding does correspond with the earlier subgroup analysis indicating that students with lower ELA achievement at baseline did not tend to experience large declines in performance associated with LSP scholarship usage.

In general, the results presented in this section do not suggest that differential attrition has strongly biased the primary results presented in Table 3. Specifically, unless we make fairly restrictive assumptions, LSP scholarship usage continues to be associated with negative impacts in both math and ELA achievement. Furthermore, given that we do not know the reason for the higher attrition rates observed among the control group, we generally prefer the Lee bounds over the ABK bounds. While our upper bound estimate of the LSP effect on ELA achievement using Lee's bounds is indeed statistically insignificant, we caution the reader against using this extreme estimate to serve as the program's effect on student achievement after one year.

Table 6

Accounting for Differential Attrition

	N	without covariates			with covariates		
	(Lower bound / Upper bound)	Primary LATE	Lower bound	Upper bound	Primary LATE	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ELA achievement							
Primary LATE	1,766	-0.21***			-0.22***		
		(0.06)			(0.06)		
Lee bounds	1,733 / 1,731		-0.28***	-0.05		-0.29***	-0.06
			(0.06)	(0.05)		(0.09)	(0.06)
ABK bounds							
30th percentile	528 / 610		-0.11	0.15		-0.16	0.12
			(0.15)	(0.12)		(0.15)	(0.14)
25th percentile	431 / 506		-0.10	0.20		-0.13	0.17
			(0.16)	(0.12)		(0.15)	(0.13)
			(1544.82)	(0.89)		(7.12)	(0.83)
Math achievement							
Primary LATE	1,766	-0.65***			-0.65***		
		(0.10)			(0.09)		
Lee bounds	1,733 / 1,727		-0.72***	-0.43***		-0.72***	-0.44***
			(0.08)	(0.07)		(0.08)	(0.07)
ABK bounds							
30th percentile	522 / 641		-0.94***	-0.30**		-0.97***	-0.33**
			(0.18)	(0.13)		(0.18)	(0.16)
25th percentile	433 / 554		-0.96***	-0.26*		-0.98***	-0.27
			(0.24)	(0.13)		(0.27)	(0.18)

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Lee bounds lower bounds are based on removing the top 4.8 percent of performers from the treatment group and Lee upper bounds are based on removing the bottom 4.8 percent. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

Sensitivity of results to definitions of enrollment and lotteries. Our second check examines the sensitivity of results to our chosen definitions of LSP scholarship usage and what constitutes a lottery. Regarding scholarship usage, we have up to this point largely focused on enrollment exposure—or the percentage of quarters an individual used an LSP scholarship to enroll in a private school—to capture the effects of the LSP in the first year of statewide expansion. Alternatively, we could identify the effects of LSP scholarship usage using a dichotomous variable simply indicating if a student ever attended a private school (Wolf et al., 2010). In addition, we have chosen to define lotteries as cases in which a group of individuals that had the same priority category and who applied to the same grade in the same school had an observed chance of winning a lottery greater than 0 but less than 100. Thus, as another robustness check, we examine the extent to which our estimated effects differ in a subsample based on a restricted definition of what constitutes a lottery. Given the unexpected direction and magnitude of our findings, we would hope that our results are not sensitive to such definitions.

Table 7 presents the results from models using alternative definitions of enrollment and lotteries. For the sake of comparison, Columns (1) and (4) duplicate the results for the most specified versions of our primary models presented in Table 3. In columns (2) and (5), we present the results of models substituting a variable indicating if a student ever used their LSP to enroll in a private school for our variable indicating the percentage of private school enrollment. Columns (3) and (6), on the other hand, present the results of analyses that mirror our primary models, but in which we have restricted the analytical sample to individuals who have participated in lotteries with winning percentages varying between 5 and 95 percent.

Table 7

Estimated LSP Usage Effects Using Alternative Definitions of Enrollment and Lotteries

	ELA			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Enrollment	-		-	-		-
Exposure	0.22***		0.22***	0.65***		0.65***
	(0.08)		(0.07)	(0.09)		(0.10)
		-			-	
LSP Ever Enrolled		0.21***			0.63***	
		(0.07)			(0.09)	
Baseline Achievement	0.65***	0.65***	0.65***	0.62***	0.62***	0.62***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Female	0.15***	0.15***	0.15***	0.10**	0.11**	0.10**
	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
	-		-			
FRL	0.18***	-0.18**	0.18***	0.03	0.03	0.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Race/Ethnicity						
Hispanic	0.13	0.12	0.13	0.25*	0.24**	0.26**
	(0.17)	(0.15)	(0.16)	(0.14)	(0.12)	(0.12)
White	0.11	0.11	0.12	0.24***	0.23***	0.25***
	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.07)
Other	-0.04	-0.04	-0.04	0.04	0.04	0.04
	(0.11)	(0.14)	(0.12)	(0.17)	(0.21)	(0.17)
In New Orleans School	0.16*	0.16	0.16	0.04	0.03	0.04
	(0.09)	(0.10)	(0.12)	(0.10)	(0.11)	(0.12)
N	1766	1766	1748	1766	1766	1748
Risk Sets	194	194	193	194	194	193
Within R-squared	0.43	0.43	0.43	0.39	0.39	0.39
Between R-Squared	0.41	0.40	0.41	0.31	0.31	0.31
Overall R-Squared	0.42	0.42	0.42	0.38	0.38	0.38

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

The results presented in Table 7 clearly indicate that our results are sensitive neither to our definition of enrollment or lottery. While we can continue to explore the sensitivity of these results to alternative definitions of enrollment or even more restrictive definitions of lotteries, the stability of the results presented in Table 7 lends support to our primary analyses.

Accounting for differences in the number of school preferences offered. The baseline equivalency results presented in Table 2 indicate that eligible LSP applicants who won an LSP to their first choice school offered slightly fewer school choices on average relative to applicants who did not receive a scholarship to their first choice school. The analyses presented thus far have not directly accounted for this difference. Instead, we have effectively assumed that the statistical significance of this finding is essentially due to random chance.

In Table 8, we examine the extent to which our primary estimates are affected by this assumption by including fixed effects capturing the number of school preferences offered by a student. Columns 1 through 3 present findings for ELA achievement and columns 4 through 6 present our findings for math. The models employ more control variables as one moves from left to right within each performance grouping. The excluded comparison group for school choices is one. By using fixed effects, as opposed to a continuous variable identifying the number of school preferences offered, we are allowing for maximum flexibility in capturing differences in LSP applicants offering different numbers of school options.

In general, the results presented in Table 8 strongly mirror those presented in Table 3: LSP scholarship users are found to be 22 percent of a standard deviation behind the control group in ELA achievement after one year and 66 percent of a standard deviation behind in math. Furthermore, the coefficient estimates for the included school preference grouping fixed effects are insignificant in nearly every case, indicating that students offering more than one school preference generally do not differ in final achievement relative to applicants offering only one option. Thus, the results presented in Table 8 generally suggest that the significant difference between treatment and control group members in the number of school preferences offered is unlikely to have strongly biased our estimates of the effect of using an LSP scholarship on student achievement after one year.

Table 8

Estimated Effects of LSP Usage Comparing Individuals Offering the Same Number of School Preferences

	ELA achievement			Math achievement		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Enrollment Exposure	-0.22**	-0.22***	-0.22***	-0.65***	-0.65***	-0.66***
	(0.08)	(0.07)	(0.07)	(0.10)	(0.11)	(0.09)
School preferences offered†						
Two	-0.07	-0.06	-0.07	-0.12*	-0.10	-0.10
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)
Three	-0.04	-0.04	-0.04	-0.06	-0.04	-0.04
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Four	-0.03	-0.03	-0.03	-0.06	-0.05	-0.05
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Five	-0.11	-0.10	-0.10	-0.08	-0.06	-0.06
	(0.09)	(0.08)	(0.10)	(0.11)	(0.08)	(0.10)
Model Summary						
Demographic controls		X	X		X	X
New Orleans control			X			X
N	1,766	1,766	1,766	1,766	1,766	1,766
Risk sets	194	194	194	194	194	194

*** - p<.01, ** - p<.05, * - p<0.10
† - Excluded comparison group is students who only offered one school preference.
Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.
Source. Authors' calculations.

Curricular advantage? Prior experimental analyses examining the achievement impacts of public and private voucher programs have not found any statistically significant negative program impacts. Thus, our findings of substantial negative first year impacts of statewide expansion of the LSP are surprising. At the same time, it is important to recognize that our analyses are based on achievement on the Louisiana state assessments, rather than nationally representative exams. These results may simply reflect the fact that public schools are operating with curricula that are already aligned with the state assessments, while private schools have yet to align their curricula.

While we cannot provide an exhaustive examination of the teaching methods of the private schools in our sample, our testing data allow us to partially examine this question. We test for a curricular advantage by making use of the fact that some of the Louisiana state assessments include both criterion-referenced and norm-referenced exam questions. In particular, while the Louisiana assessments in grades four and eight only include criterion-referenced items, the iLEAP assessments offered in grades three, five, six, and seven include both criterion- and norm-referenced exam questions. If public school students experience a disproportionate curricular advantage, one would expect smaller negative LATE impact estimates on the iLEAP exams than on the LEAP exams.

Table 9 presents results from models examining the extent to which LSP usage effects differ across test type. For both ELA and math, we find that students taking the LEAP exam do appear to perform worse than iLEAP takers; however the estimated differences are statistically insignificant in all but one case. While not definitive, this pattern of results suggests that the substantial negative LATE impact estimates could be partially driven by the stronger alignment of the public school curricula to the state assessments. At the same time, it is important to note that LSP scholarship users still performed quite poorly on the hybrid iLEAP exams. Thus, while these findings may provide some insight into the substantial magnitude of our estimated impacts, they nevertheless support the general finding of a negative overall effect of the program after one year.

Table 9

Estimated LSP Scholarship Usage Effects by Test Type

	ELA			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP User taking iLEAP (CRT-NRT Hybrid)	-0.18***	-0.18***	-0.19***	-0.57***	-0.56***	-0.56***
	(0.07)	(0.10)	(0.06)	(0.08)	(0.09)	(0.09)
LSP User taking LEAP (CRT only)	-0.31*	-0.33*	-0.34**	-0.90***	-0.92***	-0.92***
	(0.19)	(0.19)	(0.18)	(0.16)	(0.16)	(0.16)
Difference	-0.11	-0.14	-0.14	-0.33*	-0.35	-0.35
	(0.17)	(0.20)	(0.22)	(0.20)	(0.23)	(0.26)
Model Summary						
Demographic Controls		X	X		X	X
New Orleans Control			X		X	X
N	1,766	1,766	1,766	1,766	1,766	1,766
Risk Sets	194	194	194	194	194	194
First stage F						
Exposure	686.1	259.5	231.3	684.1	258.4	230.3
Interaction	409.7	155.3	137.9	409.5	155.2	137.9

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

Conclusion

This paper presents the preliminary analyses of the first year participant effects of the statewide expansion of the Louisiana Scholarship Program (LSP), one of the newest and largest school voucher programs in the U.S. This study contributes to the existing literature on the participant effects of publicly funded voucher programs for two reasons. First, it uses a highly rigorous experimental design to estimate treatment effects while avoiding self-selection bias concerns. Second, it is the first evaluation of a statewide school voucher program. These contributions will add to the existing knowledge on the effects of private school choice programs.

The results presented in this paper indicate significant and substantial negative achievement impacts associated with using an LSP scholarship. In general, we find that LSP scholarship usage is associated with a quarter of a standard deviation decline in ELA achievement and nearly two-thirds of a standard deviation decline in math. These findings are the first of their kind among random assignment evaluations of school voucher programs and are robust to several alternative specifications.

At the same time, it is important to keep in mind that our analyses are based on a small subsample of participants in the first year of the program with performance data on the Louisiana state assessments. Specifically, our analysis sample represents approximately twenty percent of the 2012 cohort of eligible applicants. Thus, in a real sense, this paper is not an evaluation of the entire program, but an evaluation of the experiences of students in grades three through seven at baseline, who participated in actual lotteries, with testing outcomes in year 1. The educational impact of the LSP on the many thousands of program participants who do not satisfy those criteria remains, at this point, unknown. Readers are encouraged not to draw firm conclusions from this initial analysis due to the severe threats to external validity posed by those limitations of the sample.

At this point we can only speculate as to why our results differ so dramatically from the voucher experiments conducted previously. Our working hypothesis at this point is that a higher-quality set of private schools participated in earlier voucher and scholarship programs in Washington, DC; New York City; Dayton, Ohio; and Charlotte, North Carolina; in which more positive voucher experimental impacts were reported. Less than one-third of the private schools in Louisiana chose to participate in the LSP in its first year, possibly because of the extensive regulations placed on the program by government authorities (Kisida, Wolf, & Rhinesmith, 2015). Although it is only speculation at this point, the Louisiana Scholarship Program regulatory requirements may have played a role in preventing the private school choice program from delivering better outcomes to its participants.

References

Abdulkadiroglu, A., Pathak, P. A., & Roth, A. E. (2005). The New York City high school match. *American Economic Review, Papers and Proceedings*, *95*, 364-367.

Angrist, Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.

Angrist, J. D., Bettinger, E. P., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Columbia. *The American Economic Review*, *93*(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*, 299-323.

Barrow, L. & Rouse, C. E. (2008). School vouchers: Recent findings and unanswered questions. *Economic Perspectives*, 32, 2–16.

Bitler, M. P., Domina, T., Penner, E. K., & Hoynes, H. W. (2013). *Distributional effects of a school voucher program: Evidence from New York City* (NBER Working Paper No. 19271). National Bureau of Economic Research.

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

Bozeman, B. (1987). *All organizations are public*. San Francisco, CA: Jossey-Bass.

Cassell, M. (2003). *How governments privatize: The politics of divestment in the United States and Germany*. Washington, DC: Georgetown University Press.

Catt, A. D. (2014). Public rules on private schools: Measuring the regulatory impact of state statutes and school choice programs. Indianapolis, IN: Friedman Foundation for Educational Choice. Retrieved from http://www.edchoice.org/CMSModules/EdChoice/FileLibrary/1056/Public-Rules-on-Private-Schools-Measuring-the-Regulatory-Impact-of-State-Statutes-and-School-Choice-Programs.pdf

Chingos, M. M., & Peterson, P. E. (2012, August). The effects of school vouchers on college enrollment: Experimental evidence from New York City. Washington, DC: Brown Center on Education Policy at Brookings.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets & America's schools*. Washington, DC: Brookings Press.

Cowen Institute for Public Education Initiatives (2013, January). Spotlight on choice: Parent opinions on school selection in New Orleans. Tulane University. Retrieved from http://www.coweninstitute.com/wp-content/uploads/2013/01/Choice-Focus-Groups-FINAL-small.pdf

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

Cullen, J. B., Jacob, B. A., & Levitt, S. (2003, November). The effect of school choice on student outcomes: Evidence from randomized lotteries (NBER Working Paper No. 10113).

Dewey, J. (1916). *Democracy and education*. New York: Macmillan.

Doolittle, F., & Connors, W. (2001). Designing education voucher experiments: Recommendations for researchers, funders, and users. *Privatizing education* (Levin, H. M., ed.). Boulder, CO: Westview Press.

Figlio, D. (2011) Evaluation of the Florida Tax Credit Scholarship Program participation, compliance, and text scores in 2009-10, Report to the Florida State Department of Education, August.

Friedman Foundation (2014). School choice programs. Retrieved from http://www.edchoice.org/School-Choice/School-Choice-Programs

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

Greene, J. P. (2001). Vouchers in Charlotte. *Education Matters*, 1, 55–60.

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

Gutmann, A. (1987). *Democratic education*. Princeton, NJ: Princeton University Press.

Hammons, C. (2002). The effects of town tuitioning in Maine and Vermont. Indianapolis, IN: Friedman Foundation. Retrieved from http://www.edchoice.org/Research/Reports/The-Effects-of-Town-Tuitioning-in-Maine-and-Vermont.aspx

Henig, J. R. (1994). *Rethinking school choice: Limits of the market metaphor*. Princeton, NJ: Princeton University 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, 191–217.

Howell, W. G., & Peterson, P. E. (with Wolf, P. J., & Campbell, D. E.) (2006). *The educational gap: Vouchers and urban schools* (Rev. ed.). Washington, DC: Brookings.

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.

Kisida, B., Wolf, P. J., & Rhinesmith, E. (2015). *Views from private schools: Attitudes about school choice programs in three states*. Washington, DC: American Enterprise Institute.

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

Louisiana Department of Education (2013a, Spring). iLEAP Interpretive Guide, Grades 3, 5, 6, and 7. Retrieved from http://www.louisianabelieves.com/docs/assessment/ileap-interpretive-guide.pdf

Louisiana Department of Education (2013b, Spring). LEAP Interpretive Guide, Grades 4 and 8. Retrieved from http://www.louisianabelieves.com/docs/assessment/leap-interpretive-guide.pdf

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76: 1071-1102.

Levin, H. M., ed. (2001). *Privatizing education: Can the marketplace deliver choice, efficiency, equity, and social cohesion?* Boulder, CO: Westview Press.

Macedo, S. (2000). *Diversity and distrust: Civic education in a multicultural democracy*. Cambridge, MA: Harvard University Press.

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. MPR Reference No. 8404-045. Cambridge, MA: Mathematica Policy Research.

Mill, J. S. (1962) *Utilitarianism, on liberty, essay on Bentham*. (Warnock, M. ed.) New York: Meridian.

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

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

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,* 699–717.

Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, 113, 553–602.

Salisbury, D., & Tooley, J. (eds.) (2005). *What America can learn from school choice in other countries*. Washington, DC: CATO.

Sikkink, D. (2012). Religious school differences in school climate and academic mission: A descriptive overview of school organization and student outcomes. *Journal of School Choice*, *6*(1), 20–39.

Trivitt, J. R., & Wolf, P. J. (2011). School choice and the branding of Catholic schools. *Education Finance and Policy*, *6*(2), 202–245.

What Works Clearinghouse (2014). *Procedure and Standards Handbook, Version 3.0*. Washington, DC: Institute of Education Sciences.

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

Wolf, P. J. (2008). Vouchers. *The International Encyclopedia of Education* (McCulloch, G., & Crook, D., eds.). London: Routledge, pp. 635-636.

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

Wolf, P., Gutmann, B., Puma, M, Kisida, B., Rizzo, L., & Eissa, N. O. (2009) Evaluation of the DC Opportunity Scholarship Program: Impacts after three years, U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Washington, DC: U.S. Government Printing Office, NCEE 2009-4050, March. Retrieved from http://ies.ed.gov/ncee/pubs/20094050/

Wolf, P., Gutmann, B., Puma, M, Kisida, B., Rizzo, L., & Eissa, N. O., & Carr, M. (2010) Evaluation of the DC Opportunity Scholarship Program: Final report, U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Washington, DC: U.S. Government Printing Office, NCEE 2010-4018, June. Retrieved from http://ies.ed.gov/ncee/pubs/20104018/pdf/20104018.pdf

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

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

Appendix

Table A1

First Stage Regression Results Using LSP Receipt to Predict Usage

	ELA			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Awarded	0.70***	0.70***	0.70***	0.70***	0.70***	0.70***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Demographic controls		X	X		X	X
New Orleans School Control			X			X
N	1766	1766	1766	1766	1766	1766
Risk Sets	194	194	194	194	194	194
Joint F-statistic	1014.39	292.25	256.43	1011.56	291.17	306.75
Adj. R-Squared	0.66	0.66	0.66	0.66	0.66	0.65

*** - p<.01, ** - p<.05, * - p<0.10
Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.
Source. Authors' calculations.

Table A2

Estimated effects of LSP Usage on Science and Social Studies Achievement

	Science			Social Studies		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Enrollment Exposure	-0.35***	-0.36***	-0.35***	-0.42***	-0.42***	-0.43***
	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.08)
Baseline Achievement	0.53***	0.52***	0.52***	0.46***	0.45***	0.45***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Female		0.01	0.01		0.02	0.02
		(0.04)	(0.03)		(0.04)	(0.05)
FRL		-0.20**	-0.20**		-0.18	-0.18**
		(0.10)	(0.09)		(0.12)	(0.09)
Race/Ethnicity						
Hispanic		0.30**	0.30***		0.07	0.07
		(0.12)	(0.12)		(0.15)	(0.14)
White		0.33***	0.33***		0.28***	0.28***
		(0.08)	(0.09)		(0.08)	(0.08)
Other		0.23	0.22		-0.09	-0.09
		(0.19)	(0.17)		(0.16)	(0.16)
In New Orleans School			-0.07			0.08
			(0.11)			(0.13)
N	1,744	1,744	1,744	1,743	1,743	1,743
Risk Sets	194	194	194	194	194	194
Within R-squared	0.31	0.32	0.32	0.24	0.24	0.24
Between R-Squared	0.31	0.32	0.32	0.17	0.17	0.17
Overall R-Squared	0.31	0.33	0.32	0.24	0.24	0.24

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets.

Source. Authors' calculations.

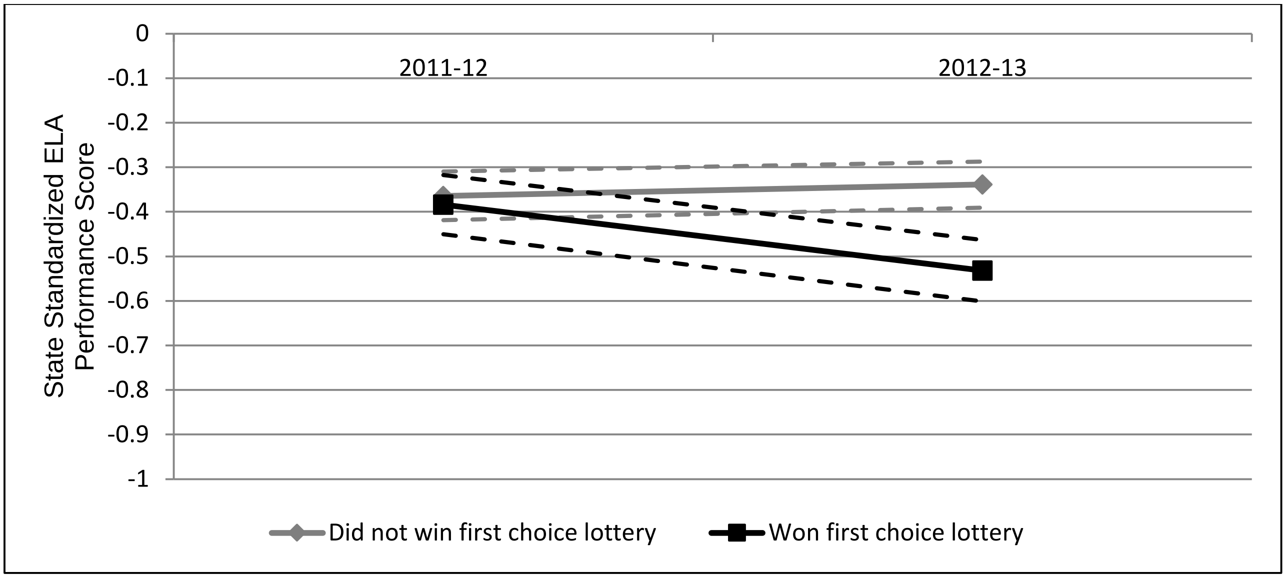

Figure A1. Comparison of treatment and control group average ELA performance over time. Achievement has been standardized by grade and year to the Louisiana state test taking distribution. Dotted lines represent 95% confidence intervals for the performance averages. These results indicate that control group students did experience a mild improvement relative to the state over time; however treatment group students experienced a large decline in performance between 2011-12 and 2012-13.

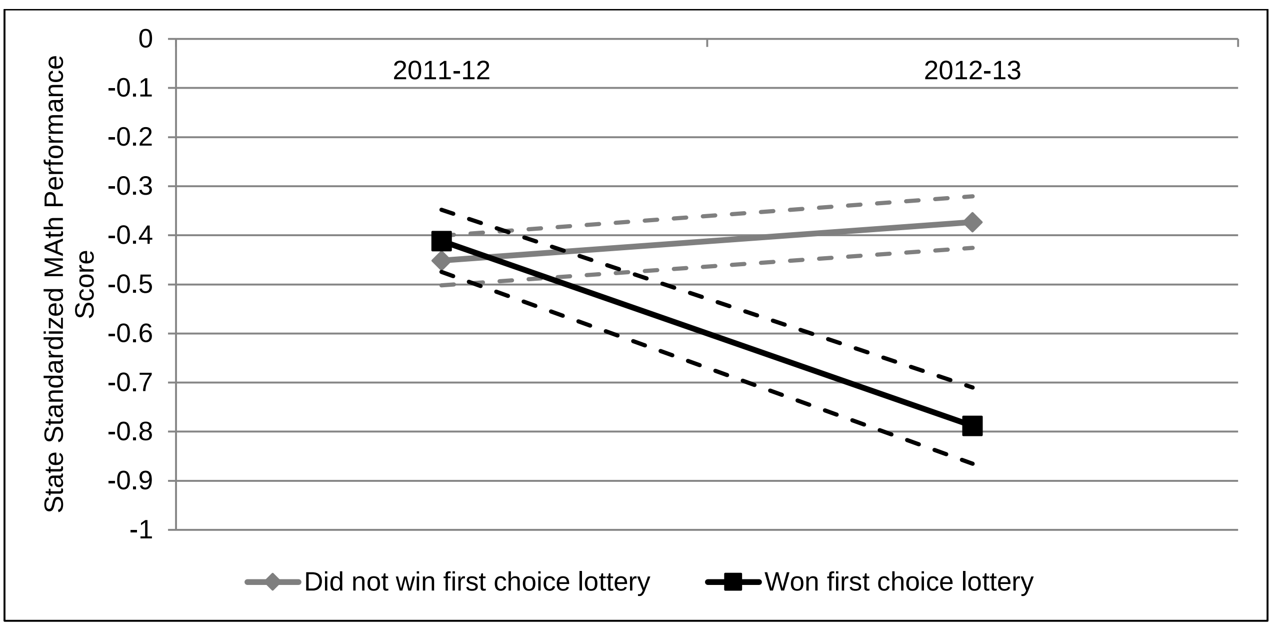

Figure A2. Comparison of treatment and control group average math performance over time. Achievement has been standardized by grade and year to the Louisiana state test taking distribution. Dotted lines represent 95% confidence intervals for the performance averages. These results indicate that control group students did experience a mild improvement relative to the state over time; however treatment group students experienced a large decline in performance between 2011-12 and 2012-13.

Chapter 2[31]

Measures of Non-cognitive Skills and Political Tolerance after Two Years

Introduction

While the majority of quantitative studies examining the effects of school choice programs on participating students have tended to focus on academic and attainment outcomes, a growing body of research suggests that other skills should also play an important role in such evaluations (Mills, 2013). Specifically, studies have found that character skills such as conscientiousness and self-control are predictive of individual academic and workforce success (Almlund et al. 2011; Heckman & Kautz, 2012; Tangney, Baumeister, & Boone, 2004). More importantly, evidence from some studies suggests that different education environments can influence these skills (Cunha & Heckman, 2008; Dee & West, 2011; Jackson, 2012; West et al., 2014). While two recent studies have examined the relationship between enrollment in charter schools and non-cognitive skills (Dobbie & Fryer, 2014; West et al., 2014); no such research exists for private school voucher programs. This paper aims to address this gap by providing the first descriptive analysis of differences in non-cognitive skills and measures of political tolerance associated with the Louisiana Scholarship Program.

Specifically, we examine differences in student responses to several measures of non-cognitive skills and political tolerance collected through a phone survey conducted between November, 2014 and February, 2015. The specific non-cognitive skills measures included in our analysis are the Grit scale (Duckworth & Quinn, 2012), the Locus of Control scale (Rotter, 1966), and the Rosenberg's (1965) Self-esteem scale. In general, the findings presented in this

[31] This paper was co-authored with Albert Cheng, Colling E. Hitt, Patrick J. Wolf, and Jay P. Greene.

paper suggest no strong differences between students who received an LSP scholarship and those who did not across all measures. These findings hold across several specifications and robustness checks.

The remainder of this paper proceeds as follows. In the next section, we describe the existing literature examining the development of non-cognitive skills and civic attitudes in private school choice systems. We then outline the methodology used to estimate differences in these skills in the first group of students to experience the expansion of the Louisiana Scholarship Program statewide. Next, we present our primary results and a series of checks designed to estimate the robustness of our findings. Finally, we conclude with a summary of our findings and implications for future research.

Alternative outcomes and school choice

Evaluations of school voucher programs have generally focused on estimating voucher impacts on student achievement and attainment (Mills, 2013). The focus on academic outcomes is intuitive, as student achievement is strongly linked to the economic success of individuals (Heckman, 2008) and countries (Hanushek & Woessman, 2009). Moreover, recent moves to test-based accountability systems in the United States have made measures of student achievement increasingly available to researchers (West et al., 2014).

At the same time, there is a growing body of research demonstrating that measures of student non-cognitive skills—such as self-control and conscientiousness—are also predictive of short- and long-run life outcomes (Almlund et al. 2011; Heckman & Kautz, 2012; Tangney, Baumeister, & Boone, 2004). Furthermore, there has been a long-standing view in the United States of the role of education in developing future citizens, in addition to positive economic agents (Dewey, 1916; Gutmann, 2003). Taken together, these points suggest a need for school voucher evaluations to expand beyond measures of student achievement and attainment in assessing program effectiveness. In the following sections, we review literature supporting the importance of non-cognitive skills and civic engagement, as well as existing evidence of the effectiveness of school choice programs in promoting these outcomes. In general, the literature on the effects of voucher programs in developing these skills is quite limited; a finding which we argue supports the importance of our work.

The importance of non-cognitive skills and the role of school choice in their development

While evidence indicates positive relationships between student academics and later life outcomes (Heckman, 2008), recent empirical research has also demonstrated the important roles other characteristics play in life outcomes. These skills fall under the broad classification of "non-cognitive skills" in the economics literature, which West et al. (2014) note "has become a catchall tem for traits of skills not captured by assessments of cognitive ability and knowledge." (p. 1) Many of these skills have been found to be linked to positive short- and long-run life outcomes. For example, Tangney, Baumesiter, and Boone (2004) find in an analysis of college students that measures of self-control are strongly related to college GPA, adjustment to college, and better emotional responses to stressful situations. Conscientiousness, defined as strong attention to detail and vigilance—has been found to be a strong predictor of school success and attainment (Farkas, 2003; Lleras, 2008; Almlund et al. 2011), as well as future employment and earnings (Farkas, 2003). In addition, Heckman, Stixrud, and Urzua (2006) have found evidence of negative relationships between non-cognitive skills and the likelihood of incarceration and teenage pregnancy using data collected by the National Longitudinal Survey of Youth (NLSY) of 1979. Heckman and Kautz (2012) note that the positive relationships between life outcomes and non-cognitive skills are somewhat to be expected, given that traits like conscientiousness and self-control tend to be particularly valuable in the workplace.

This evidence of the potential for positive economic returns to non-cognitive skills development has recently captured the attention of policymakers, the media, and the general public (Tough, 2012; Whitman, 2008). As West and colleagues (2014) note, policymakers and practitioners are becoming increasingly interested in the development and use of measures of student non-cognitive skills. Furthermore, while the malleability of non-cognitive skills is the focus of contentious debate in the psychological literature (Almlund et al. 2011, Baumesiter, Vohs, & Trice, 2007; Duckworth, Peterson, Matthews, & Kelly, 2007), studies have found that different education environments may influence these character skills (Cunha & Heckman, 2008; Dee & West, 2011; Jackson, 2012; West et al., 2014).

Unfortunately, the literature estimating the effects of school voucher programs on the development of non-cognitive skills is very limited. Specifically, none of the existing experimental evaluations of school vouchers have included a non-cognitive component in their analysis beyond the proxies of high school attainment and college enrollment (Wolf et. al. 2013; Cowen et al. 2013; Chingos & Peterson 2015; and Warren 2011). This lack of research is somewhat surprising, given the specific emphasis placed by many private schools on discipline (Figlio & Ludwig, 2012), explicit attempts at character building (Bryk, Lee, & Holland, 1993), as well as existing evidence from longitudinal studies finding that private school attendance is associated with significantly fewer risky behaviors on average (Figlio & Ludwig, 2012; Mocan & Tekin, 2007).

While the literature on the effects of voucher programs on the development of non-cognitive skills is essentially non-existent; a growing number of studies examining charter schools have included non-cognitive skills measures in their broader evaluation of program effectiveness. In general, however, the findings from these studies suggest that charter schools may harm the development of non-cognitive skills. For example, Dobbie and Fryer (2014), find that students winning admission via lottery to the Promise Academy middle school in Harlem, NY exhibited lower rates of self-esteem and "grit" relative to lottery losers—albeit only the latter results were statistically significant. Furthermore, West and colleagues note a paradoxical finding in their 2014 evaluation of Boston public and charter schools: student self-assessments of non-cognitive skills are strongly and positively related to attendance, behaviors, and achievement gains; however relationships at the school level are insignificant or negatively related to school level measures. Furthermore, students attending high performing charter schools report particularly low scores on non-cognitive skills measures. After providing a series of specification checks, the authors argue that their results are likely driven by a reference group bias, whereby students in more disciplined schooling environments may be holding themselves to higher standards than students in other schools.[32]

In summary, the existing literature on non-cognitive skills development in schools of choice is very limited. Among the 11 existing random assignment evaluations of school voucher programs in the United States, none have examined how vouchers impact student non-cognitive skills development. Furthermore, while existing evidence from charter school evaluations suggest the potential for negative impacts, there is strong reason to believe that these results are driven by school environment (West et al., 2014).[33] Nevertheless, given increasing evidence of a

[32] Dobbie and Fryer (2014) similarly hypothesized that such reference group bias could be driving their negative results; however they do not investigate further.
[33] This sentiment is shared by Tuttle and colleagues in their 2013 experimental evaluation of KIPP middle schools. In particular, they find that students randomly admitted to KIPP via oversubscription lotteries are more likely to report lying to their parents and losing their temper

link between non-cognitive skills and positive life outcomes, as well as the explicit focus of many private schools on such alternative educational outcomes, it is important to document the development of non-cognitive skills in school choice evaluations. The present study, which includes survey measures of non-cognitive skills, represents a first step in addressing this gap in the school voucher literature.

Civic skills and school choice

Supporters of traditional public schools often argue that one of the primary benefits of democratically governed education is that it can educate children to successfully promote our democratic values (Dewey, 1916; Gutmann, 2003; Henig, 1999). For example, Amy Gutmann (2003) claims "[a] central part of the historic mission of a democratically accredited school system is to educate citizens who are capable of sitting on juries, assessing public proposals (about schools, for example), exercising their rights, fulfilling their responsibilities, and seizing their opportunities to live a good life as they see fit." (p. 126) School choice, opponents argue, will lead to a fractured system of education and will therefore fail to pass on our collective values (Berliner & Biddle, 1996).

Wolf (2005) examines the evidence on the effects of school choice and civic values in a systematic review, focusing on findings from experimental studies as well as rigorous quasi-experimental methods that approximate random assignment. In general, he notes that 20 studies with 48 separate estimates of civic effects of private school choice meet his selection criteria. Of the 48 total estimates, he finds only three indicating that private school choice negatively affects

in school while also being significantly more likely to complete their homework on time. The authors note that these seemingly contradictory findings may be reflective of KIPP's "no excuses" school environment, which places a strong emphasis on school discipline and hard work.

civic values. In contrast, 29 findings (60 percent) are either positive or contingently positive.[34] In the present study, we focus exclusively on estimating differences in measures of political tolerance between students awarded an LSP scholarship and those not receiving a scholarship. As such, it is worth highlighting Wolf's (2005) findings regarding private schooling and political tolerance. Wolf includes seven studies estimating the effects of private school attendance on political tolerance in his 2005 review. These studies were included because they either used random assignment (Campbell, 2002; Howell & Peterson, 2002; Wolf, Peterson, & West, 2001) or were published in peer-reviewed journals. In general, Wolf finds private school attendance associated with higher levels of tolerance, although some studies found no difference between public and private school students.[35]

In general, the evidence presented in this section indicates a large gap in the school voucher literature, whereby none of the existing voucher evaluations have examined either non-cognitive skills development or measures of civic engagement. This paper therefore aims to provide a first attempt at addressing this gap in the literature by describing differences in measures of non-cognitive skills and political tolerance among students who received and did not receive an LSP scholarship in the first year of statewide expansion. The next section outlines the methodology used to study these topics.

Methodology

This section introduces our methodology for investigating differences in non-cognitive skills and political tolerance levels between students who were awarded and not awarded an LSP in the 2012-13 school year. We begin by describing the phone survey data collection process and

[34] Wolf (2005) categorizes a finding as contingently positive if it reports statistically significant positive findings for a type of school rather than all schools.
[35] Interestingly, one study found that attending a secular or Catholic private school was beneficial but attending a non-Catholic religious school undermined political tolerance (Campbell, 2001)

then move on to a description of the primary non-cognitive and tolerance measures used in this study. The section concludes with description of the final sample of survey respondents as well as a comparison of our sample to the full population of eligible applicants in the first year of the program's expansion statewide.

Data collection

Our study is based on the results of phone surveys of a selected subsample of the nearly 10,000 LSP eligible applicants in the first year of the program's expansion statewide. After first developing a survey incorporating several well-known scales designed to capture student non-cognitive skills, our research team worked closely with an independent research group specializing in phone survey administration to complete data collection. Data collection began on November 18, 2014 and was concluded on February 7, 2015 after a total of 1,000 records were collected.

Our research team provided the independent survey group with a randomly ordered list of LSP eligible applicants that was divided into two strata. The first strata consisted of students who received no exemptions in the LSP scholarship application process; whereas the second strata included students who had participated in the New Orleans pilot program and students with special education exemptions.[36] These strata were ordered in the data to prioritize students who were not subject to exemptions in the LSP award process and therefore were more likely to have been randomly assigned to a scholarship placement in a private school or to the control group.

[36] We excluded 159 students with severe disabilities from our call sample because their listed disabilities likely precluded their participation in the phone survey. Specifically, we excluded the following disability categories: Autism, Developmental Delay, Intellectual Disability (mild through severe), and Multiple Disabilities.

Upon contacting a household, surveyors first asked to speak with a parent or guardian to verify they had reached the intended family, describe the purpose of the study, and request consent to administer the survey to the child. After receiving consent, the surveyor asked to speak with the child, verified that the child's name matched the name on the intended record[37], and then administered the survey to the child[38]. In general, surveys lasted between 10 and 15 minutes. At the conclusion of the survey, the surveyor asked to speak again with the student's parent or guardian. The surveyor thanked the parent for their participation and provided the family with a toll-free number to call in case the family had any additional questions about the study.

Our final survey sample consists of 999[39] students, two thirds of whom received an LSP scholarship. This sample represents slightly more than 10 percent of the eligible applicants in the 2012-13 school year. These data have been merged with administrative data on student achievement and demographics provided by the Louisiana Department of Education (LDE). In addition, we have supplemented these data with information on school-level characteristics publicly available through the National Center for Education Statistics's Common Core of Data (CCD) and Private School Universe Survey (PSUS).

Measures of non-cognitive skills and civic attitudes

This section describes four measures of non-cognitive skills and civic attitudes that are the basis for our study. We have chosen the measures outlined in this section due to their use in

[37] Several contacted households had more than one child in the LSP eligible applicant sample. In order to preserve the random ordering of the call list, surveyors validated that the child was the child of interest and explained to families that they may be receiving a call at a later point to interview the other children in the household.

[38] Parents were allowed to listen in as the survey was conducted if requested. Surveyors were asked to identify such records in the collected data.

[39] Our final analytical sample excludes one of the original 1,000 respondents because the child's guardian later contacted the research team and asked that the child be removed from the study.

other studies of school choice. For example, Dobbie and Fyer (2014) and West and colleagues (2014) have included the Grit scale in evaluations of the effects of charter school attendance on student outcomes. Dobbie and Fryer (2014) additionally include the Locus-of-control and Self-esteem scales. Finally, the Political Tolerance scale described in this section has been used in several studies of private schools (Wolf, 2005).

In addition, there are a few features of our design and implementation of the scales described below which will likely add noise to our final results. For example, none of the scales used in this study have been validated for phone surveys, nor have they been validated in populations as young as the study sample. In addition, the research team made a small number of changes to some of the survey items after consulting with the independent survey group to improve language clarity.

Finally, items were randomly ordered within instruments across individuals in an attempt to ensure that individual responses were not biased by the presentation order of questions.

These notes aside, we proceed to a description of the scales used in our analysis.

Grit. The first non-cognitive skill measured in this study is grit, or an individual's "perseverance and passion for long-term goals" (Duckworth et al., 2007). Our measure of grit is based on the 8-item Short Grit Scale developed by Duckworth and Quinn (2012) that has been adapted for young children.[40] An individual's grit score is based on their average responses to eight five-point Likert scale items[41] that include questions like "New ideas and projects sometimes distract me from previous ones" and "I am a hard worker".

[40] The adapted 8-item grit scale is available on Dr. Duckworth's website: https://upenn.app.box.com/8itemgritchild

[41] Students are asked to choose among the following options: "Very much like you", "Mostly like you", "Somewhat like you", "Not much like you", and "Not like you at all".

Studies using different versions of the scale have found that grit is predictive of several positive outcomes. Duckworth et al. (2007) find that grit is positively associated with career stability in a sample of adults, positively related to GPA among undergraduates at an elite Northeastern university, is a better predictor of retention among West Point first years than either a measure of self-control or an assessment administered by West Point, and that individuals with higher grit scores were more likely to outlast their competitors in the National Spelling Bee. Furthermore, Duckworth and Quinn (2012) find that grit is positively related to student GPA, independent of IQ. On the other hand, two recent studies using the Grit scale in evaluations of charter schools have identified negative relationships between charter school attendance and Grit (Dobbie & Fryer, 2014; West et al., 2014). Both studies note, however, that the negative relationships may be driven in part by reference group bias resulting from differences in individual schooling environments. The Grit scale has a .53 internal reliability score across our whole sample, with a reported .52 for students in grades 2 through 6 and a reported .58 among students in grades 7 through 12. This is substantially lower than the generally accepted threshold of .75 for internal reliability scores (Croker & Aligna, 1986).

In a separate section appearing after the Grit scale items, we included a series of vignettes describing fictitious individuals with different levels of Grit. While these vignettes have been used by some to attempt to address reference group bias (King, Murray, Salomon, & Tandon, 2004; Zamarro, Vonkova, and DeBerg, 2014), we instead use individual responses to these vignettes in a robustness check designed to determine if reference group bias is present in our analysis.

Locus of Control. The second scale included in our survey is the locus of control scale developed by Rotter (1966) which is designed to capture the extent to which an individual believes rewards are the result of their own actions. We record an individual's locus of control based on their responses to six four-point Likert scale items.[42] The specific items are taken from the High School and Beyond Third Follow-up (1986); and include questions like "Good luck is more important than hard work for success" and "Every time I try to get ahead, something or somebody stops me". The Locus of Control scale has an internal reliability score of .47 across all phone survey respondents, with a .44 among students in grades 2-6 and a .54 among students in grades 7 through 12 at data collection.

Self-esteem. We capture individual self-esteem levels using Rosenberg's (1965) popular self-esteem scale. A respondent's self-esteem score is calculated as their average response across10 four-point Likert scale items. Each of the 10 items are designed to capture an individual's view of their self-worth; including questions like "I am able to do thinks as well as most other people" and "I certainly feel useless at times". In a 2003 review of studies using the self-esteem scale, Baumeister, Campbell, Krueger, and Vohs note that self-esteem is only moderately related to school performance, is a strong predictor of individual happiness, is associated with a stronger likelihood of speaking up in a group, among other findings. The reported internal reliability score is .77 for the Self-esteem scale across all respondents, with a .73 reported for students in grades 2 through 6 and a .83 reported for students in grades 7 through 12.

Political Tolerance. The final scale examined in this study attempts to capture participant civic attitudes by providing a measure of their political tolerance. The political tolerance protocol first asks individuals to identify a group that "has beliefs that [they] oppose the most" and then asks a series of questions regarding the level of political freedoms the

[42] Individuals are asked to select among four responses to each question: "Strongly Disagree", "Disagree", "Agree", "Strongly Agree".

individual would allow this group to enjoy. For example, individuals are asked to indicate if they "Strongly Disagree", "Disagree", "[are] Neutral", "Agree", or "Strongly Agree" that "The government should be able to secretly listen in on the telephone conversations" of their least preferred group. Unlike the three previous scales, the political tolerance scale was only administered to students without disabilities in grades 5 through 10 at baseline[43] due to the sensitive nature of the topic.[44] In a 2005 review of the literature on private school choice and civic outcomes, Patrick Wolf notes that several studies have used this scale to compare political tolerance levels between public and private school students, largely finding higher levels of tolerance among private school students. The internal reliability score for this scale for this group of students is .77.

Sample Description

Data collection began in November of 2014 and continued for nearly four months until a sample of 999 records were collected. This group of respondents, representing slightly more than 10 percent of all eligible LSP applicants in 2012, will provide the basis for our primary analysis. It is important to recognize that comparisons presented in this paper rely on a selected sample, based on the small number of families that opted into the phone survey. In particular, the small response rate raises some concerns that our results may not be representative of the broader population of eligible LSP applicants.

Table 1 presents descriptive statistics for several student characteristics collected at baseline for two groups of students: the students responding to our phone survey and all other

[43] These students should be in grades 7 through 12 as the time of survey administration unless they were held back during the time period examined.
[44] In addition, the phone survey included a prompt before and during the questions noting, "If you are at all uncomfortable answering any of these questions, you may choose not to answer. That is completely ok."

eligible LSP applicants in 2012-13. The data presented in Table 1 is based either on student characteristics collected in the 2011-12 school year or from their application.[45]

In general, the results presented in Table 1 do not suggest that the phone survey sample differs strongly from the population of eligible applicants. The survey respondent sample is slightly less likely to be African American and to be identified as Free- or Reduced-Price Lunch eligible relative to the population of eligible LSP applicants. In addition, survey respondents were slightly more likely to have applied for the program in third grade and slightly less likely to have applied in either seventh or eighth grade. The remaining differences, including those observed among students with achievement data in 2011-12, are fairly negligible. In particular, while students in both groups on average scored below the state average in all subjects, the two groups do not differ strongly in their observed performance. While there is no way to determine if the students differ from the general population of eligible applicants on unobservable dimensions, the similarly of the distributions presented in Table 1 suggests that such selection bias may not be a major concern for our analysis.

[45] Student grade, for example, were collected from a student's application--we would roughly expect these students to be two grades higher at the time of the survey if they were admitted to the grade applied for and progressed at a normal pace through grades. FRL and achievement data are only available for students in grades 3-7 who took either the iLEAP or LEAP exams. Finally, small percentage of students is missing data required to identify if they were living in a metropolitan statistical area at the time of application.

Table 1

Comparison of characteristics at baseline across multiple samples

	Phone survey sample		Other eligible applicants		Diff.	Std. Err.
	N	Mean	N	Mean		
Female	999	0.53	8,530	0.50	0.02	0.02
African American	999	0.84	8,530	0.87	-0.03**	0.01
White	999	0.09	8,530	0.08	0.01	0.01
Hispanic	999	0.03	8,530	0.02	0.01	0.01
Free- or reduced-price lunch eligible[a]	356	0.78	3,395	0.82	-0.04*	0.02
Living in Metropolitan Statistical Area[b]	921	0.96	7,588	0.96	0.00	0.01
School preferences listed	999	1.97	8,530	2.06	-0.09*	0.04
Grade at application						
Kindergarten	999	0.21	8,530	0.21	0.00	0.01
First	999	0.14	8,530	0.13	0.01	0.01
Second	999	0.11	8,530	0.11	0.01	0.01
Third	999	0.14	8,530	0.11	0.04***	0.01
Fourth	999	0.09	8,530	0.10	-0.01	0.01
Fifth	999	0.09	8,530	0.08	0.01	0.01
Sixth	999	0.09	8,530	0.08	0.00	0.01
Seventh	999	0.05	8,530	0.08	-0.03***	0.01
Eighth	999	0.04	8,530	0.05	-0.01**	0.01
Ninth	999	0.04	8,530	0.05	-0.01	0.01
Tenth	999	0.02	8,530	0.02	0.00	0.00
Standardized state achievement[c]						
Math	355	-0.51	3,330	-0.52	0.00	0.05
ELA	356	-0.45	3,328	-0.43	-0.02	0.05
Science	355	-0.54	3,322	-0.55	0.01	0.06
Social Studies	355	-0.46	3,322	-0.47	0.01	0.06

*** - p<.01, ** - p<.05, * - p<0.10

[a.] Student FRL status is only available for students appearing in the state's testing data.

[b.] Data on Metropolitan Statistical Area is taken from the American Community Survey.

[c.] Student achievement data are restricted to students taking the standard state assessments (iLEAP or LEAP) in grades 3 through 7 in the 2011-12 school year.

Note. Student achievement data has been standardized within subject and grade to the state's testing distribution. Source: Authors' calculations.

Results and Discussion

In the following sections, we present the primary results from our analyses examining differences in measures of student non-cognitive skills and political tolerance; as well as a series of checks designed to investigate the robustness of our results. In general, the evidence presented here largely suggests that the two groups of students did not differ across any of the four measures of interest two years after initial LSP scholarship assignment.

Primary observational analysis

Table 2 presents pairwise correlations between the three non-cognitive skills measures, tolerance measure, and four estimates capturing student achievement growth between the 2011-12 and 2012-13 school years.[46] Panel A presents results for the full set of respondents with complete responses for all measures, excluding Political Tolerance; while Panel B presents results for a subset of students who additionally provided responses for the Political Tolerance scale.[47] We include achievement scores gains in Table 1 to examine the relationship between the included non-cognitive skills measures and student achievement gains; however in doing so, we have substantially restricted the sample for which we can estimate these relationships. Nevertheless, the relationships observed in Table 1 among the non-cognitive skills measures generally hold in the full sample of survey respondents.

In both Panel A and Panel B, the group of non-cognitive skills measures are strongly correlated with one another, as are the group of achievement gains measures. On the other hand, the two groups—non-cognitive skills and achievement gains—are not strongly correlated. These findings corroborate the work of West et al. (2014), who found significant, but very weak relationships between Grit and math and ELA achievement gains. Interestingly, Political Tolerance does not appear to be related to either Grit or Self-esteem (Panel B); but is significantly and positively related to Locus of Control.

---

[46] Keeping in line with the work of West and colleagues (2014), we calculate mean performance gain as the average residual resulting from a regression of standardized achievement in 2012-13 on a cubic function of achievement in 2011-12.

[47] Due to the sensitive nature of the items on the Political Tolerance survey, we only administered the scale to students in grades 7 through 12 in the fall of 2014 who did not indicate a disability on their original LSP application.

Table 2

Correlation matrices of non-cognitive skills, tolerance, and achievement growth measures, by age group

	Grit	Locus of Control	Self-esteem	Political Tolerance	Res. math gain	Res. ELA gain	Res. science gain
Panel A: Phone survey sample with complete responses (N=229)							
Locus of Control	0.43***						
Self-esteem	0.37***	0.53***					
Political Tolerance	---	---	---				
Res. math gain	-0.02	-0.02	-0.03				
Res. ELA gain	0.06	0.08	0.05	---	0.46***		
Res. science gain	0.04	0.00	-0.02	---	0.30***	0.44***	
Res. social studies gain	0.09	-0.07	-0.03	---	0.27***	0.39***	0.31***
Panel B: Including political tolerance (N=177)							
Locus of Control	0.47***						
Self-esteem	0.34***	0.56***					
Political Tolerance	-0.04	0.17**	0.09				
Res. math gain	-0.07	-0.05	-0.09	0.03			
Res. ELA gain	0.09	0.09	0.06	0.02	0.40***		
Res. science gain	0.10	0.07	-0.05	-0.03	0.28***	0.39***	
Res. social studies gain	0.06	-0.06	-0.05	-0.11	0.21***	0.42***	0.32***

*** - p<.01, ** - p<.05, * - p<0.10

Note. Samples restricted to students with complete responses across all measures.

Source. Authors' calculations.

Before moving on to more complex analysis, it is helpful to examine the raw distributions of our measures between students awarded an LSP scholarship and those not awarded to get a sense of how the two groups compare. Figure 1 separately plots kernel density estimates of the distributions for each of our four measures for students awarded and not awarded an LSP in 2012-13. While the plots in Figure 1 do not control for student demographics and achievement, they are nevertheless informative. In particular, the similarity between the two distributions in each graph is quite striking, suggesting little differences between the two groups. This is confirmed by Kolmogrov-Smirnov tests, which fail to reject the null of similar distributions in each case (Grit: $p = .29$; Locus of Control: $p = .35$; Self-esteem: $p = .41$; Political Tolerance: $p = .55$).

In general, the results presented in Figure 1 do not suggest strong differences in non-cognitive skills and political tolerance between the two groups of students after two years of potential program participation. Nevertheless, these findings are based on simple comparisons between the two groups. Next, we examine whether the null findings presented in Figure 1 persist when controlling for observational differences between the two groups using multiple regression analysis.

Tables 3 and 4 present results of regression models designed to improve model precision by controlling for various baseline characteristics. Table 3 presents results for models focusing on Grit and Locus of Control and Table 4 presents models focusing on Self-esteem and Political Tolerance. In both tables, columns 1 and 4 present simple models analogous to the distributional analysis presented in Figure 1. Columns 2 and 5 include controls for student demographics along

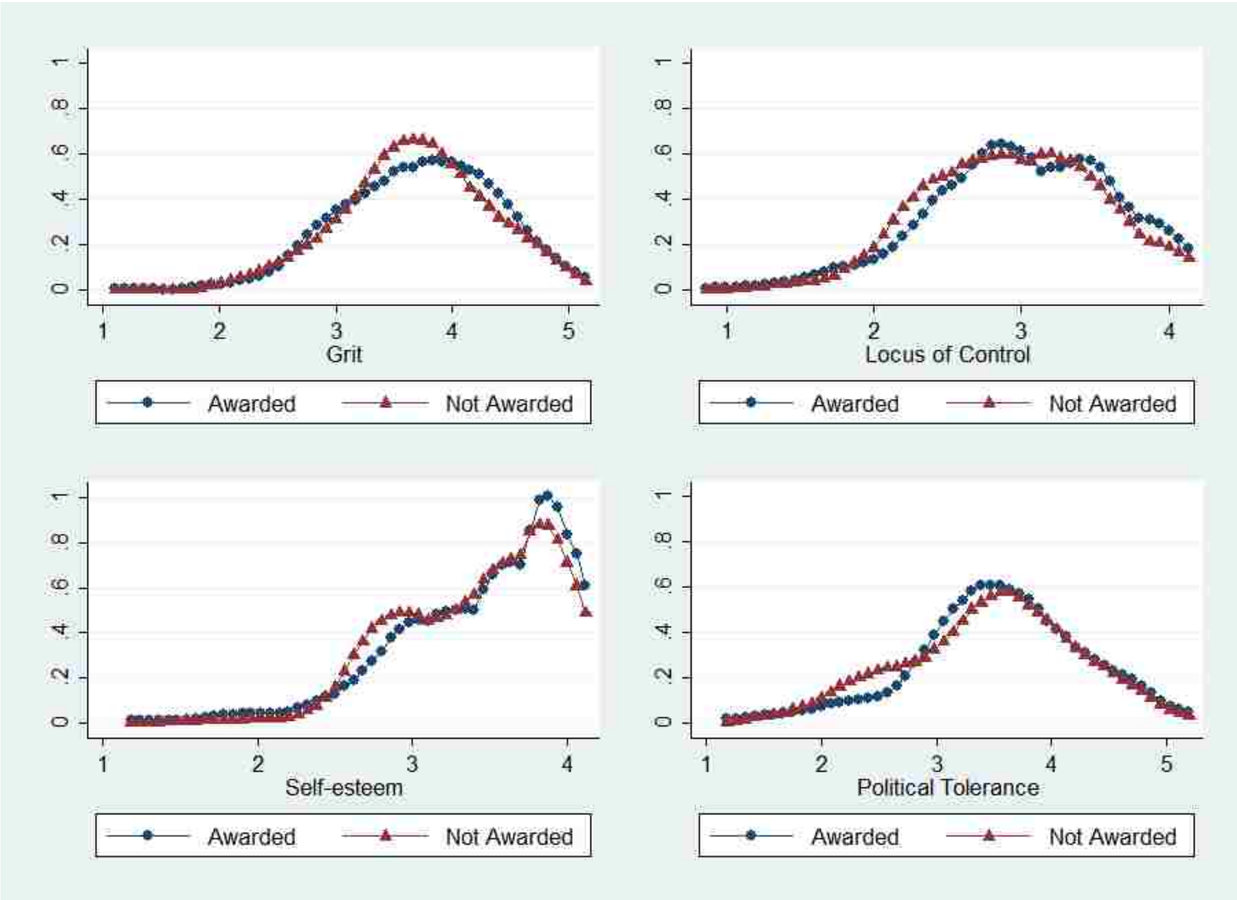

Figure 1. Kernel density estimated distributions of non-cognitive skills and tolerance measures comparing students receiving an LSP scholarship and those who did not receive a scholarship.

with fixed effects for grade and the number of school preferences offered at application.[48] Columns 3 and 6 additionally include controls for student math and ELA achievement in the 2011-12 school year. These analyses are limited to the subset of students in our sample who took either the Louisiana LEAP or iLEAP exam in grades 3 through 7 in that year.

In general, the results presented in Tables 3 and 4 suggest limited differences between students receiving and not receiving an LSP scholarship on all measures. Even after controlling for several baseline covariates, the general story of insignificant differences between the two

[48] Families could offer up to 5 school preferences on their application. In order to control for unobservable differences between families offering more or fewer school preferences, we include the total number of choices offered as a vector of dummy variables.

groups suggested in Figure 1 persists. At the same time, it is important to note that the results presented in Tables 3 and 4 are quite noisy. The reported standard errors--especially for the LSP award identifier--are fairly large across all models, typically equal in magnitude to the coefficient estimate itself. Similarly, while overall model precision generally improves with the inclusion of additional covariates, all models perform quite poorly in parsing away error variance as none of the adjusted R-squared values surpass .09. While we expected somewhat noisy results, given the lack of studies validating the included scales via phone surveys or in samples of children as young as some of those included in our sample, the results presented in Tables 3 and 4 do not give us strong confidence in these models.

Finally, an examination of the estimated coefficients for the baseline covariates in Tables 3 and 4 reveals some interesting relationships in our sample. For example, females report higher levels of GRIT, but do not differ substantially from males on the remaining measures. Moving in the last two years is associated with lower levels of Grit and Self-esteem, but higher levels of Political Tolerance. Finally, student achievement has little predictive value for the set of non-cognitive skills measures; however students with higher baseline math achievement appear to be less tolerant than other students and students with higher baseline ELA achievement appear to be relatively more tolerant.

In summary, the results presented in this section do not suggest that students awarded an LSP scholarship differed from students who did not receive a scholarship on measures of non-cognitive skills and political tolerance two years later. These results hold both in sample comparisons of scholarship receipt status, as well as in more complex analyses controlling for several covariates collected at baseline.

Table 3

Regression adjusted relationships between Grit and Locus of Control and LSP scholarship receipt

	GRIT			Locus of Control		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Awarded	0.04	-0.02	0.01	0.06	0.04	0.04
	(0.04)	(0.05)	(0.07)	(0.04)	(0.04)	(0.06)
Female		0.14***	0.06		0.00	-0.05
		(0.04)	(0.07)		(0.04)	(0.06)
Black		0.09	0.31		0.24*	0.41*
		(0.10)	(0.27)		(0.13)	(0.24)
White		-0.11	-0.14		0.17	0.15
		(0.12)	(0.30)		(0.15)	(0.25)
Hispanic		0.27*	0.31		0.33**	0.48
		(0.15)	(0.32)		(0.17)	(0.29)
Special Ed		-0.23**	-0.69***		-0.11	-0.16
		(0.11)	(0.18)		(0.10)	(0.19)
Moved		-0.10**	-0.10		0.03	0.09
		(0.05)	(0.09)		(0.04)	(0.07)
Mom's education						
Finished high school		0.12	-0.11		-0.03	-0.10
		(0.11)	(0.15)		(0.11)	(0.10)
Went to college but did not finish		0.21*	-0.05		0.12	0.07
		(0.11)	(0.15)		(0.11)	(0.10)
Finished college		0.18	-0.08		0.05	0.08
		(0.11)	(0.15)		(0.11)	(0.10)
Std. math achievement, 2011-12			0.05			0.05
			(0.05)			(0.05)
Std. ELA achievement, 2011-12			0.02			0.07
			(0.05)			(0.05)
Grade FE		X	X		X	X
Choices offered FE		X	X		X	X
N	999	924	330	999	924	330
Adj. R-squared	0.00	0.04	0.06	0.00	0.04	0.09

*** - p<.01, ** - p<.05, * - p<0.10

Note. Math and ELA achievement has been standardized to the state testing distribution by grade for students taking the iLEAP or LEAP exams in grades 3 through 7 in 2011-12. Heteroskedasticity robust standard errors are presented in parentheses.

Source. Authors' calculations.

Table 4

Regression adjusted relationships between Self-esteem and Political Tolerance and LSP scholarship receipt

	Self-esteem			Political Tolerance		
	(1)	(2)	(3)	(4)	(5)	(6)
LSP Awarded	0.03	-0.03	0.00	0.08	0.05	0.11
	(0.03)	(0.04)	(0.05)	(0.09)	(0.11)	(0.13)
Female		0.02	0.01		-0.05	-0.12
		(0.03)	(0.05)		(0.10)	(0.11)
Black		0.13*	0.25*		0.05	0.06
		(0.07)	(0.13)		(0.16)	(0.25)
White		0.03	0.06		-0.19	-0.07
		(0.09)	(0.15)		(0.23)	(0.31)
Hispanic		0.21**	0.39**		-0.06	-0.03
		(0.09)	(0.16)		(0.44)	(0.47)
Special Ed		-0.15*	-0.26**			
		(0.08)	(0.11)			
Moved		-0.09**	0.00		0.18*	0.13
		(0.04)	(0.06)		(0.11)	(0.11)
Mom's education						
Finished high school		0.00	0.00		-0.12	-0.17
		(0.08)	(0.11)		(0.17)	(0.22)
Went to college but did not finish		0.07	0.06		-0.18	-0.23
		(0.07)	(0.11)		(0.16)	(0.21)
Finished college		0.06	0.18*		0.00	-0.01
		(0.07)	(0.10)		(0.15)	(0.20)
Std. Math			0.04			-0.25***
			(0.04)			(0.08)
Std. Math			-0.02			0.25***
			(0.04)			(0.09)
Grade FE		X	X		X	X
Choices offered FE		X	X		X	X
N	999	924	330	247	238	211
Adj. R-squared	0.00	0.08	0.09	0.00	-0.01	0.04

*** - p<.01, ** - p<.05, * - p<0.10

Note. Math and ELA achievement has been standardized to the state testing distribution by grade for students taking the iLEAP or LEAP exams in grades 3 through 7 in 2011-12. Heteroskedasticity robust standard errors are presented in parentheses.

Source. Authors' calculations.

Robustness checks

The results presented thus far suggest that receiving an LSP scholarship was not related to student non-cognitive skills and levels of political tolerance among students participating in the first year of the program's statewide implementation. In this section, we present two analyses designed to check the robustness of our findings. First, we examine if the estimated relationships hold in the subsample of students whose scholarship award was determined by binding lotteries.[49] Next we investigate the extent to which reference group bias may be present in our Grit scale results. In general, we find limited evidence of different results in our subsample of LSP applicants participating in binding lotteries as well as little evidence suggesting response bias is of concern in our study.

Focusing on subsample of phone survey respondents who participated in binding lotteries. Our first robustness check examines if the null results identified in the observational analyses presented in the preceding section hold among a subsample of phone survey participants whose LSP scholarship award was determined by a binding lottery. As is detailed in Chapter 1, LSP scholarships were awarded to students through a matching algorithm designed to take into account student school preferences as well as a set of priorities established by the Louisiana Department of Education (LDE). While all students were subject to the matching algorithm, LSP scholarships were only awarded by lottery in cases when there were more students applying in the same priority category than seats available to the same grade in the same school. Thus, as a first check of the robustness of the null results presented in Tables 3 and 4, we examine the extent to which these findings persist in the subsample of eligible applicants participating in binding lotteries. By focusing on binding lotteries, we will be limiting our

[49] See Chapter 1 for a detailed description of the LSP scholarship award process.

sample; however we will be providing a better control for unmeasurable factors driving selection into private schooling. While we should not expect to find substantially different results in this group, such a finding would raise concerns regarding our primary analyses.

Our focus on binding lotteries requires a change in the model used to estimate differences in students receiving and not receiving an LSP scholarship. Specifically, we employ a similar two-stage least squares (2SLS) model to that presented in Chapter 1 which allows an estimation of the effect of LSP usage on the non-cognitive skill and tolerance measures of interest. In particular, we first predict the likelihood that a student enrolls in a private school using their LSP scholarship lottery outcome as a predictor along with a series of controls for demographics, baseline achievement, and individual risk set. This predicted usage variable is used as a substitute for observed usage in a model predicting one of the given dependent variables: Grit, Locus of Control, Self-esteem, and Political Tolerance. The specific 2SLS model is:

1. $E_i = \sum \pi_j R_{ji} + \delta T_i + \boldsymbol{X\beta} + \boldsymbol{A\rho} + u_i$

2. $Y_i = \sum \alpha_j R_{ji} + \tau \widehat{E}_i + \boldsymbol{X\gamma} + \boldsymbol{A\theta} + \epsilon_i$

Where:

- *R* is a fixed effect for a student's first choice school lottery or "risk set"
- *E* is a variable indicating if a student used an LSP scholarship to enroll in a private school
- *T* is a variable indicating if a student received an LSP scholarship to their first choice school
- *Y* is one of the four outcome measures of interest in this study: Grit, Locus of Control, Self-esteem, and Political Tolerance

- X is a vector of student demographics[50]
- A is a vector of variables capturing student achievement in 2011-12[51]

Table 5 presents the results of the 2SLS estimations of the differences between LSP scholarship users and other students on our non-cognitive skills and political tolerance measures. Column 1 presents results for simple models that only include risk set fixed effects; column 2 provides the results from specifications that additionally control for student demographics; and column 3 presents results for models additionally controlling for student baseline achievement. The latter models are restricted to the subset of students who took the Louisiana state assessments in grades 3 through 7 in 2011-12. Across all models, the results from first stage regressions suggest winning an LSP scholarship is highly predictive of use: LSP winners, on average, are over .85 percentage points more likely to enroll in a private school across all models; and the reported joint-F statistics meet Staiger and Stock's (1997) recommended threshold of 10.

The results presented in Table 5 do not strongly differ from those presented in Tables 3 and 4: students using an LSP scholarship to their first choice school do not differ for the most part from other students in nearly every model. The lone exception is that we find that LSP scholarship users on average report significantly higher scores on the Locus of Control scale in a model accounting for lotteries, student demographics, and student baseline achievement. This finding is somewhat surprising, given the overall insignificant results observed for the companion model presented in Table 4.

Table 5

[50] Demographic controls include gender, race/ethnicity, an indicator of student mobility, mother's education, and variables capturing the number of school preferences offered at application.
[51] Regressions including student achievement are restricted to students who took the iLEAP or LEAP exam in math and ELA in grades 3 through 7 in 2011-12.

Student-level relationships between non-cognitive skills and tolerance and LSP scholarship receipt in binding lotteries

	without covariates	+ demographic controls	+ student achievement
Dependent Variable	(1)	(2)	(3)
Grit	0.05	-0.02	0.12
	(0.09)	(0.09)	(0.16)
Locus of Control	0.04	0.03	0.18*
	(0.08)	(0.08)	(0.11)
Self-esteem	-0.06	-0.09	-0.02
	(0.07)	(0.08)	(0.12)
Model summary			
N	639	587	202
Risk sets	280	264	105
First stage joint F	974.3	78.6	23.8
Political tolerance	0.00	-0.08	-0.03
	(0.20)	(0.23)	(0.27)
N	157	150	136
Risk sets	83	79	76
First stage joint F	229.3	17.9	16.4

*** - p<.01, ** - p<.05, * - p<0.10

Note. All models include risk set fixed effects. Across all models, winning an LSP scholarship is highly predictive of use: all estimated coefficients on LSP awarded are over .85 and have reported p-values of less than .001.
Source. Authors' calculations

There are at least two reasons for this discrepancy. First, it is important to note that the focus on binding lotteries, in addition to the requirement of baseline achievement data in the model in question, restricts the sample on which this result is based (sample size of 202 compared to a sample size of 330 in the primary analysis). In addition, it is important to recognize that the results presented in Table 5 are based on local average treatment effects—or the estimated effects of LSP usage for those students whose treatment assignment influences their take up. In contrast, the observational analyses presented in Table 3 estimate relationships based on the mere scholarship award outcome. Nevertheless, while these points may explain the significant finding for Locus of Control, the generally insignificant results presented in Table 5 generally corroborate the insignificant findings presented in Tables 3 and 4.

Reference group bias and the Grit scale. Two recent analyses of charter schools (Dobbie & Fryer, 2014; West et al., 2014) have provided evidence suggesting that observed differences in levels of Grit among charter school students may be explained by students facing school environments with higher behavioral expectations. In both cases, reference group bias was motivated by divergent findings between the included non-cognitive skills measures and other positive outcomes.[52] In contrast, our results largely suggest insignificant differences between the two groups of students after two years. It is possible, however, that reference group bias has potentially played a role in the observed findings: with a positive or negative shift occurring that may exactly offset any gains or losses associated with the program. While this seems unlikely, we explore this possibility in regards to Grit.

King et al. (2004) note that reference group bias has long been an important issue facing survey data. In particular, individuals in different groups may potentially respond differently to the same set of questions because their group-based experiences have led them to comprehend the same question in different ways. Naumann and John (2013), for example, compare conscientiousness ratings with undergraduate GPAs in a sample of Anglo-American and Asian-American undergraduates at UC Berkley. In general, they find Anglo-American undergraduates reporting significantly higher conscientiousness scores than Asian-American students while also posting lower GPAs on average; however this relationship disappears when the researchers asked

[52] Dobbie and Fryer (2014), for example, find large positive effect estimates for achievement and significant negative findings on the Grit scale. West et al. (2014), find positive relationships between student level measures of Grit and achievement that do not appear at the school level.

students to compare themselves to an explicit reference group (eg: "a typical Asian-American student").

We incorporated a series of vignettes designed to capture the construct of Grit in our survey in order to explore the extent to which reference bias could be an issue in our study.[53] Specifically, survey participants were asked to respond if the individual in the following vignettes was “Very much like [the participant]”, “Mostly like [the participant]”, “Somewhat like [the participant]”, “Not much like [the participant]”, or “Not like [the participant] at all”:

1. *Collin almost never finishes his homework. He stops working on a problem to watch TV or to play with his friends. How much like you is Collin?*
2. *Riley always gets his homework done, even if he has to stay up late or avoid playing with his friends. He even spends a lot of time trying to answer hard questions that confuse him. How much like you is Riley?*

By asking individuals the extent to which they self-identified with the vignettes, we are unable to use these responses as anchors in an attempt to address selection bias (King et al., 2004). On the other hand, by examining the distribution of responses to these vignettes between students who received an LSP and those who did not, we can use their responses to establish the extent to which reference bias may be an issue in our study. For example, if we were to find LSP awardees with low Grit scores tend to strongly identify with Collin (least gritty) while students who do not win a scholarship tend not to identify with Collin, there would be evidence suggesting that the two groups are judging themselves according to different standards. More

[53] In their 2004 paper, King and colleagues develop a method to address potential reference group bias by using individual responses to a series of vignettes on the topic to adjust an individual’s responses to account for their differing environments. As the authors note, “the idea is to recode the categorical self-assessment relative to the set of vignettes.” (p. 195) This method is growing in popularity among researchers administering surveys across many groups of individuals (Zamarro, Vonkova, & DeBerg, 2015).

generally, large differences in distributions of responses to these vignettes between LSP receivers and non-receivers with similar Grit scores would be indicative of reference group bias. Table 6 compares the distribution of responses between LSP scholarship receivers and non-receivers for the most and least gritty survey respondents. Panel A focuses on the least gritty students—or those in the bottom quartile of observed Grit scores—and Panel B focuses on the most gritty students—those in the top quartile of Grit scores. For each group, we present the distribution of responses to the "least gritty" vignette (Collin) and the "most gritty" vignette (Riley). Differences in the response distributions between the students receiving and not receiving LSP scholarships would provide evidence suggestive of a response bias effect. In contrast, we observe little evidence of divergent distributions in Table 6.

The distributions in Panel B (student with the highest Grit scores) for the most part are shifted in the expected direction: both groups of students strongly identify with Riley (most gritty) and rarely identify with Collin (least gritty). The distributions for students with the lowest reported Grit scores, in contrast, are relatively flatter. Somewhat surprisingly, a large percentage of both LSP awardees and non-awardees indicate that they are very unlike Collin (least gritty). This potentially highlights the need to be wary of social desirability bias when examining survey response data (Paulhus, 1991). Nevertheless, the most important feature of Table 6 is that the distributions of responses between students who received an LSP scholarship and those who did not receive a scholarship do not differ strongly across all models. This is confirmed with Chi-square tests of distributional independence, which fail to reject the null hypotheses in almost every case. The lone exception for responses to the least gritty vignette (Collin) students with the lowest Grit scores, in which the null of independence is rejected at the 5 percent confidence level

Table 6

Distribution of responses to vignettes among top and bottom quartile of LSP scholarship awardees and non-awardees

Vignette	LSP status	N	*"Very much like me"* (1)	*"Mostly like me"* (2)	*"Somewhat like me"* (3)	*"Not much like me"* (4)	*"Not like me at all"* (5)
Panel A: Bottom Grit quartile							
Collin - Least Gritty	Not awarded LSP	53	25%	8%	21%	17%	30%
	Awarded LSP	150	11%	11%	23%	8%	47%
Riley - Most Gritty	Not awarded LSP	53	26%	17%	32%	8%	17%
	Awarded LSP	150	33%	23%	24%	7%	13%
Panel B: Top Grit quartile							
Collin - Least Gritty	Not awarded LSP	74	1%	0%	8%	11%	80%
	Awarded LSP	188	1%	0%	7%	7%	86%
Riley - Most Gritty	Not awarded LSP	74	72%	12%	15%	1%	0%
	Awarded LSP	188	68%	15%	9%	3%	6%

Note. Chi-square tests of independence of distributions fail to reject the null hypothesis in all but one case: students scoring in the bottom Grit quartile responses to the least gritty vignette (Collin) are significantly different at the 5 percent level ($\chi^2(4)$=11.5, p = .02).
Source. Authors' calculations.

($\chi^2(4)$=11.5, p = .02). Thus, while other studies have suggested that reference group bias may influence survey-based measures of student non-cognitive skills, we find little evidence that our earlier analyses suffer from such bias in our data.

Conclusion

The results presented in this paper represent the first attempt to examine differences in non-cognitive skills and political tolerance in a private school voucher program. In general, our findings do not suggest that students who were awarded an LSP scholarship differed significantly from students who did not receive a scholarship award two years after initial assignment on scales measuring individual Grit, Locus of Control, Self-esteem, and Political Tolerance. These findings are robust to several alternative specifications; and we furthermore find little evidence suggestive of reference group bias.

This research comes at an important time for the evaluation of educational interventions. Non-cognitive measures are increasingly being used in education evaluations (Dobbie & Fryer, 2014; Tough, 2012; West et al. 2014); however researchers continue to recommend caution in expanding their use in policy evaluation citing the possibility of reference bias and other issues that may produce misleading results (Dobbie & Fryer, 2014; Duckworth & Yaeger, 2015; West et al. 2014). While we do not find strong evidence of reference group bias in our results, we nevertheless highlight several caveats that should be taken into account when analyzing our findings. First, it is important to note that these findings are based on a subset of individuals volunteering to participate in the phone survey, ultimately representing little more than 10 percent of eligible applicants in 2012. While we present evidence indicating that survey participants do not differ strongly from the full population of LSP applicants, we cannot rule out that our sample differs from other program participants on unobservable dimensions. Second, and perhaps most importantly, we again note that the scales used in this analysis have been validated neither for use in a phone survey nor in samples of children as young as some of those included in our survey. As we argue in the results section, these factors likely play a non-trivial role in the relatively large amounts of noise observed in our models.

Nevertheless, we believe this paper advances the knowledgebase on the effects of private school choice programs on participating students by providing the first comparison of non-cognitive skills and political tolerance measures in a school voucher program. With a growing research base demonstrating the important role of non-cognitive skills in lifelong outcomes, as well as a long standing view of the important role of education in developing civic skills, it is important for future evaluations of choice programs to explore outcomes beyond achievement and attainment. At the same time, we caution future researchers to take into account the potential threats to analysis of survey response bias, reference group bias, and issues that may inflate error variance into their evaluation design.

References

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. (2011). *Personality psychology and economics* (NBER Working Paper No. 16822). Cambridge, MA: National Bureau of Economic Research.

Baumeister, R. F., Vohs, K. D., & Trice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355.

Berliner, D. C., & Biddle, B. J. (1996). *The manufactured crisis: Myths, frauds, and the attack on America's public schools*. New York, NY: Basic Books.

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

Campbell, D. E. (2002). *The civic side of school reform: How do school vouchers affect civic education?*. Princeton, NJ: Working paper of the Center for the Study of Democratic Politics.

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., 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.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43*(4), 738–782.

Dee, T. S., & West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, *33*(1), 23–46.

Dewey, J. (1916). *Democracy and education*. New York, NY: The Macmillian Company.

Dobbie, W., & Fryer, R. G. (2014). *The medium-term impacts of high-achieving charter schools* (Working Paper).

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). GRIT: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101.

Duckworth, A. L., & Quinn, P. D. (2012). Development and validation of the Short Grit Scale. *Journal of Personality Assessment*, *91*(2), 166–174.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251.

Farkas, G. (2003). Cognitive skills and non-cognitive traits and behaviors in stratification processes. *Annual Review of Sociology*, *29*, 541–562.

Figlio, D., & Ludwig, J. (2012). Sex, drugs, and Catholic schools: Private schooling and non-market adolescent behaviors. *German Economic Review*, *13*(4), 385–415.

Gutmann, A. (2003). Assessing arguments for school choice: Pluralism, parental rights, or educational results? In A. Wolfe (Ed.), *School Choice: The Moral Debate* (pp. 126–148). Princeton, NJ: Princeton University Press.

Hanushek, E. A., & Woessmann, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation* (NBER Working Paper No. 14633). Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3).

Heckman, J. J., & Kautz, T. D. (2012). *Hard evidence on soft skills* (NBER Working Paper No. 18121). Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.

Henig, J. R. (1999). Call for choice and radial reform. In *Rethinking School Choice: Limits of the Market Metaphor* (pp. 3–25). Princeton, NJ: Princeton University Press.

Howell, W. G., & Peterson, P. E. (2002). School choice and American democracy. In *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institution.

Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–205.

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of non-cognitive factors in explaining differences in educational attainment and earnings. *Social Science Research*, *37*, 888–902.

Mocan, H. N., & Tekin, E. (2007). *Catholic schools and bad behavior: A propensity score matching analysis* (NBER Working Paper No. 9172). Cambridge, MA: National Bureau of Economic Research.

Naumann, L. P., & John, O. P. (2013). Toward a domain-specific approach to cultural differences: The influence of cultural values and reference-group standards on self-reported personality. *Unpublished Manuscript*.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–322.

Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. Boston, MA: Houghton Mifflin Harcourt.

United States Department of Education. (1986). *High School and Beyond, 1980: Sophomore and Senior Cohort Third Follow-up*. Washington, DC: National Center for Education Statistics.

Warren, J. R. (2011). Graduation rates for choice and public school students in Milwaukee, 2003-2009. School Choice Wisconsin.

West, M. R., Kraft, M. K., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). *Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling* (Working Paper).

Whitman, D. (2008). *Sweating the small stuff: Inner-city schools and the new paternalism*. Washington, DC: Thomas B. Fordham Institute Press.

Wolf, P. J. (2005). School choice and civic values. In *Getting Choice Right* (pp. 210–244). Washington, DC: Brookings Institution Press.

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., Peterson, P. E., & West, M. R. (2001). *Results of a school voucher experiment: The case of Washington, DC after two years* (Working Paper No. 01-05). Cambridge, MA: Program on Education Policy and Governance, Harvard University.

Zamarro, G., Vonkova, H., & DeBerg, V. (2015). *International comparisons of student perceptions on teacher's classroom management: Improving comparability with the anchoring vignettes method* (Paper presented at the 40th annual conference of the Association for Education Finance and Policy). Washington, DC.

Chapter 3[54]

Effects of the Arkansas Academic Challenge Scholarship on College Outcomes

Introduction

In February 2010, President Obama argued for a shift in focus in the reauthorization of the ESEA towards a K-12 education system geared at ensuring career- and college-readiness (The White House, 2010). More recently, President Obama has set a vision for America to have the highest proportion of college graduates in the world by 2020.[55] This focus on college success is merited, as economists Claudia Goldin and Lawrence Katz (2008) argue that America's higher education system has played a pivotal role in America's emergence as one of the leading economies in the world. A number of policies have been developed over the last half-century to increase college access and success. These policies include federal aid programs such as the GI Bill for military veterans and the Pell Grant program for low-income families.

Enrollment in higher education and degree completion has dramatically increased recently. According to the US Department of Education, the percentage of 18- to 24-year-olds enrolled in a degree-granting institution has increased from 36 percent in 2001 to 41 percent in 2012 (Snyder, 2014). Meanwhile, the percentage of 25-to 29-year-olds who have earned at least a bachelor's degree has increased from 28 percent in 2003 to 34 percent in 2013 (National Center for Education Statistics, 2014).

Nevertheless, while college access and graduation rates have improved over the last half-century, these improvements have not been experienced equally among all segments of the

---

[54] This paper was co-authored with Albert Cheng.
[55] Source: The White House. Available: http://www.whitehouse.gov/issues/education/higher-education/building-american-skills-through-community-colleges

population. Disparities along demographic characteristics such as income and race remain. A recent Pell Institute (2015) report found that while the college access gap between the richest and poorest American families has dropped significantly since the 1970s, there remains a sizeable gap in bachelor's degree completion. Specifically, while over 90 percent of students entering colleges from the highest income families completed a 4-year degree, only 21 percent of low-income family college entrants completed college in 2012. Likewise, by 2013, 40 percent of White 25- to 29-year olds have earned at least a bachelor's degree. Yet only 21 and 16 percent of their Black and Hispanic counterparts have done the same.

A number of policies aim to improve both the rate at which individuals attend college and the rate of successful completion. For example, recent research suggests that simple "nudges" can improve the rate at which individuals matriculate and persist in college. Castleman and Page (2013) find that college intending high school graduates randomly assigned to receive text message reminders of important college enrollment deadlines were significantly more likely to matriculate in college than their counterparts who did not receive the texts. Similarly, Castleman and Page (2014) find that freshman at two-year community colleges who were randomly assigned to receive text messages reminding them about renewing their financial aid were significantly more likely to remain enrolled through their sophomore year of college. In addition, Bettinger, Long, Oreopoulos, and Sanbonmatsu (2012) found that students of families randomly assigned assistance in completing a FAFSA application were more likely to attend and persist in college than students in families who did not receive assistance. Nevertheless, while these studies suggest simple nudges may be effective, the most prevalent policy aimed at increasing college access and success is financial aid.

Financial aid for higher education can work to improve college attendance by reducing the cost of college (Dynarski, 2008). Indeed, the cost of higher education has significantly risen over the past decade. In constant 2011-2012 dollars, the price for tuition, room, and board at public four-year institutions increased from $12,000 in 2001 to $17,000 to 2011. Prices for private four-year institutions rose from $29,000 to $34,000 over the same period (National Center for Education Statistics, 2014). Unsurprisingly, over 80 percent of students in four-year institutions between 2008 and 2012 reported receiving some type of financial aid (National Center for Education Statistics, 2014).

Aid can take several forms, but the most prevalent are loans and grants (College Board, 2013). Whereas loans provide funds upfront for later payments, grant aid represents a direct financial subsidy from the perspective of the recipient (College Board, 2013). In the 2012-13 school year, Federal loans and grants accounted for 61 percent of total aid received by undergraduate students. Institutional grants constituted another 19 percent of total aid received. In contrast, state grants accounted for only 5% of total aid (College Board, 2013). The majority of grant aid has a need-based component (over 70%), but several states have grant aid programs with eligibility requirements largely linked to student performance on standardized college readiness assessments and high school performance. In this paper, we examine the effectiveness of one such program, the Arkansas Academic Challenge Scholarship (ACS), in improving student outcomes in college.

As Cornwell, Lee, and Mustard (2005) note, there are at least three motivations for states to offer merit-aid programs: (1) increasing college enrollment, (2) incentivizing high performing high schoolers to stay in state, and (3) promoting and rewarding academic achievement. Regarding the latter point, Scott-Clayton (2012) notes that merit scholarships could improve student outcomes through two channels. First, by reducing the cost of college access, merit scholarships could help to minimize non-academic stresses in students' lives which could then translate into higher achievement. Second, merit-scholarships may improve outcomes by directly incentivizing students to maximize behaviors that are associated with college success. For example, ACS requires students to maintain a GPA of 2.5 while enrolling in at least 15 credit hours a semester in order to continue qualifying for the program. Such requirements may incentivize individuals at the margin to work harder to meet their GPA requirements in order to continue receiving ACS. As of 2012, 14 states offered merit-based scholarships with eligibility requirements covering a large portion of high school graduates (Scott-Clayton, 2012). This number does not include state programs with more rigorous academic requirements—such as Missouri's Bright Flight scholarship which provides scholarships to the top 3 percent of ACT test takers in the state.

While state financed merit-aid programs have been around since the 1980s, these programs largely came under empirical scrutiny in the early 2000s due to the increased availability of administrative data. Largely relying on empirical techniques exploiting abrupt policy changes or strong institutional knowledge of program award mechanisms, researchers have found such programs to improve the likelihood of college attendance (Cornwell, Mustard, and Sridhar, 2006; Dynarski, 2003; Kane, 2003; Scott-Clayton, 2012), persistence (Bettinger, 2004), cumulative GPA (Scott-Clayton, 2012), and likelihood of graduation (Dynarski, 2008; Scott-Clayton, 2012). At the same time, research examining these programs has not reported consistently positive findings. For example, Cornwell, Lee, and Mustard (2005) find a decreased likelihood of taking a full-time course load and an increased likelihood in enrollment in summer school classes, both among entering freshman, as students attempted to navigate the Georgia

HOPE scholarship program. In addition, while Scott-Clayton (2012) finds that West Virginia's PROMISE scholarship positively influenced entrance and graduation rates, she finds no significant impacts on four-year college persistence. Finally, several studies have found that the positive effects of state-funded merit-aid programs are largely concentrated around the programs' eligibility thresholds (Cornwell, Lee, and Mustard, 2005; Scott-Clayton, 2012).

This paper follows in the vein of prior studies utilizing knowledge of the assignment policy to examine the effects of the Arkansas Academic Challenge Scholarship (ACS) on student outcomes. Specifically, we use administrative data on current college students who became eligible for the ACS because of their academic credentials to estimate the effects of receiving the scholarship on short- and long-term college outcomes via regression discontinuity design. This study contributes to the literature on state-based merit scholarships for three reasons. First, while prior research has largely focused on the effects of merit-aid programs on entering freshman, we examine how receipt of the ACS affected college outcomes for students who were currently enrolled in college at the introduction of the program (hereafter referred to as "current students"). While we intend to examine the effects of the ACS on student outcomes for entering freshman cohorts in future iterations of this research, the current paper nevertheless adds to the literature by providing insight into how merit-aid policies affect students in the short-run. In addition, by focusing on students who were currently enrolled in college at the time of the program's expansion, as opposed to incoming Freshman, our study represents the first attempt to estimate the motivational effects of a merit-aid scholarship on student outcomes that is unaffected by any motivational effects occurring prior to enrollment in college.[56] Second, and perhaps most importantly, this paper compliments exiting research by focusing on a program with relatively

[56] For example, the mere existence of merit-aid scholarships may incentivize students to work harder in high school to meet the program eligibility requirements (Dynarski, 2004).

low eligibility requirements. In particular, while other programs impose eligibility requirements covering up to 40 percent of in-state college enrollees (Scott-Clayton, 2012), the students maintaining a 2.5 cumulative college GPA and continuously enrolled are eligible for the ACS scholarship. Finally, given the mixed findings of exiting research on state-financed merit-aid, our analysis benefits the literature by examining the effects of such programs in a new state and program context.

The remainder of this paper proceeds as follows. In the following section, we provide a more detailed description of the Arkansas Academic Challenge Scholarship. Next, we outline the quasi-experimental strategy and data we use to estimate the effects of the ACS on college outcomes. We then move on to our primary findings and a series of robustness checks. Finally, we conclude with a discussion of our findings and their implications.

The Arkansas Academic Challenge Scholarship

In November of 2008, Arkansas voters approved the first statewide lottery in Arkansas: the Arkansas Scholarship Lottery. Proceeds from the lottery were used to dramatically expand the existing Arkansas Academic Challenge Scholarship Program. Arkansas Scholarship Lottery tickets first went on sale in the fall of 2009; and the first round of lottery-funded scholarships were awarded to students in the fall of 2010 (Arkansas Department of Higher Education, 2010).

The goal of the ACS is "to provide meaningful financial help to those qualifying" (Arkansas Secretary of State, 2011). When the program was first enacted, the ACS provided relatively generous scholarships of $5,000 to students attending 4-year colleges.[57] As is shown in Table 1, the award amounts have changed over time to account for changing revenue streams for

[57] For comparison: the published tuition for the University of Arkansas-Fayetteville—the state's flagship institution-was $5,010 in the 2010-11 school year (source: National Center for Education Statistics, Integrated Postsecondary Education Data Systems (IPEDS): http://nces.ed.gov/ipeds/).

the lottery and the increasing application rate in recent years. To date, the Arkansas Scholarship Lottery has funded scholarships for over 130,000 Arkansas residents to attend 2- or 4-year colleges.[58]

Table 1

ACS award payouts by cohort year

Year First Awarded	4 Year School	2 Year School
2010-11	$5000	$2500
2011-12	$4500	$2250
2012-13	$4500	$2250
2013-14	$2000 (1st year)	$2000
2013-14	$3000 (2nd year)	$2000
2013-14	$4000 (3rd year)	$2000
2013-14	$5000 (4th year)	$2000

Source: Arkansas Department of Higher Education (http://scholarships.adhe.edu/scholarships-and-programs/a-z)

At its inception—and the period under scrutiny in this analysis—there were three categories of students eligible for the ACS.

1. Students who received the ACS prior to the fall of 2010 ("Prior Recipients")
2. First-time freshman entering college in the fall of 2010 or afterwards ("Traditional Recipients")
3. Students who entered college before the fall of 2010 as first-time freshman who were continuously enrolled in consecutive semesters prior to the fall of 2010 ("Current Achiever Recipients")

In this paper, we focus on estimating the impacts of ACS on college outcomes for the final category of eligible students: Current Achievers. To be eligible for the ACS, Current Achievers had to satisfy several requirements. First, they had to fill out a Free Application for Federal Student Aid (FAFSA) application. Second, they had to be continuously enrolled for at

[58] Source: http://myarkansaslottery.com/about/scholarships

least 12 hours each semester since their freshman year. Finally, Current Achievers had to have a cumulative GPA of 2.5 at the time of their application for the program. As we describe in the following section, we leverage detailed data on students at a large Arkansas university during this time period as well as knowledge of the strict eligibility requirements for students to examine the effects of the ACS on Current Achiever college outcomes using a regression discontinuity design.

While our analysis focuses on the final category of students, it is important to point out that the requirements for Traditional Recipients of the ACS are particularly low among state-based merit scholarships. In particular, Traditional Recipients must [a] complete a FAFSA, [b] have a high school GPA of 2.5 or higher, and [c] scored a 19 or higher on the ACT (or achieved an equivalent score on another college readiness assessment) (Arkansas Secretary of State, 2011). In the future, we intend to expand on the current analysis by examining the effects of ACS receipt on students around these relatively low performance thresholds.

Method

Our goal is to estimate the impacts of receiving the Arkansas Academic Challenge Scholarship (ACS) on college-going outcomes for students who were currently enrolled in college during the expansion of the program in the fall of 2010. As Dynarski (2008) notes, it is generally challenging to estimate the causal effects of financial aid on college enrollment, persistence, and attainment because of the likely unobservable factors influencing these outcomes. Fortunately, ACS's strict eligibility criteria for Current Achievers—full-time enrollment and cumulative GPA of 2.5—allows us to estimate the causal impact of the program on students near the eligibility thresholds. Specifically, we estimate ACS impacts by utilizing knowledge of the program's eligibility requirements in a regression discontinuity design (Kane, 2003; Scott-Clayton, 2012; van der Klaauw, 2002).

In this section, we describe the data and analytical strategy used to estimate ACS impacts on college outcomes. In addition, we provide initial graphical analyses supporting our RDD method.

Before moving on to a description of our data and analytical strategy, it is important to highlight the following points. First, while the ACS uses strict eligibility requirements to allocate scholarships, students were still required to apply for an ACS. Thus, we estimate ACS impacts using a fuzzy regression discontinuity design (FRDD) by first predicting the likelihood of receiving an ACS given one's eligibility status. Second, while the dual eligibility requirements allow for comparisons of several types of ACS receivers and non-receivers, our primary analysis focuses exclusively on a comparison of individuals who have met the full-time enrollment qualification.

Data

We estimate the impacts of the ACS on college outcomes using detailed administrative data on students at a large Arkansas university (LAU). These data include student level demographics, high school qualifications, information on credit accumulation, cumulative GPA, and major by semester, and family financial data.

While our data include information on all students enrolled at LAU by freshman cohort-year, our analysis examines a restricted sample of students. First, because we are interested in ACS impacts on Current Achievers, we focus on students entering as freshman in years 2007-08 through 2009-10. Students in these cohorts would be entering their senior, junior, and sophomore years when the ACS was expanded by the lottery scholarship in the fall of 2010-11. Second, we restrict our analysis to students identified as applying to LAU from within Arkansas as out-of-state applicants are not eligible for the scholarship. Finally, we restrict our sample to students who filled out a FAFSA at the time of their initial application because, while the ACS does not have a set income threshold, students are still required to submit a FAFSA to be eligible for the program.[59] After making these selections, we are left with an analytical sample comprising 331 students from cohort year 2007-08, 464 from cohort year 2008-09, and 745 from cohort year 2009-10.

Analytical strategy

This section introduces our analytical strategy for estimating the effects of the Arkansas Academic Challenge Scholarship on college outcomes. Our broad goal is to utilize the eligibility requirements of the ACS to estimate the program's impact on Current Achievers—or students who were already enrolled at LAU in the fall of 2010. In this section, we describe the particular comparison group used to estimate the ACS's effects, outline the fuzzy regression discontinuity model used to estimate those effects, and describe the specific sample under scrutiny in our analysis.

Selecting a comparison group. The ACS eligibility criteria for Current Achievers are suggestive of a dual discontinuity in qualification status. Specifically, Current Achievers are eligible for the ACS if they met the continuing full-time enrollment requirement (enrolled at least 12 hours in each semester since their initial matriculation) and the cumulative GPA requirement (2.5 GPA). The dual nature of the ACS eligibility requirements suggests at least three potential comparison groups for ACS recipients. Specifically, one can compare ACS qualifiers to (1) students satisfying the GPA requirement but who failed to meet the credit hours requirement, (2)

[59] We identify students as having filled out a FAFSA if their record indicates an expected family contribution. These data are populated by LAU using FAFSA data.

students satisfying the credit hours requirement but who did not meet the GPA requirement, and (3) students who did not meet either requirement.[60]

These potential comparison groups are displayed in Figure 1, which graphs individual credit hours earned in the spring semester of 2010[61]—the last semester before the expansion of the ACS—and cumulative GPA at the end of the spring semester for in-state enrollees who had completed a FAFSA and were continuously enrolled at LAU since their initial matriculation. There are three types of LAU students presented in Figure 1: individuals who did not receive the ACS (blue crosses), individuals who received the ACS (red circles), and individuals who had received an earlier version of the ACS at any point in their college career (small black circles). While our analysis could potentially examine each of the three comparisons outlined above, we instead focus on comparison (2): students satisfying the credit hours requirement who failed to meet the GPA requirement. This comparison is highlighted with a black box in Figure 1. We have chosen this comparison because, while credit hours may appear to be a continuous variable, it is at best an ordinal variable when we restrict our analysis to a small band around the credit hour cutoff. This violates the continuity requirement of assignment variables in an RDD (Imbens & Leimux, 2008). Thus, our estimates of the effects of the ACS on college outcomes will be

[60] As Reardon and Robinson (2012) note, one can accomplish the final comparison by combining the two continuous assignment variables into a single continuous variable using a Euclidean distance transformation. For the purposes of our analysis, we could use the variable $d_i = sign(cGPA_i)\sqrt{cGPA_i^2 + cHrs_i^2}$, where cGPA is the ACS GPA requirement centered at the cutoff score of 2.5 and cHrs is the credit hour requirement centered at 12 hours.

[61] The ACS required Current Achievers to be continuously enrolled full-time (12 hours) in every semester prior to the fall of 2010. The sample presented in Figure 1 has first been restricted to only those individuals who have met the continual enrollment requirement in every semester before the spring of 2010. This allows us to effectively turn the continual enrollment requirement into a single continuous variable: credit hours earned in the spring of 2010.

driven by a comparison of individuals meeting the ACS credit hours requirements with cumulative GPAs within a small range around 2.5 points.[62]

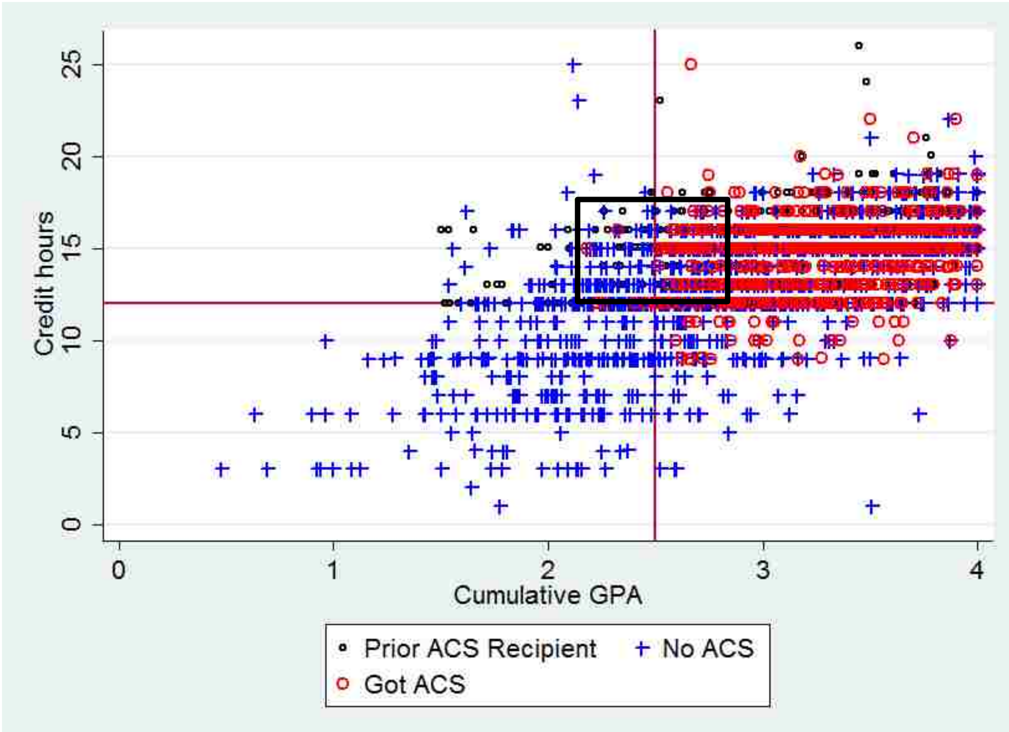

Figure 1. Dual eligibility requirements for the ACS. Current Achievers qualifying for the ACS had to have at least 12 credit hours and a cumulative GPA of 2.5 in the semester prior to the introduction of the ACS. This figure highlights the dual nature of the ACS eligibility requirements. The black box depicts the comparison under study in this analysis: individuals near the pre-ACS GPA threshold who have satisfied the credit hours requirement.

Before moving on, there are several important points to make about Figure 1. First, while they constitute a small portion of the data, prior ACS recipients (small black circles) are clearly represented in the sample presented in Figure 1. As we describe later, our analytical models control for these individuals using a dummy variable indicating if an individual ever received an ACS prior to the 2010-11 expansion. Second, there are a small number of individuals who received the ACS scholarship in the fall of 2010 who appear to have only met one of the two

[62] Previous RDD studies of the effects of financial aid on students have similarly examined impacts while conditioning on one or more assignment variables (ie: Kane, 2003; Scott-Clayton, 2012).

eligibility requirements. While our fuzzy regression discontinuity will partially account for these individuals in the first stage of our instrumental variables model, it is important to note that these individuals highlight potential issues with our current assignment variables.[63] As such, we would like to stress that the findings presented here are preliminary and should be taken with caution.

The fuzzy regression discontinuity model. Our analytical strategy is to leverage our knowledge of the ASC eligibility criteria to examine the effects of the program on Current Achievers via a regression discontinuity design. This technique has been commonly used in papers examining the effects of financial aid on college enrollment (Kane, 2003; Scott-Clayton, 2012; van der Klaauw, 2002) and outcomes (Scott-Clayton, 2012). Nevertheless, while ACS eligibility is determined by well-defined cutoffs, it is not generally true that ACS receipt is solely a function of one's qualification status. Specifically, although students may meet each of the eligibility requirements outlined above, it is not guaranteed that all qualifying students will actually apply for the program. As such, we identify the impact of the ACS on students using a fuzzy regression discontinuity design in which we predict ACS receipt using one's qualification status.

Specifically, we employ the following two-stage least squares (2SLS) model to estimate the effects of ACS using predicted—rather than observed—ACS receipt as our independent variable of interest:

[63] Only five individuals who do not meet both qualifications but nevertheless received an ACS scholarship make it in to our final analytical sample. We have tested the extent to which these observations influence the results presented in Tables 4 through 6—our primary ACS effect estimates—by re-running the models while excluding these observations. In all cases, the estimated coefficients are in the same direction, but in a small number of cases, removal of these observations nudges the coefficients estimates over the statistically significant threshold. In general, however, the estimates are not substantially different and we therefore conclude that inclusion of these observations is not problematic for our estimation.

1. $R_i = a + bQualify_i + f(cGPA_i)'c + f(cGPA_i xQualify_i)'d + X_i'g + e_i$

2. $Y_i = \alpha + \beta\widehat{R_i} + f(cGPA_i)'\delta_1 + f(cGPA_i xQualify_i)'\delta_2 + X_i'\gamma + \epsilon_i$

where R_i indicates observed ACS receipt, $Qualify_i$ indicates if an individual qualifies for ACS, $f(.)$ is a second-order polynomial function of the centered pre-ACS GPA assignment variable (*cGPA*), and X is a vector of demographic control variables capturing student gender, ethnicity, and financial resources. As was described in the previous section, all models first condition on having met the ACS credit hours requirement of 12 hours. If one's qualification status successfully predicts the probability that they take up a scholarship and our model sufficiently captures the underlying relationship between the assignment variable and our outcomes of interest and qualification status, then the parameter estimate on $\widehat{R_i}$ represents the causal effect of receiving an ACS for those individuals near the assignment variable's threshold.

Outcome variables of interest. We are interested in estimating the impact of receiving the ACS on both short- and long-term college outcomes. In particular, we are interested in two broad categories of outcomes: measures collected after one year of program participation and measures collected at the end-of-college. We briefly describe the seven measures we use to capture these outcomes here.

GPA after 1 year. Our first outcome measure is a student's cumulative GPA at the end of the spring semester of 2011. While we refer to this variable as "GPA after 1 year", it actually reflects one's GPA after two semesters of program participation. Following Scott-Clayton (2012), we impute for missing values of GPA in this semester using previously observed cumulative GPA values for the student.

Likelihood of persisting 1 year. We identify a student as having persisted 1 year in college if they are still enrolled in the spring semester of 2011. Because this is a binary variable, we estimate the effects of ACS on student persistence using instrumental variables in a probit specification.

Credit hours accumulated after 1 year. We calculated credit hours accumulated as the difference between one's credit hours earned at the end of the 2011 spring semester and their credit hours accumulated at the end of the 2010 spring semester. As with GPA, we impute for missing credit hour values using credit hours accumulated in earlier semesters.

Credit hours accumulated after 2 years. As an intermediary step between our year 1 outcomes and our end-of-college outcomes, we examine the effects of the ACS on student credit hour accumulation after two years of ACS availability. This variable is calculated as the difference between one's credit hours accumulated at the end of the 2012 spring semester and their hours accumulated in the spring of 2010. Students entering college in the fall semester of 2007—the oldest cohort in our data—are expected to be starting their fourth year of college in the fall of 2010. Analyses examining the effects of the ACS on credit hour accumulation after two years will exclude these students because there is no expectation that they will still be enrolled at LAU two years after the fall of 2010.

Final GPA. In addition to examining one's GPA after one year of ACS availability, we estimate the impact of ACS receipt on a student's last-observed GPA. As with the short-run GPA measure, we impute for missing values of GPA using earlier observed values (Scott-Clayton, 2012).

Likelihood of graduating within 4 and 5 years. Our final two outcome measures estimate the impact of ACS receipt on graduating within four and five years. These are binary variables collected from LAU's administrative data indicating if a student received a diploma by their 9th or 11th semester[64], respectively. As with persistence, we estimate the effects of ACS receipt on these measures using instrumental variables with a probit specification to account for the variables' binary nature.

Figure 2 provides a first look at how ACS qualification is related to our outcome variables of interest. Specifically, Figure 2 presents graphs of simple regressions of the seven outcome measures against cumulative GPA in the spring of 2010 (hereafter pre-ACS GPA), which has been centered at the ACS cutoff of 2.5 GPA points. All models condition on meeting the ACS credit hours requirement and control for the underlying relationship between outcomes and pre-ACS GPA using a quadratic specification—the same specification that we employ in our primary analytical models. In addition, the graphs are restricted to a pre-ACS GPAs ranging between 2.2 and 2.8 points (or a band of 0.30 GPA points). Because ACS qualification does not perfectly predict receipt, these graphs represent intent-to-treat estimates.[65]

The results presented in Figure 2 provide some evidence suggesting negative impacts on persistence, credit hour accumulation, and the likelihood that one graduates within 4 and 5 years. On the other hand, the graphs for both GPA measures do not reflect strong ACS impacts. Finally, it is important to note that the data presented in these graphs are quite noisy—especially for the binary outcomes of interest. While our more sophisticated regression models will help to address some of this noise using control variables, we nevertheless want to again highlight the need for applying caution when interpreting our results.

[64] Analyses focusing on the probability of graduating within 5 years exclude students in the 2009-10 cohort because we cannot observe this outcome for these students as our data end in 2013-14.

[65] The difference between receipt and qualification suggests that one can get a good approximation of the treatment-on-treated impact estimates by dividing the intent-to-treat estimates by 0.40.

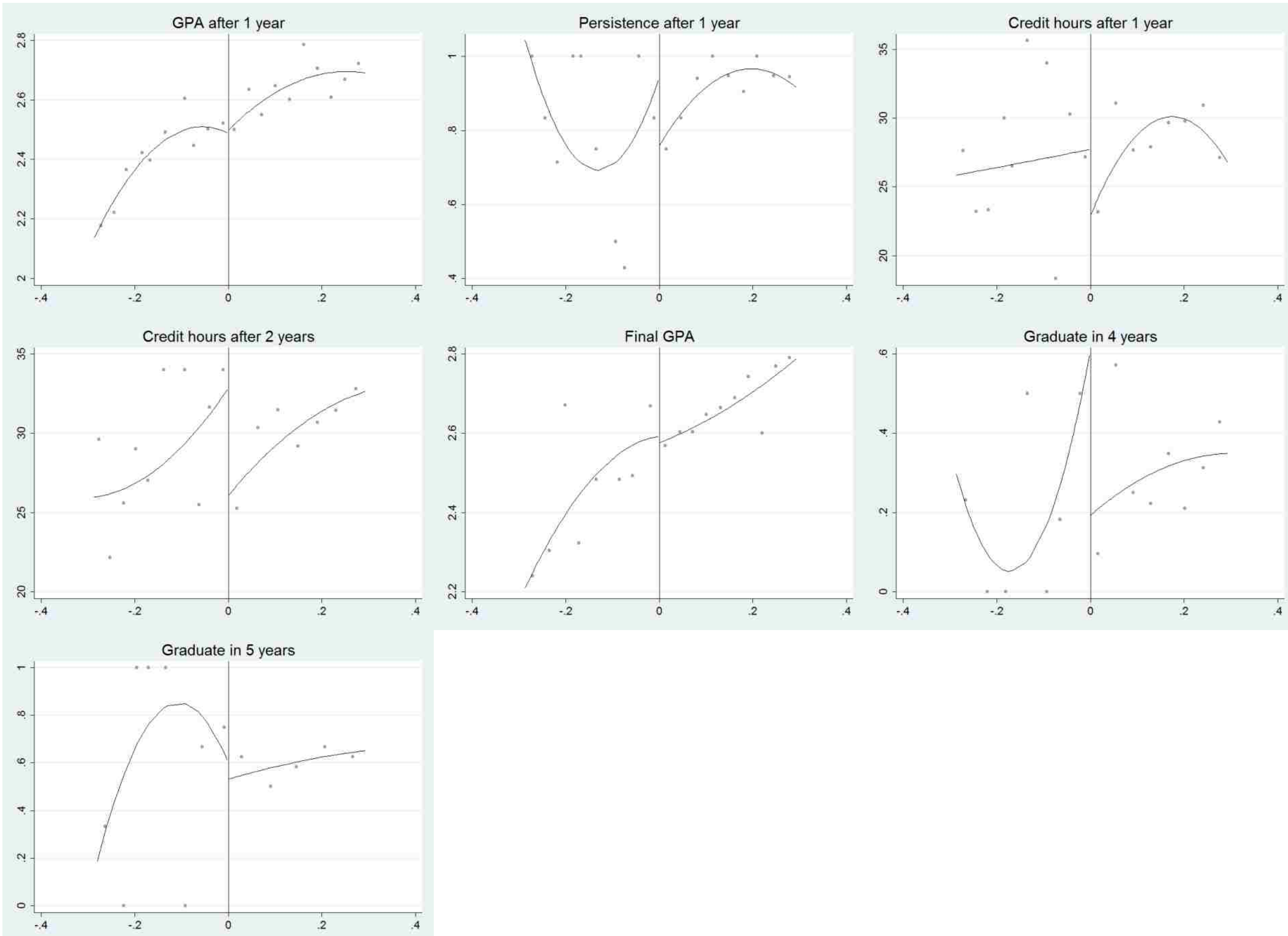

Figure 2. Outcome variables by centered pre-ACS GPA assignment variable, conditional on meeting the ACS hours requirement. All graphs employ a quadratic specification for the assignment variable and are restricted to our primary analytical range of 2.20 to 2.50 pre-ACS GPA points.
Source: Authors' calculations.

Density of pre-ACS GPA assignment variable. As a last check of our regression discontinuity specification, we examine in Figure 3 the distribution of our pre-ACS GPA assignment variable in the 2007-08, 2008-09, and 2009-10 LAU cohorts. Specifically, Figure 3 presents the density of individuals within .05 GPA point bins to the left and right of the 2.5 point cutoff. Ideally, one would examine a relatively smooth density to the left and right of the cutoff; as a discontinuous density is suggestive of gaming of the assignment variable (Imbens & Lemieux, 2008; Scott-Clayton, 2012). Unfortunately, the results presented in Figure 3 indicate a relatively large increase in the grouping of individuals scoring at or slightly above a 2.5 in the spring of 2010. This is largely accounted for by nine students with cumulative GPAs exactly equaling 2.5 (seven in the 2009-10 cohort and 2 in the 2008-09 cohort). This discontinuous jump in pre-ACS GPA to the right of the performance threshold may indicate that students worked harder in the spring of 2010 to increase their GPAs in order to meet the ACS requirements. Nevertheless, while such gaming would be problematic for our analysis, we believe this is unlikely to be a problem as the ACS eligibility requirements were officially passed in April, 2010, leaving little opportunity for students to strategically manipulate their GPAs (Arkansas Department of Higher Education, 2010). In short, the discontinuity observed in Figure 3 likely represents a random distortion in the data rather than individual gaming.

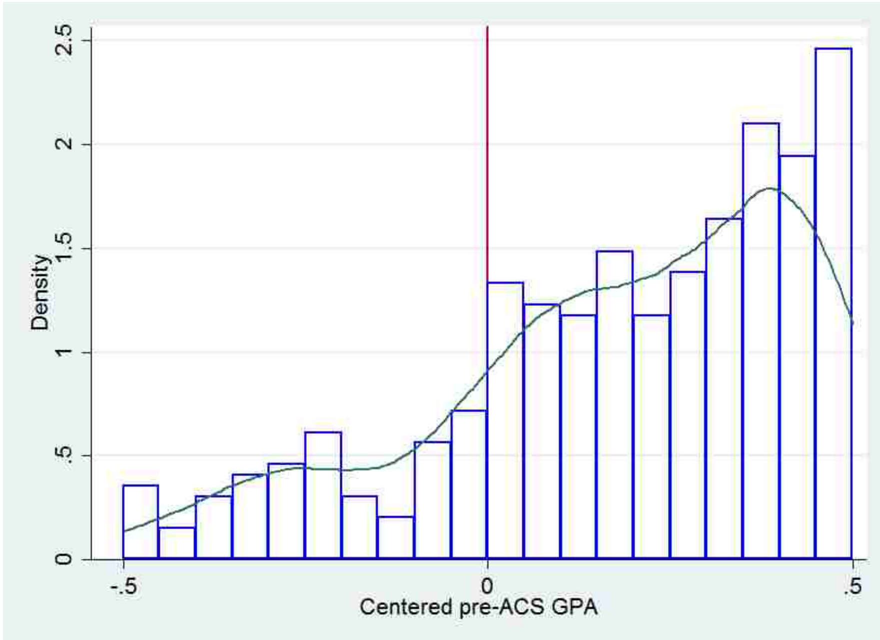

Figure 3. Graph of density by Centered pre-ACS GPA assignment variable with kernel density overlay. Bins represent .05 GPA point gaps. All individuals have met the ACS hours threshold. *Source.* Authors' calculations.

Sample

Table 2 presents descriptive statistics for three groups of students for a number of demographic characteristics. The first three columns of Table 2 present data for our primary analytical sample: students entering LAU in the 2007-08, 2008-09, and 2009-10 school years that had applied from within the state of Arkansas, had submitted a FAFSA application, met the ACS credit hours requirement, and had pre-ACS GPAs ranging between 2.2 and 2.8 GPA points. The next three columns ("Relaxed GPA and Hours Requirements") represent all in-state applicants in the 2007-08 through 2009-10 cohort years who submitted a FAFSA application. The final three columns ("Relaxed Residency Requirement") present descriptive data for all applicants who submitted a FAFSA, independent of their original state of residence.

In general, the students included in our analytical sample are slightly more likely to be Black, slightly less likely to have a parent who attended college, and have lower pre-college

Table 2

Descriptive statistics for our analytical sample and other comparison groups

	Analytical Sample			Relaxed GPA and Hours Requirements			Relaxed Residency Requirement		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Male	321	0.53	0.50	3430	0.50	0.50	5310	0.49	0.50
Ethnicity									
Black	321	0.13	0.33	3430	0.09	0.28	5310	0.08	0.27
Hispanic	321	0.06	0.23	3430	0.04	0.20	5310	0.04	0.20
White	321	0.76	0.43	3430	0.80	0.40	5310	0.80	0.40
Other	321	0.24	0.43	3430	0.20	0.40	5310	0.20	0.40
Parent went to college	321	0.53	0.50	3425	0.57	0.49	5303	0.63	0.48
High School GPA	321	3.38	0.33	3429	3.56	0.42	5308	3.58	0.42
ACT Composite	320	23.65	3.14	3419	25.50	3.91	5290	25.59	3.85
Expected Family Contribution Percentile									
0-24	321	0.30	0.46	3430	0.27	0.44	5310	0.25	0.43
25-49	321	0.33	0.47	3430	0.30	0.46	5310	0.25	0.43
50-74	321	0.17	0.38	3430	0.19	0.40	5310	0.25	0.43
75-100	321	0.20	0.40	3430	0.24	0.43	5310	0.25	0.43
Cohort Year									
2007-08	321	0.17	0.37	3430	0.31	0.46	5310	0.31	0.46
2008-09	321	0.32	0.47	3430	0.34	0.48	5310	0.35	0.48
2009-10	321	0.51	0.50	3430	0.34	0.48	5310	0.35	0.48

Note. Individuals included in the analytical sample have submitted a FAFSA, applied to LAU from within Arkansas, met the ACS credit hours requirement of 12 hours, and had a pre-ACS cumulative GPA between 2.20 and 2.80 GPA points.
Source. Authors' calculations.

academic credentials on average. In addition, students in our analytical sample tend to have lower expected family contribution (EFC) values—a variable indicating the expected amount of school tuition a family's income can support—than students in the other groups. Finally, it is important to note that younger individuals (those from the 2009-10 cohort) compose a relatively larger portion of the sample. This is not particularly surprising, as more students in the older cohort have had more opportunities to leave LAU and more opportunities to fail to qualify due to the continual enrollment requirement. Nevertheless, it is important to note that this younger cohort composes a larger portion of our sample.[66]

Results

In this section, we present the preliminary estimates of the effects of the Academic Challenge Scholarship on college outcomes for students near the program's GPA eligibility requirements. In general, our results suggest the scholarship had limited impacts on Current Achiever GPA after one year, persistence likelihood, and credit accumulation after both one and two years. On the other hand, ACS recipients appear to have a significantly lower final GPA and likelihood of graduating within four years relative to non-recipients. Nevertheless, ACS recipients appear to catch up to non-recipients over time, as our models indicate insignificant differences in the probability of graduating within five years. In the following sections, we present the results from our preliminary analyses along with several specification tests examining the robustness of our results.

Primary RDD Analysis

As discussed above, we estimate the effects of the ACS on college outcomes using a fuzzy regression discontinuity design. In particular, we are using one's qualification status to

[66] We include a moderator analysis examining if ACS effects are differentially experienced among the cohorts as a specification check.

108

first predict ACS receipt in a two-stage least squares (2SLS) framework; and then using this predicted likelihood of receipt to estimate the impact of the scholarship on student outcomes.

Before moving on to our primary results, we present in Table 3 the results for first stage regressions using ACS qualification to predict scholarship receipt. Columns 1-4 present results for our preferred pre-ACS GPA band of .30 GPA points[67] (or GPAs ranging between 2.2 and 2.8 points), with model specifications increasing in complexity as one moves from left to right in Table 3. Columns 5 and 6 present results from fully specified models employing larger and smaller pre-ACS GPA bands which are used to check the stability of our results in a following section.

In general, the results presented in Table 3 suggest that ACS qualification is a relevant predictor of ACS receipt, with take-up probabilities ranging between 30 and 40 percentage points. In addition, we note that the first stage joint F-statistics are greater than 10 in models presented in columns 1 through 5, satisfying Staiger and Stock's (1997) recommended threshold for instrumental variable relevance. The final model, which relies on a smaller pre-ACS GPA band (.20)—and therefore a smaller sample—does not meet this threshold. We therefore again recommend caution when interpreting these results.

[67] After visually inspecting the data, we chose .30 as our pre-ACS GPA bandwidth because it provided a sufficiently large sample with which to estimate effects while also allowing for us to restrict our estimates to a small distance from the ACS GPA cutoff.

109

Table 3

First Stage Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)
Qualify for ACS	0.32**	0.40***	0.31**	0.38***	0.36***	0.40***
	(0.10)	(0.09)	(0.08)	(0.09)	(0.05)	(0.09)
Controls						
Student demographics		X		X	X	X
Family income			X	X	X	X
Observations	322	321	322	321	451	205
Joint F-Statistic	15.86	16.31	12.21	12.64	20.71	7.50
R-squared	0.24	0.29	0.26	0.30	0.29	0.25
ACS GPA Band	0.30	0.30	0.30	0.30	0.40	0.20

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA.

Source. Authors' calculations.

110

The models presented in Table 4 examine the effects of ACS receipt on student cumulative GPA and likelihood of persisting one year after ACS receipt. The parameter estimates for the persistence models (columns 5 through 8) represent marginal probabilities estimated at mean values for the included covariates. In general, the results presented in Table 4 suggest that ACS receipt has statistically insignificant, yet positive, effects on Current Achievers after one year. Specifically, our preferred models—which include demographic and financial background control variables—suggest that ACS recipients scored on average .13 GPA points higher than their counterparts after one year and were 5 percentage points more likely to be enrolled at LAU after one year. Nevertheless, the relatively large standard errors presented in Table 4 indicate that we cannot state with confidence that the true effect estimate differs from zero.

Table 5 presents results for models estimating the ACS effects on credit hour accumulation after one (columns 1-4) and two years (columns 5-8). The latter analyses are restricted to the 2008-09 and 2009-10 cohorts because the 2007-08 cohort are expected to be in

Table 4

RD Estimated ACS Effects on GPA and Likelihood of Persisting One Year After Receipt

	Outcome: Year 1 GPA				Outcome: Pr(Persist 1 Year)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted ACS Receipt	0.12	0.15	0.11	0.13	0.00	0.07	-0.02	0.05
	(0.12)	(0.13)	(0.10)	(0.11)	(0.22)	(0.19)	(0.25)	(0.20)
Male		-0.04**		-0.04**		-0.01		-0.01
		(0.02)		(0.01)		(0.03)		(0.04)
Black		-0.03		-0.01		-0.59**		-0.53***
		(0.06)		(0.07)		(0.25)		(0.19)
Hispanic		0.07		0.09		-0.61***		-0.55***
		(0.11)		(0.11)		(0.23)		(0.16)
Other race		-0.09		-0.10		0.58**		0.53***
		(0.06)		(0.07)		(0.23)		(0.17)
Parent attended college		0.00		0.01		0.04		0.03
		(0.04)		(0.04)		(0.02)		(0.03)
EFC Percentile : 25-49			-0.11***	-0.10***			-0.03	-0.02
			(0.03)	(0.02)			(0.03)	(0.02)
EFC Percentile : 50-74			-0.09*	-0.11***			0.03	0.01
			(0.05)	(0.03)			(0.06)	(0.05)
EFC Percentile : 75-100			0.01	-0.01			0.01	-0.01
			(0.04)	(0.05)			(0.03)	(0.02)
Prior ACS recipient			0.01	0.02			0.04	0.02
			(0.03)	(0.03)			(0.04)	(0.03)
N	322	321	322	321	322	321	322	321
ACS GPA Band	.30	.30	.30	.30	0.30	0.30	0.30	0.30

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. See Table A1 for remaining coefficient estimates.

Source. Authors' calculations.

Table 5

RD Estimated ACS Effects on Credit Hour Accumulation After Receipt

	Outcome: 1 Year Credit Accumulation				Outcome: 2 Year Credit Accumulation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted ACS Receipt	-6.42	-3.59	-7.47	-4.72	-8.39	-6.35	-10.16	-8.38
	(13.86)	(11.50)	(15.56)	(13.69)	(13.09)	(12.30)	(15.60)	(15.61)
Male		-0.93		-0.99		-0.90		-1.02
		(1.04)		(1.30)		(0.92)		(1.28)
Black		1.20		1.74		3.75		4.61
		(3.74)		(4.52)		(4.57)		(5.90)
Hispanic		-0.74		-0.32		2.14		3.06
		(2.60)		(3.19)		(4.35)		(5.80)
Other race		-1.21		-1.34		-3.63		-4.29
		(3.04)		(3.46)		(3.85)		(4.81)
Parent attended college		1.43		1.41		1.48		1.90
		(1.20)		(1.11)		(1.62)		(1.81)
EFC Percentile : 25-49			-1.61	-1.68*			1.41	0.72
			(1.26)	(0.98)			(1.73)	(1.62)
EFC Percentile : 50-74			0.96	-0.23			0.97	-0.38
			(3.64)	(3.03)			(3.50)	(3.17)
EFC Percentile : 75-100			1.33	0.59			-0.89	-0.87
			(2.35)	(2.01)			(1.66)	(1.34)
Prior ACS recipient			1.49	1.03			2.00	1.94
			(2.09)	(1.96)			(2.05)	(2.45)
N	322	321	322	321	251	251	251	251
ACS GPA Band	0.30	0.30	0.30	0.30	.30	.30	.30	.30

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. See Table A1 for remaining coefficient estimates.

Source. Authors' calculations.

their fourth year when they could receive the expanded ACS and are therefore less likely to persist for two years. In contrast to the results presented in Table 4, the credit accumulation models suggest that ACS recipients earned fewer credit hours—roughly five fewer hours over a two semester period and nine fewer hours over a four semester period. Nevertheless, as before, all effect estimates are not statistically significant.

Finally, Table 6 presents ACS effect estimates for three end-of-college outcomes: final GPA, likelihood of graduating within four years, and likelihood of graduating within five years. While the results in Tables 4 and 5 are all statistically insignificant, ACS recipients appear to score about a quarter of a GPA point lower than non-recipients at the end of their college career and are significantly less likely to graduate within four years. While the latter estimates are particularly large—suggesting ACS recipients are 50 percentage points less likely to graduate within four years—they do align with the simple graphical analysis presented in Figure 2. On the other hand, ACS recipients appear to catch up to their counterparts by their fifth year in college, as the effect estimates presented in columns 9 through 12 are statistically insignificant. It is important to note, however, that the 2009-10 cohort cannot contribute to the latter findings due to data limitations. We explore the extent to which the 2009-10 cohort is driving the other observed findings in the following section.

In summary, the results presented in Tables 4 through 6 suggest that Current Achievers receiving the ACS do not look significantly different than their counterparts after one year, but have significantly lower final GPAs and likelihood of graduating within four years. At the same time, ACS recipients are no less likely to graduate within five years. In the following section, we examine the extent to which our results stand up against two robustness checks.

Table 6

RD Estimated ACS Effects on End-of-College Outcomes

	Outcome: Final GPA				Outcome: Pr(Grad. in 4 years)				Outcome: Pr(Grad. in 5 years)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Pred. ACS Receipt	-0.31**	-0.25**	-0.31*	-0.26*	-0.54***	-0.51***	-0.56***	-0.53***	0.17	0.27	-0.08	0.08
	(0.15)	(0.10)	(0.18)	(0.13)	(0.05)	(0.06)	(0.04)	(0.07)	(0.34)	(0.23)	(0.33)	(0.30)
Male		-0.04		-0.03		-0.01		-0.01		0.12**		0.12*
		(0.02)		(0.03)		(0.04)		(0.05)		(0.06)		(0.07)
Black		0.02		0.03		0.12*		0.13		-0.23		-0.21
		(0.09)		(0.11)		(0.07)		(0.09)		(0.14)		(0.23)
Hispanic		0.17		0.19		-0.08		-0.05		-0.35		-0.39
		(0.14)		(0.16)		(0.15)		(0.15)		(0.26)		(0.28)
Other race		-0.17*		-0.18*		-0.07*		-0.07		0.17		0.18
		(0.09)		(0.10)		(0.04)		(0.05)		(0.13)		(0.19)
Parent att. college		0.08		0.09*		0.08*		0.06		-0.04		-0.02
		(0.05)		(0.05)		(0.04)		(0.05)		(0.07)		(0.07)
EFC Percentile : 25-49			-0.07**	-0.05**			-0.04	-0.04			-0.07	-0.11*
			(0.03)	(0.02)			(0.07)	(0.07)			(0.09)	(0.06)
EFC Percentile : 50-74			0.01	-0.05			0.11*	0.08			0.14	0.05
			(0.09)	(0.08)			(0.07)	(0.09)			(0.15)	(0.18)
EFC Percentile : 75-100			0.09	0.06			0.09	0.07			0.03	-0.02
			(0.06)	(0.05)			(0.06)	(0.06)			(0.11)	(0.07)
Prior ACS recipient			0.04	0.04			0.07***	0.05			0.18**	0.14
			(0.04)	(0.05)			(0.02)	(0.03)			(0.08)	(0.08)
N	322	321	322	321	322	321	322	321	159	158	159	158
ACS GPA Band	.30	.30	.30	.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. See Table A2 for remaining coefficient estimates.

Source. Authors' calculations.

Specification Checks

In this section, we present several checks of our results. We begin by expanding our analytical sample by relaxing some of the requirements for ACS qualification in order to increase the statistical power of our analysis. Next, we examine the extent to which our results are sensitive to varying pre-ACS GPA bandwidths. Two following analyses examine the extent to which model misspecification may be driving our primary results. Finally, we provide a robustness check in which we attempt to estimate ACS effects using earlier LAU cohorts for whom the Arkansas scholarship lottery financed expansion of the ACS was unavailable. It would be problematic for our analysis if we find that ACS receipt positively predicts college outcomes in a sample of students who could not receive the scholarship. In general, the results presented in this section do not harm the credibility of our primary findings.

Including non-FAFSA filers to increase power. As mentioned earlier, the large standard errors present in Tables 4 through 6 indicate substantial noise in our models. This is partially a function of our analytical strategy: by focusing on a subset of LAU students within a small pre-ACS GPA range, we are significantly limiting the power of our study. In addition, it may be the case that college outcomes measures are relatively noisy for students with low academic credentials. Given these issues, it would be reasonable to assume that the null findings largely presented in Tables 4 through 6 are the result of low statistical power rather than a true equivalence of outcomes.

Fortunately, we can increase our model power by relaxing some of the requirements for inclusion in our analytical sample. For example, the analytical sample has until now excluded non-FAFSA filers in order to strictly adhere to the ACS qualifications for Current Achievers. Instead, we can expand the analytical sample—while not strongly adjusting the primary estimates—by including in-state students who did not complete a FAFSA using the following model:

3. $R_i = a + hF_i + bF_ixQualify_i + f(F_ixcGPA_i)'c + f(F_ixcGPA_ixQualify_i)'d + X_i'g + e_i$

4. $Y_i = \alpha + \tau F_i + \beta F_ix\widehat{R_i} + f(F_ixcGPA_i)'\delta_1 + f(F_ixcGPA_ixQualify_i)'\delta_2 + X_i'\gamma + \epsilon_i$

where equation (3) represents a first stage regression for the model outlined in equation (4) and F indicates if an individual has completed a FAFSA. The models presented in equations (3) and (4) differ from our primary models only in that they include the FAFSA indicator along with its interaction with the treatment and assignment variables of interest to our RDD. The coefficient on FAFSA gives the difference between FAFSA non-recipients and FAFSA non-filers. The coefficient on the interaction expresses the difference of interest: the estimated difference between qualified ACS applicants who are predicted to receive an ACS scholarship and eligible applicants who are not predicted to receive the scholarship. Given the model specification, we should expect the estimates to strongly mirror those presented in Tables 4 through 6 while the overall precision of the model will be improved.

Tables 7a and 7b present analyses based on the model outlined in equations 3 and 4. Two models are presented for each outcome: a simple model and a model that includes demographic controls.[68] As expected, the estimated coefficients on the interactions largely mirror those presented in Tables 4 through 6. Furthermore, the estimated relationships for the baseline characteristics are similar. Interestingly, despite the increase in model precision—as noted by the

[68] Our data limits us to only knowing family income via the FAFSA application. We cannot include controls for family income because they are perfectly collinear with FAFSA. Thus, while the model precision is improved in tables 7a and 7b, we caution the reader to take the results presented in tables 4 through 6 as our primary results.

Table 7a

RD Effect Estimates for Short- and Medium-run Outcomes, including Non-FAFSA Filers

	Year 1 GPA		Pr(Persist 1 Year)		Year 1 Credit Acc.		Year 2 Credit Acc.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FAFSA & ACS	0.13	0.15	0.04	0.09	-7.12	-4.65	-10.68	-9.12
Interaction	(0.11)	(0.11)	(0.28)	(0.26)	(13.96)	(12.19)	(13.06)	(13.26)
FAFSA Filer	-0.11**	-0.10**	-0.02	-0.03	-0.39	-0.60	2.38	2.49
	(0.04)	(0.04)	(0.07)	(0.07)	(3.77)	(3.55)	(2.96)	(3.39)
Male		-0.04***		0.00		-0.71		-0.44
		(0.01)		(0.04)		(0.72)		(0.71)
Black		-0.06		-0.65***		1.35		3.94
		(0.06)		(0.23)		(3.75)		(4.43)
Hispanic		0.06		-0.67***		-0.56		2.37
		(0.11)		(0.19)		(2.21)		(4.30)
Other race		-0.06		0.63***		-1.36		-3.45
		(0.06)		(0.20)		(2.82)		(3.69)
Parent attended college		-0.00		0.00		1.05		1.37
		(0.04)		(0.03)		(0.95)		(1.28)
N	489	488	489	488	480	479	373	373
ACS GPA Band	.30	.30	.30	.30	.30	.30	.30	.30

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA.

Source. Authors' calculations.

Table 7b

RD Effect Estimates for Short- and Medium-run Outcomes, including Non-FAFSA Filers

	Final GPA		Pr(Graduate in 4 Years)		Pr(Graduate in 5 Years)	
	(1)	(2)	(3)	(4)	(5)	(6)
FAFSA & ACS Interaction	-0.33**	-0.26***	-0.66***	-0.62***	0.17	0.24
	(0.14)	(0.09)	(0.07)	(0.07)	(0.41)	(0.27)
FAFSA Filer	-0.01	0.02	0.11***	0.11***	-0.11	-0.10
	(0.03)	(0.03)	(0.03)	(0.04)	(0.17)	(0.14)
Male		-0.04**		-0.03		0.10*
		(0.02)		(0.05)		(0.05)
Black		-0.02		0.17***		-0.24**
		(0.08)		(0.06)		(0.11)
Hispanic		0.15		-0.06		-0.31
		(0.12)		(0.13)		(0.21)
Other race		-0.13		-0.13***		0.19
		(0.09)		(0.03)		(0.13)
Parent attended college		0.07*		0.07		-0.03
		(0.04)		(0.05)		(0.08)
N	484	483	489	488	240	239
ACS GPA Band	.30	.30	.30	.30	.30	.30

*** p<0.01, ** p<0.05, * p<0.1

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA.

Source. Authors' calculations.

smaller observed standard errors in Tables 7a and 7b—the general relationships remain: ACS receipt is associated with insignificant differences in the short-run; but significant negative relationships between final GPA and likelihood of graduation. Nevertheless, the differences again dissipate after 5 years.

Thus, despite the increased statistical power of the analyses presented in Table 7a and 7b, the general story remains the same: ACS receipt is associated with insignificant differences on short-run outcomes, negative results for final GPA and likelihood of graduating within four years, and yet an insignificant difference in the likelihood of graduating within five years.

Varying band widths. Table 8 presents results from models examining the relationship between ACS receipt and our seven college outcomes using varying pre-ACS GPA bandwidths. All models are fully specified, employing demographic and financial history controls. Columns 1, 4, and 7 reproduce the results for our analyses using our preferred 0.30 bandwidth. Columns 2, 5, and 8 present results for a larger bandwidth of 0.40 (GPA ranging from 2.1 to 2.9 points). Columns 3, 6, and 9 present results for a smaller bandwidth of 0.20 (GPA ranging from 2.3 to 2.7 points). Unfortunately, we were unable to achieve convergence for our probit models estimating ACS effects for year 1 persistence for pre-ACS GPA bands of .40 and .20. Instead, we present results from Linear Probability Models (LPM), which effectively treat the dichotomous persistence variable as a continuous variable.[69] While LPM estimation is not the preferred method[70], it is the best means available to estimate these relationships.

[69] This assumption leads both to heteroskedasticity in the model (Angrist & Pischke, 2009) and bias in estimated effects (Wooldridge, 2002).
[70] Roughly 17 percent of observations in the .40 GPA band LPM estimate have predicted probabilities falling outside the plausible range of 0 and 1. The corresponding percentage in the .20 GPA band model is 22 percent.

Table 8

RD Effect Estimated with Varying Band Widths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Year 1 GPA			Pr(Persist 1 Year)ǂ					
Predicted ACS Receipt	0.13	0.08	0.00	0.05	0.05	-0.15			
	(0.11)	(0.18)	(0.17)	(0.20)	(0.28)	(0.35)			
N	321	451	205	321	451	205			
ACS GPA Band	0.30	0.40	0.20	0.30	0.40	0.20			
	Year 1 Credit Acc.			Year 2 Credit Acc.					
Predicted ACS Receipt	-4.72	-7.04	-7.74	-8.38	-10.15	-11.23			
	(13.69)	(11.48)	(15.38)	(15.61)	(12.36)	(14.28)			
N	321	451	205	251	358	160			
ACS GPA Band	0.30	0.40	0.20	.30	.40	.20			
	Final GPA			Pr(Graduate in 4 years)			Pr(Graduate in 5 years)		
Predicted ACS Receipt	-0.26*	-0.25*	-0.38	-0.53***	-0.48***	-0.53***	0.08	-0.05	-.01
	-0.13	-0.13	-0.24	(0.07)	(0.08)	(0.05)	(0.30)	(0.23)	(0.41)
N	241	345	146	321	451	205	158	221	101
ACS GPA Band	0.30	0.40	0.20	0.30	0.40	0.20	0.30	0.40	0.20

*** p<0.01, ** p<0.05, * p<0.1

ǂ Effects on persistence are estimated using linear probability models for .40 and .20 ACS GPA Bands due to a failure to achieve convergence in probit models.

Note. Standard errors (parentheses) account for clustering of individuals in major. All models are fully specified: they include controls for student demographics, family finance, entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA.

Source. Authors' calculations.

In general, the effect estimates presented in Table 8 are not particularly sensitive to band width: ACS receipt is insignificantly related to year 1 GPA, credit hour accumulation, and the likelihood of graduating within five years. Similarly, ACS is strongly and negatively related to both final GPA and the likelihood of gradating in four years. In addition, the coefficient estimates are fairly similar across models in Table 8. The consistency of these findings is promising for our analysis.

Differential effects among prior ACS recipients. Currently, prior ACS recipients—or students who received an ACS in at least one semester after initial matriculation prior to the expansion of the program in the fall of 2010—constitute nearly 40 percent of our analytical sample. The models presented in Tables 4 through 6 account for differences in college outcomes experienced among this group using a simple dummy variable; effectively assuming that the ACS expansion treatment effects on college outcomes are experienced similarly among old and new ACS recipients.[71] If, instead, prior ACS recipients differ from new ACS recipients, our primary results could suffer from bias due to misspecification. For example, the insignificant results observed in Tables 4 through 6 could be explained by prior ACS recipients having already experienced the benefits of the program, thereby effectively masking any positive results experienced by new recipients. It is important to examine the extent to which prior ACS recipients are driving our overall estimates because they constitute such a large portion of the analytical sample.

[71] At the same time, the qualification and receipt rates are actually fairly similar between the two groups. Specifically, 74 percent of students who did not receive an ACS in the past qualified in the fall of 2010 compared to 66 percent of prior ACS recipients. Of these, 51 percent of students who did not receive the ACS in the past received the scholarship in the fall of 2010 compared to 61 percent of prior ACS recipients.

Table 9

Comparing Prior ACS Recipients with New ACS Recipients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year 1 GPA		Pr(Persist 1 Yr.) ǂ		Year 1 Credit Acc.		Year 2 Credit Acc.	
Predicted ACS Receipt	0.12	0.13	0.02	0.07	-6.80	-4.75	-8.97	-7.08
	(0.11)	(0.10)	(0.37)	(0.31)	(14.38)	(13.67)	(12.34)	(12.50)
ACS Receipt & Prior ACS Interaction	0.05	0.01	-0.10	-0.10	1.04	0.62	7.23*	6.49*
	(0.11)	(0.13)	(0.09)	(0.08)	(2.37)	(1.82)	(4.39)	(3.70)
Prior ACS Recipient	-0.02	0.01	0.06*	0.07***	0.02	0.76	-2.26	-1.43
	(0.04)	(0.05)	(0.04)	(0.02)	(0.98)	(1.79)	(1.56)	(2.28)
Controls for demographics and income		X		X		X		X
N	322	321	322	321	322	321	251	251
	Final GPA		Pr(Grad. in 4 Yrs.) ǂ		Pr(Grad. in 5 Yrs.) ǂ			
Predicted ACS Receipt	-0.32**	-0.26**	-1.11***	-0.93***	0.02	-0.14		
	(0.15)	(0.13)	(0.38)	(0.29)	(0.49)	(0.48)		
ACS Receipt & Prior ACS Interaction	0.14	0.06	0.29**	0.30	0.99***	0.97***		
	(0.21)	(0.23)	(0.15)	(0.25)	(0.33)	(0.31)		
Prior ACS Recipient	-0.06	0.02	-0.11	-0.04	-0.31***	-0.21***		
	(0.07)	(0.11)	(0.08)	(0.11)	(0.07)	(0.07)		
Controls for demographics and income		X		X		X		
N	322	321	322	321	159	158		

*** p<0.01, ** p<0.05, * p<0.1

ǂ Effects on persistence are estimated using linear probability models for .40 and .20 ACS GPA Bands due to a failure to achieve convergence in probit models.

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. All models with discrete dependent variables are estimated using linear probability regression models due to a failure to achieve convergence in binary probit models.

Source. Authors' calculations.

Table 9 presents results from analyses exploring the extent to which prior ACS recipients experience different treatment effects than new ACS recipients.[72] Specifically, the estimates in the first row of Table 9 describe the estimated difference between ACS recipients and non-recipients who have not won an ACS in the past. The second row, titled "ACS Receipt & Prior ACS Interaction" presents the estimated ACS treatment different between new and prior ACS recipients. Finally, the coefficient estimates in row three represent the differences in outcomes between students who received the ACS in the past and those who did not among non-ACS recipients.

The results presented in Table 9 provide some assurance that our primary estimates are not strongly confounded by the presence of prior-ACS recipients. First, the coefficients on the majority of interaction terms are insignificant: indicating that new- and prior-ACS recipients did not experience dramatically different outcomes following the expansion of the program in the fall of 2010. The exceptions are the models examining credit accumulation after two years and the two models estimating the ACS effects on the likelihood of graduation. The significant interactions in the latter models are less concerning because the coefficient estimates are based on linear probability regressions, which tend to be biased when continuous covariates are included in the model (Wooldridge, 2002). [73] At the same time, LPM estimation is the best available analytical option in these cases, as our probit estimations failed to achieve convergence. The significant interactions in the year two credit accumulation models are more concerning. Moreover, while insignificant, new ACS receivers earned fewer credit hours than

[72] As with the analyses presented in the prior section, all models examining discrete dependent variables are estimated using linear probability models.

[73] Indeed, over between 24 and 33 percent of observations in the models estimating the likelihood of graduation within 4 years have predicted probabilities of graduating outside the plausible range of 0 or 1. Similarly, around 15 percent of observations are outside this range for the LPMs estimating the likelihood of graduating within 5 years.

students who did not qualify for the ACS and did not receive it in the past. This finding is concerning for the models focusing on credit hour accumulation after two years, and we therefore recommend caution in interpreting these results.

In addition, none of the estimated differences between prior and new ACS recipients (row 3) are largely statistically significant, further highlighting the similarity of the two groups. Again, the exceptions to this are largely associated with linear probability regressions, which are not generally the preferred method for estimating effects with binary dependent variables. These models indicate that, among students who did not receive the ACS in fall 2010, former ACS recipients were slightly more likely to persist through the spring of 2011 and were less likely to graduate within five years.

Finally, the results presented in Table 9 confirm the earlier finding that students who are predicted to receive the ACS have significantly lower final GPAs than other students. Given the relatively small differences in estimated effects between Table 9 and Tables 4 through 6, as well as the substantively small differences estimated by the interaction effects, we conclude that our overall effects are not strongly biased by prior ACS recipients.

Cohort interactions. As noted earlier, our primary analyses are based on an uneven panel. Currently, our data do not allow us to estimate 5-year graduation for all cohorts because the 2009-10 cohort—who were sophomores when they could win the scholarship—could at most be enrolled in their fifth year at the conclusion of our data. Given the strong negative four year graduation rate results and the insignificant five year graduation results; there is a concern that the effects of the ACS on college outcomes may vary by cohort. In Table 10 we present results

.

Table 10

RD Effect Estimates Separated by Cohorts

	(1)	(2)	(3)	(4)	(5)	(6)
	Year 1 GPA		Pr(Persist 1 Yr.)		Year 1 Credit Acc.	
ACS Recipient in '07 Cohort	0.18	0.16	0.00	0.03	-7.96	-6.77
	(0.20)	(0.20)	(0.36)	(0.29)	(12.75)	(12.55)
ACS Recipient in '08 Cohort	0.42***	0.26*	-0.09	-0.03	-3.72	-3.32
	(0.16)	(0.13)	(0.37)	(0.30)	(16.84)	(16.05)
ACS Recipient in '09 Cohort	0.03	0.01	0.07	0.17	-4.69	-2.00
	(0.08)	(0.04)	(0.35)	(0.27)	(14.93)	(13.46)
Cohort 2007	-0.00	0.00	0.06	0.09**	3.31	4.03
	(0.03)	(0.03)	(0.05)	(0.04)	(3.11)	(3.15)
Cohort 2008	-0.16***	-0.10**	0.12*	0.14**	1.62	2.59
	(0.05)	(0.05)	(0.07)	(0.06)	(2.46)	(2.65)
Controls for demo. and income		X		X		X
N	322	321	322	321	322	321
	Year 2 Credit Acc.		Final GPA		Pr(Grad. in 4 Yrs.)	
ACS Recipient in '07 Cohort			-0.25	-0.23	-1.08***	-0.95***
			(0.16)	(0.19)	(0.34)	(0.23)
ACS Recipient in '08 Cohort	-6.31	-8.16	-0.12	-0.22**	-0.97***	-0.87***
	(15.06)	(17.41)	(0.09)	(0.10)	(0.34)	(0.26)
ACS Recipient in '09 Cohort	-8.57	-8.45	-0.41**	-0.32**	-1.10**	-0.90*
	(12.99)	(15.26)	(0.17)	(0.13)	(0.50)	(0.54)
Cohort 2007			-0.10***	-0.06	0.03	0.07
			(0.04)	(0.04)	(0.18)	(0.23)
Cohort 2008	-0.54	0.32	-0.14**	-0.06	-0.04	0.01
	(2.00)	(2.07)	(0.07)	(0.06)	(0.14)	(0.16)
Controls for demo. and income		X		X		X
N	251	251	322	321	322	321

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note. Standard errors (parentheses) account for clustering of individuals in major. All models include controls for entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. Persistence and Graduation models are estimated using LPM regressions.

Source. Authors' calculations.

from models exploring the possibility of differential treatment effects across cohorts through a series of interactions.

Specifically, we split out the receipt variable across the three cohorts. The coefficient estimates associated with the rows titled “Cohort 2007” and “Cohort 2008” represent the estimated difference between in outcomes between non-ACS recipients in the given cohorts and the 2009 cohort. The coefficient estimates on the interactions presented in rows one through three represent the estimated treatment differential for each cohort. As before, we are only able to present results of linear probability models for our binary outcome models due to a failure to achieve convergence in probit models

For the most part, the results presented in Table 10 do not suggest that ACS treatment effects differed largely across cohorts. At the individual coefficient level, there is evidence that some relationships are driven more strongly by particular cohorts. For example, the 2008-09 cohort may be driving the results for Year 1 GPA while the 2007-08 cohort appears to contribute little to the results observed for Final GPA. Nevertheless, it is important to note the great degree of noise present in Table 10, which is symptomatic of the small sizes of the analytical samples. Indeed, while some cohort-specific treatment coefficients are statistically significant, t-tests examining the extent to which all three coefficient estimates are the same fail to reject the null hypothesis across all models. Thus, in general, while the evidence presented in Table 10 suggests that certain cohorts may be driving our results, we are unable to support this claim with statistical tests due to power limitations.

Placebo analysis. As a final check of our analytical strategy, we estimate ACS effects in a cohort of students who could not receive the lottery scholarship financed ACS. Specifically, we examine a “placebo” sample of students who matriculated at LAU in cohort years 2004-05, 2005-06, and 2006-07. All students included in our analyses meet the same requirements of our analytical sample: they are in-state applicants who have filled out a FAFSA, have been continuously enrolled, meet the ACS credit hours requirement, and have cumulative GPAs between 2.2 and 2.8 points. In addition, we restrict the cohorts to mirror their counterparts in our analytical sample: the 2004-05 cohort is examined in their 7th semester, 2005-06 cohort is examined in their 5th semester, and the 2006-07 cohort is examined in its 3rd semester.

The results of our placebo analysis are presented in Table 11. Because we do not have data on scholarship receipt, all models estimate the intent-to-treat effect of ACS qualification. In addition, all models are fully specified. As expected, ACS qualification is not significantly related to any of the seven college outcomes. More importantly, all estimated effects are substantively small; providing strong evidence that ACS qualification was not related to outcomes in these earlier cohorts. While not definitive, these findings again lend credence to our analysis

Discussion

In this paper, we examine the effect of the Arkansas Academic Challenge Scholarship—a broad-based state-financed merit-aid scholarship—on college outcomes at a Large Arkansas University (LAU) by using scholarship eligibility requirements in a regression discontinuity framework. In general, our results indicate that currently enrolled LAU students who received the ACS did not earn higher cumulative GPAs and were no more likely to be enrolled after one year. On the other hand, ACS recipients had significantly lower final GPAs and were significantly less likely to graduate within four years. Nevertheless, our models indicate that ACS recipients were no less likely to graduate within five years than comparison group students. Taken along with the negative—but statistically insignificant—findings on credit hours, our results may suggest that

Table 11

RD Effect Estimates for Students Entering LAU in 2004-05, 2005-06, and 2006-07

	(1)	(2)	(3)	(4)
	Year 1 GPA	Pr(Persist 1 Year)	Year 1 Credit Acc.	Year 2 Credit Acc.
Predicted ACS Receipt	0.06	0.16	-0.03	0.40
	(0.05)	(1.46)	(0.12)	(2.39)
N	266	255	233	217
ACS GPA Band	.30	.30	.30	.30
	Final GPA	Pr(Graduate in 4 years)	Pr(Graduate in 5 years)	
Predicted ACS Receipt	0.08	-0.03	0.12	
	(0.05)	(0.12)	(0.27)	
N	258	260	103	
ACS GPA Band	.30	.30	.30	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note. Standard errors (parentheses) account for clustering of individuals in major. All models are fully specified: they include controls for student demographics, family finance, entering cohort year, pre-ACS hours below 15 hours, and quadratic functions of the assignment variable, centered pre-ACS GPA. To mirror the samples available for our primary analyses, the “Year 2 Credit Acc.” model (row 1; column 4) excludes the 2004-05 cohort and the “Pr(Graduated in 5 years)” (row 2; column 3) model excludes the 2006-07 cohort.

Source. Authors’ calculations.

ACS recipients were more likely to delay graduation.

These findings differ from earlier RDD analyses of state-financed merit-aid programs. For example, Scott-Clayton finds in her 2012 examination of the West Virginia PROMISE scholarship that recipients earned significantly higher final GPAs, accumulated significantly more credit hours, and were significantly more likely to graduate within four years. While we would like to stress that the findings presented in this paper are preliminary, we offer two comments regarding our divergent findings.

First, we are examining a substantively different student population compared to prior studies. Specifically, our study is focused on students who were currently enrolled in college when they became eligible for the ACS (as opposed to entering freshmen) meeting relatively weak academic credential requirements (enrolling for 12 hours a semester and earning a cumulative GPA of at least 2.5 points). While West Virginia's PROMISE program is similar to the Arkansas ACS, it should not be unexpected to find that these different student populations would have different experiences.

Second, while our results differ from other studies, it is important to note that they present a consistent story. Specifically, the statistically insignificant negative credit hour accumulation findings and evidence suggesting that ACS recipients are just as likely as non-recipients to graduate within five years are consistent with students delaying their time to graduation after receiving a scholarship. While this may not be the intended impact of the ACS, it nevertheless helps to explain why our findings differ from other studies.

Finally, we would like to highlight the preliminary nature of the research presented in this paper. For example, our current analysis does not yet employ a sophisticated procedure for determining band width nor are all models complete. We intend to address these issues in future versions of this research.

References

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Arkansas Department of Higher Education (2010, August 16). ADHE Response to FOI Requests Concerning the Arkansas Academic Challenge Scholarship for the 2010-11 Academic Year. Retrieved from http://www.adhe.edu/DIVISIONS/FINANCIALAID/Pages/ADHEResponsetoFOIRequests.aspx

Arkansas Secretary of State (2011). Arkansas Academic Challenge Scholarship Rules and Regulations. Retrieved from http://www.sos.arkansas.gov/rulesRegs/Pages/default.aspx

Bettinger, E. P. (2004). How financial aid affects persistence. In C. M. Hoxby (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*. Chicago, IL: University of Chicago Press.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, 1205–1242.

Castleman, B. L., & Page, L. C. (2013). *Summer nudge: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates?* (EdPolicy Works Working Paper No. 9). University of Virginia: Curry School of Education. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/9_Castleman_SummerTextMessages.pdf

Castleman, B. L., & Page, L. C. (2014). *Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence* (EdPolicy Works Working Paper No. 29). University of Virginia: Curry School of Education. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/29_Freshman_Year_Financial_Aid_Nudges.pdf

College Board. (2013). *Trends in student aid*. New York, NY. Retrieved from http://trends.collegeboard.org/

Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *The Journal of Human Resources*, *40*(4), 895–917.

Cornwell, C. M., Mustard, D. B., & Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics*, *24*(4), 761–786.

Dynarski, S. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *The American Economic Review*, *93*(1), 279–288.

Dynarski, S. (2008). Building the stock of college-educated labor. *The Journal of Human Resources, 43*(3), 576–610.

Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: Belknap Press of Harvard University Press.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*, 615–635.

Kane, T. J. (2003). *A quasi-experimental estimate of the impact of financial aid on college-going* (NBER Working Paper No. 9703). Cambridge, MA: National Bureau of Economic Research.

National Center for Education Statistics. (2014). *The condition of education*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://nces.ed.gov/programs/coe/indicator_caa.asp

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness, 5*(1), 83–104.

Scott-Clayton, J. (2012). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, *46*(3), 614–646.

Snyder, T. D. (2014). *Mobile digest of education statistics, 2013* (NCES 2014-085). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Staiger, D. & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, pp. 557-586.

The Pell Institute. (2015). *Indicators of higher education equity in the United States: 45 year trend report*. Washington, DC: Council for Opportunity in Education. Retrieved from www.pellinstitute.org/downloads/publications-Indicators_of_Higher_Education_Equity_in_the_US_45_Year_Trend_Report.pdf

The White House. (2010). President Obama calls for new steps to prepare America's children for success in college and careers [Press release]. Retrieved from https://www.whitehouse.gov/the-press-office/president-obama-calls-new-steps-prepare-america-s-children-success-college-and-care

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review*, *43*(4), 1249–1287.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: Massachusetts Institute of Technology.

Appendix

Table A1

Remaining Estimated Coefficients for Second Stage Regressions, Simple and Full Models Only

	(1)	(2)	(3)	(4)
	Year 1 GPA		Pr(Persist 1 Year)	
Cohort 2007-08	0.06	0.07	0.02	0.03
	(0.05)	(0.07)	(0.02)	(0.02)
Cohort 2008-09	0.00	0.00	0.06***	0.06**
	(0.03)	(0.03)	(0.02)	(0.03)
Pre-ACS hours <15	-0.01	0.01	-0.04	-0.03
	(0.04)	(0.03)	(0.08)	(0.06)
cGPA	0.44	0.39	0.01	0.10
	(0.58)	(0.80)	(0.36)	(0.19)
cGPA^2	-1.82	-1.70	1.32	1.82*
	(2.48)	(3.16)	(1.39)	(1.04)
Qualify x cGPA	-0.07	-0.22	0.36	0.19
	(1.29)	(1.52)	(1.00)	(0.69)
Qualify x cGPA^2	2.62	3.27*	-1.83	-2.30
	(1.59)	(1.80)	(2.44)	(2.01)
Observations	322	321	322	321
	One Year Credit Accumulation		Two Year Credit Accumulation	
Cohort 2007-08	1.85	1.91	---	---
	(1.99)	(2.42)	---	---
Cohort 2008-09	1.92***	1.94***	0.39	0.44
	(0.49)	(0.62)	(1.08)	(1.38)
Pre-ACS hours <15	-3.54	-2.91	-2.28	-1.56
	(2.94)	(2.45)	(1.68)	(1.45)
cGPA	-11.38	-12.75	0.49	5.98
	(27.69)	(25.71)	(21.20)	(19.76)
cGPA^2	-75.49	-72.84	-63.16	-40.53
	(93.32)	(85.25)	(79.35)	(61.26)
Qualify x cGPA	40.80	36.63	21.42	14.97
	(32.98)	(28.52)	(47.28)	(48.93)
Qualify x cGPA^2	28.94	38.24	57.02	37.25
	(150.05)	(144.98)	(123.00)	(151.74)
Observations	322	321	251	251
Student demographics		X		X
Family income		X		X

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note. Standard errors (parentheses) account for clustering of individuals in major. Estimated coefficients correspond with the results presented in Tables 4 and 5. All models are restricted to samples within the pre-ACS GPA band of .30. The 2007-08 cohort is excluded from the two year credit accumulation models because this members of this cohort are expected to be in their fourth year of college when they could receive the expanded ACS; and are therefore less likely to be enrolled two years after potential receipt.

Source. Authors' calculations.

Table A2.

Remaining Estimated Coefficients for Second Stage Regressions, Simple and Full Models Only

	(1)	(2)	(3)	(4)
	Final GPA		Pr(Graduate in 4 years)	
Cohort 2007-08	-0.03	-0.02	0.01	0.02
	(0.04)	(0.05)	(0.05)	(0.06)
Cohort 2008-09	-0.02	-0.02	0.01	0.01
	(0.04)	(0.04)	(0.03)	(0.02)
Pre-ACS Hours <15	-0.08*	-0.05	-0.15***	-0.14***
	(0.04)	(0.04)	(0.03)	(0.03)
cGPA	1.95***	1.78***	1.65***	1.77**
	(0.74)	(0.64)	(0.48)	(0.75)
cGPA^2	3.06	2.82	4.51**	4.80
	(3.29)	(2.75)	(1.90)	(3.07)
Qualify x cGPA	-0.75	-0.83	-0.08	-0.21
	(1.67)	(1.50)	(0.93)	(1.03)
Qualify x cGPA^2	-3.36	-2.51	-6.98**	-7.39*
	(3.64)	(2.51)	(2.81)	(4.28)
Observations	322	321	322	321
	Pr(Graduate in 5 years)			
Cohort 2007-08	0.08**	0.07		
	(0.04)	(0.05)		
Pre-ACS Hours <15	-0.13	-0.17		
	(0.15)	(0.13)		
cGPA	-2.47*	-0.84		
	(1.35)	(1.66)		
cGPA^2	-13.37***	-7.55		
	(4.66)	(6.47)		
Qualify x cGPA	1.48	-0.88		
	(2.01)	(0.72)		
Qualify x cGPA^2	17.34**	13.42		
	(8.73)	(11.40)		
Observations	159	158		
Student demographics		X		X
Family income		X		X

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note. Standard errors (parentheses) account for clustering of individuals in major. Estimated coefficients correspond with the results presented in Table 6. All models are restricted to samples within the pre-ACS GPA band of .30.

Source. Authors' calculations.

Conclusion

Cash transfers are an increasingly popular policy instrument both abroad and in the United States. In this dissertation, I investigate how two statewide educational interventions using conditional cash transfer programs affected student outcomes.

Chapter 1 presents the first experimental study of a statewide school voucher program in the United States: the Louisiana Scholarship Program. Unlike similar random assignment evaluations of school voucher programs, the effect estimates presented in Chapter 1 indicate that students who used the LSP scholarships to enroll in private schools experienced significant declines in both ELA and math achievement. These estimates are quite substantial and are robust to several specification checks.

Chapter 2 also examines the experiences of students participating in the first year of the LSP's statewide expansion. Unlike other school voucher evaluations, which have largely focused on student achievement and attainment, Chapter 2 represents the first attempt to examine how students participating in a voucher program differ on non-cognitive skills and political tolerance measures from other applicants. In general, the results presented in Chapter 2 do not indicate that the two groups differed strongly on these measures after two years. Furthermore, while researchers attempting to estimate the effects of charter schools on Grit have cautioned against the presence of reference group bias (Dobbie & Fryer, 2014; West et al., 2014), we find little evidence of such bias in our estimates.

Finally, Chapter 3 investigates the effects of a broad-based aid scholarship in Arkansas, the Academic Challenge Scholarship, on college outcomes for students who were currently enrolled in college at the time of the program's expansion. The estimates presented in Chapter 3 indicate that students who barely qualified for the scholarship performed no differently than students who barely missed the ACS cutoffs after one year, but had significantly lower final GPAs and four year graduation rates. Nevertheless, ACS recipients appear to have caught up to their counterparts after five years; suggesting that students receiving the ACS may use the additional funds to delay their graduation.

While Chapters 1 through 3 all include several robustness checks designed to examine result sensitivity, it is important to consider the following caveats when interpreting these findings. First, it is important to recognize that the results presented in Chapter 1 are based on a subset of eligible LSP applicants who had baseline achievement and whose scholarship receipt was determined by oversubscription lotteries. These restrictions effectively limit the analysis to an analytical sample representing slightly less than 20 percent of eligible applicants. It is therefore important to recognize that the estimated results may not be indicative of the experiences of other students in the program. Furthermore, the estimates in Chapter 1 are based on little more than 6 months of program participation. Future versions of this research will explore if these negative effects persist as students and schools become acclimated to the program.

While the estimated differences between LSP receivers and non-receivers are insignificant in Chapter 2, they are also particularly noisy. This is likely due to two factors: the scales used are validated neither for phone surveys nor for children as young as some of the children included in our survey sample. In addition, it is important to note that the final analytical sample examined in Chapter 2 represents only 10 percent of the population of eligible LSP applicants in the first year of the program's expansion. As is often the case with survey-based research, the estimates presented in this chapter likely suffer from sample selection issues.

Chapter 3 additionally suffers from statistical power concerns due to both the small sample of students with academic qualifications around the ACS cutoffs as well as the general power issues involved with fuzzy regression discontinuity estimations. In addition, the analysis presented in Chapter 3 is currently limited by an inability to estimate program effects on five year graduation rates for students in the 2009-10 cohort due to data limitations. Given the substantially large negative effects estimated for four year graduation rates, it is possible that the overall ACS effect may be negative when these students are included in the final analysis. This question will be addressed in future research once we obtain an additional year of data.

Despite these caveats, the research presented in this dissertation makes several contributions to the school choice and higher education financial aid literatures. Chapter 1 represents the first experimental evaluation of a statewide voucher program as well as the first such study to estimate statistically significant achievement effects. Chapter 2 is the first attempt by researchers to compare students participating in a school voucher program on non-cognitive skills and measures of political tolerance. Finally, Chapter 3 represents the first attempt to estimate the impact of a broad-based merit-aid scholarship with low academic qualifications on college outcomes for students who were currently enrolled in college at the time of scholarship receipt.

On the whole, however, the most important takeaway from this research is the possibility of unintended, and potentially perverse, outcomes resulting from policy designs. The significant negative achievement effect estimates presented in Chapter 1 represent the first of their kind among random assignment evaluations of school voucher programs in the U.S. Moreover, the magnitude of the estimated effects is highly concerning. Policymakers considering voucher programs in other states should consider recent research suggesting that the program's regulatory framework creates disincentives for private school participation (Kisida, Wolf, & Rhinesmith, 2014). Similarly, the evidence presented in Chapter 3 suggesting that students currently enrolled in college who received the ACS may have delayed graduation may be worrisome to policymakers. In particular, this evidence suggests that similar programs should consider how students currently enrolled in college, as well as students with particularly low academic credentials, may be impacted by similar scholarship programs.

Thus, while each of the analyses presented in this dissertation benefit their respective literatures due to the novelty of their subjects, this research is a reminder that unintended consequences are likely to arise in any educational intervention. This suggests a continued need for policymakers and researchers to work together in an effort to design, monitor, and re-design programs for the benefit of all involved.

References

Dobbie, W., & Fryer, R. G. (2014). The medium-term impacts of high-achieving charter schools (Working Paper).

Kisida, B., Wolf, P. J., & Rhinesmith, E. (2015). Views from private schools: Attitudes about school choice programs in three states. Washington, DC: American Enterprise Institute.

West, M. R., Kraft, M. K., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling (Working Paper).

Appendix

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

February 2, 2015

MEMORANDUM

TO:	Patrick Wolf Anna Jacob Egalite Daniel Bowen Collin Hitt Thomas Waggoner	Jay Greene Jonathan Mills Albert Cheng Martin Lueken
FROM:	Ro Windwalker IRB Coordinator	
RE:	PROJECT CONTINUATION	
IRB Protocol #:	13-02-501	
Protocol Title:	*State Mandated Evaluation of the Louisiana Students Scholarships for Excellence Program and Course Choice Program*	
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB	
Previous Approval Period:	Start Date: 02/22/2013 Expiration Date: 02/21/2015	
New Expiration Date:	02/21/2016	

Your request to extend the referenced protocol has been approved by the IRB. If at the end of this period you wish to continue the project, you must submit a request using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. Failure to obtain approval for a continuation on or prior to this new expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

This protocol has been approved for 910,000 total participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, 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.

109 MLKG Building • 1 University of Arkansas • Fayetteville, AR 72701
Voice (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

October 24, 2014

MEMORANDUM

TO:	Gary Ritter Michael Crouch
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT CONTINUATION
IRB Protocol #:	13-08-059
Protocol Title:	*Profiles of Successful College Students and Best Predictors of College Success*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Previous Approval Period:	Start Date: 10/01/2013 Expiration Date: 08/29/2014
New Expiration Date:	08/29/2015

Your request to extend the referenced protocol has been approved by the IRB. If at the end of this period you wish to continue the project, you must submit a request using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. Failure to obtain approval for a continuation on or prior to this new expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

This protocol has been approved for 45,000 total participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, 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 210 Administration Building, 5-2208, or irb@uark.edu.

210 Administration Building • 1 University of Arkansas • Fayetteville, AR 72701
Voice (479) 575-2208 • Fax (479) 575-3846 • Email irb@uark.edu