UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Essays in the Economics of Education

Permalink

https://escholarship.org/uc/item/70h5071f

Author

Young, Sam Ming

Publication Date

2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in the Economics of Education

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Economics

by

Sam Ming Young

Committee in charge:

Professor Julian Betts, Co-Chair
Professor Julie Cullen, Co-Chair
Professor Gordon Dahl
Professor Sally Sadoff
Professor Krislert Samphantharak

2017

Copyright

Sam Ming Young, 2017

All rights reserved.

The Dissertation of Sam Ming Young is approved and is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California, San Diego

2017

DEDICATION

For Honesty, who inspired my interest in economics

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures vii

List of Tables viii

Acknowledgements xi

Vita xii

Abstract of the Dissertation xiii

Chapter 1 Do Colleges Favor Students from Lower-Achieving High Schools? Evidence from the University of California 1
1.1 Introduction 1
1.2 Background and Related Literature 5
1.2.1 UC-Specific Background 7
1.3 Conceptual and Empirical Framework 8
1.4 Data 10
1.5 Results 13
1.5.1 Nonlinearity Across API Deciles 15
1.5.2 Enrollment Composition 16
1.5.3 Possible Explanations: Academic Ability Versus Other Factors . 19
1.6 Discussion 21
1.7 Tables and Figures 23

Chapter 2 The Long-Term Effects of Public School Choice: Lottery Evidence from San Diego 34
2.1 Introduction 34
2.2 Related Literature 36
2.2.1 Achievement of Students Participating in Choice 36
2.2.2 School Choice in San Diego 38
2.3 Data and Methods 44
2.3.1 Sample and Lottery Basics 44
2.3.2 Outcomes of Interest 46
2.3.3 Estimating Treatment Effects 48

2.4 Results 50
2.4.1 Testing for Lottery Fairness and Differential Attrition 50
2.4.2 Verification That the Intensity of the Treatment Contrast Is Large 52
2.4.3 Intent to Treat Estimates 55
2.4.4 Estimated Impacts of Treatment on the Treated 58
2.4.5 Test for Variations in the Effect by School Characteristics 61
2.4.6 Test for Variations in the Effect by Student Subgroup 62
2.5 Implications and Conclusion 64
2.6 Tables and Figures 66

Chapter 3 How Sensitive Are School Value-Added Estimates to Methodology? 86
3.1 Introduction 86
3.2 Value-Added Overview and Related Literature 88
3.2.1 General Overview 88
3.2.2 Similar Papers 90
3.3 Methods 94
3.4 Data 96
3.5 Results 97
3.5.1 Distribution of Value-Added Coefficients and Their Precision . . 97
3.5.2 Sensitivity to Covariates and Model Class 99
3.5.3 Correlations Between Value-Added and Level of Test Score . . . 103
3.6 Discussion 105
3.7 Tables and Figures 109

Appendix A Combined Appendix 119
A.1 Appendix to Chapter 1 119
A.1.1 Index Score 119
A.1.2 Standard Errors 121
A.2 Appendix to Chapter 2 129
A.3 Appendix to Chapter 3 136
A.3.1 Arellano-Bond Estimation 136

Bibliography 146

LIST OF FIGURES

Figure 1.1. Admissions Criteria for University of California Colleges 23

Figure 1.2. UCSD Admissions Formula Used Through 2011 25

Figure 1.3. Histogram of API Decile Among Applicants 26

Figure 1.4. Graphical Depiction of Table 1.4 Results 26

Figure 2.1. Simple Example of Lottery Admissions 66

Figure 2.2. Priority Codes for Magnet, VEEP, and Choice Programs 67

Figure 3.1. Histograms of Value-Added Coefficients, Math 115

Figure 3.2. Histograms of Value-Added Coefficients, Reading 116

Figure 3.3. Scatterplots of Value-Added Coefficients, Math 117

Figure 3.4. Scatterplots of Value-Added Coefficients, Reading 118

Figure A1. Histogram of Index Score 120

Figure A2. Scatterplots of Shrunk Value-Added Coefficients, Math 144

Figure A3. Scatterplots of Shrunk Value-Added Coefficients, Reading 145

LIST OF TABLES

Table 1.1. Descriptive Statistics 27

Table 1.2. Probit Average Marginal Effects with Varying Controls 28

Table 1.3. Probit Average Marginal Effects by Campus 29

Table 1.4. Probit Average Marginal Effects with Each API Decile 30

Table 1.5. Actual, Predicted, and Counterfactual Admission by API Decile (Percentage Shares) 31

Table 1.6. Actual, Predicted, and Counterfactual Enrollment by API Decile (Percentage Shares) 32

Table 1.7. Regression Results for Cumulative College GPA 33

Table 2.1. Applications to True Lotteries, and Characteristics of Applicants . . 68

Table 2.2. The Relationship Between Winning a Lottery and Student Characteristics 69

Table 2.3. Comparison of School Characteristics: Choice Program Schools Versus Local Schools 70

Table 2.4. Intent to Treat Estimates of the Impact of Winning a Lottery on High School Outcomes 72

Table 2.5. Intent to Treat Estimates of the Impact of Winning a Lottery on Postsecondary Enrollment Outcomes 73

Table 2.6. Intent to Treat Estimates of the Impact of Winning a Lottery on Postsecondary Graduation Outcomes 74

Table 2.7. Estimates of the Impact of Winning a Lottery and Enrolling on High School Outcomes 75

Table 2.8. Estimates of the Impact of Winning a Lottery and Enrolling on Postsecondary Enrollment Outcomes 76

Table 2.9. Estimates of the Impact of Winning a Lottery and Enrolling on Postsecondary Graduation Outcomes 77

Table 2.10. Estimates of the Impact of Enrolling in Any School to Which Applied on High School Outcomes 78

Table 2.11. Estimates of the Impact of Enrolling in Any School to Which Applied on Postsecondary Enrollment Outcomes 79

Table 2.12. Estimates of the Impact of Enrolling in Any School to Which Applied on Postsecondary Graduation Outcomes 80

Table 2.13. Significance of Interactions Between Winning a Lottery and Differences in School Characteristics 81

Table 2.14. Significance of Interactions Between Winning a Lottery and Student Characteristics 84

Table 3.1. Descriptive Statistics 109

Table 3.2. Summary Information About Coefficients and Standard Errors, Before Shrinkage 110

Table 3.3. Spearman Rank Correlations of Value-Added Coefficients Across Control Specifications 111

Table 3.4. Spearman Rank Correlations of Value-Added Coefficients Across Model Classes 112

Table 3.5. Relative Ranking of Schools, Compared Across Models, for Math . 113

Table 3.6. Relative Ranking of Schools, Compared Across Models, for Reading 114

Table 3.7. Spearman Rank Correlations Between Value-Added and Level of Test Score 114

Table A1. Within-High School Error Correlation Thresholds 123

Table A2. Descriptive Statistics, Alternate View 124

Table A3. Selected Means by UC Campus 125

Table A4. OLS Admission Results by Campus 126

Table A5. Probit Average Marginal Effects with API Decile as Spline 127

Table A6. Average Marginal Effects for Probit Regression of Yield, Each API Decile 128

Table A7. The Relationship Between Winning a Lottery and Attrition 129

Table A8. More Detailed Version of Intent to Treat Estimates, High School Outcomes 130

Table A9. More Detailed Version of Intent to Treat Estimates, Postsecondary Enrollment Outcomes 132

Table A10. More Detailed Version of Intent to Treat Estimates, Postsecondary Graduation Outcomes 134

Table A11. Intent to Treat Estimates on Postsecondary Graduation Outcomes with 2-Year and 4-Year Degrees Combined 136

Table A12. Variable Classification in Arellano-Bond Regressions 138

Table A13. Spearman Rank Correlations of Shrunk Value-Added Coefficients Across Control Specifications 139

Table A14. Spearmank Rank Correlations of Shrunk Value-Added Coefficients Across Model Classes 140

Table A15. Relative Ranking of Schools Using Shrunk Estimates, Compared Across Models, for Math 141

Table A16. Relative Ranking of Schools Using Shrunk Estimates, Compared Across Models, for Reading 142

Table A17. Spearman Rank Correlations Between Shrunk Value-Added and Level of Test Score 143

ACKNOWLEDGEMENTS

I would like to acknowledge Professors Julian Betts and Julie Cullen for their support as co-chairs of my committee. They showed more patience and devoted more time and energy to me than I could have asked for. I would also like to acknowledge Andy Zau for his constant support with the school district data. Finally, I would like to thank the many friends who helped throughout graduate school in both direct and indirect ways.

Chapter 1, in full, is material currently being prepared for submission for publication. Young, Sam M. The dissertation author was the sole investigator and author of this paper.

Chapter 2, in full, is an unpublished manuscript that contains the following co-authors. Betts, Julian R.; Young, Sam M.; Zau, Andrew C.; Bachofer, Karen Volz. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is an unpublished manuscript that contains the following co-authors. Betts, Julian R.; Young, Sam M. The dissertation author was the primary investigator and author of this paper.

VITA

2008	Bachelor of Arts, University of California, Davis
2012	Master of Arts, University of California, San Diego
2010–2016	Teaching Assistant, Department of Economics University of California, San Diego
2011–2016	Research Assistant, University of California, San Diego
2017	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

"College Prep for All: Will San Diego Students Meet Challenging New Graduation Requirements?" San Francisco: Public Policy Institute of California, April 2016

FIELDS OF STUDY

Major Field: Economics

ABSTRACT OF THE DISSERTATION

Essays in the Economics of Education

by

Sam Ming Young

Doctor of Philosophy in Economics

University of California, San Diego, 2017

Professor Julian Betts, Co-Chair
Professor Julie Cullen, Co-Chair

This dissertation consists of three papers on the economics of education. All pertain to the notion of higher- or lower-achieving schools, in some way.

Chapter 1 estimates the magnitude by which the achievement level of a student's high school affects college admissions once students make the decision to apply. Using administrative data from the University of California, I find a significant negative association between a school's overall achievement level and the probability of admission at 5 out of 8 UC campuses between 2001–2006, including 5 of the 6 most selective. The average magnitude across all UC campuses is about -0.8 percentage points for a

one-decile increase in school-average achievement.

Chapter 2 uses lottery data from San Diego Unified School District to study the long-term effects of school choice on students' postsecondary outcomes. This paper contributes to the school choice literature by being among the first to focus on postsecondary enrollment and graduation, rather than nearer-term outcomes such as standardized test scores in the immediate years following the lottery. Despite large effects on what secondary schools students attend, we find mostly insignificant effects on postsecondary enrollment and graduation outcomes from winning a school choice lottery.

Chapter 3 examines the sensitivity of school value-added estimates to the methodology used. Education scholars frequently use "value-added" models to assess schools' effectiveness. Yet there are many variations of the model that scholars sometimes use, which may affect corresponding results. Using a menu of three model classes and three control specifications each — for a total of 9 specifications — we find a relatively modest sensitivity to the control specification, and a somewhat greater sensitivity to the class of value-added model. Precision appears sensitive to the inclusion of school-level controls and whether researchers use panel instrument techniques to address dynamic panel bias. We also observe a moderate positive correlation between value-added estimates and schools' average levels of test scores.

Chapter 1

Do Colleges Favor Students from Lower-Achieving High Schools? Evidence from the University of California

1.1 Introduction

In addition to race and socioeconomic status, the high school that a student attends is another important dimension of an advantaged versus disadvantaged past. Coursework, peers, and overall resources vary across schools, giving rise to vastly different experiences between students of different secondary schools. Many of these differences can be systematically approximated using a school's overall level of achievement. Generally speaking, the higher a school's average level of achievement, the more resources and advantages it typically offers.

In light of this, it is possible that colleges may incorporate the overall achievement or "quality" of a student's school into their admissions decisions. Many colleges, for instance, value equity or diversity along a variety of dimensions, and may wish to increase the share of students from disadvantaged schools. School quality may also help put students' past achievements into context, thereby conveying additional information about

student ability.

Whether driven by diversity or other reasons, examples can be found in which school quality directly enters into admissions. For instance, section 51.805 of the Texas Education Code lists 18 factors that determine admissions for students who do not otherwise gain automatic admission. These include "the financial status of the applicant's school district" and "the performance level of the applicant's school." Similarly, colleges in the University of California system follow a set of 14 admissions criteria designed to evaluate students relative to their life circumstances. Coming from a "disadvantaged social or educational environment" is listed among various personal background factors taken into account. To help achieve this goal, at least two UC campuses used (and displayed on their websites) admissions rubrics giving explicit bonuses to students from lower-achieving schools.

Moreover, these bonuses occur even net of the admissions consideration of class rank — which is otherwise often discussed as a way in which the high school affects admissions. Outside of class rank, there has been little academic attention paid to the direct use of school quality in admissions, and little is known about its magnitude. This paper attempts to help fill this gap.

Using administrative data from the University of California, I estimate the admission rules of UC colleges between 2001 and 2006, paying particular attention to how a school's average achievement level affects admission probability net of all other observed information. This analysis is done from a partial equilibrium perspective, focusing on the admissions stage itself, after students have already completed high school and submitted their applications. Absent any barriers to identification, the effect of interest answers "How much more or less likely is an otherwise-identical applicant (ex post) to be admitted when coming from a higher-achieving school?"[1]

[1]To clarify, 'ex post' means the comparison of interest is between students who are identical *after*

The administrative data allow me to control for many essential determinants of college admissions, including high school GPA, SAT scores, intended major, and socioeconomic status. However, these data ultimately make up only a subset of the information observed during actual admissions. A general threat to identification is that, among the remaining items that are observed *during admissions but not in the data*, there may be other factors that vary systematically with school quality. Effects of these factors may be incorrectly attributed to school quality, resulting in biased parameter estimates.

Taking results at face value, I find school quality to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006 while controlling for other factors — including status within the UC's system-level percent plan. Aggregating across campuses, the average association is about -0.8 percentage points for a one-decile increase in high school achievement, or roughly equivalent to a 0.04 decrease in high school GPA. The estimated marginal effect appears to be somewhat larger at the lowest-achieving schools. Back-of-the-envelope calculations suggest that this admissions practice increases the share of attending students from below the median high school by 14 percent (from 24.7 to 28.2 percent), while decreasing the share from above the median high school by 5 percent.

Taking results at face value, I find school quality to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006 while controlling for other factors — including status within the UC's system-level percent plan. Aggregating across campuses, the average association is about -0.8 percentage points for a one-decile increase in high school achievement, or roughly equivalent to a 0.04 decrease in high school GPA. The estimated marginal effect appears to be somewhat larger at the lowest-achieving schools. Back-of-the-envelope calculations suggest that

having gone through high school, and who both decided to apply to a given college. Except for their high school, these students look identical from a college's perspective. These students may not have been identical 'ex ante,' or before entering high school.

this admissions practice increases the share of attending students from below the median high school by 14 percent (from 24.7 to 28.2 percent), while decreasing the share from above the median high school by 5 percent.

Acknowledging the possibility of omitted variables bias, these results provide an initial estimate, or a starting point, for an effect that has had little past benchmark for its magnitude. This magnitude matters from a policy perspective, since it ultimately affects students' access to postsecondary education. Any effects on either the extensive or intensive margin of college attendance may have long-term impacts on students' earnings and other life outcomes.[2] Admissions practices may also affect students' incentives for which high schools to attend, with both partial and general equilibrium impacts on students.[3] While these latter topics fall outside the scope of the current piece, they provide reasons to study the topic at hand and gain a more complete understanding of how high schools affect postsecondary education access.

The rest of the paper proceeds as follows. Section 1.2 reviews background and related literature. Section 1.3 provides a conceptual and empirical framework. Section 3.4 describes the administrative data used. Section 3.5 presents empirical results. Section 1.6 concludes and offers additional discussion.

[2]The evidence on returns to college brand name or selectivity is slightly mixed, but more commonly indicates an effect than no effect. Brewer, Eide and Ehrenberg (1999), Behrman, Rosenzweig and Taubman (1996), Hoekstra (2009), and Black and Smith (2006) are among the papers that find significant effects, while Dale and Krueger (2002; 2011) have two notable exceptions that find no significant effect.

[3]Evidence from Texas suggests that its top 10% policy — in which students ranking in the top 10% of their class are guaranteed admission to any state university — induced some students to attend lower-achieving schools (Cullen, Long and Reback, 2013; Cortes and Friedson, 2014). This behavior may have mixed impacts on other students. Students who "trade down" may provide a positive peer externality on raw achievement, but may also confer a negative class rank externality by displacing incumbent students from the top 10%.

1.2 Background and Related Literature

In the broadest sense, this paper fits within the overall study of steps colleges take to increase access for students from disadvantaged groups. Such efforts have been common in the U.S. since at least the 1960s, and in recent decades have expanded to include a widening set of personal background factors. This can be partly seen in Long (2007). Beginning in the 1990s, several states began moving away from (and even banning) traditional race-based affirmative action at public universities. Subsequently, many of these states began considering a broader set of contextual factors, such as family income, the high school setting, and other aspects of hardship.

On the role of the high school, there currently exist two segments of the literature that explore how high schools affect access to postsecondary education. Each of these literatures, however, approaches the issue of access from a slightly different angle than the current piece. As such, neither literature is entirely parallel, nor provides definitive answers to the key question of this paper.

In one literature, a handful of papers examine the relationship between a high school's overall achievement and college attendance. From this group of studies, there is evidence that attending a higher-achieving school increases rates of college attendance (Deming et al., 2014; Berkowitz and Hoekstra, 2011), and even the selectivity of the college attended (Berkowitz and Hoekstra, 2011). These analyses, however, are generally interested in ex ante effects — comparing students who are identical *before* entering high school. Intermediate effects of the high school upon other aspects of a student's record — including the decision of whether to apply to college — are thus included as part of the overall effect. So even as these studies have found beneficial effects from attending a higher-achieving school, it is difficult to disentangle how colleges themselves respond to school quality (once students apply) from the effects of these other intermediate

outcomes.

A second literature focuses on how a school's achievement affects admissions *via class rank*. To the extent that peer achievement matters greatly for class rank, it is sometimes argued that class rank preferences implicitly favor students from lower-achieving schools.[4] Such arguments are often raised in the context of percent plans, in which students ranking in the top $x\%$ of their high school class are guaranteed admission to at least one state university.[5]

By construction, these plans offer at least one form of access to students from disadvantaged schools. However, in practice they may have only limited reach. An important point is that their net effects depend on how students inside the top $x\%$ fare *without* the percent plan in place. In the case of Texas, where the percent threshold occurs at the top 10%, students within the top 10% had already been admitted with *near* certainty prior to the percent plan (Tienda et al., 2003), even to flagship campuses. Conditional on applying, admission outcomes typically remained unchanged for students inside these plans' direct reach.[6]

With this being the case, there is no obvious reason why school quality cannot enter more directly into the admissions process. In fact, both the Texas Education Code and the University of California's list of admissions criteria indicate that this sometimes

[4]The influence of peer achievement on class rank is sometimes called the "frog pond" effect. Espenshade, Hale and Chung (2005) study this phenomenon in the context of college admissions. There is also evidence from lotteries that attending a higher-achieving school may lower class rank. See, for example, Cullen, Jacob and Levitt (2006).

[5]Percent plans began in the late 1990s, with Texas, Florida, and California being the most prominent states with such a plan in place. Most academic studies of these plans have looked at Texas, where the percent plan applies at individual campuses. This, however, is often not the case in other settings. It is thus somewhat unclear how the results from Texas generalize to other settings.

[6]It is still possible that the policy may have affected access by inducing changes to application behavior. The total number of applicant-sending high schools increased after the top 10% policy (Long, Saenz and Tienda, 2010), and top 10% students appeared to focus their applications more heavily toward their preferred campuses (Long, Saenz and Tienda, 2010; Long and Tienda, 2010). These channels, however, are considered "general equilibrium" from the perspective of the current piece, which focuses only on the admissions stage, and takes applicant behavior as given.

occurs even in the presence of a percent plan. What remains missing is that there has been little formal study of this more direct usage. Thus, the overall ways in which high schools affect access to colleges are not fully understood.

1.2.1 UC-Specific Background

For the University of California specifically, Figures 1.1 and 1.2 provide some additional insights.

Figure 1.1 shows a list of 14 admissions criteria that all UC campuses use (University of California, 2012), with the aim of evaluating students' merits relative to the context of their life circumstances. Most relevant for this paper are items 5, 7, and 13, which involve students' high schools.

Item 5 pertains to Eligibility in the Local Context (ELC), or the UC's system-level percent plan. Beginning with the Fall 2001 entering cohort, originally students ranking in the top 4% of their high school class were guaranteed admission to at least one UC campus, though not a campus of the student's choosing.[7] Compared to Texas, the guaranteed admission is considerably weaker, as it does not apply to individual campuses. Nevertheless, it offers one form of access in a way that takes the high school into account.

Even with ELC, however, items 7 and 13 hint at further consideration of students' high schools. Item 7 calls for students' academic performance to be evaluated (at least in part) relative to the peers at their school, while item 13 lists a "disadvantaged social or educational environment" as a specific hardship for UC colleges to take into account. Item 13 in particular hints that there may be a direct bonus associated with attending a lower-achieving school, ceteris paribus. At least two examples can be found in which this was the case; for many years during the 2000s, both UC Davis and UC San Diego

[7]The eventual campus of admission is not required to be among those to which students originally applied. In this way, students cannot functionally select the campus of admission by only applying to their preferred campus.

published detailed admission rubrics on their websites, indicating such a bonus.

UC San Diego's rubric is displayed in Figure 1.2.[8] Up through Fall 2011 admissions, UC San Diego used this formula to score students on an overall level of quality, taking into account both individual achievement and background/contextual factors.[9] Applicants whose overall score exceeded some threshold were subsequently offered admission.

Looking at the formula, bonuses can be seen for various forms of hardship or disadvantage. These include socioeconomic factors, such as low family income or being in the first generation to attend college, but also (separately) the achievement level of a student's school. Within step II, one sees a bonus of 300 points — worth 300 points of combined SAT score or 0.3 points of high school GPA — given to students from the bottom 40% of California public high schools. Furthermore, this bonus was not mutually exclusive from the consideration of ELC. While ELC status did not guarantee admission (to the specific campus), ELC received its own bonus that was also worth 300 points.

1.3 Conceptual and Empirical Framework

Using the UC San Diego example as motivation, consider a stylized model of college admissions. Suppose that, at a given college, students are rated on an overall level of desirability (A_i^*) based on the contents of their application. Applicants whose desirability exceeds some cutoff c are then offered admission. This can be represented

[8]The particular screenshots displayed were retrieved July 2, 2015 using an archived version of UC San Diego's admissions website. The source URL is http://web.archive.org/web/20041205230635/http://admissions.ucsd.edu/dev3/info/comreview.html.

[9]UC Davis' formula was similar, but with slightly different numbers. In recent years, many UC campuses have begun switching to a system of 'holistic review,' which no longer uses a set formula, but similarly tries to combine the many aspects of a student's application into a single review score.

with the following system:

$$A_i^* = \beta Q_i + x_{1i}'\gamma_1 + x_{2i}'\gamma_2 + u_i \tag{1.1}$$

$$Admit_i = 1\{A_i^* > c\} \tag{1.2}$$

where i is an index for students who applied, Q_i denotes the average achievement level or "quality" of a student's high school, x_1 and x_2 denote all other items contained in the application, and u_i denotes an idiosyncratic error, perhaps from the specific admissions officers assigned to each student. Let x_1 contain items of the application that are also available to the researcher (in a dataset), and x_2 contain the remaining items that are missing from the data.

β represents the consideration given to school quality, on the margin, as its own piece of information. Holding fixed all other aspects of the application, β is the degree to which a students is viewed more or less favorably when coming from a slightly higher-achieving school. The goal of this paper is to estimate $\hat{\beta}$ and its corresponding average marginal effect on admission probability across a set of institutions.

For the case where u_i is normally distributed with unit variance, a student's probability of admission, conditional on applying, is $\Phi(\beta Q_i + x_{1i}'\gamma_1 + x_{2i}'\gamma_2)$. With access to the same information used in admissions, β, γ_1, and γ_2 can be consistently estimated using a probit regression, recovering the complete admissions rule. The fact that students sort nonrandomly to high schools does not otherwise prevent identification.

In practice, a challenge arises from the fact that datasets are typically incomplete. That is, they do not contain all items observed during admissions. Some aspects of the application, such as the transcript or personal statements, have many nuances that are difficult to fully capture. When only a portion of this information is available, a practical

version of Equation 1.1 can be written as

$$A_i^* = \beta Q_i + x_{1i}'\gamma_1 + \eta_i, \tag{1.1'}$$

where $\eta_i = x_{2i}'\gamma_2 + u_i$ makes up a composite error term from the researcher's perspective. When the researcher now estimates

$$P(Admit_i = 1) = \Phi(\beta Q_i + x_{1i}'\gamma_1), \tag{1.3}$$

$\hat{\beta}$ and its corresponding marginal effect may be biased in either direction from the truth, depending on what items are in x_2 and how they correlate with Q and x_1. Omitted variables bias is thus a principal threat to identification.

The next section gives a more specific sense of the degree to which omitted variables are a concern, describing the administrative data and what variables are present versus missing.

1.4 Data

This paper uses administrative data on freshman undergraduate applicants to the University of California between 2001 and 2006. The data are gathered by the UC Office of the President (UCOP), which maintains a master database of all applicants to the UC system. In total, 460,704 students appear during these years applying to eight different UC campuses.[10]

The dataset includes individual-level data on students' UC applications and admission outcomes, high school GPA, SAT scores, ELC status, race, and parental income and education. In addition, there is also a summary measure of achievement

[10]Overall there are nine UC campuses with undergraduates. The current analysis omits UC Merced, which had only two admission cohorts during this time frame (Fall 2005 and 2006).

for the high school attended, known as the Academic Performance Index (API). This measure is recorded at decile level in the data, with the highest-achieving schools making up the 10th decile. API decile is subsequently used as the value of Q_i when estimating Equation 1.3.

Prior to decile conversion, API is originally computed using each school's average performance on state standardized tests.[11] Specifically, each year a school's average test scores are mapped to a numeric value between 200 and 1000. These values are computed annually by the California Department of Education, and within California are often used to describe school-level achievement. API is not computed for private or out-of-state schools, due to the measure's origin from the California Public Schools Accountability Act of 1999. Moreover, students from those schools may also face a different set of admissions criteria. To keep things cohesive, all subsequent analysis restricts attention to the 320,721 applicants (about 70% of the total) from California public high schools.

Table 3.1 summarizes the data for this group of students, with additional information in Tables A2 and A3 of the appendix.

A few key patterns can be seen across API deciles. First, the vast majority of applicants come from the state's higher-achieving schools.[12] About 41 percent of applicants come from API deciles 9–10, or the top 20% of public high schools. Only 10 percent of applicants come from the bottom two API deciles, with an additional 13 percent from deciles 3–4. Overall, this translates to an average API decile of about 7.0 among applicants, and between 6.6 and 7.4 at each campus.

Second, API decile shows a pronounced positive correlation with socioeconomic

[11] Until 2014, API used results of the Standardized Testing and Reporting (STAR) Program and the California High School Exit Exam. STAR was then replaced by a different testing program, though that does not affect the admission years in this paper. For high school students, STAR covered English-language arts, mathematics, history, and science. Additional details about API can be found in California Department California Department of Education (2012).

[12] This can be seen visually in a histogram in Figure 1.3.

status. Whether looking at race, parental income, or parental education, applicants from higher API deciles tend to come from more advantaged backgrounds.

Third, students' past academic achievements vary systematically between API deciles. Students from higher API deciles have higher average SAT scores and (to a lesser extent) high school GPA, while also taking more Advanced Placement courses on average. Yet the students from *lower* API deciles are somewhat more positively selected from within their schools. Looking at the rates of ELC eligibility, about 30 percent of applicants from the lowest-achieving schools (API deciles 1–2) rank within the top 4% of their class, compared to 12 percent from the highest-achieving schools.

Having summarized the dataset as a whole, it is worth noting a few limitations about the UCOP data. While the dataset contains many of the most essential determinants of admissions, there are undoubtedly many additional items in students' applications that are not present. For instance, there is no information about personal statements, extracurricular achievements, the exact transcript, or class rank (other than ELC status). Admission effects stemming from these items cannot be directly controlled during estimation, and become possible sources of omitted variables bias.

Another limitation is that, aside from API decile, no additional information is provided about students' high schools. School name and location are not stated, nor are schools given unique identifiers. (These and other steps have been taken to guard student anonymity.) This prevents school fixed effects from being used during estimation, and also prevents instrumenting for API decile, since most instruments require knowledge about location.

A final limitation is that, for anonymity purposes, some variables are blurred from their exact values and are instead reported within a bin range. The most notable of these are high school GPA and SAT scores. Much of this issue can ultimately be resolved using an additional measure of student achievement, known as the Index score. Section A.1.1

of the appendix discusses this measure in more detail.

1.5 Results

Tables 1.2–1.4 show the marginal effects from estimating Equation 1.3 using various specifications. Before presenting these results, I note that standard errors may be somewhat understated in these tables, due to an inability to cluster error terms at the desired level.[13] Appendix A.1.2 discusses this issue in more detail, and computes the degree of within-cluster error correlation needed to make results insignificant.

Table 1.2 shows results with varying sets of controls. The top panel shows results with all campuses pooled, while the bottom panel shows the estimated marginal effect of API decile at each campus. For brevity, I primarily comment on the pooled marginal effect estimate, but these patterns are also roughly true at individual campuses.

Pooled, the estimated marginal effect appears as large as -0.021 when only controlling for the most essential student achievement measures and year. That is, attending a higher-achieving school by one decile associates with a 2.1 percentage point decrease in admission probability in column 1. Additional academic controls other than ELC (column 2) do not affect this estimate by much.

Column 3 next adds controls for race, parental income, and parental education — all of which are strongly correlated with API decile, but also bring many advantages/disadvantages of their own.[14] By all indications, UC colleges actively take these contextual factors into account. Once controlling for these demographics, the estimated marginal effect of API decile shrinks noticeably to -0.013, but remains significant.

[13] Intuitively, error terms should be grouped at the high school level due to various similarities in experience between students of the same school. This clustering cannot currently be achieved, since individual high schools are not indicated in the data.

[14] Under California Proposition 209, race cannot be explicitly considered during UC admissions. However, race is ultimately included in these regressions since it may affect other aspects of a student's experience that do enter into admissions.

Column 4 adds dummies for whether students applied to each specific UC campus, and the mean SAT score of attending students among applied campuses. Borrowing somewhat from Dale and Krueger (2002; 2011), students' decisions about which colleges to apply to contain important information about their overall quality — in many cases incorporating information from items in x_2 (e.g. extracurriculars) that colleges see but are not in the data. These additional variables help control for otherwise-unobserved aspects of student quality. While these variables themselves are usually significant, suggesting they do convey additional information about students, the average marginal effect for API decile remains mostly unchanged at -0.012.

Finally, column 5 adds Eligibility in the Local Context, or a dummy for ranking within the top 4% of one's high school class. This guarantees eligible students admission to at least one UC campus, but not a campus of their choosing.

While ELC is not legally binding for individual campuses, most UC campuses still appear to respond to ELC status. Across all campuses, ELC is associated with an 11 percentage point increase in admission probability, on average, even after controlling for various other measures of achievement. This has important implications on access for students from lower-achieving schools. Since many applicants from lower API deciles are ELC eligible (up to 30 percent from the lowest-achieving schools), the consideration toward ELC also provides a degree of access to students from these schools. Yet this does not appear to be a complete substitute for more direct use of school quality. After controlling for ELC in column 5 of Table 1.2, the marginal effect of API decile shrinks considerably at some campuses, and by about a third in the pooled case compared to column 4. But the pooled marginal effect remains significantly negative, suggesting some additional direct consideration. The final pooled estimate of -0.8 percentage points is, on average, equivalent to about a 3.5 point decrease in Index score, or a 0.04 decrease in high school GPA.

Turning attention to individual campuses, Table 1.3 shows results for each campus using the full set of controls. In this table and others, campuses are ordered by average SAT score among attending students, so that campuses with similar selectivity appear adjacent to one another.

Campus-specific marginal effects range from $+0.002$ to -0.024, and are significantly negative at five out of eight campuses — including five of the six most selective. While the estimated marginal effect is nominally positive and significant at the remaining three campuses (Irvine, Santa Cruz, and Riverside), it is worth noting that none of these campuses has an estimated marginal effect of more than $+0.002$, or 0.2 percentage points. Moreover, analysis of standard errors (see Appendix A.1.2) suggests that the positive significance disappears at these campuses under corrected standard errors clustered by high school. Magnitudes are much more pronounced at the five campuses where the marginal effect appears negative.

1.5.1 Nonlinearity Across API Deciles

Table 1.3 reports the average marginal effect at each campus as a single parameter. Yet it is possible that the marginal effect is nonlinear across API deciles. For instance, some campuses may view the difference between a 2nd and 3rd decile school (e.g.) to be more or less meaningful than the difference between a 9th and 10th decile school.

To allow for nonlinearity, I also estimate each admissions rule using a specification that includes each API decile individually. These results are displayed in Table 1.4. For all campuses, the 7th API decile (roughly the average among applicants) is omitted as a baseline. Table cells report associated differences relative to the 7th API decile, while marginal effects of a one-decile change in API can be computed by comparing adjacent estimates.

The main patterns from this specification can be seen in Figure 1.4. The marginal

change in admission probability appears to be nonlinear and generally largest at lower-achieving high schools. While the exact marginal effect varies from decile to decile, and across campuses, results that include API decile using a spline function (see Table A5) indicate that the average marginal effect is generally around -1.5 percentage points in the bottom half of the API distribution, and fairly close to zero in the upper half.

Compared to a counterfactual where all students are treated as coming from the 7th API decile (roughly the average among applicants at each campus), these results suggest that, for an average UC campus, school quality can contribute up to a 7.5 percentage point increase in admission probability for students from the lowest-achieving schools. Relative to those students' counterfactual probability of admission (41.1% across all campuses), this represents an 18 percent proportional increase.[15]

For students from the highest-achieving schools, meanwhile, the decrease in their admission probability appears to be about 0.6 percentage points (or less) at an average campus. This translates to about a 1 percent proportional decrease, relative to those students' counterfactual probability of admission.[16] Thus, whether measured by raw or proportional change, the admissions use of school quality seems to affect students in the lower API deciles much more than students in the upper API deciles.

1.5.2 Enrollment Composition

Using these results, it is possible to construct back-of-the-envelope estimates of how the (apparent) use of school quality affects the resulting pool of admitted and attending students. These results are displayed in Tables 1.5 and 1.6.

Table 1.5 first shows the actual, predicted, and counterfactual composition of admitted students by API decile. In all cases, the predicted probability of admission is the

[15]Here, the counterfactual probability of admission is computed supposing all admission rules are as estimated in Table 1.4, but that all students are evaluated as coming from the 7th API decile.

[16]Computing counterfactual admission probability the same way as before, these students have a counterfactual probability of admission of 53.2% across all campuses to which they apply.

probability implied by the regression in Table 1.4, while the counterfactual probability is computed supposing that all admission rules are as estimated in Table 1.4, but that all students are treated as coming from the 7th API decile.[17] This generates a predicted and counterfactual probability of admission for each student at each campus to which he/she applies. Summing these probabilities within each API decile and within each campus gives the predicted and counterfactual compositions displayed in Table 1.5.

For enrollment composition, I next combine a student's probability of admission with the probability of enrollment *conditional on admission*, at a given campus, to estimate the overall probability that each student enrolls. Mathematically,

$$\hat{E}(Enroll_i) = \hat{P}(Admit_i = 1) \cdot \hat{P}(Enroll_i = 1 | Admit_i = 1), \tag{1.4}$$

with the counterfactual version replacing $\hat{P}(Admit_i = 1)$ with $\hat{P}(Admit_i = 1)_{counter} = \Phi(x'_{1i}\hat{\gamma}_1)$. To obtain $\hat{P}(Enroll_i = 1 | Admit_i = 1)$, I estimate a probit regression of enrollment among the admitted students at each campus, using the same explanatory variables as when modeling admissions in Table 1.4. These conditional enrollment patterns are then assumed true of all applicants to that campus, including those who were not admitted.[18] Summing the resulting values of $\hat{E}(Enroll_i = 1)$ within each API decile and within each campus gives the predicted and counterfactual compositions displayed in Table 1.6.

Looking at the numbers in Tables 1.5 and 1.6, one sees that the predicted admission and enrollment shares (Panel B) generally come quite close to matching their actual

[17] To be precise, the counterfactual probability of admission replaces each student's API decile with 7 and re-computes the probability of admission given by the probit function. Mathematically, $\hat{P}(Admit_i = 1)_{counter} = \Phi(x'_{1i}\hat{\gamma}_1)$, where the estimated effect due to API decile has been nullified. Depending on students' proximity to the admission margin, this may affect some students' admission probability more than others, even within the same API decile and campus.

[18] Table A6 displays the results of these conditional enrollment or "yield" equations. The yield equations can only be estimated on the pool of admitted students, since the researcher never observes whether students who were not admitted would have enrolled if given the chance.

compositions. The most common error is that the model sometimes overestimates the share of attending or admitted students from the very top API deciles. Correspondingly, comparisons between the actual and counterfactual shares in Tables 1.5 and 1.6 may somewhat overstate the impact on students in the upper API deciles.

Looking first at the shares of admits, between 2001 and 2006, about 27.0 percent of admits came from API deciles 1–5, or high schools below the California median. Under the constructed counterfactual, this share may have been only 25.0 percent if not for the admissions consideration given to school quality. At least according to this exercise, this admissions practice increases the share of admits from API deciles 1–5 by about 8 percent (25.0 to 27.0 percent), while decreasing the share of admits from deciles 6–10 by about 3 percent (75.0 to 73.0 percent). As with the estimated marginal effects earlier, the biggest gains (in a proportional sense) appear to accrue at the lowest API deciles. While only about 8.6 percent of admits came from the API deciles 1–2, this is about a 13 percent gain over the estimated counterfactual of 7.6 percent.

Once factoring in students' predicted probability of enrollment, the effects described earlier become somewhat more pronounced among students who ultimately attend. According to Table 1.6, about 28.2 percent of attending students between 2001 and 2006 came from the API deciles 1–5. This is about 14 percent higher than the estimated counterfactual of 24.7 percent. For students from the very lowest-achieving schools (API deciles 1–2), the proportional increase is about 26 percent (7.0 to 8.8 percent). Meanwhile, the decrease in the share of attending students from API deciles 6–10 is about 5 percent (75.3 to 71.8 percent).

1.5.3 Possible Explanations: Academic Ability Versus Other Factors

Preferences for students from lower-achieving high schools, ceteris paribus, may be driven by either academic or non-academic reasons. On one hand, many colleges have preferences for a diverse student body. But school quality may also convey additional information about students' ability, (e.g.) by helping to put students' past achievement into context.

Using college performance measures from the UCOP data, it may be possible to test which of these explanations has more empirical support. If the admissions behavior is driven primarily by academic considerations, one generally expects to see school quality predict college GPA in the same direction as its admissions use. To see whether this is the case, Table 1.7 regresses students' cumulative college GPA on API decile and other observables from the admissions equation.[19]

The top panel shows results using ordinary least squares (OLS). Here, although API decile is often a negative predictor of admission probability, it positively predicts college GPA at each campus. This is consistent with a number of other studies of how high school characteristics predict college grades.[20]

However, selection may play an important role in these OLS results. College GPA, at a given campus, is only observed for students who ultimately enroll — itself a subset of students who are first admitted. College GPA results obtained on this subsample may not be true of the original pool of applicants. Making matters worse, the ways in which students are admitted and enroll may itself create correlations between API

[19] I use cumulative college GPA, rather than first-year GPA, in case the influence of high schools upon college performance weakens over time. However, results using first-year college GPA are very similar to those in Table 1.7. Results using bachelor's degree completion as the outcome are included in the appendix.

[20] Some examples include Betts and Morell (1999), Black et al. (2015), and Rothstein (2004). These papers often do not include a school's overall achievement directly as a regressor, but their results are broadly consistent with a positive relationship.

decile and unobservables, among enrollees, even if no such correlation existed among all applicants.[21]

To try to correct for selection, panels B and C estimate the college GPA equation using a Heckman (1979) correction. The two panels differ in how they model the selection equation. Panel B uses no exclusion restrictions, relying only the nonlinearity of the inverse Mills ratio for identification, while panel C uses admission offers to other UC campuses as selection shifters.

Results in panel B offer somewhat less evidence of API decile being a positive predictor of college GPA, with significance disappearing at four campuses, and the pooled coefficient decreasing by about 60 percent compared to OLS. Yet no campuses show API decile to be a negative predictor of college GPA. Results in panel C, meanwhile, are mostly similar to those obtained using OLS.

These Heckman (1979) specifications have their own challenges to contend with, and ultimately may not fully account for selection.[22] However, all three of the current specifications suggest API decile to be either a neutral or positive predictor of college GPA, ceteris paribus. At least in this sense, it appears the consideration given to school quality during admissions may be motivated more by non-academic considerations. Increasing campus diversity is one possible explanation, though it is

[21] Cushing and McGarvey (2004) consider the case where students must be above some quality threshold to gain admission, but only students *below* some other quality threshold enroll. This may arise if students apply to many colleges, with differing selectivity, and then enroll at the "best" college to which they are admitted. In such a setting, enrolling students who are positively selected on a particular trait are generally *negatively* selected on unobservables, and vice versa. The end result is that coefficients in OLS regressions of college GPA will be biased in the opposite direction from their true sign, or "toward zero."

Such a result may apply here, as the general setup of Cushing and McGarvey (2004) appears broadly true for UC colleges. Once students are admitted, nearly all traits that positively predict admissions negatively predict enrollment, and vice versa. Students with higher predicted probabilities of admission generally have lower (conditional) rates of enrollment.

[22] In the first case (no exclusion restrictions), the inverse Mills ratio remains *close* to a linear combination of other observables. Identification is often tenuous in such cases (Puhani, 2000). In the second case, although admission offers to other UCs are predictive of enrollment, they also correlate with unobserved student ability, or part of the error term in the college GPA equation.

also possible that students from lower API deciles are positively selected on leadership, community involvement, or other non-academic contributions — with these benefits outweighing any costs to academic performance from the colleges' perspective.

1.6 Discussion

This paper examines the direct use of a school's overall achievement or "quality" in college admissions decisions. Though this is potentially a common practice, it so far has received little attention in the literature, and its magnitude is largely unknown. This paper provides an initial attempt to estimate this magnitude across a set of institutions, using administrative data from the University of California.

Controlling for students' high school GPA, SAT scores, race, socioeconomic status, and ELC status, I find a school's overall achievement (measured by API decile) to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006, including five of the six most selective campuses. Averaging across all campuses, a one-decile increase in API is associated with about a 0.8 percentage point decrease in admission probability. This marginal effect appears to be larger at the bottom end of the API distribution, perhaps closer to -1.5 percentage points. Back-of-the-envelope calculations suggest that this admissions practice increases the share of attending students from schools below the California median high school by 14 percent, while decreasing the share from above the median high school by 5 percent.

Due to an inability to observe all items used during admissions, I cannot rule out the possibility of omitted variables bias. These results are thus presented as more of a starting point than a definitive answer for the effect of interest.

This paper is otherwise among the first to consider this particular aspect of how high schools affect access to postsecondary education — an area that is not fully understood. Future work may further explore this area and build upon these results. For

instance, even with the current results, it is unclear what overall share of colleges uses a similar admissions practice, and whether there exist systematic differences by region, by selectivity, between public and private colleges, or other dimensions.

Future work may also estimate results from an ex ante perspective. For many parents and policymakers, the greater interest may be in how school quality affects admissions/enrollment *when families still have a chance to move*. While there exists some evidence of positive ex ante effects (Deming et al., 2014; Berkowitz and Hoekstra, 2011; Dynarski et al., 2013), and general findings of strong demand for higher-achieving schools (Black, 1999; Rothstein, 2006), there also appear to be instances of families "trading down" under specific circumstances (Cullen et al., 2013; Cortes and Friedson, 2014). Incentives, as well as families' *perceptions* of incentives (perhaps influenced by salience), may both affect school choice responses in practice. Other authors may investigate these and other possible behavioral responses, and the resulting impacts on families.

Chapter 1, in full, is material currently being prepared for submission for publication. Young, Sam M. The dissertation author was the sole investigator and author of this paper.

1.7 Tables and Figures

1. Academic Grade Point Average (GPA) calculated on all academic courses completed in the subject areas specified by the University's admission requirements (the a-g* subjects), including additional points for completion of University certified honors courses (see 4, below). It is recommended that the maximum value allowed for the GPA shall be 4.0. *An additional subject requirement, labeled "g" Visual & Performing Arts was added in 2002.

2. Scores on the following tests: the Scholastic Assessment Test I or the American College Test. Effective with applicants applying for Fall 2012 and thereafter, the SAT Subject Tests are not required but may be recommended by some programs on some campuses.

3. The number, content of, and performance in courses completed in academic subjects beyond the minimum specified by the University's eligibility requirements.

4. The number of and performance in University approved honors courses, College Board Advanced Placement courses, International Baccalaureate courses, and transferable college courses completed. It is recommended that caution be exercised in order not to assign excessive weight to these courses, especially if considerable weight already has been given in the context of 1, above. Additionally, in recognition of existing differences in availability of these courses among high schools, it is recommended that reviewers assess completion of this coursework against the availability of these courses at the candidate's secondary school.

5. Being identified as eligible in the local context, by being ranked in the top 9% of the class at the end of the junior year, as determined by academic criteria established by the University of California.**Effective with applicants applying for Fall 2012 or thereafter, ELC increased from the top 4% to the top 9%.

6. The quality of the senior year program, as measured by type and number of academic courses (see 3 and 4, above) in progress or planned.

7. The quality of academic performance relative to the educational opportunities available in the applicant's secondary school.

8. Outstanding performance in one or more specific academic subject areas.

9. Outstanding work in one or more special projects in any academic field of study.

10. Recent, marked improvement in academic performance, as demonstrated by academic grade point average and quality of coursework (see 3 and 4, above) completed and in progress, with particular attention being given to the last two years of high school.

Figure 1.1. Admissions Criteria for University of California Colleges

11. Special talents, achievements, and awards in a particular field, such as in the visual and performing arts, in communication, or in athletic endeavors; special skills, such as demonstrated written and oral proficiency in other languages; special interests, such as intensive study and exploration of other cultures; or experiences that demonstrate unusual promise for leadership, such as significant community service or significant participation in student government; or other significant experiences or achievements that demonstrate the applicant's promise for contributing to the intellectual vitality of a campus.

12. Completion of special projects undertaken either in the context of the high school curriculum or in conjunction with special school events, projects or programs co-sponsored by the school, community organizations, postsecondary educational institutions, other agencies, or private firms, that offer significant evidence of an applicant's special effort and determination or that may indicate special suitability to an academic program on a specific campus.

13. Academic accomplishments in light of the applicant's life experiences and special circumstances. These experiences and circumstances may include, but are not limited to, disabilities, low family income, first generation to attend college, need to work, disadvantaged social or educational environment, difficult personal and family situations or circumstances, refugee status, or veteran status.

14. Location of the applicant's secondary school and residence. These factors shall be considered in order to provide for geographic diversity in the student population and also to account for the wide variety of educational environments existing in California.

Figure 1.1 Admissions Criteria for University of California Colleges, Continued

Selection Process

Drawing upon the broad guidelines set forth by the Board of Admissions & Relations with Schools (BOARS), the UCSD Faculty Committee on Undergraduate Admissions has approved the following procedures for freshman selection which will be implemented by the Office of Admissions & Relations with Schools.

Step I - Academic Review	Maximum Consideration
Uncapped Grade Point Average (GPA)	4,500
Scores of All Required Exams	3,200
Number of "A-G" Courses Beyond the Minimum	500
Step II - Additional Academic Factors	
Eligibility in the Local Context (ELC)	300
Educational Environment	300
Step III - Socioeconomic Factors	
Low Family Income	300
First Generation College Attendance	300
Step IV – Personal Characteristics and Achievement Factors	
Demonstrated Leadership	300
Special Talents/Achievements/Awards	300
Community and Volunteer Service	300
Participation in Pre-Collegiate/Motivational and Enrichment Programs	300
Special Circumstances/Personal Challenges	500

Step V – Computing a Comprehensive Review Score

Eligible applicants will be assigned a comprehensive review score by totaling the scores from each category listed in steps 1 through 4. Eligible applicants are then ranked based upon that assigned score. Applicants with the strongest combination of academic, personal characteristics and achievement factors will be admitted in sufficient numbers to meet campus enrollment goals.

Inset: Description of Educational Environment

Educational Environment

A disadvantaged educational/school environment indicates that the applicant attends a California high school that is among the 4th or 5th quintile of all California public high schools using the following academic indicators: high school completion rate, percentage of students enrolled in college preparation classes, percentage of students enrolled in Advanced Placement/Honors courses, percentage of students admitted to the UC/CSU, and the percentage of students taking the Scholastic Aptitude Test (SAT) or the American College Test (ACT).

Figure 1.2. UCSD Admissions Formula Used Through 2011

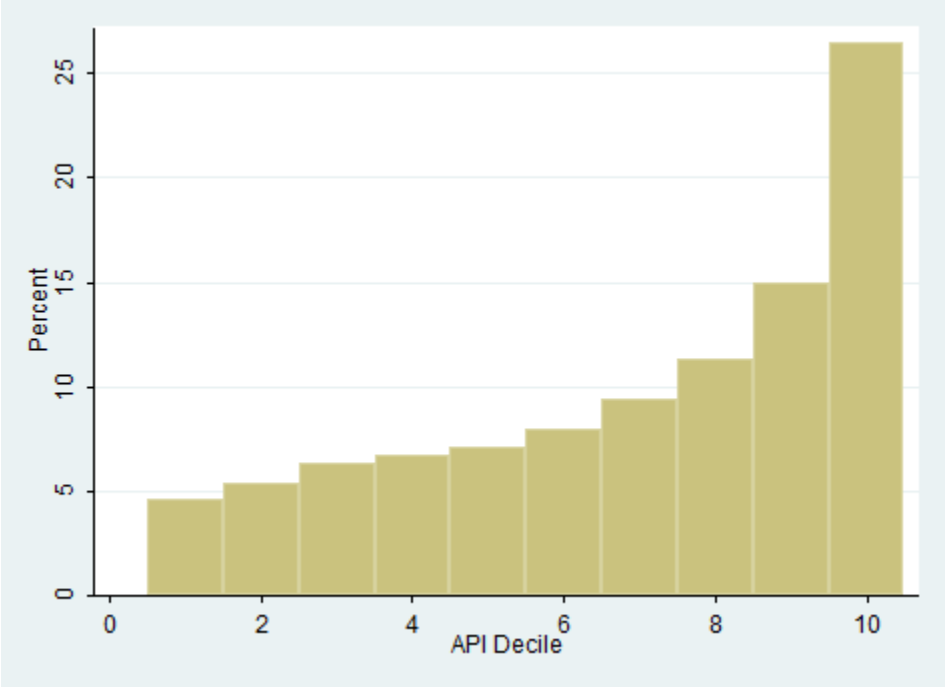


Figure 1.3. Histogram of API Decile Among Applicants

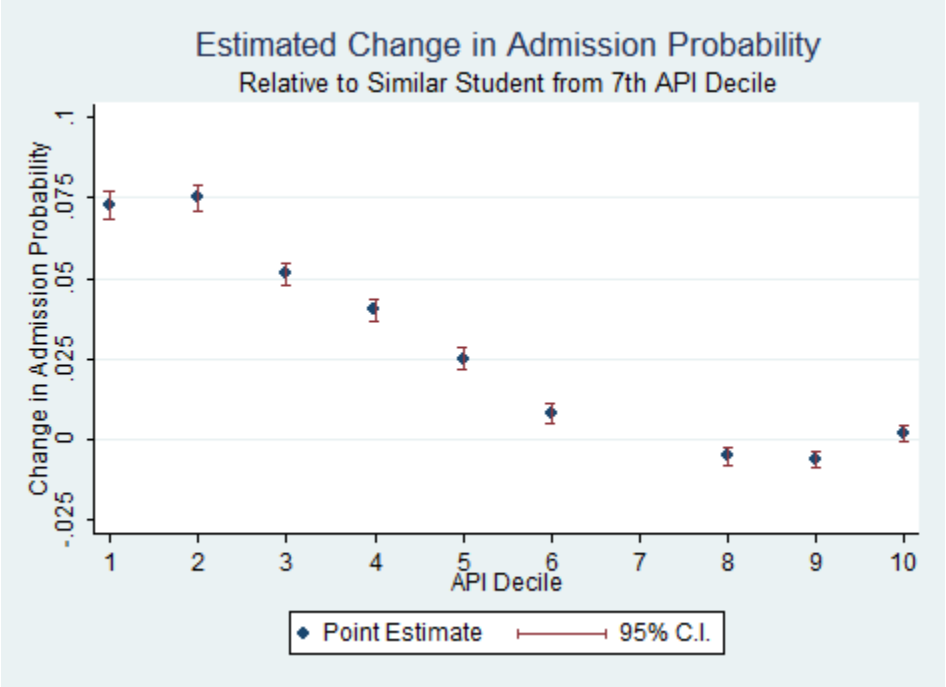


Figure 1.4. Graphical Depiction of Table 1.4 Results

Table 1.1. Descriptive Statistics

	API Deciles					
	1-2	3-4	5-6	7-8	9-10	All
N	31,847	41,695	47,952	66,255	132,972	320,721
Share of Applicants	0.10	0.13	0.15	0.21	0.41	1
Student Achievement Measures						
HS GPA	3.27	3.33	3.40	3.42	3.41	3.38
	(0.50)	(0.49)	(0.47)	(0.46)	(0.45)	(0.47)
SATI Total	982	1071	1132	1173	1235	1161
	(152)	(170)	(164)	(158)	(147)	(175)
AP Courses	6.65	7.07	7.08	7.26	7.39	7.20
	(4.98)	(5.44)	(5.28)	(5.45)	(5.28)	(5.31)
ELC[a]	0.30	0.27	0.25	0.20	0.12	0.19
Socioeconomic Background						
Parental Income Percentile	44.5	56.0	66.7	72.8	78.6	68.4
	(26.2)	(29.3)	(28.9)	(27.0)	(24.8)	(29.1)
Parental Education Percentile	32.3	49.6	61.5	69.6	77.8	65.4
	(29.9)	(32.0)	(29.8)	(26.6)	(22.5)	(30.4)
Underrepresented Minority	0.67	0.40	0.25	0.18	0.09	0.23
White	0.07	0.21	0.35	0.42	0.42	0.35
Asian	0.23	0.33	0.33	0.32	0.40	0.35
Other/Unknown Race	0.04	0.06	0.07	0.08	0.09	0.08
Admission Rates						
Admitted to Berkeley	0.27	0.26	0.26	0.26	0.27	0.27
Admitted to Los Angeles	0.24	0.25	0.24	0.24	0.29	0.26
Admitted to San Diego	0.42	0.45	0.45	0.43	0.45	0.44
Admitted to Irvine	0.44	0.50	0.57	0.59	0.64	0.58
Admitted to Santa Barbara	0.57	0.59	0.60	0.54	0.50	0.54
Admitted to Davis	0.66	0.65	0.62	0.60	0.64	0.63
Admitted to Santa Cruz	0.68	0.75	0.80	0.81	0.82	0.79
Admitted to Riverside	0.78	0.82	0.88	0.89	0.92	0.88

Notes: Standard deviations are in parentheses. Admission rates are conditional on applying.

a — 'ELC' is short for Eligibility in the Local Context, which denotes whether a student ranks in the top 4% of his/her high school class by GPA. These students are guaranteed admission to at least one UC campus, but not a campus of the student's choosing.

Table 1.2. Probit Average Marginal Effects with Varying Controls

Panel A: Pooled Regression

	(1)	(2)	(3)	(4)	(5)
API Decile (10=Highest Achiev.)	-0.021***	-0.020***	-0.013***	-0.012***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ELC Status					0.110***
					(0.001)
Observations	1,195,656	1,195,656	1,195,656	1,195,656	1,195,656
Pseudo R-squared	0.477	0.498	0.510	0.520	0.527

Panel B: Marginal Effect of API Decile by Campus

Berkeley	-0.020***	-0.019***	-0.013***	-0.013***	-0.009***
Los Angeles	-0.022***	-0.021***	-0.013***	-0.013***	-0.011***
San Diego	-0.029***	-0.029***	-0.018***	-0.017***	-0.011***
Irvine	-0.010***	-0.009***	-0.008***	-0.007***	0.001***
Santa Barbara	-0.039***	-0.038***	-0.028***	-0.026***	-0.024***
Davis	-0.025***	-0.025***	-0.014***	-0.012***	-0.008***
Santa Cruz	-0.005***	-0.004***	0.000	0.000	0.001*
Riverside	0.001***	0.001***	0.001***	0.002***	0.002***
Application Year (3-Year Bin)	Y	Y	Y	Y	Y
HSGPA and SATI[a]	Y	Y	Y	Y	Y
Other Academic Controls[b]		Y	Y	Y	Y
Race, Parental Income and Ed.			Y	Y	Y
UC Campuses Applied[c]				Y	Y

Notes: The dependent variable is admission, conditional on having applied. The pooled regression uses separate intercepts for each campus, but otherwise a single set of coefficients. To maintain the full sample across specifications, explanatory variables add an indicator for "missing value" when needed. Except for income percentile, this is typically only a small percentage of cases. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

a – Includes dummies for bin ranges of high school GPA and SAT Math and Verbal scores, along with an additional measure known as the Index score, which is a linear combination of high school GPA and SAT scores. Index score is included using a cubic polynomial.
b – Includes SATII Writing and Third Subject scores, the number of Advanced Placement courses taken, and intended college major.
c – Includes dummies for whether a student applied to each UC campus, along with the mean SAT score of attending students among campuses applied.

Table 1.3. Probit Average Marginal Effects by Campus

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	-0.009***	-0.011***	-0.011***	0.001***	-0.024***	-0.008***	0.001*	0.002***	-0.008***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0002)
ELC Status	0.092***	0.032***	0.156***	0.298***	0.131***	0.202***	0.063***	0.052***	0.110***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.007)	(0.004)	(0.001)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.453	0.439	0.594	0.583	0.519	0.476	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: The dependent variable is admission, conditional on having applied. All regressions include the full set of controls from Table 1.2. The pooled regression uses separate intercepts for each campus, but otherwise a single set of coefficients. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.4. Probit Average Marginal Effects with Each API Decile

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
1st API Decile	0.094***	0.113***	0.069***	0.031***	0.110***	0.104***	0.007	-0.014***	0.073***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.002)
2nd API Decile	0.076***	0.119***	0.081***	0.019***	0.134***	0.115***	0.009	-0.005	0.075***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.004)	(0.002)
3rd API Decile	0.037***	0.078***	0.069***	0.003	0.106***	0.083***	0.005	-0.011***	0.051***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.002)
4th API Decile	0.034***	0.049***	0.051***	-0.005	0.095***	0.059***	0.007	-0.006	0.040***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.002)
5th API Decile	0.019***	0.018***	0.033***	0.005	0.069***	0.030***	0.000	0.000	0.025***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.002)
6th API Decile	0.005	-0.001	0.011***	0.003	0.038***	0.003	0.001	0.005	0.008***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
7th API Decile	—	—	—	—	—	—	—	—	—
8th API Decile	-0.007*	-0.006*	-0.009***	0.010***	-0.022***	-0.000	-0.002	-0.001	-0.005***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)
9th API Decile	-0.010***	-0.005	-0.015***	0.015***	-0.046***	-0.001	0.011***	0.009***	-0.006***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.001)
10th API Decile	-0.004	0.004	-0.010***	0.020***	-0.058***	0.030***	0.012***	0.005*	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.001)
ELC Status	0.094***	0.036***	0.155***	0.293***	0.132***	0.197***	0.062***	0.052***	0.110***
	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.007)	(0.004)	(0.001)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.595	0.584	0.520	0.480	0.505	0.478	0.528
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517
P-value for Linearity[a]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000

Notes: The 7th API decile is omitted as a baseline. All regressions include the full set of controls from Table 1.2. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

a — Reports the p-value on an F test that the probit coefficients between consecutive API deciles are all equidistant.

Table 1.5. Actual, Predicted, and Counterfactual Admission by API Decile (Percentage Shares)

Panel A: Actual Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	4.6	4.4	3.0	3.5	4.2	3.6	3.7	5.0	3.9
2	4.9	5.4	4.2	4.5	5.0	4.2	3.8	6.2	4.7
3	5.5	6.0	5.5	5.4	5.8	5.2	4.9	7.0	5.7
4	6.2	6.1	6.1	5.7	6.4	6.0	5.5	7.2	6.2
5	6.3	5.6	6.5	6.3	6.9	6.5	6.1	7.3	6.5
6	7.2	6.1	7.4	7.0	7.8	8.3	7.9	7.4	7.5
7	8.6	7.7	8.9	8.6	9.1	8.5	9.6	9.2	8.9
8	10.4	9.6	10.8	10.6	11.2	10.7	12.2	10.6	10.8
9	14.5	14.6	15.2	16.5	14.9	13.8	15.2	15.0	15.0
10	31.9	34.4	32.3	31.9	28.7	33.2	31.2	25.2	30.7

Panel B: Predicted Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	4.6	4.3	3.0	3.5	4.2	3.6	3.7	4.9	3.9
2	4.8	5.3	4.2	4.5	4.9	4.2	3.7	6.2	4.7
3	5.4	6.0	5.5	5.3	5.8	5.1	4.9	7.0	5.7
4	6.2	6.1	6.1	5.7	6.4	6.0	5.4	7.2	6.2
5	6.3	5.7	6.5	6.3	6.8	6.5	6.1	7.3	6.5
6	7.2	6.1	7.4	7.0	7.8	8.3	7.9	7.4	7.5
7	8.6	7.8	8.9	8.6	9.2	8.5	9.6	9.2	8.9
8	10.4	9.6	10.8	10.6	11.2	10.8	12.2	10.6	10.9
9	14.5	14.6	15.2	16.5	14.9	13.8	15.2	15.1	15.1
10	31.9	34.4	32.3	32.0	28.8	33.2	31.2	25.2	30.8

Panel C: Counterfactual Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	3.0	2.7	2.6	3.4	3.3	3.1	3.7	5.1	3.4
2	3.6	3.3	3.5	4.4	3.8	3.6	3.7	6.2	4.2
3	4.8	4.6	4.8	5.4	4.7	4.7	4.9	7.1	5.2
4	5.6	5.3	5.6	5.9	5.4	5.7	5.4	7.2	5.8
5	6.1	5.6	6.1	6.4	6.0	6.5	6.1	7.3	6.4
6	7.4	6.7	7.3	7.1	7.3	8.6	7.9	7.3	7.5
7	9.0	8.4	9.1	8.8	9.2	8.9	9.7	9.2	9.1
8	11.1	10.7	11.3	10.7	11.7	11.3	12.3	10.6	11.2
9	15.7	16.1	16.1	16.5	16.3	14.5	15.2	14.9	15.6
10	33.7	36.6	33.7	31.6	32.2	33.2	31.0	25.1	31.6

Notes: Cells display the overall percentage across admission cohorts 2001 to 2006. Predicted admission probabilities are generated by the regression in Table 1.4. Counterfactual admission probabilities are estimated using those same regression results, but supposing all students are treated as coming from the 7th API decile.

Table 1.6. Actual, Predicted, and Counterfactual Enrollment by API Decile (Percentage Shares)

Panel A: Actual Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	5.0	4.9	2.2	2.9	4.5	3.5	3.4	5.1	3.9
2	4.8	6.2	4.1	3.4	5.1	4.3	3.1	8.3	4.9
3	5.5	7.2	5.6	4.5	6.1	5.8	4.4	9.1	6.0
4	6.2	6.8	6.3	5.0	6.6	7.2	4.7	8.6	6.5
5	6.3	5.9	6.8	6.3	7.3	7.4	6.4	8.9	6.9
6	7.8	6.0	7.8	6.6	8.4	9.8	8.9	7.5	7.8
7	8.5	7.6	9.4	8.8	9.9	9.2	10.7	9.7	9.2
8	10.4	9.6	11.4	11.3	12.3	11.5	14.5	10.4	11.3
9	13.8	15.0	15.7	19.6	16.3	12.9	16.1	13.9	15.4
10	31.7	30.7	30.8	31.6	23.4	28.4	27.7	18.4	28.0

Panel B: Predicted Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	5.1	4.9	2.2	2.8	4.4	3.4	3.4	5.1	3.9
2	4.8	6.1	4.0	3.4	5.0	4.3	3.1	8.2	4.9
3	5.5	7.1	5.5	4.5	6.1	5.7	4.3	9.0	6.0
4	6.2	6.6	6.2	4.9	6.6	7.1	4.8	8.7	6.4
5	6.3	5.8	6.7	6.3	7.2	7.3	6.5	9.0	6.9
6	7.8	5.9	7.6	6.6	8.5	9.8	8.9	7.6	7.8
7	8.5	7.6	9.4	8.9	10.1	9.2	10.7	9.8	9.2
8	10.3	9.6	11.4	11.3	12.4	11.5	14.4	10.5	11.3
9	13.8	15.1	15.7	19.6	16.4	13.0	16.2	13.9	15.4
10	31.5	31.3	31.1	31.7	23.5	28.7	27.7	18.3	28.2

Panel C: Counterfactual Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	3.2	2.8	1.7	2.6	3.1	2.8	3.4	5.2	3.1
2	3.4	3.5	3.1	3.3	3.3	3.5	3.1	8.3	3.9
3	4.8	5.1	4.5	4.6	4.3	4.9	4.3	9.2	5.2
4	5.5	5.6	5.3	5.2	5.0	6.5	4.8	8.8	5.8
5	6.1	5.8	6.1	6.4	5.9	7.3	6.5	9.0	6.6
6	8.0	6.6	7.6	6.8	7.6	10.4	9.0	7.5	7.9
7	9.0	8.4	9.6	9.2	10.1	9.9	10.8	9.7	9.6
8	11.2	11.1	12.1	11.3	13.3	12.3	14.6	10.5	12.0
9	15.2	17.2	17.0	19.4	18.9	14.0	16.0	13.7	16.5
10	33.7	33.9	33.1	31.0	28.5	28.4	27.4	18.1	29.4

Notes: Cells display the overall percentage across admission cohorts 2001 to 2006. Each student's probability of enrollment is computed as the product of admission probability and probability of enrollment conditional on admission. Conditional enrollment probability is estimated using a probit regression of enrollment on the same explanatory variables in Table 1.4, among the admitted students at each campus. Predicted and counterfactual admission probabilities are computed the same way as described in Table 1.5.

Table 1.7. Regression Results for Cumulative College GPA

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
Panel A: Ordinary Least Squares									
API Decile	0.017***	0.023***	0.022***	0.023***	0.029***	0.028***	0.024***	0.025***	0.024***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005
R-squared	0.275	0.305	0.274	0.220	0.321	0.320	0.219	0.202	0.327
Panel B: Heckman Correction With No Exclusion Restrictions									
API Decile	0.002	-0.004	0.003	0.019***	0.024	0.024***	0.024***	0.022***	0.010**
	(0.005)	(0.007)	(0.007)	(0.006)	(0.016)	(0.003)	(0.002)	(0.004)	(0.004)
Inverse Mills Ratio	0.774***	0.956***	1.433***	-0.182	0.162	1.383***	-0.080	0.322	2.474***
	(0.231)	(0.236)	(0.461)	(0.252)	(0.463)	(0.416)	(0.050)	(0.378)	(0.660)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005
Panel C: Heckman Correction With Other UC Admission Offers									
API Decile	0.020***	0.021***	0.021***	0.025***	0.029***	0.028***	0.024***	0.025***	0.024***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Inverse Mills Ratio	-0.154***	0.069***	0.085***	0.086***	0.022	0.088***	0.062***	0.057**	-0.143***
	(0.032)	(0.017)	(0.010)	(0.013)	(0.016)	(0.015)	(0.019)	(0.022)	(0.012)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005

Notes: The dependent variable is cumulative college GPA. All regressions include the full set of controls from Table 1.2. Heckman results are estimated using the two-step procedure. Panel B uses no additional selection shifters in the Heckman first stage, while Panel C uses dummies of admission offers to other UC campuses. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 2

The Long-Term Effects of Public School Choice: Lottery Evidence from San Diego

2.1 Introduction

Although the public debate over whether to expand school choice in primary and secondary schools tends to focus on private schools, school choice programs limited to public school options have proliferated. Public school choice is an increasingly widespread policy tool aimed at improving public schools. Grady and Bielick (2010) calculate that between 1993 and 2007 the percentage of students attending public schools through choice programs rose from 11 to 16%, and that by 2007-2008, 48% of school districts offered some kind of school choice. These trends are likely to continue in the wake of ambitious expansions of charter school options in many states, and concomitant efforts by public school districts to offer alternative forms of school choice such as magnet schools in response.

Proponents of school choice argue that by giving students and their families a range of school options they will be able to identify schools that best meet the needs of those children. They also argue that choice exerts competitive pressures on other public schools to improve their performance. Critics of choice programs worry that the

programs will lead schools to segregate along racial, ethnic, and socioeconomic lines. They also question whether choice programs will truly improve student outcomes.

School districts and the federal government invest substantial sums of money on school choice. In 2010 the federal government devoted $509 million to various school choice initiatives, including $100 million to the Magnet Schools Assistance Program and $256 million to charter schools, among other initiatives (U.S. Department of Education, 2011). Individual school districts collectively spend far more than this on supporting schools of choice and providing transportation to students. For example, in 2009-2010, a single school district, San Diego Unified School District, budgeted $9.4 million for its magnet school program and $39.4 million for transportation services, much but not all of which was devoted to busing for its various school choice programs (San Diego Unified School District, 2010).

Given the wide adoption of school choice and the public expenditures it creates, it is important to know whether school choice improves student outcomes. This study contributes to a surprisingly small literature looking beyond the immediate impact on test scores and instead examines high school graduation and postsecondary outcomes. We argue that these longer term measures of educational attainment are more vital to adult success than student scores on tests that students will view as low stakes. We study a variety of high school and postsecondary outcomes using data from seven years of applications to choice programs in the San Diego Unified School District (SDUSD), which is the second largest school district in California.

Importantly, schools in these choice programs use a lottery when oversubscribed. The lottery is crucial because its randomization procedure creates the equivalent of experimental treatment and control groups, removing statistical bias in non-experimental studies that may arise due to unobserved differences between those who do and do not enter school choice programs.

2.2 Related Literature

There is a large and growing literature on the relation between various forms of school choice and student achievement in the United States. However, much of the research uses non-experimental methods which necessitate finding a non-experimental control group. We briefly review findings of the impact of three forms of school choice examined in this paper — magnet schools, open enrollment, and voluntary busing aimed at integrating districts — with an emphasis on lottery-based studies that can exploit randomization among students.[1]

Among the papers studying these forms of school choice, we know of only one paper, by Deming, Hastings, Kane and Staiger (2014), that uses lottery data to rigorously analyze the effect of school choice on postsecondary outcomes. The general literature instead focuses mostly on test scores.[2] The shortage of studies looking beyond test scores is unfortunate, as a good school can likely help students in ways that test scores cannot measure. Indeed, Heckman, Stixrud and Urzua (2006) make a persuasive case that what they label non-cognitive skills may be as important as cognitive skills in determining adult outcomes.

2.2.1 Achievement of Students Participating in Choice

Magnet Schools

Ballou (2009) reviews the literature on magnet schools and achievement, and finds four studies that use lotteries to estimate causal impacts, and overall he reports mixed results. (One of these papers, Betts, Rice, Zau, Tang and Koedel (2006), henceforth Betts

[1] Separate literatures exist on charter schools and voucher programs. For recent reviews of the literature on charter schools, see Betts and Tang (2014) and Epple, Romano and Zimmer (2015b). For a recent meta-analysis of the U.S. and international literatures on vouchers, see Shakeel, Anderson and Wolf (2016). Other useful reviews of vouchers include Figlio (2009), Epple, Romano and Urquiola (2015a) and Egalite and Wolf (2016).

[2] Unlike the case for the types of school choice we study, a number of papers on charter schools have looked at postsecondary outcomes. See Betts and Tang (2014) for a review.

et al. (2006), finds mostly no effects of winning a magnet school lottery in San Diego Unified School District, but a moderate positive impact on mathematics achievement in magnet high schools. They study only a single year of school choice applications, compared to the seven years studied here.) More recently, Betts, Kitmitto, Levin, Bos and Eaton (2015) study a national sample of elementary schools that converted to magnet schools and found no overall impact on the achievement of resident students, although results varied school by school.

Open Enrollment

Four recent studies use lotteries to evaluate the impact of open enrollment policies.

In the paper most closely related to the current piece, Deming, Hastings, Kane and Staiger (2014) study postsecondary outcomes in Charlotte-Mecklenburg. They find evidence that students whose neighborhood public school was of low quality had better postsecondary outcomes if they won a lottery to attend another high school through an open enrollment program, with girls benefiting more than boys.

Several papers have examined K-12 school outcomes. Cullen, Jacob and Levitt (2006) provide analysis of the Chicago Public Schools' (CPS') open enrollment program. Focusing on applications to the district's high schools, they find that lottery winners and losers fare almost identically on subsequent tests of math and reading achievement, and other measures of academic outcomes such as absences and courses completed. However, after the lottery, lottery winners are slightly less likely to be arrested than lottery losers. Betts et al. (2006) provide an experimental analysis of the impact of open enrollment programs on math and reading achievement in SDUSD, using data for roughly 2,000 students who applied to various lotteries in 2001. They find, in almost all cases, that lottery winners in the district's open enrollment program fare no better or worse in math and reading than did lottery losers one, two and three years after the lottery. Hastings,

Kane and Staiger (2009) study school choice in Charlotte-Mecklenburg, North Carolina and conclude that there are heterogeneous effects on test scores, with bigger academic gains to students whose families seek high value-added schools.

Traditional Busing

Bradley and Bradley (1977), Crain and Mahard (1978, 1981) and Cook et al. (1984) summarize the voluminous early literature on the effect of busing programs designed to integrate schools racially. Early studies tended to find no (or in a few cases positive) effects on achievement. Betts et al. (2006) find no impact on test scores from SDUSD's traditional busing program, VEEP, based on a one-year sample of lottery applications.

In conclusion, we note that although there exists a non-lottery based literature that focuses mostly on the impact on test scores, the lottery-based literature is smaller. Only one of the lottery-based papers on the impact of magnet schools, open enrollment and voluntary busing school choice examines longer-term student educational outcomes, which is the main goal of the current paper.

2.2.2 School Choice in San Diego

School choice is popular in SDUSD. At the start of our study period (2001-02), roughly 25% of students attended a school outside their local attendance zone through one of the district's choice programs, with this number growing to 40% in 2009-10. Furthermore, because SDUSD uses a random drawing process to admit applicants to schools, the district provides an excellent opportunity to causally analyze the impact of school choice on student outcomes.

During the period we study, SDUSD students who wished to attend a public school other than their neighborhood school had four main options available: magnet

schools, the Voluntary Enrollment Exchange Program (VEEP), the state-mandated open enrollment program (known locally as School Choice), and charter schools. In addition, Program Improvement School Choice (PISC), initiated in fall 2002, responded to the No Child Left Behind (NCLB) requirements to provide choice options to students at schools that fail to make Adequate Yearly Progress in a given subject for two consecutive years. SDUSD met this challenge by using or adapting its bus routes under the VEEP and Magnet programs to provide school choice to such students. Students who were eligible for PISC were also given higher priority within the VEEP and Magnet programs, and were more likely to gain admission than other students.

SDUSD's magnet school and Voluntary Enrollment Exchange Program (VEEP) have their roots in a court desegregation order in 1977, and have operated for almost forty years. Since 1993, SDUSD has also had an open enrollment program mandated under California state law, known as Choice. (See Zau and Betts (2005) for more information on the history of school choice in SDUSD.) Thus, the programs are well-established, with histories tracing back two to four decades. Charter schools also have become an important and well established component of SDUSD. By 2009-2010, 41 charters schools accounted for 11% of district enrollment.

In fall 2009 the percentages of students in each program were 9% in magnet schools, 6% in VEEP, 14% in Choice, and 11% in charters. PISC enrollment totaled 7% of students, spread across the Magnet, VEEP and Choice programs, plus a small percentage of students in transportation routes specifically designed for PISC.

During the application years we study, fall 2001 to fall 2007, the district distributed a detailed brochure to each family early in the school year, advertising the various choices available. Parents then had until mid-winter to apply to any of these choice programs. Where necessitated by over-subscription, the district held a centralized random drawing to determine which students were accepted to a given school and

program. (Except for three charter schools converted from regular public schools, the district did not process applications for charter schools.)

Students could apply to any VEEP, Magnet, Choice, and charter schools in a given year. All four options were readily available to elementary, middle and high school students. To be eligible for PISC, however, a student had to be attending a school designated as needing program improvement because it had failed to meet requirements of California's definition of Adequate Yearly Progress (under NCLB) for two years in a row.

Magnet Schools

At its inception, the goal of the magnet school program was to attract students from primarily white areas into non-white areas by offering specialized curricula and additional resources such as reduced teacher/student ratios, teaching labs, and field trips. In some cases magnet schools attempted to create flows in the opposite direction, into schools in predominantly white areas, in order to provide students with opportunities they would not otherwise have at their neighborhood school and to improve the diversity of the student body. As of the 2010-11 school year, 32 magnet schools operated in SDUSD, up slightly from 29 magnet schools in the 2001-02 school year.

Any student who resided within the district's boundaries could apply to a magnet school. Admissions otherwise used lotteries with two dimensions of priority sorting, before random numbers entered. The first dimension is unique to magnet schools (not present in the other choice types), and is intended to encourage greater integration. Students were grouped into one of four "clusters" based on their local default school. These clusters were sorted by racial composition, with cluster 1 having the largest percentage of white students, and cluster 4 having the largest percentage of non-white students. Magnet schools with a large percentage of white students first accepted applications

from students living in cluster 4, then 3, 2, and 1. Conversely, schools with a large percentage of non-white students accepted applications in cluster order 1, 2, 3, and 4. (Within each cluster, however, no distinction was made between applicants of different races and ethnicities.) After this consideration of students' school cluster, priority was then assigned based on factors like whether a student had a sibling already attending the school. This second dimension of sorting is common to all types of school choice that we study, and not just magnet schools.

The district provided transportation for district students who attend a magnet school outside of their neighborhood and had a walking distance greater than two miles. These students were termed "non-resident" students. Non-resident attendees of magnets accounted for 8.5% of district enrollment in 2009-2010, up from 6.2% in 2001-2002. In the last few years, non-residents have accounted for one-third to one-half of total enrollment at these magnets. Demand for these programs was high during the years we study, and not all students could be accommodated. In 2001-2002, there were 17,621 applications to magnet schools from 11,368 applicants, for 2,217 open slots.

The Voluntary Enrollment Exchange Program (VEEP)

The VEEP program typically attracted students from primarily non-white neighborhoods to schools in primarily white neighborhoods. VEEP, formerly known as the Voluntary Ethnic Enrollment Program, for many years supported a court desegregation order. Throughout the period that we study, however, the district was no longer under court supervision and did not focus solely on racial and ethnic integration, precipitating the name change. The program sought to increase diversity among students in the receiving schools, and to improve achievement among VEEP students by enabling them to attend higher-performing schools.

As with magnet schools, the district provided transportation for VEEP. Unlike

the magnet program, however, VEEP limited the choices available to each student, so as to contain the costs of transportation. It did so by offering students in "VEEP-sending" schools a short list of VEEP-receiving schools to which they could apply.

The district allowed students of any race to apply to the VEEP program. The nature of the VEEP sending/receiving schools still tended to lead to racial and ethnic desegregation, but not as strongly as in the past because, for instance, white students at VEEP-sending schools could now apply to leave for schools that typically have a greater share of white students than their current schools. The program was open to all students who resided in the local attendance zone of a school with a VEEP pattern. VEEP students at the elementary level followed the feeder pattern of their new school, rather than having to return to their neighborhood for middle and high schools.

Receiving schools, in conjunction with the district's Enrollment Options Office, determined the number of slots available at the school each year. As with the magnet program, demand for this program was strong. In 2001-2002, there were 8,099 applications from 5,636 applicants, for a total of 2064 available slots. VEEP's share of district enrollment fell slightly between fall 2001 and fall 2009, from 6.4% to 5.9% of district enrollment.

The application process for VEEP was similar to magnet schools: students who applied before a mid-winter cutoff date were given first priority, and siblings of students currently in the program were at the top of that list. Wait lists did not carry over, so if a student was not accepted (did not "win" the random lottery), he or she would have to re-apply the following year. Admission was also based on availability, with the highest-achieving schools typically having the fewest openings due to demand.

School Choice (Open Enrollment) Program

The third choice program, the district-wide School Choice program, provided any student with the opportunity to attend any school within the district's boundaries. It is mandated by state law passed in 1993 as a way of increasing the options for every student. The main drawback of this program is the lack of district-provided transportation. In 2001-2002, there were 12,705 applications from 8,311 applicants for 3045 new slots. This program has grown markedly, doubling its share of district enrollment between fall 2001 and fall 2009, to 13.6%.

Charter Schools

Charter schools — or publicly funded schools that operate more independently of the school district than traditional public schools — are an increasingly popular alternative to the traditional school. These schools obtain their charters from the district, which in turn provides oversight. As of 2009-2010, 41 charter schools accounted for 10.8% of district enrollment, up from 5.5% in 2001-2002.

Apart from a few settings such as New York City, it is a commonplace finding that charter schools are not oversubscribed, and thus do not conduct lotteries. Gleason et al. (2010) conducted the first nationwide lottery-based study of charter schools. They focus on middle schools, and report having exceptional difficulty finding charter schools that were oversubscribed. Indeed, only 130 out of 492 such schools nationwide used admission lotteries, and only a fraction of these schools agreed to participate, compounding potential selection effects. Thus the problem we face in San Diego with a lack of lottery data mirrors the (lack of) use of lotteries nationwide. Due to the lack of lottery data for charter schools, we do not examine this type of school choice.

2.3 Data and Methods

2.3.1 Sample and Lottery Basics

SDUSD began to use a lottery to admit students to school choice programs in 2001. We have access to lotteries from fall 2001 through fall 2007, a total of 7 years. We restrict our sample to applications to middle and high schools, in grades 6 to 12, because most students applying to the elementary grades (grade 5 and below) have either not graduated from high school, or graduated too recently for our desired postsecondary outcomes to be well-defined.[3] We also omit all applicants who applied from outside the district, regardless of whether they won or lost their lotteries. Our concern was that we would typically only observe follow-up data for these out-of-district lottery winners, and not the lottery losers, making for an unfair comparison.[4]

In any lottery-based study, it is crucial to know about the rigor with which the lottery is conducted, and whether it is conducted consistently over time. The same district staff member in the central district offices conducted the lotteries each of the years we studied.

After the student applied to a specific school and grade, applications were assigned to priority groups (such as sibling reunification) within school and grade. Lotteries were then conducted, with separate lotteries for each school, grade and priority group. Applications received random numbers within each school $\times$ year $\times$ grade $\times$ priority group cell. Based on their random number, students were contacted in strict rank order and offered admission. An identical approach was taken for all three of the choice programs. Figure 1 illustrates using a simple hypothetical example. In this example, a true lottery occurs for group B. Figure 2 shows the full list of priority groupings for all

[3]Results were similar when we added elementary school applications to our analysis.

[4]That is, the admission offer itself makes it much more likely for out-of-district lottery winners to transfer to SDUSD than out-of-district lottery losers, who do not receive such an offer.

three choice types, which sometimes vary by year.

In many cases, all students in the top priority group(s) for a given school and grade were admitted, since the enrollment limit was large enough to accommodate all of these students. However, the enrollment limit would often "hit" within some lower priority group. Students in this priority group experienced a "true" lottery, in that some students were admitted and others not, depending on their random number draw. There also exist cases where students in all priority groups were admitted to a given school and grade, signaling that the school truly was not oversubscribed in that grade. However, since the randomization from the lottery is essential for obtaining exogenous variation, we do not include such cases because there is no randomly created control group.

Analytically, we treat each lottery as a separate experiment. Thus, in addition to applying to the same school in the same year, students must also be in the same grade level and priority group to be compared against one another. We do so because students who apply to (e.g.) grade 6 at a magnet school may not be identical on average to students who apply to the same magnet in grade 7. Further, among the applications to a given school and grade, students in the (e.g.) sibling reunification group are likely to differ from students in the other priority groups.

Overall, we identified, for students applying to choice programs for grades 6-12, the number of true lotteries and applications listed in Table 2.1. The bottom part of the table displays mean characteristics of the students making the applications to each program. For both the Magnet and VEEP programs, students' baseline reading and math scores are below the district average (scores are scaled to have a mean of zero and standard deviation of one within the district), and Hispanic and black students represent the two largest racial/ethnic groups applying. Very few students applying to these programs have a parent who holds a Bachelor's degree or higher. The open enrollment program, known as Choice, attracts an applicant pool that is quite different. Their baseline math and

reading scores are above the district average, the two largest racial/ethnic groups among applicants are Hispanics and whites. Notably, 40 percent of students applying to Choice have parents with a Bachelor's degree or higher, compared to just 21 and 11 percent for Magnet and VEEP programs, respectively. The intuition for this difference is that unlike the Magnet and VEEP programs, Choice does not provide busing, thus requiring families to drive their students to choice schools, or to make other transportation arrangements. Providing this transportation may prove more difficult for families with fewer resources.[5]

2.3.2 Outcomes of Interest

We examine the impact of winning a lottery to attend a middle or high school through a school choice program on two types of outcomes: (1) outcomes which are measured during the high school years but which are likely to be related to longer term outcomes such as college graduation and earnings, and (2) postsecondary outcomes. The specific outcomes are detailed below.

High School Outcomes

Because we focus on those applying to school choice programs in grades 6 to 12, we include a number of secondary school outcomes:

- An indicator for student being on track in grade 9, meaning that they passed at least ten units, had no more than one semester F in a college preparatory course, and maintained a Grade Point Average (GPA) of at least 2.0[6]
- An indicator for student passing the California High School Exit Examination in their first attempt in grade 10

[5]Koedel, Betts, Rice and Zau (2009) show that the Magnet and VEEP programs tend to integrate SDUSD racially and in terms of socioeconomic status, but that the Choice program slightly increases segregation. They argue that the lack of busing for Choice explains why its impact on segregation is different.

[6]We include the GPA provision because throughout the period studied SDUSD required students to have a GPA of at least 2.0 in order to graduate.

- The percentage of days attended in grade 12
- The GPA in grade 12
- The citizenship grade in grade 12 (based on teachers' report card reports on behavior in each class, this measure is an average across all courses, scaled to 0 to 4)
- An indicator for whether the student graduated on time

Postsecondary Outcomes

We examine two types of postsecondary outcomes — enrollment and degree attainment — at various points after high school graduation. For students indicated as dropouts in the data, prior to high school graduation, we maintain these students in our analysis but set all postsecondary outcomes to zero.

i) Enrollment

- Three separate indicators for postsecondary enrollment in the first year after high school graduation: enrollment in any postsecondary institution; enrollment in two-year institutions; and enrollment in four-year institutions.
- Three separate continuous measures for the number of years of postsecondary enrollment in the first four years after high school graduation: the number of years in any postsecondary institution; the number of years in two-year institutions; and the number of years in four-year institutions.

ii) Degree Attainment

- Three indicators for Bachelor's degree within four, five or six years of high school graduation

- Three indicators for an Associate's degree or other community college graduation event within four, five and six years of high school graduation

The postsecondary data are derived from the National Student Clearinghouse (NSC). The NSC is a nationwide student-level database of postsecondary enrollment and degrees granted for postsecondary institutions, which jointly account for about 93% of total postsecondary enrollment nationwide. In the NSC data, many of the San Diego community colleges did not name the degrees awarded, and so for two-year colleges, our outcome consists of a measure of an Associate's degree or any graduation event recorded by a two-year college. This measure thus combines Associate degrees and sub-Associate certificates.

2.3.3 Estimating Treatment Effects

The first central part of our analysis consists of generating intent-to-treat (ITT) estimates of the impact of winning a lottery to either the Magnet, VEEP, or Open Enrollment ("Choice") programs. Our ITT equations typically take the form:

$$Y_{ijgst} = \sum_{j=1}^{J} \alpha_j + WIN_{ijsg}\beta + \delta_M S^M_{i,0} + \delta_R S^R_{i,0} + MALE_i\gamma + \epsilon_{ijgst} \tag{2.1}$$

where i indexes students, j indexes lotteries, and g and s denote the intended grade level and school of the lottery application, with $g \in \{6, 7, ..., 12\}$.[7] The base year of the application is defined as time 0, such that Y_{ijstg} denotes an outcome measured t years after the application year, and all other subscripts continue to refer to the original lottery application.

We model these outcomes as a function of, most importantly, a dummy variable WIN_{ijsg} indicating whether the student won the given lottery, and lottery fixed effects,

[7]The lottery indicator j makes s and g subscripts redundant, since lotteries are specific to a given school and grade (and lottery group). However, we include s and g for clarity.

α_j. Other terms in the equation include math and reading achievement from state-administered standardized tests, expressed as Z-scores, from the baseline school year in which the student applies (the $S_{i,0}$ terms); a dummy variable for males; and an error term ϵ_{ijstg}.[8] The coefficient of interest is $\hat{\beta}$, which states the estimated impact of winning a lottery on the given outcome. We also cluster standard errors by lottery in all regressions.

It is important to obtain ITT estimates, as described above, because they tell policymakers the gains from making one more school choice offer. However, policymakers are also interested in knowing the causal impact on a student of entering a school choice program. This impact of treatment on the treated will generally be bigger than the intent to treat estimate because not all lottery winners take the offer (the take-up effect). We can estimate the impact of winning a lottery and attending the lottery school using two stage least squares (2SLS), in which we instrument for the new regressor of interest — enrolling at the school to which the student applied the next year — with a dummy variable indicating whether the student won the lottery. This approach simply scales up the intent-to-treat estimate by the reciprocal of the of the take-up rate.[9] This approach assumes the entire intent-to-treat effect derives from students who win the lottery and actually enroll at that school the following year. It otherwise assumes no causal effect on lottery winners who do not take up the offer.

There is a second and broader adjustment that one can make to the ITT estimator to estimate the impact of treatment on the treated. This second approach not only accounts for lottery winners who do not switch to the school, but for students who lose a lottery but manage to gain entry to another school of choice by winning a different lottery (the

[8] If the lottery is conducted fairly, then we should obtain unbiased causal estimates of the impact of winning a lottery from equation (1) even if we exclude the baseline achievement and male indicator. However, their inclusion can increase precision. We replicated all models without the baseline test scores and results were generally similar but less precise.

[9] Bloom (1984) is one of the early papers suggesting this scaling factor, although not in an instrumental variables framework. Thus, the estimator is essentially what has become known as the single Bloom adjustment, although the 2SLS approach also makes appropriate adjustments to the standard errors.

substitution or cross-over effect). In this version of estimating the average impact of treatment on the treated, we can use an instrumental variables (IV) approach in which we first alter equation (1) so that the regressor of interest is not the indicator for winning a lottery, but whether the student the following year switched into any school participating in the given choice program. We then instrument for this endogenous variable using the indicator for whether the student won the current school choice lottery (lottery j). Mechanically, the IV estimator is identical to dividing the ITT estimates by $(b - a)$, where b is the proportion of lottery winners who accept the admission offer and a is the proportion of lottery losers who nonetheless enter another school of choice (e.g. by winning in another lottery).[10] Some researchers refer to this as the double Bloom (1984) adjustment.

2.4 Results

2.4.1 Testing for Lottery Fairness and Differential Attrition

We define the lottery process as "fair" if there is no statistically significant difference in observable characteristics between lottery winners and lottery losers. Because all of our outcome variables are related in some way to academic achievement, it is particularly important to test whether the initial test scores of applicants who win a given lottery and those who lose a given lottery are identical. If we retain the null of no difference in baseline test scores between lottery winners and losers, it greatly buttresses the argument that the lottery is conducted fairly.

Since lotteries differ greatly in their percentages of winners and losers, we do not show a summary statistics table comparing the average characteristics of winners and losers, which in this context compares more across lotteries than within a given lottery.

[10]This statement is in fact contingent upon identical treatment effects across lotteries. See section 5 of Heckman, LaLonde and Smith (1999), for a description of the underlying assumptions.

Instead we use a regression approach: regressing a given characteristic on lottery fixed effects plus an indicator for whether a student won a given lottery. These results are displayed in Table 2.2.

The first two rows of Table 2.2 show the coefficient for lottery winners in models of baseline Z-scores in math and reading — that is, in the year that students applied — for true lotteries. For none of the three types of school choice is there a significant difference between lottery winners and losers in baseline test scores. The third row examines whether the presence of these lagged test scores, which we include as controls in our impact analysis below, are equally likely to be missing for lottery winners and losers. No significant differences emerge. The next six rows of Table 2.2 test for any differences in basic demographics between lottery winners and losers. We find no meaningful differences here, except that lottery winners are slightly more likely to be white (about 2.0 percentage points more likely, when pooling across all three choice types). Despite generally few differences in observed characteristics between lottery winners and losers, we control for baseline achievement in math and reading, as well as gender, in all outcome models as a precaution.

The final row of Table 2.2 does not test for lottery fairness so much as differential attrition. The outcome here is whether the student leaves the district the year after the school choice application, based on enrollment records. We find no significant differences for Magnet or VEEP programs, but a small but significant difference for the Choice program, and in the pooled case. This result hints at differential attrition between lottery winners and losers. Table A7 of the appendix presents additional results about differential attrition, on an outcome-by-outcome basis. For most outcomes that we study, we see a slightly higher rate of attrition among lottery losers than lottery winners, though this difference often is not statistically significant. Differential attrition appears most pronounced for postsecondary graduation outcomes, specifically, which contain all of the

significant cases in Table A7.

2.4.2 Verification That the Intensity of the Treatment Contrast Is Large

One concern is whether the intensity of the treatment contrast is large between students' local schools and the schools to which they apply. Because many of the factors that influence a student's educational trajectory are not readily observed, we cannot measure the differences in all of the relevant measures of school and classroom environment. However, we can test for differences in observable characteristics.[11]

Therefore, we compare lottery winners' local school with the school to which they applied. The comparisons include 14 variables — nine related to academic and behavioral outcomes for students attending each school, and five measures of the socioeconomic and racial/ethnic mix of the student body. All these variables are averaged for each school over the school years 2001-2002 through 2007-2008.

We derive many of the school academic measures from the statewide achievement test, the California Standards Test (CST), which during the period we study tested students annually in reading and mathematics between grades 2 and 11. We transformed test scores for students into Z-scores by grade and year. For English we averaged the Z-scores across the middle or high school for all students enrolled (up to grade 11). A limitation of the math scores is that, after grade 7, students were given math tests that varied depending on the math course taken, making it very difficult to compare achievement among all students in any grade above grade 7. Therefore, for math CST results we were able to characterize middle schools by math performance in grades 6 and 7 only, omitting grade 8. For the same reason we lacked math scores for high schools,

[11] Research by Betts, Zau and Rice (2003) suggests that San Diego schools, especially at the middle and high school levels, vary considerably in terms of peer achievement and the qualifications of teachers, and that these variables are associated with the student's own subsequent gains in achievement.

and so in the tables described below school math performance results include only middle schools. We characterized schools by four test-score measures: average Z-scores in reading and math, and average school value-added. The latter we calculated by regressing test scores on students' lagged test scores, as well as controls for gender, race/ethnicity, parental education level and grade level, and then averaging the residuals for each school.

Other academic measures of the schools included the percentage of students who attended that school who ultimately graduated on time, the percentage of days absent, the percentage of students on track in grade 9, and a citizenship grade. For middle schools, the measures on on-time graduation and being on track in grade 9 are observed during high school but calculated among all students who attended the given middle school. But the absence and citizenship variables refer to outcomes while the students were attending the given middle or high school. Being on track in grade 9 means that a student in his or her first time in grade 9 completed at least the expected ten units, with no more than one F in a college-preparatory course, and with a cumulative GPA of at least 2.0. (A cumulative GPA of 2.0 has been a graduation requirement in the district for the entire period we study.) The citizenship variable is averaged over the citizenship grade that every teacher gives to students in their classes in the respective middle and high schools. It provides important insights into teachers' views of how well behaved students are in the classroom, and thus provides a proxy for the non-cognitive behavior of students at a school.

School comparisons also include socioeconomic measures including the proportion of students from each of four racial/ethnic groups, and the proportion of students whose parents hold a Bachelor's or higher degree.

Tables 2.3A–2.3C compare the local schools and the schools to which students applied in the Magnet, VEEP and Choice programs, respectively, while Table 2.3D pools all three programs. The sample consists of applications to true lotteries, in which students

were both lotteried in and lotteried out. Tables show the mean of the given characteristic in the school to which the student applied, the mean difference between the choice school and the student's local default school, along with the p-value for a test that the local school and the choice school were identical on the given characteristic.

Several striking patterns emerge. First, across all three types of choice, and in the pooled analysis in Table 2.3D, the mean difference between the school in the choice program and the student's local default school is highly significant, with p-values well below 0.01.

Second, the differences mostly indicate that students were applying to schools that outperformed the local schools across the panoply of academic measures, as well as the two behavioral outcomes: the citizenship grades and percentage of days absent. The schools to which students applied also tended to have student bodies with a higher socioeconomic status and a greater proportion of white students, and a lower proportion of black and especially Hispanic students. The lone exception is magnet schools, for which the schools to which students applied had about 2 percentage points more black students. This could reflect both the result of past moves to these schools by non-resident black students as well as the fact that most of the magnet schools operated in lower-income areas of the city and offered specific curricular emphases in the hope of attracting students from higher income areas of the district. Results on the proportion Asian are mixed but the differences were closer to zero than for the other groups.

A third striking result from the tables is that the value-added in test scores at the schools to which students applied was near zero, meaning that student gains in achievement were on average similar to the district as a whole. (Value-added here is computed based on all students at each school, rather than just lottery applicants.) Nonetheless, value-added was significantly higher at the choice schools relative to the local schools in all but one of eight cases. (The exception was middle school math at

magnet schools, where math value-added was lower by 0.05 standard deviation than at local public schools.) That is, in all but one case, students were applying to schools where students made more progress on average in reading and math than at the student's local school. Of course, parents cannot readily observe value-added but do observe average achievement through annual publication of CST results. These test score levels were consistently higher at the choice schools, ranging from 0.16 to 0.65 of a standard deviation.

The overall conclusion is that in our sample of applications to true lotteries, students were on average applying to schools that differed significantly from their local schools. Schools of choice had higher test scores and value-added in test scores, better behavioral outcomes and better outcomes in high school such as graduating on time. They also tended to have greater proportions of students who were white and whose parents held a Bachelor's degree or higher.

2.4.3 Intent to Treat Estimates

We begin with an intent-to-treat analysis of the impact of winning a lottery to any of the programs. Tables 2.4–2.6 summarize the key coefficient ($\hat{\beta}$) and its standard error for high school outcomes, postsecondary enrollment outcomes, and postsecondary graduation outcomes, respectively.[12] To help provide context for the size of estimated impacts, the bottom row of these tables shows the mean value of the outcomes for the students in the pooled set of applications.

In Table 2.4, we see that there were no statistically significant differences between lottery winners and losers for any of the six high school outcomes, for any of Magnet, VEEP, or Choice. The final row pools all of the lotteries across the three types of choice and again finds no significant differences. Further, the coefficients were typically small,

[12] Appendix tables A8 to A10 show slightly more detailed versions of these results, providing additional information about R-squared, the number of observations, and number of clusters.

with roughly as many negative as positive.

Table 2.5 shows results when we model the number of years enrolled in postsecondary education by given numbers of years after high school graduation. As mentioned earlier, postsecondary outcomes are set to 0 for high school dropouts. Again, with one exception, none of the ITT estimates is statistically significant. The exception is that when we pool all types of school choice, the number of years enrolled in a two-year college four years after graduation is 0.08 higher for lottery winners. This pooled result probably derives from the VEEP and Choice lotteries, where positive estimates were close to becoming significant. Note however that there is no overall significant effect on overall time enrolled in postsecondary education, and that the estimated effect on enrollment on four-year college enrollment is negative, although not significant. This pattern of findings hints weakly at the possibility that VEEP and Choice programs could induce a slight substitution towards two-year and away from four-year colleges.

Table 2.6 shows impacts of winning a lottery on degrees obtained four, five and six years after high school graduation. As discussed earlier, our community college graduation measure combines Associate degrees and sub-Associate certificates. For Magnet, VEEP, and the pooled case, there is a significant 1-3 percentage point increase in the likelihood that lottery winners complete an Associate or other two-year college certification four or five years after high school graduation. However, no significant effect is found six years after high school graduation. One interpretation is that winning a school choice lottery may accelerate time to degree for community college degrees and certificates, while having no impact on the overall probability of obtaining a degree or certificate by six years after graduation.

For the most part, Table 2.6 shows no significant relationship between winning a lottery and whether students subsequently obtain a Bachelor's degree. One exception is that winners of magnet school lotteries appear to be about 5 percentage points less

likely to obtain a Bachelor's degree within six years of high school graduation. Another exception is that Choice lottery winners are about four percentage points less likely to obtain a Bachelor's within four years of graduation, but for these students, no significant difference remains between lottery winners and losers five or six years after graduation.

Table A11 of the appendix provides an alternative view, combining 2-year and 4-year college certifications into a single outcome measured 4, 5, and 6 years after high school graduation. Using this version of the outcomes, ITT estimates of the effects on postsecondary degree attainment remain mostly insignificant. At four years after high school graduation, lottery winners in the VEEP program have about a 4 percentage point higher chance of having either a 2-year or 4-year college certification compared to lottery losers. But this effect subsequently dissipates and becomes insignificant. For none of the choice programs, nor the pooled case, is there a significant impact on the combined outcome of having either a 2-year or 4-year college certification either five or six years after high school graduation.

Overall, our ITT analysis of the impact of choice on postsecondary enrollment and degrees suggests some additional interest in 2-year colleges, but perhaps partly as a result of students shifting away from 4-year colleges, rather than all new interest in postsecondary education. We find some evidence of positive and significant effects on both enrollment and degree completion at 2-year colleges, while estimated impacts for 4-year colleges are typically negative (though usually not significant). Results are usually insignificant when looking at degree completion or the total number of years enrolled in either type of program, combined.

For students who applied to only one school, the intent-to-treat estimate is straightforward to interpret: it is the impact of being offered a chance to enroll in the choice school, instead of enrolling in the default local school. Interpretation is somewhat more difficult in cases where students applied to multiple schools. Now, we are estimating

the impact of making one more offer to attend a choice school than the student would otherwise have received. It could be that the first lottery won has the biggest impact. (If all the schools to which a student applied offered similar advantages over the local school, then the impact of winning a lottery could be smaller if the student had already won a lottery to another the schools to which she applied.)

To examine this possibility, we re-ran all the intent-to-treat estimates for each type of choice, as well as pooled, after restricting the sample to students who had applied to only one school. If we found much bigger estimated impacts, and impacts that were significant, in this exercise, it would support the notion that most of the gains from school choice accrue with the first lottery won. However, we found almost identical patterns of statistical significance to those reported above. In some cases coefficients were slightly bigger, but in almost all cases the estimated impacts remained insignificant.

2.4.4 Estimated Impacts of Treatment on the Treated

Tables 2.7–2.9 show estimates of the impact of treatment on the treated for the sets of high school outcomes, measures of postsecondary enrollment and college graduation, respectively. Table 2.7 shows no significant effect of enrolling in the school to which a student applied on various high school outcomes. The precision varies by type of school choice and outcome. While there are some quite precise "zero" estimates, the magnet school estimate of the impact on the probability of graduating on time is both large and imprecise. The estimated impact is about 0.10, with a 95% confidence interval of roughly -0.07 to 0.27.

Table 2.8 turns to the impact of winning a lottery and enrolling on six different measures of the number of years of postsecondary enrollment. In general there are no significant estimated impacts. The only exception is that the impact of winning a Choice (open enrollment) lottery on the number of years enrolled in a two-year college within

one and four years of high school graduation are significant, with effects of 0.14 and 0.44 respectively. When we pooled across types of school choice, we found a significant 0.27-year increase in the number of years enrolled in a two-year college four years after high school graduation. This pooled effect probably derives mainly from the large and significant effect for Choice. Countering these positive effects on two-year college enrollment, the table provides some evidence that Choice may cause substitution away from four-year colleges to two-year colleges: the impact of winning a Choice lottery and enrolling on number of years of postsecondary enrollment is never significant, but the negative point estimates suggest that Choice is associated with lower enrollment in four-year colleges.

Table 2.9 examines the impact of winning a lottery and enrolling on postsecondary degrees and certificates. Again, in most cases the impacts are not statistically significant, but exceptions emerge. For all cases except Choice, the impact of winning a lottery and enrolling is linked to a modest but significant boost in the probability of earning a community college Associate degree or certificate within four and/or five years, but the effects become insignificant after six years. As for the Bachelor's outcomes, most of the estimates are insignificant. Both exceptions point to negative impacts. The more notable finding is that, for those who won a Magnet lottery and enrolled at the magnet, there is an estimated drop in the probability of earning a Bachelor's degree six years after graduating from high school of 0.22. This is a large effect, given that the proportion of all students in the pooled set of true lotteries who obtained a Bachelor's degree after six years was 0.183.[13] Second, winning a Choice lottery and enrolling is associated with a drop of 0.12 in the probability of obtaining a Bachelor's degree in four years, but the impacts after five and six years are very close to zero and are not significant.

[13]We find some corroboration for this effect in Table 2.8, where the estimated number of years enrolled in a four-year college after six years was negative for magnet schools, although very imprecisely estimated.

Table 2.10 shows the impact of winning a lottery and enrolling the following year in any school participating in the given type of school choice. In general the coefficients are higher than in Table 2.7 but not markedly — meaning that it is not common for a student to win one lottery, turn that school down, and then enroll in another — but it certainly does happen. None of the estimated impacts is significant, but the precision is somewhat lower for this version of the impact of treatment on the treated.

Turning to postsecondary enrollment, Table 2.11 shows highly similar patterns of statistical significance compared to the single Bloom (1984) adjustment. Again there are significant effects for Choice on years enrolled in a two-year college one and four years after graduation, and when pooling all three types of choice for impacts on years enrolled four years after high school graduation. One newly significant effect is for VEEP, where enrolling in any of the VEEP schools to which a student applied is significantly negatively related to enrollment in a four-year college in the first year after completing high school. However, the effect becomes insignificant when looking at years enrolled in the first four years after high school.

Table 2.12 shows the impacts of enrolling in a choice program on our measures of postsecondary degree and certificate attainment. With the exception of the impact of enrolling in a magnet school on obtaining a two-year college degree or certificate within four years of graduation, the patterns of statistical significance are identical to the results in Table 2.9 from the impact of enrolling in the specific school to which one applied. The most striking finding is that enrolling in a magnet school has a large negative impact on the probability of obtaining a Bachelor's degree within six years of graduating from high school. There are positive impacts of enrolling in a VEEP school on obtaining a degree or credential from a two-year college within four or five years, but again the effects dissipate by year six.

2.4.5 Test for Variations in the Effect by School Characteristics

Schools vary dramatically in demographics and academics, and it may not be sensible to assume that the impact of winning a lottery to one school should be the same as for another school. The overall tenor of the intent-to-treat estimates is that winning a lottery has no significant effect on high school and postsecondary outcomes, with a few exceptions, typically indicating negative effects. It could be that there are significant effects, positive or negative, for certain types of choice schools. For example, if a student wins a lottery to attend a school having much higher value-added in test scores than the local school, one could imagine enrolling in such a school could have a positive impact. Indeed, Hastings, Kane and Staiger (2009) found, in one district, that school choice led to test score gains for those applying to schools with higher average test scores.

Theoretically, what should matter for impacts are the *differences* in characteristics between the school to which the student applied and the default local school which the student might otherwise attend. Accordingly, we re-estimated the intent-to-treat models after adding to equation (1) the difference in a given characteristic between the applied versus local school, and an interaction of this difference with the dummy for winning the lottery. We measure differences for the same 14 school characteristics discussed in the earlier tables, covering measures of students' average academic achievement, behavior, and demographics.

We estimated these interactions after pooling all true lotteries across the three types of school choice, and also separately by each program. We emphasize the pooled results because we primarily seek patterns that are consistent across types of choice, rather than idiosyncratic to just one choice program.

Table 2.13 summarizes the level of significance of the interactions, as well as the sign of the interaction in cases where the interaction is significant at the five percent

level. Panel A shows results for high school outcomes, with Panels B and C showing results for postsecondary enrollment and postsecondary degree attainment. What is most apparent across the three panels is how rarely interactions between winning a lottery and differences in school characteristics are statistically significant. Further, when the impact of winning a lottery changes with the gap in a given type of school trait, it is usually for isolated outcomes.

We estimated these interaction models separately for the three school choice programs as well. Results are not shown but are available on request. It was more frequently the case that interactions were significant in these models but the patterns varied across outcomes and types of school choice in ways that suggested at most idiosyncratic effects. The percentage of interactions that were significant at the five percent level were 6.3% for Magnet, 11.5% for Choice and 21.8% for VEEP. These first two percentages are close to the type I error rate expected. VEEP was the outlier, and had two outcomes that accounted for about half of the significant effects. These two grade 12 outcomes — the percentage of days attended and citizenship grades — showed significant positive interactions between winning the lottery and the differences between the choice school and the local school in reading scores, math scores, reading value-added, mean GPA, the proportion of students who were white and the proportion of parents with a Bachelor's or higher. While all of these relationships are easy to motivate, for example in terms of positive peer effects, these patterns did not show up for other high school or postsecondary outcomes, with a few exceptions.

2.4.6 Test for Variations in the Effect by Student Subgroup

It seems plausible that students from different subgroups could benefit differentially from winning a lottery into a given choice program. For instance Deming et al. (2014), in the only lottery-based study of the impact of open enrollment on postsecondary

outcomes, found that girls benefited from school choice, but boys did not. It is crucial to test for these interactions from a policy perspective. It could be the case that choice programs have no impact on many or even most students, but could lead to better outcomes for some students, and potentially a Pareto improvement.

We tested for variations in the effects of school choice by student gender, race/ethnicity, English Learner status, special education status, and parental education. Additionally we tested whether the grade to which a student applied was related to the impact. Presumably, the earlier a student applies to the school choice programs the greater the impact. (This logic follows because the choice programs allow students to articulate from middle school to the high school in the same attendance area, or in the case of magnets, to magnet high schools with a similar theme, if available. So applying to a middle school choice program could offer a choice path all the way through high school graduation.)

Table 2.14 summarizes the interactions between these student characteristics and winning a lottery for the pooled sample, using similar notation to that in Table 2.13. The only difference in the way that we characterize significant interactions was that when interactions were significant but the impact of winning a lottery did not increase or decrease monotonically we indicate significance at the 1% and 5% levels using ** and * respectively. (Non-monotonicity could arise for personal characteristics that were measured not by binary or continuous variables but rather by a series of dummy variables. Such was the case for the grade applied to, parental education and race/ethnicity.) The top part of each panel shows interactions with the student's own lagged achievement, with the bottom panel showing various demographic characteristics along with English Learner and special education status.

The three panels of Table 2.14 suggest that overall few personal characteristics significantly mediated the impact of winning a lottery. The grade level for which students applied was often significant, but rarely in a monotonic way. For example, the impact

of winning a lottery was in some cases larger in grade 8 but smaller in grade 10, both relative to the impact in grade 6.

2.5 Implications and Conclusion

The results suggest that on average granting a slot in a school of choice to a student does not significantly alter his or her high school or postsecondary trajectories. Some exceptions do emerge. Winning a lottery may shorten the time taken to obtain a community college degree or certificate, but with no effect on the long run likelihood of obtaining that degree or certificate. The only effect that may be lasting is that six years after high school graduation, Magnet lottery winners may be less likely to have obtained a Bachelor's degree than lottery losers.

Policymakers should care about the intent-to-treat results because they speak to the value of expanding a school choice program by offering one more spot. But policymakers may care even more about the impact on a student of actually enrolling in a school of choice. We estimated two types of impacts of treatment on the treated. Using winning a lottery as an instrument, we estimated the impact of enrolling in the school applied to, as well as that of enrolling in any school in the given program, such as the Magnet program. Overall, the results were similar to the intent-to-treat estimates.

These results may mask larger impacts in specific cases. Theoretically, the bigger the difference between a student's default local school and the school to which she applies, the bigger the impact of winning a lottery to attend the school of choice. We tested whether the intent-to-treat effects varied by the difference between the two schools in terms of achievement, student behavior and student demographics. Although on average these inter-school differences were quite large, we typically did not find that any of these differences mediated the impact of winning a lottery. The biggest exception was that winning a VEEP lottery had bigger positive impacts on the two grade 12 behavioral

outcomes — attendance and citizenship grades — if the VEEP school to which one applied enrolled students with higher achievement and/or better behavior, or had a greater share of white students.

Overall, these results do not suggest that school choice has generated meaningful gains for students in San Diego. But this conclusion must be tempered for several reasons. First, there could be other unmeasured outcomes on which students gain if admitted to a choice program. Second, there could be larger effects, either positive or negative, for specific types of students applying to specific types of schools. We hope to study this in the future by looking for interactions between student background and the difference in school characteristics between the local school and the school of choice. Third, further robustness checks are planned. Among these planned extensions are checks that differential attrition has not biased the results. Future extensions will likely include estimation of another type of impact of treatment on the treated, which instruments for the number of years attending a school of choice, and robustness checks to make sure that substitution towards charter schools has not biased the results.

Chapter 2, in full, is an unpublished manuscript that contains the following co-authors. Betts, Julian R.; Young, Sam M.; Zau, Andrew C.; Bachofer, Karen Volz. The dissertation author was the primary investigator and author of this paper.

2.6 Tables and Figures

Priority	Description	Who Is Admitted?	
A	Sibling	All	
B	Specialized Course	Some	⟵ True Lottery
C	Before Priority Deadline	None	
D	After Priority Deadline	None	
E	Gifted Cluster	None	

Notes: For a given school and grade, applications are grouped in descending order by the above priority groups. In this example, there are adequate seats to admit all students applying from the Sibling priority group, A. Enough seats are then available to admit a portion, but not all, of the students in group B. A true lottery occurs for group B, with only those students with the lowest random numbers in the Specialized Course category being admitted. No applications from groups C, D or E are then admitted, due to a lack of seats, in this hypothetical example.

Figure 2.1. Simple Example of Lottery Admissions

Magnet

Priority Type	
0	Sib/Continuity
1	Continuity
2	Sib/Before 9/3*
3	Sibling
4	Before 9/3*
5	Before Priority Deadline
6	After Priority Deadline
7	Before 9/3 OD*
8	Before Priority Deadline OD
9	After Priority Deadline OD

*Eliminated

Priority Description 04-05 to 05-06	
0	Sib/Continuity
10	Continuity
30	Sibling
31	Priority for PISC-see attached
32	Priority for PISC
33	Priority for PISC
34	Priority for PISC
35	Priority for PISC
36	Priority for PISC
50	Before Priority Deadline
60	After Priority Deadline
80	Before Priority Deadline OD
90	After Priority Deadline OD

Priority Description 06-07 to 07-08	
0	Sib/Continuity
100	Continuity
300	Sibling
301	Priority for PISC-see attached
302	Priority for PISC-see attached
303	Priority for PISC-see attached
304	Priority for PISC-see attached
305	Priority for PISC-see attached
306	Priority for PISC-see attached
307	Priority for PISC-see attached
308	Priority for PISC-see attached
309	Priority for PISC-see attached
310	Priority for PISC-see attached
500	Before Priority Deadline
600	After Priority Deadline
800	Before Priority Deadline OD
900	After Priority Deadline OD

VEEP

Priority Type	
A	Sibling
D	Before 9/3*
E	Before Priority Deadline
F	After Priority Deadline
H	Before Priority Deadline OD
I	After Priority Deadline OD

*Eliminated

Priority Description 04-05 to 05-06	
A0	Sibling
B1	Priority for PISC-see attached
B2	Priority for PISC-see attached
B3	Priority for PISC-see attached
B4	Priority for PISC-see attached
B5	Priority for PISC-see attached
B6	Priority for PISC-see attached
E0	Before Priority Deadline
F0	After Priority Deadline

Priority Description 06-07 to 07-08	
A00	Sibling
B01	Priority for PISC-see attached
B02	Priority for PISC-see attached
B03	Priority for PISC-see attached
B05	Priority for PISC-see attached
B05	Priority for PISC-see attached
B06	Priority for PISC-see attached
B07	Priority for PISC-see attached
B08	Priority for PISC-see attached
B09	Priority for PISC-see attached
B10	Priority for PISC-see attached
E00	Before Priority Deadline
F00	After Priority Deadline

Choice

Program Criterion	
1	Sibling-concurrenly enrolled
2	Gifted Cluster
3*	General indistrict
4	General out of district

*2001-02 was specialized course General indistrict was 4 and out of district was 5.

Priority Type	
A	Sibling
B*	Calendar Change
C*	Specialized Course
E	Before Priority Deadline
F	After Priority Deadline
H	Before Priority Deadline OD
I	After Priority Deadline OD
J	Gifted Cluster

*Eliminated after 2001-02

Figure 2.2. Priority Codes for Magnet, VEEP, and Choice Programs

Table 2.1. Applications to True Lotteries, and Characteristics of Applicants

	Magnet	VEEP	Choice	Pooled
Number of Applications	7,140	3,810	3,415	14,365
Win That Lottery	0.46	0.69	0.43	0.51
Prior Math Score (Z)	-0.13	-0.24	0.25	-0.07
Prior Reading Score (Z)	-0.12	-0.32	0.31	-0.07
Male	0.47	0.5	0.47	0.48
Parent Ed $\geq$ Bachelor	0.21	0.11	0.40	0.23
White	0.12	0.04	0.35	0.15
Black	0.27	0.16	0.14	0.21
Asian	0.13	0.07	0.14	0.12
Hispanic	0.47	0.73	0.36	0.52

Notes: Apart from the number of applications and test scores, all other rows refer to proportions of applications in which students had the given characteristic. The sample consists of all applications to true lotteries for admission in fall 2001 through fall 2007.

Table 2.2. The Relationship Between Winning a Lottery and Student Characteristics

	Magnet	VEEP	Choice	Pooled
Prior Math Score	0.000	0.039	-0.073	-0.007
	(0.026)	(0.032)	(0.051)	(0.020)
Prior Reading Score	0.014	0.040	-0.032	0.010
	(0.034)	(0.037)	(0.055)	(0.024)
Presence of Lagged Scores	0.010	0.003	-0.015	0.001
	(0.008)	(0.010)	(0.011)	(0.006)
Male	0.010	-0.018	-0.014	-0.004
	(0.019)	(0.023)	(0.017)	(0.012)
Parent Ed $\geq$ Bachelor's	0.004	0.013	0.026	0.012
	(0.014)	(0.015)	(0.029)	(0.011)
White	0.024	0.009	0.027	0.020*
	(0.015)	(0.010)	(0.018)	(0.009)
African-American	-0.021	-0.008	-0.002	-0.013
	(0.012)	(0.018)	(0.016)	(0.008)
Hispanic	0.020	0.009	-0.016	0.008
	(0.014)	(0.023)	(0.024)	(0.011)
Asian/Pacific Islander	-0.021	-0.007	-0.009	-0.014
	(0.013)	(0.011)	(0.023)	(0.009)
Left District After Lottery	-0.004	-0.015	-0.029*	-0.013*
	(0.009)	(0.010)	(0.013)	(0.006)

Notes: For each choice type, various student characteristics are used as outcomes and regressed on a dummy for winning the lottery plus lottery fixed effects. This table reports the coefficients on 'win lottery' from these regressions, along with their standard errors. Results are informative of lottery fairness. Insignificant results indicate that lottery winners and losers show no statistically significant difference on the given characteristic. Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table 2.3. Comparison of School Characteristics: Choice Program Schools Versus Local Schools

Panel A: Magnet Schools

School Characteristic	Mean at Applied School	Mean Difference (Applied−Local)	p-value Difference = 0
Reading, Z-score	0.00	0.36	<0.00005
Math, Z-score[a]	-0.12	0.15	<0.00005
School Value-Added, Reading	0.00	0.04	<0.00005
School Value-Added, Math[a]	-0.08	-0.05	<0.00005
Mean Grade Point Average	2.50	0.39	<0.00005
Proportion on Track in Grade 9	0.68	0.18	<0.00005
Proportion Graduating HS on Time	0.77	0.15	<0.00005
% Days Absent	5.24	-1.83	<0.00005
Mean Citizenship Grade	2.75	0.17	<0.00005
Proportion Students White	0.25	0.17	<0.00005
Proportion Students Black	0.21	0.02	<0.00005
Proportion Students Asian	0.14	-0.04	<0.00005
Proportion Students Hispanic	0.39	-0.15	<0.00005
Proportion Parent Ed $\geq$ Bachelor's	0.21	0.09	<0.00005

Panel B: VEEP Schools

School Characteristic	Mean at Applied School	Mean Difference (Applied−Local)	p-value Difference = 0
Reading, Z-score	0.16	0.65	<0.00005
Math, Z-score[a]	0.20	0.61	<0.00005
School Value-Added, Reading	0.00	0.04	<0.00005
School Value-Added, Math[a]	0.02	0.04	<0.00005
Mean Grade Point Average	2.63	0.54	<0.00005
Proportion on Track in Grade 9	0.66	0.20	<0.00005
Proportion Graduating HS on Time	0.79	0.21	<0.00005
% Days Absent	5.42	-1.76	<0.00005
Mean Citizenship Grade	2.96	0.46	<0.00005
Proportion Students White	0.37	0.31	<0.00005
Proportion Students Black	0.10	-0.08	<0.00005
Proportion Students Asian	0.13	0.01	0.0006
Proportion Students Hispanic	0.39	-0.25	<0.00005
Proportion Parent Ed $\geq$ Bachelor's	0.29	0.19	<0.00005

a — Restricted to middle schools and grades 6–7

Table 2.3. Comparison of School Characteristics: Choice Program Schools Versus Local Schools, Continued

Panel C: Choice Schools

School Characteristic	Mean at Applied School	Mean Difference (Applied−Local)	p-value Difference = 0
Reading, Z-score	0.32	0.53	<0.00005
Math, Z-score[a]	0.41	0.54	<0.00005
School Value-Added, Reading	0.01	0.03	<0.00005
School Value-Added, Math[a]	0.03	0.04	<0.00005
Mean Grade Point Average	2.72	0.42	<0.00005
Proportion on Track in Grade 9	0.72	0.16	<0.00005
Proportion Graduating HS on Time	0.82	0.14	<0.00005
% Days Absent	5.31	-1.13	<0.00005
Mean Citizenship Grade	3.03	0.32	<0.00005
Proportion Students White	0.42	0.22	<0.00005
Proportion Students Black	0.08	-0.07	<0.00005
Proportion Students Asian	0.13	-0.02	<0.00005
Proportion Students Hispanic	0.37	-0.14	<0.00005
Proportion Parent Ed $\geq$ Bachelor's	0.37	0.23	<0.00005

Panel D: All Schools (Magnet, VEEP, and Choice)

School Characteristic	Mean at Applied School	Mean Difference (Applied−Local)	p-value Difference = 0
Reading, Z-score	0.12	0.48	<0.00005
Math, Z-score[a]	0.09	0.40	<0.00005
School Value-Added, Reading	0.01	0.04	<0.00005
School Value-Added, Math[a]	-0.03	0.04	0.0039
Mean Grade Point Average	2.59	0.44	<0.00005
Proportion on Track in Grade 9	0.69	0.18	<0.00005
Proportion Graduating HS on Time	0.79	0.16	<0.00005
% Days Absent	5.30	-1.64	<0.00005
Mean Citizenship Grade	2.87	0.28	<0.00005
Proportion Students White	0.32	0.22	<0.00005
Proportion Students Black	0.15	-0.03	<0.00005
Proportion Students Asian	0.14	-0.02	<0.00005
Proportion Students Hispanic	0.38	-0.17	<0.00005
Proportion Parent Ed $\geq$ Bachelor's	0.27	0.15	<0.00005

a — Restricted to middle schools and grades 6–7

Table 2.4. Intent to Treat Estimates of the Impact of Winning a Lottery on High School Outcomes

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Magnet	-0.00758	0.000665	-0.174	-0.0198	-0.0160	0.0254
	(0.0123)	(0.0174)	(0.314)	(0.0372)	(0.0327)	(0.0228)
VEEP	-0.0114	-0.0136	-0.0250	0.0839	0.0276	0.0136
	(0.0264)	(0.0235)	(0.526)	(0.0553)	(0.0400)	(0.0237)
Choice	0.0241	0.0148	-0.460	-0.0472	0.00494	-0.00602
	(0.0344)	(0.0224)	(0.318)	(0.0428)	(0.0379)	(0.0176)
Pooled	-0.000268	0.00207	-0.201	-0.000635	0.00104	0.0151
	(0.0127)	(0.0120)	(0.220)	(0.0264)	(0.0215)	(0.0134)
Outcome Mean (Pooled)	0.581	0.614	93.57	3.206	2.893	0.725

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table 2.5. Intent to Treat Estimates of the Impact of Winning a Lottery on Postsecondary Enrollment Outcomes

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Magnet	-0.000545	-0.0108	-0.00427	-0.0164	0.0244	-0.0804
	(0.0212)	(0.0246)	(0.0228)	(0.0590)	(0.0506)	(0.0661)
VEEP	0.0185	0.0437	-0.0449	0.00397	0.131	-0.0786
	(0.0253)	(0.0322)	(0.0250)	(0.0772)	(0.0952)	(0.0882)
Choice	0.0294	0.0471	-0.0216	0.120	0.145	-0.0276
	(0.0167)	(0.0253)	(0.0237)	(0.0829)	(0.0793)	(0.0708)
Pooled	0.0122	0.0181	-0.0185	0.0256	0.0838*	-0.0656
	(0.0125)	(0.0155)	(0.0144)	(0.0416)	(0.0397)	(0.0438)
Outcome Mean (Pooled)	0.700	0.494	0.300	2.296	1.581	0.992

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table 2.6. Intent to Treat Estimates of the Impact of Winning a Lottery on Postsecondary Graduation Outcomes

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Magnet	0.0163*	0.0166*	0.0131	-0.00234	-0.0196	-0.0505*
	(0.00701)	(0.00716)	(0.0106)	(0.0107)	(0.0148)	(0.0213)
VEEP	0.0226**	0.0338**	0.0170	0.0171	-0.00737	0.0106
	(0.00786)	(0.0120)	(0.0144)	(0.0105)	(0.0172)	(0.0234)
Choice	0.0138	0.0217	0.0183	-0.0437*	-0.00793	0.00905
	(0.0177)	(0.0199)	(0.0278)	(0.0196)	(0.0282)	(0.0273)
Pooled	0.0170**	0.0219**	0.0157	-0.00657	-0.0135	-0.0193
	(0.00587	(0.00643)	(0.00879)	(0.00799)	(0.0105)	(0.0142)
Outcome Mean (Pooled)	0.032	0.045	0.062	0.061	0.136	0.183

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table 2.7. Estimates of the Impact of Winning a Lottery and Enrolling on High School Outcomes

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Magnet	-0.0282	0.00261	-0.649	-0.0735	-0.0602	0.101
	(0.0433)	(0.0677)	(1.142)	(0.135)	(0.123)	(0.0885)
VEEP	-0.0279	-0.0344	-0.0619	0.209	0.0688	0.0343
	(0.0627)	(0.0586)	(1.289)	(0.127)	(0.0974)	(0.0602)
Choice	0.0694	0.0458	-1.325	-0.136	0.0142	-0.0189
	(0.100)	(0.0691)	(0.938)	(0.124)	(0.107)	(0.0553)
Pooled	-0.000820	0.00668	-0.622	-0.00197	0.00322	0.0492
	(0.0386)	(0.0385)	(0.680)	(0.0815)	(0.0665)	(0.0437)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table 2.8. Estimates of the Impact of Winning a Lottery and Enrolling on Postsecondary Enrollment Outcomes

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Magnet	-0.00215	-0.0425	-0.0168	-0.0647	0.0964	-0.317
	(0.0829)	(0.0962)	(0.0893)	(0.228)	(0.199)	(0.250)
VEEP	0.0439	0.104	-0.107	0.00944	0.313	-0.187
	(0.0609)	(0.0808)	(0.0582)	(0.181)	(0.231)	(0.207)
Choice	0.0895	0.143*	-0.0658	0.365	0.444*	-0.0842
	(0.0461)	(0.0724)	(0.0735)	(0.234)	(0.208)	(0.213)
Pooled	0.0390	0.0576	-0.0590	0.0817	0.267*	-0.209
	(0.0406)	(0.0495)	(0.0455)	(0.133)	(0.125)	(0.138)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table 2.9. Estimates of the Impact of Winning a Lottery and Enrolling on Postsecondary Graduation Outcomes

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Magnet	0.0658*	0.0685	0.0574	-0.00949	-0.0811	-0.220**
	(0.0334)	(0.0353)	(0.0503)	(0.0422)	(0.0624)	(0.0714)
VEEP	0.0500**	0.0749**	0.0377	0.0381	-0.0164	0.0236
	(0.0163)	(0.0259)	(0.0314)	(0.0245)	(0.0372)	(0.0526)
Choice	0.0376	0.0565	0.0508	-0.122**	-0.0212	0.0261
	(0.0464)	(0.0511)	(0.0789)	(0.0441)	(0.0734)	(0.0772)
Pooled	0.0519**	0.0671**	0.0483	-0.0202	-0.0417	-0.0602
	(0.0185)	(0.0207)	(0.0285)	(0.0236)	(0.0323)	(0.0433)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table 2.10. Estimates of the Impact of Enrolling in Any School to Which Applied on High School Outcomes

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Magnet	-0.0346	0.00312	-0.859	-0.0925	-0.0757	0.124
	(0.0532)	(0.0809)	(1.508)	(0.171)	(0.155)	(0.110)
VEEP	-0.0401	-0.0476	-0.0864	0.298	0.0946	0.0492
	(0.0907)	(0.0814)	(1.796)	(0.188)	(0.134)	(0.0849)
Choice	0.0929	0.0629	-1.874	-0.194	0.0200	-0.0265
	(0.133)	(0.0930)	(1.338)	(0.172)	(0.152)	(0.0782)
Pooled	-0.00149	0.0114	-1.068	-0.00335	0.00547	0.0863
	(0.0703)	(0.0657)	(1.156)	(0.139)	(0.113)	(0.0744)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table 2.11. Estimates of the Impact of Enrolling in Any School to Which Applied on Postsecondary Enrollment Outcomes

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Magnet	-0.00271	-0.0535	-0.0212	-0.0815	0.121	-0.400
	(0.104)	(0.121)	(0.113)	(0.289)	(0.250)	(0.324)
VEEP	0.0634	0.150	-0.154*	0.0136	0.451	-0.270
	(0.0876)	(0.113)	(0.0778)	(0.262)	(0.319)	(0.291)
Choice	0.128	0.204*	-0.0937	0.519	0.632*	-0.120
	(0.0664)	(0.0992)	(0.104)	(0.339)	(0.293)	(0.302)
Pooled	0.0704	0.104	-0.107	0.148	0.483*	-0.378
	(0.0737)	(0.0881)	(0.0807)	(0.242)	(0.220)	(0.241)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table 2.12. Estimates of the Impact of Enrolling in Any School to Which Applied on Postsecondary Graduation Outcomes

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Magnet	0.0939	0.0987	0.0952	-0.0133	-0.114	-0.350**
	(0.0506)	(0.0534)	(0.0776)	(0.0590)	(0.0903)	(0.132)
VEEP	0.0727**	0.112**	0.0563	0.0547	-0.0242	0.0348
	(0.0269)	(0.0375)	(0.0438)	(0.0372)	(0.0555)	(0.0762)
Choice	0.0570	0.0826	0.0752	-0.183**	-0.0307	0.0386
	(0.0706)	(0.0750)	(0.119)	(0.0629)	(0.106)	(0.116)
Pooled	0.101**	0.130**	0.0928	-0.0381	-0.0787	-0.112
	(0.0381)	(0.0421)	(0.0541)	(0.0439)	(0.0610)	(0.0814)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table 2.13. Significance of Interactions Between Winning a Lottery and Differences in School Characteristics

Panel A: High School Outcomes

Outcome:	On Track in Grade 9	Passed Exit Exam in Grade 10	% Days Attended in Grade 12	Grade 12 Citizenship Grade	GPA in Grade 12	Graduated on Time
DIFFERENCES (APPLIED−LOCAL) IN:						
Reading, Z-score						
Math, Z-score[a]						
School Value-Added, Reading			++			
School Value-Added, Math[a]						
Mean Grade Point Average						
Proportion on Track in Grade 9						
Proportion Graduating HS on Time						
% Days Absent				−		
Mean Citizenship Grade						
Proportion Students White						
Proportion Students Black						
Proportion Students Asian						
Proportion Students Hispanic				−		
Proportion Parent Ed $\geq$ Bachelor's						

Notes: +, ++, − and −− indicate that the interaction between winning a lottery and the difference in the school characteristic listed was positive and significant at the 5% level, positive and significant at the 1% level, negative and significant at the 5% level, and negative and significant at the 1% level, respectively. Blank cells indicate the interaction was not significant.

a — Restricted to middle schools and grades 6–7

Table 2.13. Significance of Interactions Between Winning a Lottery and Differences in School Characteristics, Continued

Panel B: Postsecondary Enrollment Outcomes

Outcome:	# Years Enrolled in Postsecondary 1 Year After HS Graduation	# Years Enrolled in 2-Year College 1 Year After HS Graduation	# Years Enrolled in 4-Year College 1 Year After HS Graduation	# Years Enrolled in Postsecondary 4 Years After HS Graduation	# Years Enrolled in 2-Year College 4 Years After HS Graduation	# Years Enrolled in 4-Year College 4 Years After HS Graduation
DIFFERENCES (APPLIED−LOCAL) IN:						
Reading, Z-score				+		
Math, Z-score[a]			−			
School Value-Added, Reading				+		
School Value-Added, Math[a]						
Mean Grade Point Average						
Proportion on Track in Grade 9						
Proportion Graduating HS on Time						
% Days Absent						
Mean Citizenship Grade						
Proportion Students White					+	
Proportion Students Black						
Proportion Students Asian						
Proportion Students Hispanic						
Proportion Parent Ed $\geq$ Bachelor's						

a — Restricted to middle schools and grades 6–7

Table 2.13. Significance of Interactions Between Winning a Lottery and Differences in School Characteristics, Continued

Panel C: Postsecondary Graduation Outcomes

Outcome:	Associate or Other 2-Year College Cert. 4 Years After HS Graduation	Associate or Other 2-Year College Cert. 5 Years After HS Graduation	Associate or Other 2-Year College Cert. 6 Years After HS Graduation	Bachelor's 4 Years After HS Graduation	Bachelor's 5 Years After HS Graduation	Bachelor's 6 Years After HS Graduation
DIFFERENCES (APPLIED−LOCAL) IN:						
Reading, Z-score						
Math, Z-score[a]						
School Value-Added, Reading						
School Value-Added, Math[a]						
Mean Grade Point Average						
Proportion on Track in Grade 9						
Proportion Graduating HS on Time						
% Days Absent					−	
Mean Citizenship Grade						
Proportion Students White						
Proportion Students Black						
Proportion Students Asian					+	
Proportion Students Hispanic						
Proportion Parent Ed ≥ Bachelor's						

a — Restricted to middle schools and grades 6–7

Table 2.14. Significance of Interactions Between Winning a Lottery and Student Characteristics

Panel A: High School Outcomes

Outcome:	On Track in Grade 9	Passed Exit Exam in Grade 10	% Days Attended in Grade 12	Grade 12 Citizenship Grade	GPA in Grade 12	Graduated on Time
INTERACTION VARIABLE:						
Lagged Reading Z-score						
Lagged Math Z-Score						
Male						−
Grade to Which Applied				−−	++	++
Ever an English Learner						
Ever Received Special Education Services						
Race/Ethnicity			*			*
Parental Education Level						

Panel B: Postsecondary Enrollment Outcomes

Outcome:	# Years Enrolled in Postsecondary 1 Year After HS Graduation	# Years Enrolled in 2-Year College 1 Year After HS Graduation	# Years Enrolled in 4-Year College 1 Year After HS Graduation	# Years Enrolled in Postsecondary 4 Years After HS Graduation	# Years Enrolled in 2-Year College 4 Years After HS Graduation	# Years Enrolled in 4-Year College 4 Years After HS Graduation
INTERACTION VARIABLE:						
Lagged Reading Z-score					+	−−
Lagged Math Z-Score					+	
Male						
Grade to Which Applied	−		**	**		**
Ever an English Learner						
Ever Received Special Education Services					−	
Race/Ethnicity						
Parental Education Level						

Notes: +, ++, − and −− indicate that the interaction between winning a lottery and the difference in the school characteristic listed was positive and significant at the 5% level, positive and significant at the 1% level, negative and significant at the 5% level, and negative and significant at the 1% level, respectively. Blank cells indicate the interaction was not significant. For grade, race and parental education, significance is from an F-test for lack of joint significance of individual interactions. ** or * indicates that an F-test rejected the null with a p-value of < 0.01 or < 0.05 but the impact did not vary monotonically with the given variable, or the variable itself is not an ordered variable.

Table 2.14. Significance of Interactions Between Winning a Lottery and Student Characteristics, Continued

Panel C: Postsecondary Graduation Outcomes

Outcome:	Associate or Other 2-Year College Cert. 4 Years After HS Graduation	Associate or Other 2-Year College Cert. 5 Years After HS Graduation	Associate or Other 2-Year College Cert. 6 Years After HS Graduation	Bachelor's 4 Years After HS Graduation	Bachelor's 5 Years After HS Graduation	Bachelor's 6 Years After HS Graduation
INTERACTION VARIABLE:						
Lagged Reading Z-score	+	+	+	−	−	
Lagged Math Z-Score			+			
Male						
Grade to Which Applied						**
Ever an English Learner	−					
Ever Received Special Education Services					+	+
Race/Ethnicity						
Parental Education Level		*				

Chapter 3

How Sensitive Are School Value-Added Estimates to Methodology?

3.1 Introduction

The most readily available information on school performance comes in levels of test scores. Yet levels of test scores do not necessarily identify how effective schools are, as schools differ on many external factors over which they have little control — such as students' socioeconomic status or family attitudes toward schooling.

Rather than looking only at raw test scores, the effectiveness of schools is better characterized by how well they perform *given* their students' background and academic histories. Toward this goal, there has been an increase in value-added modeling (VAM) for both teachers and schools during the past 10–20 years, which measures effectiveness based on within-student gains in achievement. And while many applications of school accountability still rely on levels of performance, rather than value-added, this could change in upcoming years as VAM continues to gain popularity.

But even with growing popularity, there is not a uniform standard for how to conduct value-added modeling. This may sometimes affect results. In a real world example, in 2010 the Los Angeles Times newspaper published a relative ranking of local schools by value-added, based on research by Buddin (2010). About seven months

later, the Los Angeles Unified School District provided its own estimates of schools' value-added. Schools often did not rank similarly between the two analyses.[1]

If VAM becomes a key policy tool, it is essential to know whether its results are sensitive to the methodology used. While there have been some attempts to investigate this issue, most of these studies have looked at teacher value-added. It is unclear whether their findings necessarily carry over to school value-added, since the issues that present econometric challenges differ somewhat between the two cases. Thus, in this paper, we estimate school value-added using a variety of models to examine sensitivity to both the set of controls used and overall class of value-added model.

Our results indicate a modest sensitivity of value-added estimates to the control specification, and a somewhat larger sensitivity to the overall class of value-added model used. Correlations between individual model classes are heterogeneous, but in ways that can largely be explained by the assumed degree of persistence in each model. We also find the precision of estimates to be sensitive to the inclusion of school-level controls, and whether researchers use panel instrument techniques to address dynamic panel bias.

Finally, we compare how well school value-added estimates correlate with schools' average levels of test score. We observe an interesting pattern that although school-average levels show almost no correlation with school-average gains (when using minimal covariates), in most specifications value-added shows a moderate positive correlation with average test score levels.

The rest of the paper proceeds as follows. Section 3.2 reviews the relevant literature, Section 3.3 summarizes the methods used, while Section 3.4 summarizes the data used. Section 3.5 previews results, focusing on the relative precision of various models, sensitivity of school rankings to covariates and model class, and comparing the

[1]Imberman and Lovenheim (2016) provide additional details about this example from Los Angeles. Figure 1 in their appendix shows scatterplots of schools' value-added estimates in the various analyses.

value-added models to results from a naive analysis of test score levels, which is the typical approach used in states' accountability systems. Section 3.6 concludes.

3.2 Value-Added Overview and Related Literature

This section provides a basic overview of VAM approaches to help motivate the models that we assess. Equations are written for the context of school value-added, though the patterns that we describe are also largely true of teacher value-added. We then the review the segment of the literature that is most closely related to the current piece. A more complete literature review can be found in Koedel, Mihaly and Rockoff (2015).

3.2.1 General Overview

VAM typically models students' test score advancement over time while including a set of school indicators as regressors. An example specification is

$$\Delta S_{ist} = \alpha_s + \epsilon_{ist}, \tag{3.1}$$

where ΔS_{ist} denotes the one-year change in test score for student i at school s at time t, and α_s represents the value-added parameter for school s. This model is often referred to as the "gain-score" model.

A second common class of value-added model takes the form

$$S_{ist} = \rho S_{is,t-1} + \alpha_s + \epsilon_{ist}, \tag{3.2}$$

which we refer to as the "level-on-lag" model. Instead of using the change in score as the outcome, this model regresses the level of test score on the lagged level, along with school dummies and possibly covariates. This model is identical to the gain-score model when $\rho = 1$, but is otherwise more general, since it allows ρ to take possibly different values — as may be appropriate if past learning depreciates over time or otherwise does

not carry over perfectly between years.[2]

Researchers sometimes differ on what covariates to include. While the preceding equations are written without covariates, for simplicity, in practice these models often feature an additional set of control variables, which varies in parsimony between studies. For instance, the Tennessee Value-Added Assessment System (Sanders and Horn, 1994, 1998) advocates a parsimonious specification with few covariates — relying instead on the gain structure of VAM to (implicitly) control for student background factors through the prior test score. The argument is that, to the extent students do not sort randomly to schools, inclusion of some variables like socioeconomic status may not just capture the intended control effects, but may also proxy for school effectiveness. When this is the case, part of a school's actual effectiveness may attach to the control variables instead, resulting in biased estimates of value-added. Nevertheless, many researchers prefer to use explicit controls for student background and other factors during VAM. The most common of these controls are basic student characteristics, such as race and socioeconomic status, followed by aggregate characteristics at the school- or classroom-level.

It is also common to see Equations 3.1 and 3.2 specified with or without student fixed effects. We note, however, that the inclusion of student fixed effects may be better-suited for teacher value-added settings (which to date have been the majority of value-added studies) than school value-added settings. The reason is that, once student fixed effects are included, identification of the value-added parameters comes only from students who change teachers or schools. But whereas most students change teachers from year to year, most students do not change schools between years, except at set

[2]In recent years, a majority of studies seem to favor the "level-on-lag" specification of Equation 3.2 over the "gain" specification of Equation 3.1 (Koedel, Mihaly and Rockoff, 2015). However, there remain many earlier value-added studies that use the gain-score model.

transitional grades (like going from middle school to high school).[3] Student fixed effects thus remove a large degree of variation in school value-added settings. This issue is somewhat more innocuous in teacher value-added settings, but in both cases, the reduction in variation may result in noisier estimates of key parameters.

This highlights a few other important differences between school value-added and teacher value-added settings. While non-random sorting is a common issue to both, it likely poses a greater danger in teacher value-added models because students are sorted not only between schools, but also likely across classrooms due to ability grouping or tracking. Limited sample size is also an issue in the teacher value-added literature, where a given teacher may only be observed teaching a small number of students. Schools, on the other hand, typically generate a larger average sample of students, but may still be affected by size in a different manner, since schools themselves vary in size. Typically, changes in average test scores tend to be more volatile for smaller schools (Chay, McEwan and Urqiola, 2005; Kane and Staiger, 2002).

That the econometric pitfalls differ somewhat between the two cases is important, given that the vast majority of the value-added literature has focused on teachers, rather than schools. We revisit this issue toward the end of our literature review.

3.2.2 Similar Papers

Having described the basics of value-added, we next review the segment of the literature that is most closely related to our paper. We focus on papers that estimate value-added using a variety of model specifications and/or estimation techniques, with the specific intent comparing their performance.

One subset of papers studies the impact of student demographic controls on teacher value-added estimates. This has sometimes been a point of contention among

[3] In our data, roughly 80% of students attend only one elementary school. Moreover, the minority of students who do change schools may be fundamentally dissimilar from the rest.

researchers, due to the absence of covariates in the Tennessee Value-Added Assessment System (TVAAS). Papers studying this issue have reached somewhat mixed conclusions. Ballou, Sanders and Wright (2004) find that the inclusion of demographics controls affects teacher value-added estimates under a simple fixed effects estimator, but not under the TVAAS.[4] McCaffrey et al. (2004), meanwhile, use a simulation approach and find that the sensitivity to covariates is lower when students sort homogeneously students sort to teachers and schools, on these observed characteristics.

A second group of papers considers more general variations in the model specification, often changing both the set of covariates and overall class of value-added model. We highlight three papers in this group: Lockwood et al. (2007), Newton et al. (2010), and Tekwe et al. (2004).

These papers perform similar exercises, though the first two focus on teacher value-added, while the third focuses, like we do, on school value-added. These papers estimate value-added using a variety of specifications and/or estimation techniques, before computing correlations of how well these estimates align across cases. Lockwood et al. (2007) use the most extensive menu of models — featuring 5 sets of covariates and 4 classes of models, for 20 specifications total — to study the sensitivity of teacher value-added estimates. They find generally high correlations across models when changing either covariates, model class, or both. Newton et al. (2010) use a somewhat smaller array of models, but find similar results. Both of these preceding papers suggest only a mild degree of sensitivity to the model specification.[5] However, this is not the case

[4]Ballou, Sanders and Wright (2004) argue that, in practice, the importance of covariates may depend on how fully a model makes use of the information contained in students' test scores. The more information about a student's academic traejectory that a model gleans from test scores themselves (perhaps by using covariances between scores in different subjects and years, like in the TVAAS), the less additional information is gained by covariates, which would otherwise proxy for similar information — and the less the results change.

[5]These authors do not, however, conclude that value-added estimates are universally robust to methodology. Model specification aside, Lockwood et al. (2007) highlight sensitivity to using different math achievement measures, while Newton et al. (2010) highlight temporal instability of teacher value-added

for Tekwe et al. (2004), who study school value-added, and find results sensitive to the inclusion of student-level covariates. In their paper, school value-added estimates from the model with student controls show the lowest correlations with all the other models. This is especially true for third grade math, where the correlation with estimates from other models is typically around 0.6.

(We note that in this last finding, Tekwe et al. (2004) introduce covariates only at the same time as a change in the overall structure of the value-added model, making it unclear what role each respective component plays. But the end finding of sensitivity still differs from the earlier papers. Our paper extends upon Tekwe et al. (2004) by separately exploring sensitivity to covariates and to the class of value-added model.)

Sass, Semykina and Harris (2014) is another similar paper that estimates teacher value-added using a variety of models. Their paper initially focuses on testing the underlying assumptions of value-added, which they typically reject, before also comparing the similarity of estimates across models.[6] They do so in two ways: (1) reporting how often the same teacher is estimated as being in the top 10% or bottom 10% across separate models, and (2) computing rank correlations of the teacher value-added coefficients across models. Doing so, they find a wide range in the similarity of estimates across models. While some models give similar results to others, other combinations give very different estimates of teacher value-added. The presence of student fixed effects, in particular, seems to make a big difference. While estimates from models without student fixed effects tend to show high correlations with those of other models without student fixed effects, this often is not the case when comparing two models that both have student fixed effects, or when comparing a model with student fixed effects to one without. The authors comment that this may be due to the reduction in variation that comes from using

estimates, and that teachers can have different value-added by course taught.

[6]See Todd and Wolpin (2003) for an excellent primer on value-added that motivates and explains the underlying assumptions.

student fixed effects, which results in noisier estimates.

Guarino, Reckase and Wooldridge (2015) use a simulation approach to assess the performance of various teacher value-added models and estimation techniques while setting the data-generating process. Their simulations consider multiple variations for how students are assigned to teachers (not always randomly), the degree of persistence between years (ρ), and whether teacher effects themselves are large or small. They then assess the performance of each model/estimation approach under each of the different scenarios. Unlike the other papers mentioned, they do not compare the similarity of value-added estimates from each estimation approach against those of the other approaches, but against the underlying truth that they specify. Unsurprisingly, they find some methods of estimating value-added to be more reliable at capturing actual teacher effectiveness than others, across the various data-generating scenarios. What is more surprising is that ordinary least squares (while controlling for lagged test score level) appears to be the best overall performer — outperforming several alternatives that offer theoretical improvements based on structural considerations.

A final similar paper comes from Andrabi, Das, Khwaja and Zajonc (2011). This paper estimates school value-added using a variety of models, but differs from the others in that it does not estimate value-added for individual schools. Instead, it estimates a single treatment effect for attending a private school, cast in a value-added framework. This paper is useful for its analysis of dynamic panel bias in the context of value-added. Using both OLS and dynamic panel estimators (e.g. the Arellano-Bond (1991) estimator) for a level-on-lag equation, these authors find that OLS gives upward-biased estimates of the persistence parameter, ρ, but often still gives similar *value-added* estimates as the more data-intensive dynamic panel methods. In contrast, the gain-score model, which imposes $\rho = 1$, seems to give downward-biased estimates of value-added.

To recap, there have been somewhat mixed findings from the literature about the

sensitivity of value-added estimates to the methodology — with some papers reporting higher degrees of sensitivity to covariates and/or model class than others. Further complicating matters is the fact that most of these papers have studied teacher value-added, and not school value-added. Among the papers in our literature review, only Tekwe et al. (2004) and Andrabi et al. (2011) study school value-added. Due to some differences in the econometric challenges between the two cases, it is unclear which results also apply to school value-added.

3.3 Methods

This paper features a menu of nine (9) value-added models, consisting of three model classes and three sets of covariate specifications each. Our main analysis estimates value-added using these nine specifications, before computing correlations across models to see how sensitive results are to methodology. We then explore a number of extensions, described in the Results.

Our three classes of value-added models are listed below:

$$\Delta S_{igst} = \alpha_s + \gamma_g + \gamma_t + x'_{igst}\beta + \epsilon_{igst} \tag{3.3}$$

$$S_{igst} = \rho S_{igs,t-1} + \alpha_s + \gamma_g + \gamma_t + x'_{igst}\beta + \epsilon_{igst} \tag{3.4}$$

$$\begin{aligned}\Delta S_{igst} = \rho \Delta S_{igs,t-1} + \alpha_s \Delta 1\{\text{school } s\} + \gamma_g \Delta 1\{\text{grade } g\} \\ + \gamma_t \Delta 1\{\text{year } t\} + \Delta x'_{igst}\beta + \Delta \epsilon_{igst}\end{aligned} \tag{3.5}$$

where S_{igst} denotes the standardized test score of student i in grade g at school s at time t; ΔS_{igst} denotes the one-year change in test score; $(\alpha_s, \gamma_g, \gamma_t)$ denote school, grade, and year fixed effects respectively; x'_{igst} denotes a vector of other control variables; and ϵ_{igst}

denotes an error term.

Equations 3.3 and 3.4 correspond to the gain and level-on-lag models introduced previously. These equations are estimated using ordinary least squares (OLS). Equation 3.5 is a transformed version of the level-on-lag model, in first difference, that allows for and sweeps out a student fixed effect. We use the xtabond2 command in Stata (Roodman, 2009) to estimate this equation using an Arellano-Bond (1991) estimator that instruments the change in lagged test score, as well as the changes in the school dummies (which we treat as endogenous) using lags of the data. Additional details about how we perform this estimation are provided in the appendix. All three equations cluster standard errors by school.

The Arellano-Bond estimator helps address the issue of dynamic panel bias (Nickell, 1981), which arises in equations featuring both student fixed effects and lagged dependent variables as regressors. An undesired result is that, if left unchecked, this may prevent our uninstrumented level-on-lag model from returning consistent estimates.[7] Interestingly, this issue is only sometimes addressed by the existing value-added literature.[8] But while the Arellano-Bond model offers plausible improvements on consistency, it comes with its own downsides and tradeoffs. One is that it makes much heavier demands of the data, requiring lags to use as instruments. It also only obtains variation for identifying school value-added coefficients from students who change schools, which in our data is about one-fifth of the students. We discuss these issues in more detail when presenting results.

[7]The key issue is that in short (small T) panels, any unmodeled shocks affect the *estimated* value of the fixed effect, creating a correlation between the lagged dependent variable and the error term. Additional details are provided in the appendix.

[8]Some example papers that use panel instrument methods include Andrabi et al. (2011), Guarino, Reckase and Wooldridge (2015), and Sass, Semykina and Harris (2014). Many other papers seem to ignore or overlook issue.

Within each model class, we also feature three different specifications of controls:

Control Specification A: school, grade, and year fixed effects

Control Specification B: specification A + student characteristics

Control Specification C: specification B + school demographics

where the exact variables that comprise 'student characteristics' and 'school demographics' are listed in Table 3.1. These specifications are chosen to mirror the value-added literature, where student characteristics are the most commonly used controls, and school demographics are the next most common.

3.4 Data

To estimate these value-added models, we use data from San Diego Unified School District spanning a five-year period from the 2006-07 to 2010-11 school years. The dataset records detailed information at the student-year level, and tracks these students over time to form a panel.

Key variables include a student's grade and primary school attended each year, race, parental education, English learner status, special education status, and school-level demographics. Student achievement is measured using scores on the California Standards Test (CST) in Math and English Language Arts (henceforth 'Reading'), which we convert to Z-scores relative to the state mean and standard deviation, for the respective year and grade level. CST scores are available from grade 2 onward, when standardized testing begins. These tests are administered each year in the spring, toward the end of the school year.

Although the district records these data for students of all grade levels, for tractability reasons, we restrict our analysis to students in elementary schools.[9] Keeping

[9] A key issue for secondary grades is that not all students take the same CST Math test, even within

only cases with available outcomes, we have a total of 162,921 student-year observations, comprised of 77,647 unique students across 120 elementary schools.[10] Table 3.1 provides summary statistics for this group of students.

3.5 Results

3.5.1 Distribution of Value-Added Coefficients and Their Precision

Before showing correlations of estimates across specifications, we first examine basic patterns in the distribution of school coefficients and their standard errors, as this information may help inform some of the later results. Table 3.2 provides summary information (mean and standard deviation) about the estimated value-added coefficients and their standard errors, under each specification.

Looking first at the standard error columns, two patterns can be clearly seen. The first pattern is that, within any class of model, standard errors are much larger under control specification C, which introduces school-level demographics, than either control specifications A or B. The most likely explanation seems to be collinearity between the school demographic variables and the school fixed effects. Values of school demographics, such as the percentage of African-American students, typically do not drift much during the 5-year period that we study, such that there is not much within-school variation for these variables. Standard errors increase dramatically when the model nevertheless attempt to estimate coefficients for these variables in addition to school fixed effects.

The second pattern is that the Arellano-Bond model gives much larger standard errors than either the gain-score model or uninstrumented level-on-lag model. This pattern

a given grade. This makes it difficult to define a school's overall value-added for Math. Students in elementary grades, meanwhile, all take the same Math CST test, such that this issue does not apply.

[10] We ultimately omit six elementary schools with only a few students during this time. These omitted schools typically have fewer than 10 gain-scores available, where ΔS_{igst} is well-defined, in either subject. Of the remaining 114 schools, the average number of gain-scores is about 720 in each subject.

is not entirely surprising, since the Arellano-Bond (1991) estimator first-differences the data, meaning that school coefficients are identified by the subset of students who switch schools — only about one-fifth of students. This is a major limitation of this approach. Precision is also be affected by the use of an instrumental variables procedure.

Both of these patterns on standard errors also carry over to the coefficient estimates. Looking at the coefficient columns in Table 3.2, coefficients themselves are also most widely dispersed under control specification C (for any class of model), and more widely dispersed under the Arellano-Bond model than either the gain-score or level-on-lag model. This has systematic effects on any resulting correlations. Other things the same, we typically expect lower correlations for less precise estimates.

One way to mitigate this issue is to apply shrinkage, or adjust schools' value-added estimates toward the baseline average, depending on their precision. This is a common, though not universal, practice in the value-added literature.[11] Following Morris (1983), we use a shrinkage estimator of the form

$$\hat{\alpha}_s^{shrunk} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + V_s}\right)\hat{\alpha}_s + \left(1 - \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + V_s}\right)\bar{\alpha},$$

where $\hat{\sigma}^2$ denotes the sample variance of the estimated school coefficients, V_s denotes the variance of school s's coefficient estimate, and $\bar{\alpha}$ denotes the overall average of the school value-added coefficients, for the given specification. The more noisily that $\hat{\alpha}_s$ is estimated, the more weight is given to the baseline average, rather than the unadjusted estimate, and vice versa.

The effects of shrinkage can be seen in Figures 3.1 and 3.2, which show histograms of the value-added coefficients for Math and Reading respectively, for both

[11] Among non-economist researchers, value-added models are often estimated using hierarchical linear modeling (HLM) or empirical Bayes estimators, which intrinsically "shrink" a school's value-added parameter toward the baseline average, depending on its precision. Economists, on the other hand, typically do not use HLM as the initial estimation technique, but sometimes achieve a similar shrinkage effect after the fact. See, for example, Chetty, Friedman and Rockoff (2014) and Lefgren and Sims (2012).

shrunk and unshrunk cases. For many specifications, the shrunk estimates are nearly identical to their unshrunk counterparts. But shrinkage has a pronounced impact on specifications with larger baseline standard errors — specifically anything that uses control set C or the Arellano-Bond (1991) estimator. In these cases, the shrunk coefficients typically still have a similar mean as the unshrunk case, but are much more tightly packed around this mean.

In a few cases, the shrunk estimates also exhibit more skewness than the unshrunk estimates. (See, for example, the gain-score model and level-on-lag model for Reading under control specification C.) However, we believe these to be idiosyncratic to our results in those particular specifications, and not a general property of the shrinkage estimator.[12]

3.5.2 Sensitivity to Covariates and Model Class

We use three ways of showing the sensitivity of estimates across value-added models: (1) showing correlations between the various specifications; (2) showing scatterplots of school coefficients; and (3) showing transition matrices for how schools' quintile rankings compare under different specifications. For each of these three cases, tables and figures in the main text use unshrunk estimates throughout, though we discuss the effects of shrinkage and provide analogous tables/figures using shrunk estimates in the appendix.

Table 3.3 begins by showing Spearman rank correlations of value-added estimates across control specifications for each of the three classes of value-added models that we use.[13] We henceforth refer to these as "cross-control" correlations, while correla-

[12] Imprecise estimates (or those most affected by shrinkage) are somewhat more likely to be in the left or right tails of the initial unshrunk distribution, since the imprecision of these estimates helps them take more extreme values. Skewness may result after shrinkage if, for some reason, a disproportionate share of these imprecise values come from the right tail rather than the left, or vice versa.

[13] We use Spearman rather than Pearson correlations because in practice the main policy use of the school coefficients is to rank schools. When using Pearson correlations instead (results available upon request),

tions across model classes, for a given set of controls, are called "cross-model class" correlations.

We generally find only a mild degree of sensitivity to the covariates used. One exception is that the gain-score model appears somewhat sensitive to the use of school-level demographics — where estimated coefficients from models that include these variables show rank correlations of 0.52–0.64 compared to other versions of the gain-score model. But cross-control correlations are otherwise generally above 0.7 in all three models, for both Math and Reading. The overall average correlation when changing covariates alone is about 0.781 in Table 3.3.[14]

Table 3.4, meanwhile, shows rank correlations of value-added estimates across model classes for each of control specifications A, B, and C. As a whole, these correlations are considerably lower than those seen in Table 3.3, suggesting that value-added estimates are much more sensitive to the overall class of model than to the control specification. Strikingly, correlations between the gain-score model and the Arellano-Bond model are as low as about 0.1 for both Math and Reading, under control specification A. As an overall average, the average across-model class correlation is about 0.501 in Table 3.4, but with substantial heterogeneity between model classes.

The value of the autoregressive parameter, ρ, may explain much of the patterns observed between model classes. Whereas the gain-score model imposes $\rho = 1$, both versions of the level-on-lag model estimate ρ to be significantly lower. In our data, the uninstrumented level-on-lag model typically estimates $\hat{\rho} \approx 0.7$, while the Arellano-Bond model estimates $\hat{\rho} \approx 0.25$, for both Math and Reading. This ordering for the perceived value of ρ is consistent with the correlations between model classes. Specifically, we see

cross-control correlations are slightly higher than those displayed, while cross-model class correlations are slightly lower, on average.

[14]This calculation reports a simple average of the non-trivial correlations in Table 3.3, averaged across both subjects (Reading and Math) and all three panels. Subsequent calculations are performed similarly.

that the gain-score model and Arellano-Bond model show the lowest correlation with one another under any control specification (and about 0.222 on average within Table 3.4), while the uninstrumented level-on-lag model shows at least moderate correlation with either of the others (0.741 with the gain-score model, and 0.540 with the Arellano-Bond model when averaging across Table 3.4).

It is also plausible that imprecision may affect these correlations to a degree. Other things the same, less precise estimates typically give lower correlation values, and we have seen some specifications have systematically larger standard errors than others. We thus check how shrinkage affects the observed correlations.

Tables A13 and A14 of the appendix show cross-control and cross-model correlations using the shrunk coefficients instead. By and large, these correlations are similar to their values in the unshrunk case. Rank correlations across control specifications remain right around 0.78, on average, using the shrunk coefficients. Across-model class correlations increase slightly to about 0.514 on average, though we see both increases and decreases in the degree to which the Arellano-Bond model (which generates the least precise estimates) correlates with the others. So although we have seen that shrinkage (sometimes) has a pronounced effect on the distribution of coefficient estimates, rank correlations do not change much as a result.

Figures 3.3 and 3.4 next show scatterplots for Math and Reading respectively. Each scatterplot shows school coefficient estimates from one class of model on the X-axis, plotted against school coefficients from another class of model on the Y-axis, for the same set of covariates. Plots are arranged so that each row compares the same two classes of model throughout, while the columns show results under control specifications A, B, and C respectively.

These scatterplots make it easy to see any nonlinearities or outliers, which may affect how to view the earlier correlations. For our data, we do not observe any major

nonlinearities worth noting. However, do we observe a few outliers, especially with the Arellano-Bond model. These outliers are somewhat less pronounced after shrinkage (see Figures A2 and A3 in the appendix), but as noted, rank correlations do not change drastically.

Another pattern that stands out in these scatterplots is that the results of the gain-score model and level-on-lag model become increasingly similar as the number of controls increases. In the top row of either Figure 3.3 or 3.4, the scatterplots become much closer to a straight line as one proceeds rightward across columns. This, however, does not appear to be a general pattern for the other model classes.

Our final way of presenting these results is to show how often a school's relative ranking changes depending on the specification used. This is more than just a statistical exercise, as school districts sometimes target additional resources to bottom-performing schools based on their percentile ranking. (For example, San Diego Unified School District at one point targeted additional resources to elementary schools whose average test scores ranked within the bottom two deciles of the state as a whole.)

To do this, we rank schools in quintiles based on their value-added coefficients, under various specifications, and generate "transition matrices" for how these quintiles compare. Tables 3.5–3.6 display results in this format, comparing each of the three model classes that we use, for control specification B. In both tables, the top panel compares rankings under the gain-score model against those of the level-on-lag model; the middle panel compares rankings under the gain-score model against those of the Arellano-Bond model; and the bottom panel compares rankings under the level-on-lag model against those of the Arellano-Bond model. In all cases, the diagonal elements of each matrix show how often schools remain in the same quintile, across the two models, while off-diagonal elements show how often schools change quintiles.

Our interest in these tables is seeing what percentage of the time schools take (at

least roughly) the same relative ranking across models. To facilitate this task, all panels report the sum of diagonal entries in the upper left hand corner (just outside the transition matrix itself). Looking at these values, it can be seen that these rates are far from 100 percent, as would be the case if rankings remained unchanged between models. In even the "closest" pairing, which for both subjects is the gain-score model and the level-on-lag model, schools change quintiles slightly over half the time. Meanwhile, schools change quintiles as much as 75–80 percent of the time when comparing the least similar models (the gain-score model and the Arellano-Bond model, for both subjects).

We next consider how shrinkage affects these results. Tables A15 and A16 show analogous transition matrices using the shrunk value-added estimates. Overall, we find a similar picture as occurs in the unshrunk case. Because the gain-score and level-on-lag school coefficients are estimated fairly precisely, the school quintile assignments are virtually the same in these two models. Where shrinkage matters, for control specification B, is in the Arellano-Bond model. But even after shrinkage, schools' quintile rankings continue to differ between this model and the others at generally similar rates as before.[15] Altogether, shrinkage appears to do little to resolve the variability of schools' rankings between models.

3.5.3 Correlations Between Value-Added and Level of Test Score

Much of the motivation behind VAM stems from the idea that levels of test score alone do not necessarily convey school effectiveness, due to differences in students' socioeconomic status (e.g.) or other factors that have little to do with the schools themselves. Yet raw test scores are very readily available to the public, and moreover, many aspects of school accountability continue to rely on levels of test score rather than

[15]Consider the bottom panels of Tables A15–A16, which compare the level-on-lag and Arellano-Bond results. For Math, the sum of schools on the diagonals, representing schools that stay in the same quintile, is 36.8 percent with the unshrunk estimates and 37.8 percent with the shrunk estimates. For Reading, the percentage of schools on the diagonals is 35.1 and 38.6 in the unshrunk and shrunk cases respectively.

value-added. For example, the federal No Child Left Behind (NCLB) law, which was in place from 2002 until its replacement in 2015, required states that received federal aid for K-12 education to test students in specified grades and subject areas. Schools that had lower than required percentages of students proficient for more than two years in a row (overall or for student subgroups) were subject to an escalating set of accountability sanctions. This approach did not take into account students' past test scores, and thus had the potential to produce misleading pictures of school effectiveness.

With this being the case, the empirical degree to which value-added aligns with test score levels is a crucial policy question. To explore this question, we estimate an additional model of the form

$$S_{igst} = \alpha_s^L + \gamma_g^L + \gamma_t^L + \epsilon_{igst}^L. \tag{3.6}$$

This equation is not a value-added model, since it does not control for past achievement. Instead, it is a deliberately "naive" levels model, such that the $\hat{\alpha}_s^L$ coefficients essentially just rate schools by their average level of test scores.

Table 3.7 reports correlations between the results of this naive levels model and value-added coefficients from each of the earlier specifications. One interesting pattern is that, prior to the use of school-level controls, in covariate specifications A and B the the gain-score model shows very little correlation with school coefficients from the naive levels model, for either Math or Reading. In most other specifications, value-added shows a moderate correlation with the naive levels model: between 0.20–0.76 for Math, and 0.45–0.86 for Reading.

Earlier, we described that ρ, the value of the autoregressive parameter, was able to explain many of the correlation patterns between model classes. It is worth revisiting this issue in the context of the naive levels model.

Notice that the levels model is equivalent to the level-on-lag model when $\rho = 0$.

In contrast, the gain-score model assumes $\rho = 1$, the empirical (uninstrumented) level-on-lag model estimates ρ to be about 0.7, and the Arellano-Bond model estimates ρ to be about 0.25, for both subjects. In this light, there is a natural ordering between the naive model and each of the value-added models. If ρ is the only key difference between these models, then we should generally expect the naive levels model to show the highest correlation with the Arellano-Bond model, followed by the level-on-lag model, and finally the lowest correlation with the gain model.

The pattern does not hold in Table 3.7, but comes fairly close. The naive model actually shows the highest correlation with the level-on-lag model, rather than the Arellano-Bond model (although the Arellano-Bond model still shows a higher correlation than does the gain-score model). This discrepancy likely reflects the fact the Arellano-Bond estimator differs in two other important ways from the naive model: it instruments the changes in lagged test score and the school dummies, and it also identifies school coefficients from the subsample of students who switch elementary schools.

3.6 Discussion

This paper uses data from San Diego Unified School District to examine the sensitivity of school value-added estimates to both the control specification and overall class of value-added model. We consider three classes of value-added models — a gain-score model, a level-on-lag model, and an Arellano-Bond model that allows for a student fixed or random effect — and three control specifications each, for a total of nine specifications.

Overall, we find a relatively mild sensitivity of results to the control specification, and a somewhat greater sensitivity to the class of value-added model. Precision of estimates appears sensitive to the inclusion of school-level controls, and whether researchers use panel instrument techniques to address dynamic panel bias. Value-added

estimates also typically show a moderate positive correlation with schools' average level of test scores. We recap these results in greater detail below, while offering additional commentary. While we note that a study based on any one set of schools and testing system cannot yield definitive best practices, some tentative conclusions about preferred approaches do emerge.

An important subsidiary goal of the paper has been to compare how school rankings based on value-added compare with rankings based on average levels of test scores.[16] In most cases, we find a moderate positive correlation between the two. There are both optimistic and pessimistic ways to view this result. On one hand, it may be a reassuring sign that, at least on average, schools with higher average achievement also appear to have higher value-added. On the other hand, our analysis using transition matrices suggests that even models that show a correlation of about 0.7 with one another can rank schools in different quintiles over half the time. We thus take a somewhat more pessimistic view on this issue, that although the common practice among policymakers of ranking schools by test score levels is expedient, it is unlikely to be accurate of what these rankings are often intended to capture. Thus, the naive method of ranking schools by test scores, as the U.S. federal government required states to do for a dozen years, is unlikely to produce unbiased estimates of school effectiveness, and should probably be avoided.

This still leaves us with nine value-added models. Our results indicate a modest sensitivity of value-added estimates to the control specification. With an exception that the gain-score model appears somewhat sensitive to the use of school-level controls, we generally find correlations of 0.7 or higher across covariate specifications. Sensitivity to

[16]The latter case is a restricted version of value-added models in which $\rho = 0$. Such a restriction is strongly rejected by the data. In all of our regressions that use a level-on-lag format, whether instrumented or not, the estimated value of ρ can be statistically distinguished from both 0 and 1 at all conventional significance levels. P-values are below 0.001 in all cases.

the overall class of value-added model appears somewhat larger. For the models that we consider, the average correlation across model classes (while holding covariates fixed) is about 0.5 — albeit with substantial heterogeneity. Results of the gain-score model and Arellano-Bond model tend to show low correlation with one another, perhaps due to differences in the supposed degree of persistence (ρ), while the level-on-lag model shows moderate correlation with either of the others.

With the less-than-perfect agreement across the nine specifications, analysis of the results leads to tentative suggestions about how to narrow down the sorts of models one should estimate. The gain-score model assumes that ρ is exactly one. This assumption is strongly rejected by the data, as has been the case for some other researchers as well. Thus, the level-on-lag model (estimated by OLS) or the Arellano-Bond model (which allows for student fixed effects) are likely the most promising approaches for estimating school value-added. Due to concerns about lowered precision when one controls for school demographics, however, researchers attempting to estimate school effectiveness should exercise caution in using school-level controls. Thus, a tentative suggestion is to narrow our nine candidate specifications to four: the level-on-lag and Arellano-Bond estimators, with control specification A, which controls for school, grade and year fixed effects, or control specification B, which additionally controls for student's own demographics.

The Arellano-Bond model has the attractive property of potentially delivering more consistent estimates, due to its ability to address both individual heterogeneity and the related dynamic panel bias, but at the cost of dramatically lower precision. Thus the choice between the two models is not entirely clear, and provides yet another illustration of the trade-off between estimators on bias versus precision. Either of these methods, though, is likely to yield a more accurate picture of underlying rates of learning than naive levels models frequently used by governments or the gain-score approach that has

been widely used by researchers in the recent past.

But even after having whittled nine candidate value-added models down to four, we note that school rankings are highly sensitive to the method. For instance, using control specification B, we showed that if we rank schools into quintiles based on their estimated effectiveness using the Arellano-Bond approach, of the 20 percent of schools in the lowest quintile (for Math), only 7.9 percent — well under half — are also in the bottom quintile of schools according to the level-on-lag model. Perhaps the main lesson from the analysis is that the choice among even the most reasonable models can have dramatic effects on school rankings. Identifying bottom-performing schools for interventions remains a very difficult task, subject to considerable uncertainty.

Chapter 3, in full, is an unpublished manuscript that contains the following co-authors. Betts, Julian R.; Young, Sam M. The dissertation author was the primary investigator and author of this paper.

3.7 Tables and Figures

Table 3.1. Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max
Outcomes:					
Math CST Score (Z)	162,569	0.03	1.01	-4.79	3.72
English Language Arts CST Score (Z)	162,754	0.06	1.05	-6.40	4.85
Student Characteristics:					
Female	162,921	0.49	0.50	0	1
African-American	162,921	0.12	0.33	0	1
Hispanic	162,921	0.45	0.50	0	1
Asian/Pacific Islander	162,921	0.17	0.37	0	1
Other/Unknown Race	162,921	0.01	0.10	0	1
Parent Ed. $<$ High School Diploma	162,921	0.12	0.32	0	1
Parent Ed. = High School Diploma	162,921	0.19	0.39	0	1
Parent Ed. = Some College	162,921	0.19	0.39	0	1
Parent Ed. = College Degree	162,921	0.16	0.36	0	1
Parent Ed. = Post-Graduate	162,921	0.10	0.30	0	1
Parent Ed. = Missing/Unknown	162,921	0.24	0.43	0	1
English Learner†	162,921	0.42	0.49	0	1
Fluent English Proficient†	162,921	0.00	0.04	0	1
Special Education†	162,921	0.10	0.30	0	1
School Demographics:					
% Students on Free or Reduced Priced Lunch	162,921	63.66	32.00	0	100
% African-American	162,921	12.52	11.26	0	90.21
% Hispanic	162,921	44.54	26.28	0	100
% Asian/Pacific Islander	162,921	16.38	13.50	0	54.86
% Other/Unknown Race	162,921	1.17	1.42	0	33.33
% English Learner†	162,921	41.41	24.72	0	100
% Fluent English Proficient†	162,921	0.15	0.31	0	2.97

Notes: Each observation corresponds to a student-year. Statistics are calculated for elementary school students with either Math or English Language Arts (ELA) score observed on the California Standards Test (CST) in a given year. Math and ELA scores are in standard deviation units relative to the state mean in the given year, for the student's respective grade level. Variables marked with dagger (†) are based on students' status in their first appearance in the district. We use first appearance status since subsequent status may be endogenous to school effectiveness. The dummy for white student ($\approx$ 25%), dummy for native English speaker ($\approx$ 57%), school percentage white, and school percentage of native English speakers are omitted to avoid collinearity.

Table 3.2. Summary Information About Coefficients and Standard Errors, Before Shrinkage

	Coefficients		Standard Errors	
	Math	Reading	Math	Reading
Gain, Control A	-0.001	0.030	0.020	0.014
	(0.118)	(0.076)	(0.0017)	(0.0013)
Gain, Control B	-0.048	0.017	0.021	0.016
	(0.117)	(0.076)	(0.0017)	(0.0012)
Gain, Control C	-0.268	-0.245	0.367	0.305
	(0.172)	(0.131)	(0.1240)	(0.0956)
Level-on-Lag, Control A	-0.000	0.042	0.017	0.013
	(0.147)	(0.131)	(0.0015)	(0.0011)
Level-on-Lag, Control B	0.113	0.164	0.019	0.015
	(0.127)	(0.095)	(0.0021)	(0.0015)
Level-on-Lag, Control C	-0.044	0.015	0.322	0.299
	(0.193)	(0.138)	(0.1079)	(0.0935)
Arellano-Bond, Control A	-0.027	0.035	0.257	0.246
	(0.435)	(0.418)	(0.1714)	(0.1492)
Arellano-Bond, Control B	-0.067	0.017	0.250	0.231
	(0.615)	(0.543)	(0.1747)	(0.1278)
Arellano-Bond, Control C	-0.050	0.030	0.472	0.418
	(0.563)	(0.478)	(0.1448)	(0.1115)

Notes: Table reports the mean and standard deviation of school value-added coefficients and their standard errors, prior to shrinkage. Standard deviations are in parentheses.

Table 3.3. Spearman Rank Correlations of Value-Added Coefficients Across Control Specifications

Gain Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.993	1		Control B	0.994	1	
Control C	0.523	0.569	1	Control C	0.613	0.635	1

Level-on-Lag Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.938	1		Control B	0.946	1	
Control C	0.881	0.765	1	Control C	0.913	0.856	1

Arellano-Bond Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.815	1		Control B	0.829	1	
Control C	0.514	0.786	1	Control C	0.704	0.786	1

Table 3.4. Spearman Rank Correlations of Value-Added Coefficients Across Model Classes

Covariate Specification A

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.599	1		LevLag	0.529	1	
ABond	0.112	0.570	1	ABond	0.075	0.613	1

Covariate Specification B

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.761	1		LevLag	0.684	1	
ABond	0.302	0.571	1	ABond	0.169	0.546	1

Covariate Specification C

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.932	1		LevLag	0.940	1	
ABond	0.231	0.351	1	ABond	0.441	0.586	1

Table 3.5. Relative Ranking of Schools, Compared Across Models, for Math

Gain vs. Level-on-Lag

Sum Diagonal ≈ 46.5%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	**13.2**	5.3	0.9	0.9	0
Level-on-	2nd Quintile	4.4	**7.9**	5.3	1.8	0.9
Lag	3rd Quintile	1.8	4.4	**5.3**	7.9	0.9
Model	4th Quintile	0	1.8	**8.8**	6.1	3.5
	5th Quintile	0.9	0.9	0	3.5	**14.0**

Gain vs. Arellano-Bond

Sum Diagonal ≈ 23.6%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	4.4	5.3	**3.5**	4.4	2.6
Arellano-	2nd Quintile	6.1	**6.1**	4.4	2.6	0.9
Bond	3rd Quintile	6.1	2.6	**3.5**	4.4	3.5
Model	4th Quintile	1.8	4.4	**6.1**	2.6	5.3
	5th Quintile	1.8	1.8	2.6	**6.1**	7.0

Level-on-Lag vs. Arellano-Bond

Sum Diagonal ≈ 36.8%		Level-on-Lag Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	7.9	**3.5**	5.3	1.8	1.8
Arellano-	2nd Quintile	6.1	**7.0**	5.3	1.8	0
Bond	3rd Quintile	5.3	3.5	**5.3**	4.4	1.8
Model	4th Quintile	0.9	4.4	3.5	**6.1**	5.3
	5th Quintile	0	1.8	0.9	6.1	**10.5**

Notes: Percent frequencies are displayed. Within each row, the column entry in which the median observation occurs is in bold. The 1st quintile refers to the bottom 20 percent, while the 5th quintile refers to the top 20 percent. Results in this table use control specification B for both models.

Table 3.6. Relative Ranking of Schools, Compared Across Models, for Reading

Gain vs. Level-on-Lag

Sum Diagonal ≈ 41.2%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	**12.3**	4.4	2.6	0.9	0
Level-on-	2nd Quintile	3.5	6.1	**7.0**	3.5	0
Lag	3rd Quintile	2.6	4.4	**5.3**	4.4	3.5
Model	4th Quintile	1.8	3.5	2.6	**7.0**	5.3
	5th Quintile	0	1.8	2.6	4.4	**10.5**

Gain vs. Arellano-Bond

Sum Diagonal ≈ 22.8%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	3.5	2.6	**6.1**	3.5	4.4
Arellano-	2nd Quintile	7.0	**5.3**	2.6	2.6	2.6
Bond	3rd Quintile	4.4	4.4	**3.5**	6.1	1.8
Model	4th Quintile	2.6	4.4	**4.4**	4.4	4.4
	5th Quintile	2.6	3.5	**3.5**	3.5	6.1

Level-on-Lag vs. Arellano-Bond

Sum Diagonal ≈ 35.1%		Level-on-Lag Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	7.0	**6.1**	2.6	2.6	1.8
Arellano-	2nd Quintile	7.0	**6.1**	4.4	2.6	0
Bond	3rd Quintile	1.8	5.3	**5.3**	7	0.9
Model	4th Quintile	3.5	1.8	4.4	**5.3**	5.3
	5th Quintile	0.9	0.9	3.5	2.6	**11.4**

Notes: Percent frequencies are displayed. Within each row, the column entry in which the median observation occurs is in bold. The 1st quintile refers to the bottom 20 percent, while the 5th quintile refers to the top 20 percent. Results in this table use control specification B for all models.

Table 3.7. Spearman Rank Correlations Between Value-Added and Level of Test Score

Math				Reading			
	Gain	LevLag	ABond		Gain	LevLag	ABond
Control A	-0.043	0.732	0.628	Control A	0.075	0.856	0.680
Control B	-0.013	0.530	0.427	Control B	0.107	0.706	0.454
Control C	0.632	0.764	0.203	Control C	0.645	0.785	0.605

Notes: Covariates change for the value-added models, but not for the levels model. The levels model includes only school, year, and grade fixed effects, with no additional covariates throughout.

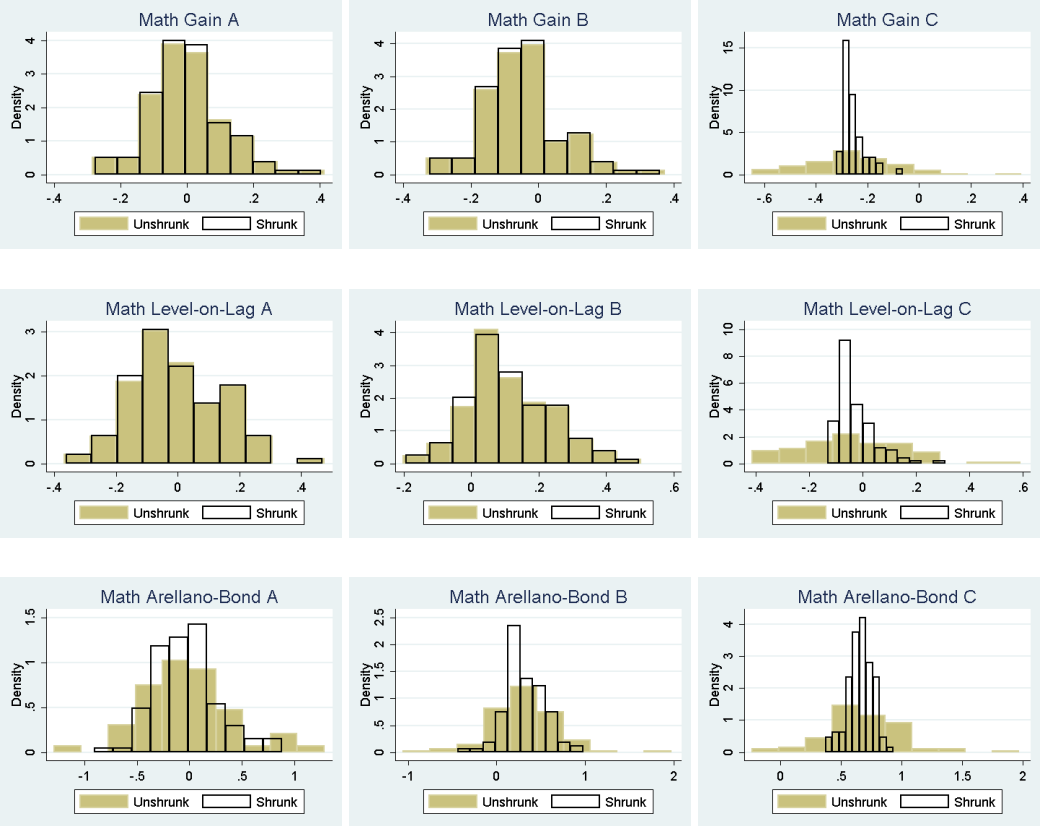


Figure 3.1. Histograms of Value-Added Coefficients, Math

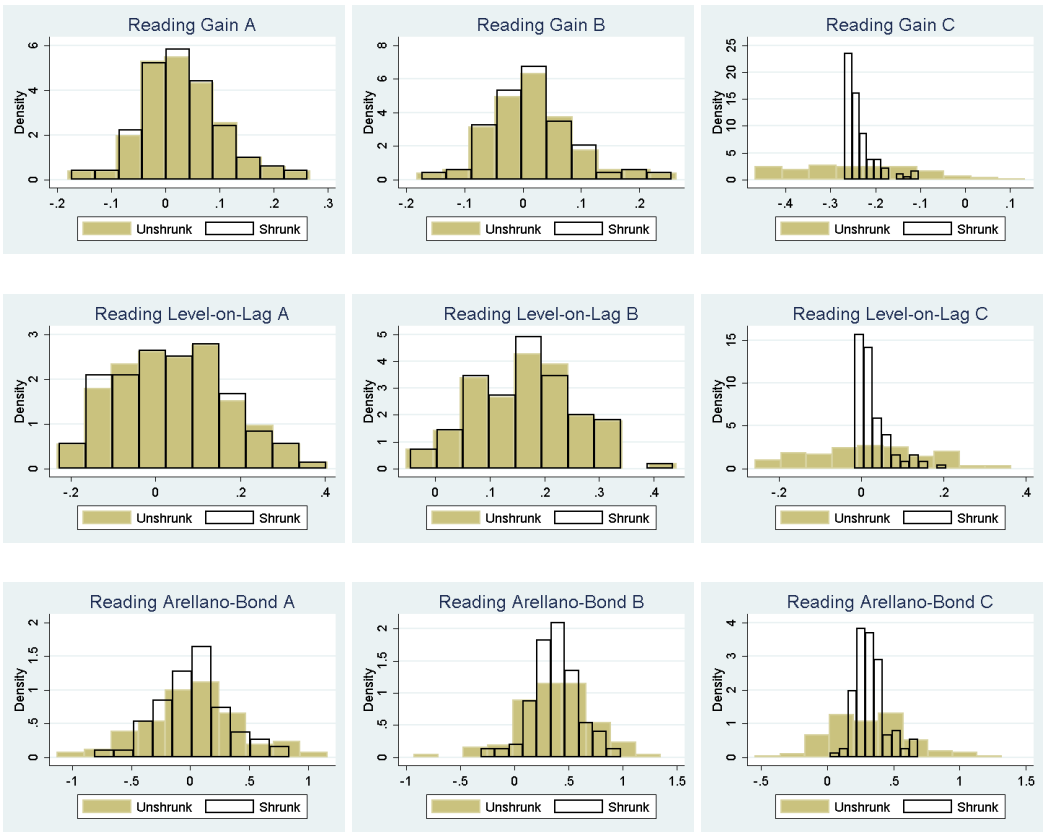


Figure 3.2. Histograms of Value-Added Coefficients, Reading

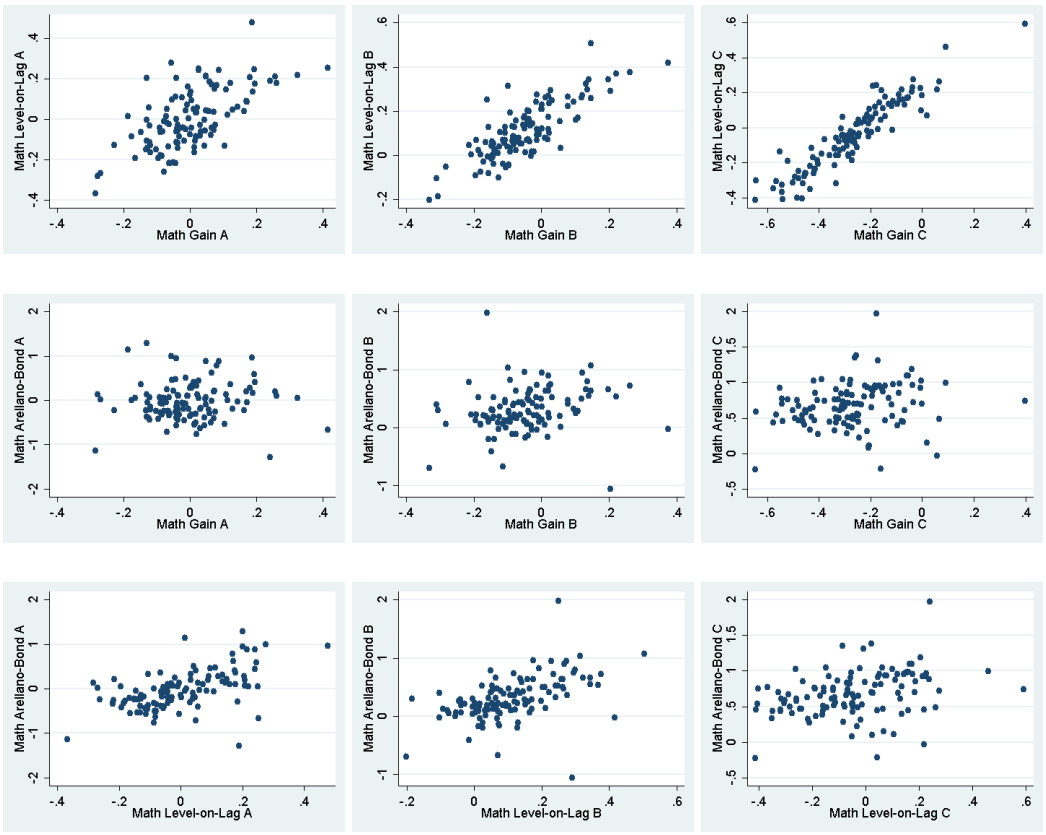


Figure 3.3. Scatterplots of Value-Added Coefficients, Math

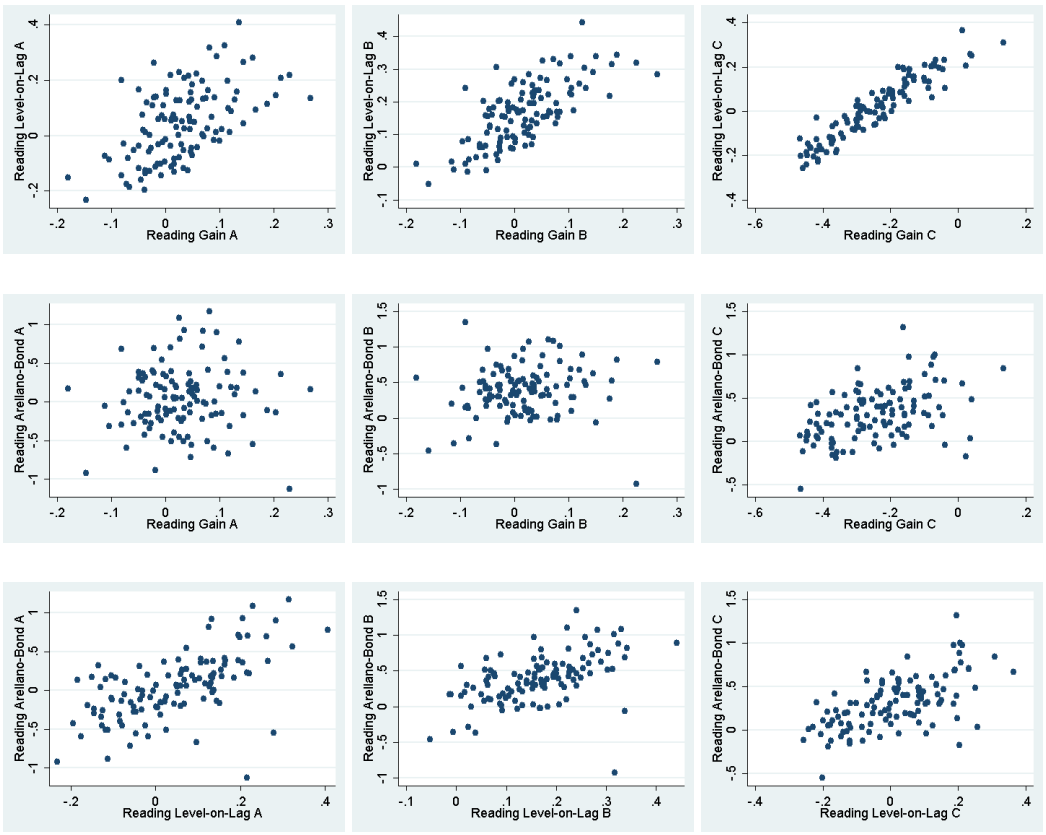


Figure 3.4. Scatterplots of Value-Added Coefficients, Reading

Appendix A

Combined Appendix

A.1 Appendix to Chapter 1

A.1.1 Index Score

To help protect student anonymity, the UCOP dataset does not report students' exact high school GPA or SAT scores, but instead reports them within a bin range. Unfortunately, these bins are somewhat coarser than one would like to fully control for academic achievement. Raw high school GPA is reported in bins of 0.25, while SAT score is reported in bins of 50 points by category (except the top bin, which goes 700–800).

Much of this problem can ultimately be resolved using an additional measure of student achievement, known as the Index score. Index score is a composite measure of student achievement, computed by UCOP as a linear combination of high school GPA and SAT scores. Unlike the individual components, this Index score is recorded precisely in the data. It is thus able to detect fine differences in achievement that are invisible to the GPA and SAT bins. Index score thus serves as the primary academic control in all regressions, and is included using a cubic polynomial for flexibility.

Figure A1 shows a histogram of Index score among applicants from California public high schools.

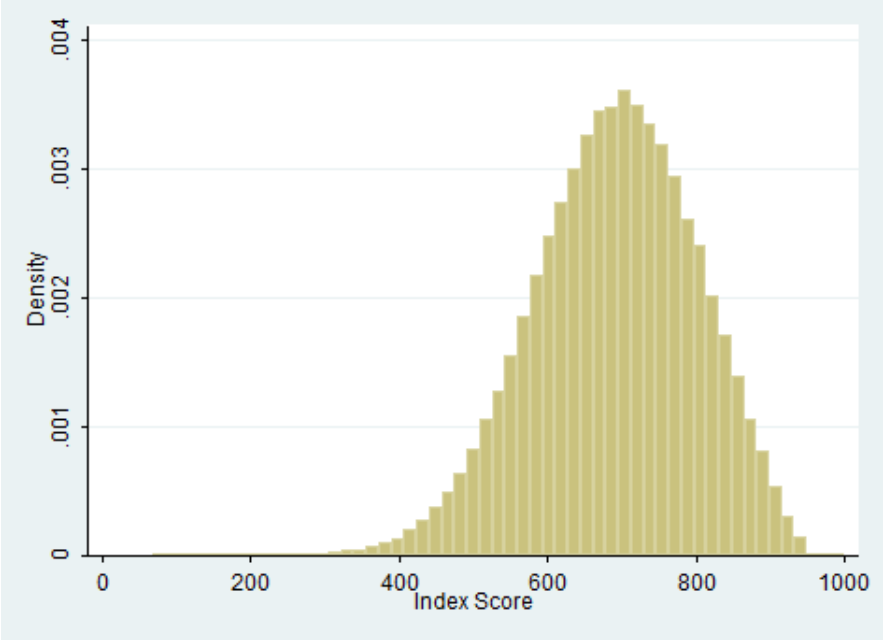


Figure A1. Histogram of Index Score

Index score has a maximum value of 1000, with up to 400 points coming from high school GPA, and 300 points each from SATI Math and Verbal scores. Regressing Index score on continuous approximations (provided by UCOP) of high school GPA and SAT scores, the approximate formula is:

$$\text{Index Score} \approx -165 + 90 \cdot \text{HSGPA} + 0.48 \cdot \text{SAT Math} + 0.48 \cdot \text{SAT Verbal}. \quad \text{(A1)}$$

One point to keep in mind, when using this measure, is that not all combinations that achieve the same Index score are necessarily equally effective for admissions. (In practice, most UC campuses seem to prefer students with higher GPA and lower SAT score than vice versa.) To help address this issue, all regressions include dummies for the bin value of high school GPA and SAT score (by category), along with the cubic polynomial of Index score.

A.1.2 Standard Errors

Due to various similarities in experience between students of the same high school, intuitively one wishes to group error terms at the high school level. This clustering cannot be achieved, however, since individual high schools are not indicated in the data. Error terms are instead clustered by student in pooled regressions, and currently unclustered for individual campuses. While a possible alternative is to cluster by API decile, that choice is also inappropriate, as ten deciles is too few to invoke asymptotic properties, and also encounters challenges using cluster-bootstrap techniques (Cameron, Gelbach and Miller, 2008).

This section instead computes the degree of within-cluster error correlation needed to make results in Table 3 insignificant based on the Moulton (1986) factor. For the case of simple regression, the OLS standard error should be inflated by a factor of

$$\sqrt{1 + \rho_x \rho_u (var(N_g)/\bar{N}_g + \bar{N}_g - 1)} \tag{A2}$$

where ρ_x denotes within-cluster correlation of x, ρ_u denotes within-cluster error correlation, N_g is the number of observations in cluster g, and $\bar{N}_g$ is the average cluster size. While this result is inexact with more than one regressor, it can still provide a reasonable approximation (Moulton, 1990). Further, although this result is for OLS, here the z statistics on average marginal effects obtained using a probit regression are very similar to those obtained under OLS (see Table A4 for OLS results).

Treating individual high schools as clusters, $\rho_x \approx 1$, as except for high schools right near decile cutoffs, which may finish above or below the cutoff in a given year, API decile is mostly stable from year-to-year. Next, based on supplemental tabulations from UCOP, a total of 1,276 California public high schools had at least one student applying to UC colleges between 2001–2006. Assuming this count also holds true for individual

campuses, and that the proportional composition by high school is also constant across campuses, one can plug in approximate values of $\bar{N}_g$ and $var(N_g)$, and thereby compute the minimum value of ρ_u that renders results insignificant.

For example, the average marginal effect for UC Berkeley has a z statistic of -21.60 in Table 3. Significance at the 5% level disappears if standard errors are understated 11.02 times relative to the truth. Under the stated assumptions, UC Berkeley has an average cluster size of 116.33 with a variance of 24,010.8. Significance at the 5% level disappears if $\sqrt{1 + \rho_u(24,010.8/116.33 + 116.33 - 1)} \geq 11.02$. This occurs when $\rho_u \geq 0.374$. That is, within-high school error correlation of at least 0.374 renders results insignificant for UC Berkeley in Table 3.

Table A1 shows the required value of ρ_u for each campus. As seen, even a small degree of within-high school error correlation makes results insignificant at UC Irvine and UC Riverside. However, these are both campuses where the average marginal effect is currently estimated to be positive. Loss of significance at these campuses does not substantively affect results. For campuses where the estimated average marginal effect is negative, ρ_u needs to be at least 0.276 before significance disappears.

Table A1. Within-High School Error Correlation Thresholds

	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside
z	-21.60	-29.69	-29.63	$+3.55$	-57.85	-18.01	$+1.95$	$+6.06$
Required Inflation on S.E.	11.02	15.15	15.12	1.81	29.52	9.19	N/A	3.09
Average Cluster Size	116.33	145.47	144.17	126.78	122.77	109.48	81.28	85.77
Var(Cluster Size)	24,010.8	37,545.7	36,876.3	28,516.5	28,966.7	21,266.0	11,720.0	13,053.5
Minimum ρ_u	0.374	0.568	0.570	0.007	> 1	0.276	N/A	0.036

Notes: Table shows the required inflation on standard errors for the coefficient on API decile to become insignificant at the 5% level in Table 3. No threshold is computed for UC Santa Cruz, whose z statistic is not significant at the 5% level.

Table A2. Descriptive Statistics, Alternate View

Variable	N	Mean	S.D.	Min	Max
API Decile	320,721	6.99	2.84	1	10
Eligibility in the Local Context	320,721	0.19	0.39	0	1
Index Score	320,443	690	109	67	1000
Raw High School GPA[a]	318,208	3.38	0.47	2.50	3.87
SATI Math[b]	315,420	597	96	400	724
SATI Verbal[b]	315,415	564	98	400	724
SATII Writing[b]	309,509	569	100	400	724
SATII Third Subject Score[b]	290,579	606	105	400	724
Number of AP Courses (in Semesters)	313,749	7.20	5.31	0	90
Parental Income Percentile[c]	255,275	68.4	29.1	6.1	96.3
Parental Education Percentile[d]	306,896	65.4	30.4	2.3	93.7
White	320,721	0.35	0.48	0	1
Asian	320,721	0.35	0.48	0	1
Underrepresented Minority	320,721	0.23	0.42	0	1
Other/Unknown Race	320,721	0.08	0.26	0	1
Science Major	320,391	0.32	0.47	0	1
Social Science Major	320,391	0.15	0.36	0	1
Other Major	320,391	0.53	0.50	0	1
Mean SAT of Applied Campuses	320,391	1205	39	1069	1296
Applied to Berkeley	320,721	0.47	0.50	0	1
Applied to Davis	320,721	0.44	0.50	0	1
Applied to Irvine	320,721	0.51	0.50	0	1
Applied to Los Angeles	320,721	0.58	0.49	0	1
Applied to Riverside	320,721	0.36	0.48	0	1
Applied to San Diego	320,721	0.58	0.49	0	1
Applied to Santa Barbara	320,721	0.51	0.50	0	1
Applied to Santa Cruz	320,721	0.33	0.47	0	1
Cumulative College GPA[e]	154,006	2.99	0.60	1.05	3.95

Notes: This table uses continuous approximations, provided by UCOP, for variables in bins. Values are typically coded at bin midpoints, with some exceptions for top and bottom end bins.

a – Recorded in bins of 0.00-2.99; 3.00-3.24; 3.25-3.49; 3.50-3.74; 3.75-4.00
b – Recorded in bins of 200-449; 450-499; 500-549; 550-599; 600-649; 650-699; 700-800
c – Parental income is originally recorded in bins of $0-9,999; $10,000-19,999; $20,000-29,999; ...; $90,000-99,999; and $100,000+ per year. These have been converted to percentiles using the 2000 U.S. Census.
d – Parental education is originally recorded in bins of no high school; some high school; high school graduate; some college; 2-year college graduate; 4-year college graduate; and postgraduate study. These have been converted to percentiles using the 2000 U.S. Census.
e – Recorded in bins of 0–1.09; 1.10–1.19; 1.20–1.29; ...; 3.90–4.00.

Table A3. Selected Means by UC Campus

Applicants

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income %-ile	Underrep. Minority
Berkeley	149,940	0.27	0.12	7.21	3.51	1217	0.26	68.4	0.20
Los Angeles	185,620	0.26	0.11	7.08	3.46	1194	0.24	67.3	0.22
San Diego	185,893	0.44	0.11	7.35	3.45	1197	0.22	70.0	0.18
Irvine	162,732	0.58	0.13	7.04	3.36	1157	0.19	66.3	0.21
Santa Barbara	163,773	0.54	0.11	7.30	3.36	1163	0.15	71.1	0.21
Davis	141,992	0.63	0.16	7.37	3.40	1173	0.17	70.8	0.17
Santa Cruz	105,203	0.79	0.14	7.27	3.28	1142	0.10	69.9	0.21
Riverside	116,222	0.88	0.15	6.64	3.22	1097	0.10	64.4	0.29

Admits

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income %-ile	Underrep. Minority
Berkeley	39,964	1	0.45	7.23	3.80	1312	0.60	69.7	0.17
Los Angeles	49,031	1	0.42	7.29	3.75	1309	0.50	68.1	0.17
San Diego	81,451	1	0.25	7.38	3.72	1280	0.44	69.0	0.15
Irvine	94,017	1	0.22	7.38	3.58	1232	0.32	70.9	0.16
Santa Barbara	87,764	1	0.21	7.13	3.59	1220	0.26	70.6	0.20
Davis	88,138	1	0.26	7.36	3.58	1223	0.26	69.9	0.16
Santa Cruz	82,265	1	0.18	7.39	3.41	1169	0.12	71.4	0.19
Riverside	95,903	1	0.18	6.84	3.30	1120	0.11	66.4	0.26

Attending Students

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income %-ile	Underrep. Minority
Berkeley	17,966	1	1	7.18	3.77	1289	0.53	68.6	0.16
Los Angeles	20,829	1	1	7.04	3.69	1267	0.41	65.4	0.18
San Diego	20,559	1	1	7.39	3.64	1245	0.32	67.4	0.13
Irvine	20,868	1	1	7.59	3.43	1191	0.17	72.6	0.14
Santa Barbara	18,815	1	1	6.95	3.47	1171	0.15	72.1	0.24
Davis	23,320	1	1	7.11	3.46	1164	0.17	66.8	0.15
Santa Cruz	14,493	1	1	7.40	3.25	1139	0.05	73.1	0.19
Riverside	17,679	1	1	6.34	3.12	1067	0.08	62.6	0.32

Notes: N reports the total number of applicants, admits, or attending students from California public high schools between 2001 and 2006. Individual students appear once at each campus applied or admitted.

Table A4. OLS Admission Results by Campus

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	−0.008***	−0.010***	−0.011***	−0.002***	−0.027***	−0.012***	−0.001	0.001***	−0.008***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0004)	(0.0002)
ELC Status	0.158***	0.074***	0.175***	0.182***	0.042***	0.064***	−0.012***	0.011***	0.108***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	148,439	185,620	183,958	161,768	163,040	139,697	103,707	109,448	1,195,677
R-squared	0.469	0.442	0.579	0.573	0.519	0.480	0.481	0.413	0.515
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

Table A5. Probit Average Marginal Effects with API Decile as Spline

Panel A: Single Knot at 5.5

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Deciles 1–5	-0.018***	-0.027***	-0.017***	-0.006***	-0.019***	-0.030***	-0.003***	0.003***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
API Deciles 6–10	-0.003***	-0.000	-0.007***	0.006***	-0.026***	0.004***	0.003***	0.001**	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.594	0.584	0.519	0.479	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Panel B: Two Knots at 4 and 7

API Deciles 1–4	-0.019***	-0.024***	-0.008***	-0.011***	-0.007***	-0.021***	-0.001	0.003**	-0.013***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
API Deciles 4–7	-0.011***	-0.019***	-0.020***	0.002**	-0.034***	-0.024***	-0.003**	0.002**	-0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
API Deciles 7–10	-0.001	0.005***	-0.002***	0.006***	-0.020***	0.013***	0.005***	0.002*	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.595	0.584	0.520	0.479	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Average Marginal Effects for Probit Regression of Yield, Each API Decile

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
1st API Decile	-0.040***	-0.065***	-0.109***	-0.063***	-0.056***	-0.115***	-0.021***	-0.049***	-0.068***
	(0.015)	(0.013)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.003)
2nd API Decile	-0.071***	-0.055***	-0.064***	-0.077***	-0.064***	-0.088***	-0.042***	-0.000	-0.053***
	(0.014)	(0.012)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.002)
3rd API Decile	-0.043***	-0.009	-0.049***	-0.056***	-0.042***	-0.067***	-0.029***	0.001	-0.033***
	(0.013)	(0.011)	(0.008)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.002)
4th API Decile	-0.040***	-0.011	-0.038***	-0.044***	-0.040***	-0.031***	-0.029***	-0.004	-0.027***
	(0.013)	(0.011)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.002)
5th API Decile	-0.023*	-0.011	-0.026***	-0.013**	-0.031***	-0.021***	-0.007	0.010	-0.012***
	(0.012)	(0.011)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.002)
6th API Decile	0.014	-0.004	-0.005	-0.014**	-0.017***	0.010	-0.004	-0.010*	-0.004*
	(0.012)	(0.011)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.002)
7th API Decile	—	—	—	—	—	—	—	—	—
8th API Decile	0.005	0.027***	0.011*	0.008	0.006	0.003	0.012**	-0.007	0.006***
	(0.011)	(0.010)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.002)
9th API Decile	-0.008	0.041***	0.006	0.038***	0.020***	-0.012*	-0.006	-0.023***	0.004**
	(0.010)	(0.009)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.002)
10th API Decile	0.001	0.011	0.001	0.013***	-0.020***	-0.006	-0.022***	-0.043***	-0.015***
	(0.009)	(0.008)	(0.006)	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.002)
ELC Status	-0.037***	-0.041***	-0.025***	-0.042***	-0.039***	-0.004	-0.011**	0.013***	-0.036***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.001)
Observations	39,964	49,031	81,447	94,014	87,764	88,138	82,264	95,901	618,530
Pseudo R-squared	0.074	0.140	0.121	0.172	0.130	0.145	0.259	0.114	0.124
Mean Dep Var	0.450	0.425	0.252	0.222	0.214	0.265	0.176	0.184	0.250

Notes: The dependent variable is enrollment, conditional on having been admitted. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1

A.2 Appendix to Chapter 2

Table A7. The Relationship Between Winning a Lottery and Attrition

	Magnet	VEEP	Choice	Pooled
On Track in Grade 9	0.007	-0.015	0.000	-0.001
	(0.010)	(0.013)	(0.013)	(0.007)
Passed Exit Exam in Grade 10	0.002	-0.006	-0.013	-0.004
	(0.009)	(0.015)	(0.011)	(0.007)
% Days Attended in Grade 12	-0.010	-0.016	-0.004	-0.010
	(0.015)	(0.024)	(0.020)	(0.011)
Grade 12 Citizenship Grade	-0.001	-0.004	-0.016	-0.006
	(0.015)	(0.022)	(0.019)	(0.011)
GPA in Grade 12	-0.004	-0.008	-0.011	-0.007
	(0.016)	(0.021)	(0.019)	(0.011)
Graduated on Time	0.002	-0.015	0.007	-0.002
	(0.009)	(0.020)	(0.016)	(0.008)
# Years Postsecondary 1 Year after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
# Years in 2-Year College 1 Year after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
# Years in 4-Year College 1 Year after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
# Years Postsecondary 4 Years after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
# Years in 2-Year College 4 Years after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
# Years in 4-Year College 4 Years after Graduation	-0.021	-0.025	0.005	-0.015
	(0.012)	(0.017)	(0.016)	(0.008)
Associate's or Other 2-Year College Certification	-0.029*	-0.021	-0.004	-0.020*
4 Years after Graduation	(0.013)	(0.017)	(0.016)	(0.009)
Associate's or Other 2-Year College Certification	-0.027*	-0.015	-0.012	-0.020*
5 Years after Graduation	(0.013)	(0.017)	(0.014)	(0.008)
Associate's or Other 2-Year College Certification	-0.023	-0.014	-0.007	-0.016*
6 Years after Graduation	(0.013)	(0.017)	(0.013)	(0.008)
Bachelor's Degree 4 Years after Graduation	-0.024	-0.019	-0.000	-0.016
	(0.014)	(0.017)	(0.014)	(0.009)
Bachelor's Degree 5 Years after Graduation	-0.022	-0.013	-0.007	-0.016
	(0.013)	(0.017)	(0.012)	(0.008)
Bachelor's Degree 6 Years after Graduation	-0.019	-0.012	-0.003	-0.013
	(0.012)	(0.017)	(0.012)	(0.008)

Notes: For each outcome of interest, an indicator for "missing" is used as the outcome and regressed on a dummy for winning the lottery plus lottery fixed effects. This table reports the coefficients on 'win lottery' from these regressions, along with their standard errors. Positive significant results indicate that lottery winners have greater attrition for the outcome of interest, while negative results indicate that lottery losers have greater attrition. Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table A8. More Detailed Version of Intent to Treat Estimates, High School Outcomes

Magnet

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Won Lottery	-0.00758	0.000665	-0.174	-0.0198	-0.0160	0.0254
	(0.0123)	(0.0174)	(0.314)	(0.0372)	(0.0327)	(0.0228)
Observations	3,476	3,885	2,865	2,979	3,062	3,662
R-squared	0.160	0.298	0.026	0.089	0.126	0.058
Number of Clusters	78	90	96	98	99	99

VEEP

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Won Lottery	-0.0114	-0.0136	-0.0250	0.0839	0.0276	0.0136
	(0.0264)	(0.0235)	(0.526)	(0.0553)	(0.0400)	(0.0237)
Observations	2,092	2,296	1,502	1,538	1,608	2,002
R-squared	0.134	0.337	0.011	0.068	0.089	0.091
Number of Clusters	41	49	49	49	50	50

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table A8. More Detailed Version of Intent to Treat Estimates, High School Outcomes, Continued

Choice

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Won Lottery	0.0241	0.0148	-0.460	-0.0472	0.00494	-0.00602
	(0.0344)	(0.0224)	(0.318)	(0.0428)	(0.0379)	(0.0176)
Observations	1,654	1,888	1,411	1,471	1,509	1,744
R-squared	0.133	0.239	0.017	0.123	0.178	0.051
Number of Clusters	34	39	46	46	46	46

Pooled

	(1) On Track in Grade 9	(2) Passed Exit Exam in Grade 10	(3) % Days Attended in Grade 12	(4) Grade 12 Citizenship Grade	(5) GPA in Grade 12	(6) Graduated on Time
Won Lottery	-0.000268	0.00207	-0.201	-0.000635	0.00104	0.0151
	(0.0127)	(0.0120)	(0.220)	(0.0264)	(0.0215)	(0.0134)
Observations	7,222	8,069	5,778	5,988	6,179	7,408
R-squared	0.142	0.287	0.019	0.088	0.126	0.063
Number of Clusters	153	178	191	193	195	195

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table A9. More Detailed Version of Intent to Treat Estimates, Postsecondary Enrollment Outcomes

Magnet

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Won Lottery	-0.000545	-0.0108	-0.00427	-0.0164	0.0244	-0.0804
	(0.0212)	(0.0246)	(0.0228)	(0.0590)	(0.0506)	(0.0661)
Observations	2,978	2,978	2,978	2,978	2,978	2,978
R-squared	0.034	0.011	0.097	0.058	0.010	0.122
Number of Clusters	96	96	96	96	96	96

VEEP

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Won Lottery	0.0185	0.0437	-0.0449	0.00397	0.131	-0.0786
	(0.0253)	(0.0322)	(0.0250)	(0.0772)	(0.0952)	(0.0882)
Observations	1,516	1,516	1,516	1,516	1,516	1,516
R-squared	0.038	0.003	0.095	0.063	0.005	0.113
Number of Clusters	46	46	46	46	46	46

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table A9. More Detailed Version of Intent to Treat Estimates, Postsecondary Enrollment Outcomes, Continued

Choice

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Won Lottery	0.0294	0.0471	-0.0216	0.120	0.145	-0.0276
	(0.0167)	(0.0253)	(0.0237)	(0.0829)	(0.0793)	(0.0708)
Observations	1,564	1,564	1,564	1,564	1,564	1,564
R-squared	0.029	0.063	0.150	0.060	0.060	0.176
Number of Clusters	44	44	44	44	44	44

Pooled

	(1) # Years Enrolled in Postsecondary 1 Year After HS Graduation	(2) # Years Enrolled in 2-Year College 1 Year After HS Graduation	(3) # Years Enrolled in 4-Year College 1 Year After HS Graduation	(4) # Years Enrolled in Postsecondary 4 Years After HS Graduation	(5) # Years Enrolled in 2-Year College 4 Years After HS Graduation	(6) # Years Enrolled in 4-Year College 4 Years After HS Graduation
Won Lottery	0.0122	0.0181	-0.0185	0.0256	0.0838*	-0.0656
	(0.0125)	(0.0155)	(0.0144)	(0.0416)	(0.0397)	(0.0438)
Observations	6,058	6,058	6,058	6,058	6,058	6,058
R-squared	0.033	0.014	0.111	0.058	0.012	0.135
Number of Clusters	186	186	186	186	186	186

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

Table A10. More Detailed Version of Intent to Treat Estimates, Postsecondary Graduation Outcomes

Magnet

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Won Lottery	0.0163*	0.0166*	0.0131	-0.00234	-0.0196	-0.0505*
	(0.00701)	(0.00716)	(0.0106)	(0.0107)	(0.0148)	(0.0213)
Observations	2,343	2,248	1,805	2,363	2,268	1,829
R-squared	0.014	0.017	0.011	0.064	0.101	0.114
Number of Clusters	89	80	69	89	80	74

VEEP

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Won Lottery	0.0226**	0.0338**	0.0170	0.0171	-0.00737	0.0106
	(0.00786)	(0.0120)	(0.0144)	(0.0105)	(0.0172)	(0.0234)
Observations	1,219	1,127	1,076	1,232	1,140	1,089
R-squared	0.008	0.016	0.014	0.049	0.080	0.089
Number of Clusters	44	44	41	44	44	41

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table A10. More Detailed Version of Intent to Treat Estimates, Postsecondary Graduation Outcomes, Continued

Choice

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Won Lottery	0.0138	0.0217	0.0183	-0.0437*	-0.00793	0.00905
	(0.0177)	(0.0199)	(0.0278)	(0.0196)	(0.0282)	(0.0273)
Observations	1,186	1,081	1,000	1,213	1,108	1,027
R-squared	0.004	0.007	0.004	0.131	0.141	0.152
Number of Clusters	41	40	38	41	40	38

Pooled

	(1) Associate or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Associate or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Associate or Other 2-Year College Cert. 6 Years After HS Graduation	(4) Bachelor's 4 Years After HS Graduation	(5) Bachelor's 5 Years After HS Graduation	(6) Bachelor's 6 Years After HS Graduation
Won Lottery	0.0170**	0.0219**	0.0157	-0.00657	-0.0135	-0.0193
	(0.00587	(0.00643)	(0.00879)	(0.00799)	(0.0105)	(0.0142)
Observations	4,748	4,456	3,881	4,808	4,516	3,945
R-squared	0.007	0.011	0.008	0.080	0.108	0.117
Number of Clusters	174	164	148	174	164	153

Notes: Robust standard errors are in parentheses, clustered by lottery. ** $p<0.01$, * $p<0.05$.

Table A11. Intent to Treat Estimates on Postsecondary Graduation Outcomes with 2-Year and 4-Year Degrees Combined

		(1) Bachelor's or Associate's or Other 2-Year College Cert. 4 Years After HS Graduation	(2) Bachelor's or Associate's or Other 2-Year College Cert. 5 Years After HS Graduation	(3) Bachelor's or Associate's or Other 2-Year College Cert. 6 Years After HS Graduation
Magnet	Win	0.0138	-0.00243	-0.0337
		(0.0142)	(0.0159)	(0.0239)
VEEP	Win	0.0370**	0.0218	0.0226
		(0.0110)	(0.0180)	(0.0265)
Choice	Win	-0.0288	0.00704	0.0180
		(0.0167)	(0.0247)	(0.0291)
Pooled	Win	0.00992	0.00601	-0.00546
		(0.00927)	(0.0107)	(0.0155)

Notes: Robust standard errors are in parentheses, clustered by lottery. ** p<0.01, * p<0.05.

A.3 Appendix to Chapter 3

A.3.1 Arellano-Bond Estimation

Roodman (2009) provides a good overview of dynamic panel bias and techniques that have developed to address this issue. We explain the basics below, and then provide additional details about how we estimate regressions using an Arellano and Bond (1991) estimator.

Dynamic panel bias arises in equations that feature both fixed effects and lagged dependent variables as regressors (Nickell, 1981). A simple such equation may take the form

$$y_{it} = \rho^* y_{i,t-1} + x'_{it}\beta^* + \mu_i + u_{it}, \tag{A3}$$

where μ_i is an unobserved heterogeneity factor for person i, and $\epsilon^*_{it} = \mu_i + u_{it}$ is a composite error term.

The key issue is that the lagged dependent variable, $y_{i,t-1}$ is mechanically an endogenous variable. If Equation A3 is estimated using a conventional fixed effects estimator, any unmodeled shocks affect the *estimated* value of the fixed effect μ_i, creating a negative correlation between $y_{i,t-1}$ and ϵ^*_{it}, and returning downward-biased estimates of ρ^*. This problem is worse the smaller is T: the number of periods in the panel.

(It is worth remembering that in our context the key interest is not on the autoregressive parameter, ρ, but on the school value-added coefficients. Nevertheless, dynamic panel bias remains a possible concern, since biased estimates of ρ also affect the corresponding school value-added coefficients.)

Arellano and Bond (1991) propose a popular method to address dynamic panel bias by using using a first difference instrumental variables (FDIV) procedure. Equation A3 is transformed into a first-differenced version:

$$\Delta y_{it} = \rho^* \Delta y_{i,t-1} + \Delta x'_{it}\beta^* + \Delta u_{it} \tag{A4}$$

which no longer contains the person fixed effect, μ_i. Under appropriate conditions — most notably an assumption about no serial correlation in u_{it} — this equation can be consistently estimated using instrumental variables, with lagged values of regressors serving as the instruments.

Which lags are valid to use depend on what variables are exogenous, endogenous, or predetermined. Exogenous variables are assumed to be uncorrelated with past, present, and future disturbances. First differences of these variables are used as instruments in Equation A4. Predetermined variables are assumed to be uncorrelated with present and future disturbances, but may be correlated with past disturbances. Lagged values of one period or more of these variables are used as instruments in Equation A4. Endogenous

variables are possibly correlated with both present and past disturbances. Lagged values of two periods or more of these variables are used as instruments in Equation A4.

Table A12 shows how we classify variable in the full specification of the Arellano-Bond model in our paper.

Table A12. Variable Classification in Arellano-Bond Regressions

Variable	Classification	Lags Used
Lagged Test Score $S_{igs,t-1}$	Predetermined	1–2
School Dummies	Endogenous	2–3
Grade Dummies	Predetermined	1–2
Year Dummies	Exogenous	N/A
Female Dummy	Exogenous	N/A
Race Dummies	Exogenous	N/A
Parent Ed Dummies	Predetermined	1–2
English Learner Status†	Predetermined	1–2
Fluent English Proficient Status†	Predetermined	1–2
Special Education Status†	Predetermined	1–2
% Free or Reduced Price Lunch	Predetermined	1–2
% African-American	Predetermined	1–2
% Hispanic	Predetermined	1–2
% Asian/Pacific Islander	Predetermined	1–2
% Native American	Predetermined	1–2
% English Learner†	Predetermined	1–2
% Fluent English Proficient†	Predetermined	1–2

Notes: Table reports the classification of variables in the full specification of the Arellano-Bond model. Exogenous variables are assumed to be uncorrelated with past, present, and future errors. Predetermined variables are assumed to be uncorrelated with present and future errors, but possibly correlated with past errors. Endogenous variables may be correlated with either present or past errors. Variables marked with dagger (†) are based on students' status in their first appearance in the district.

Table A13. Spearman Rank Correlations of Shrunk Value-Added Coefficients Across Control Specifications

Gain Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.993	1		Control B	0.995	1	
Control C	0.490	0.540	1	Control C	0.568	0.589	1

Level-on-Lag Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.937	1		Control B	0.947	1	
Control C	0.889	0.770	1	Control C	0.916	0.857	1

Arellano-Bond Model

	Math				Reading		
	Control A	Control B	Control C		Control A	Control B	Control C
Control A	1			Control A	1		
Control B	0.811	1		Control B	0.832	1	
Control C	0.524	0.790	1	Control C	0.748	0.825	1

Table A14. Spearmank Rank Correlations of Shrunk Value-Added Coefficients Across Model Classes

Control Specification A

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.600	1		LevLag	0.530	1	
ABond	0.137	0.634	1	ABond	0.099	0.654	1

Control Specification B

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.761	1		LevLag	0.683	1	
ABond	0.348	0.620	1	ABond	0.194	0.594	1

Control Specification C

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Gain	1			Gain	1		
LevLag	0.922	1		LevLag	0.929	1	
ABond	0.202	0.326	1	ABond	0.420	0.598	1

Table A15. Relative Ranking of Schools Using Shrunk Estimates, Compared Across Models, for Math

Gain vs. Level-on-Lag

Sum Diagonal ≈ 46.5%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	**13.2**	5.3	0.9	0.9	0
Level-on-	2nd Quintile	4.4	**7.9**	5.3	1.8	0.9
Lag	3rd Quintile	1.8	4.4	**5.3**	7.9	0.9
Model	4th Quintile	0	1.8	8.8	**6.1**	3.5
	5th Quintile	0.9	0.9	0	3.5	**14.0**

Gain vs. Arellano-Bond

Sum Diagonal ≈ 25.4%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	4.4	5.3	**4.4**	4.4	1.8
Arellano-	2nd Quintile	7.0	**6.1**	3.5	3.5	0
Bond	3rd Quintile	5.3	2.6	**3.5**	3.5	5.3
Model	4th Quintile	1.8	4.4	**6.1**	3.5	4.4
	5th Quintile	1.8	1.8	2.6	**5.3**	7.9

Level-on-Lag vs. Arellano-Bond

Sum Diagonal ≈ 37.8%		Level-on-Lag Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	1st Quintile	7.9	**5.3**	5.3	0.9	0.9
Arellano-	2nd Quintile	7.0	**5.3**	5.3	2.6	0
Bond	3rd Quintile	4.4	3.5	**5.3**	4.4	2.6
Model	4th Quintile	0.9	4.4	4.4	**7.0**	3.5
	5th Quintile	0	1.8	0	5.3	**12.3**

Notes: Percent frequencies are displayed. Within each row, the column entry in which the median observation occurs is in bold. The 1st quintile refers to the bottom 20 percent, while the 5th quintile refers to the top 20 percent. Results in this table use control specification B for all models.

Table A16. Relative Ranking of Schools Using Shrunk Estimates, Compared Across Models, for Reading

Gain vs. Level-on-Lag

Sum Diagonal ≈ 41.2%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Level-on-Lag Model	1st Quintile	**12.3**	4.4	2.6	0.9	0
	2nd Quintile	3.5	6.1	**7.0**	3.5	0
	3rd Quintile	2.6	4.4	**5.3**	4.4	3.5
	4th Quintile	1.8	3.5	2.6	**7.0**	5.3
	5th Quintile	0	1.8	2.6	4.4	**10.5**

Gain vs. Arellano-Bond

Sum Diagonal ≈ 21.1%		Gain Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Arellano-Bond Model	1st Quintile	4.4	2.6	**6.1**	3.5	3.5
	2nd Quintile	7.0	**5.3**	3.5	0.9	3.5
	3rd Quintile	2.6	4.4	**3.5**	7.9	1.8
	4th Quintile	4.4	4.4	**3.5**	2.6	5.3
	5th Quintile	1.8	3.5	3.5	**5.3**	5.3

Level-on-Lag vs. Arellano-Bond

Sum Diagonal ≈ 38.6%		Level-on-Lag Model				
		1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Arellano-Bond Model	1st Quintile	8.8	**6.1**	2.6	1.8	0.9
	2nd Quintile	4.4	**7.0**	4.4	3.5	0.9
	3rd Quintile	4.4	4.4	**4.4**	5.3	1.8
	4th Quintile	2.6	2.6	**5.3**	6.1	3.5
	5th Quintile	0	0	3.5	3.5	**12.3**

Notes: Percent frequencies are displayed. Within each row, the column entry in which the median observation occurs is in bold. The 1st quintile refers to the bottom 20 percent, while the 5th quintile refers to the top 20 percent. Results in this table use control specification B for all models.

Table A17. Spearman Rank Correlations Between Shrunk Value-Added and Level of Test Score

	Math				Reading		
	Gain	LevLag	ABond		Gain	LevLag	ABond
Control A	-0.043	0.731	0.680	Control A	0.074	0.855	0.720
Control B	-0.012	0.530	0.464	Control B	0.106	0.706	0.505
Control C	0.655	0.793	0.224	Control C	0.657	0.789	0.628

Notes: Covariates change for the value-added models, but not for the levels model. The levels model includes only school, year, and grade fixed effects, with no additional covariates throughout. Shrinkage has been applied to the value-added coefficients, but not the coefficients of the levels model.

Figure A2. Scatterplots of Shrunk Value-Added Coefficients, Math

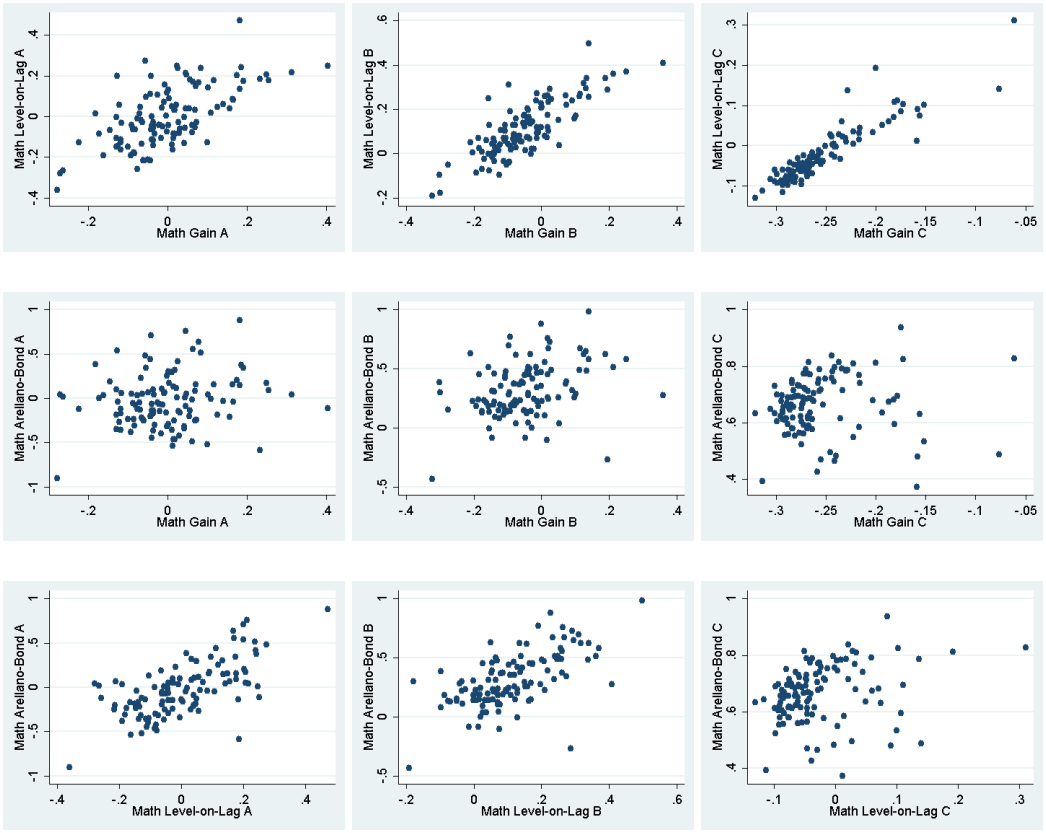
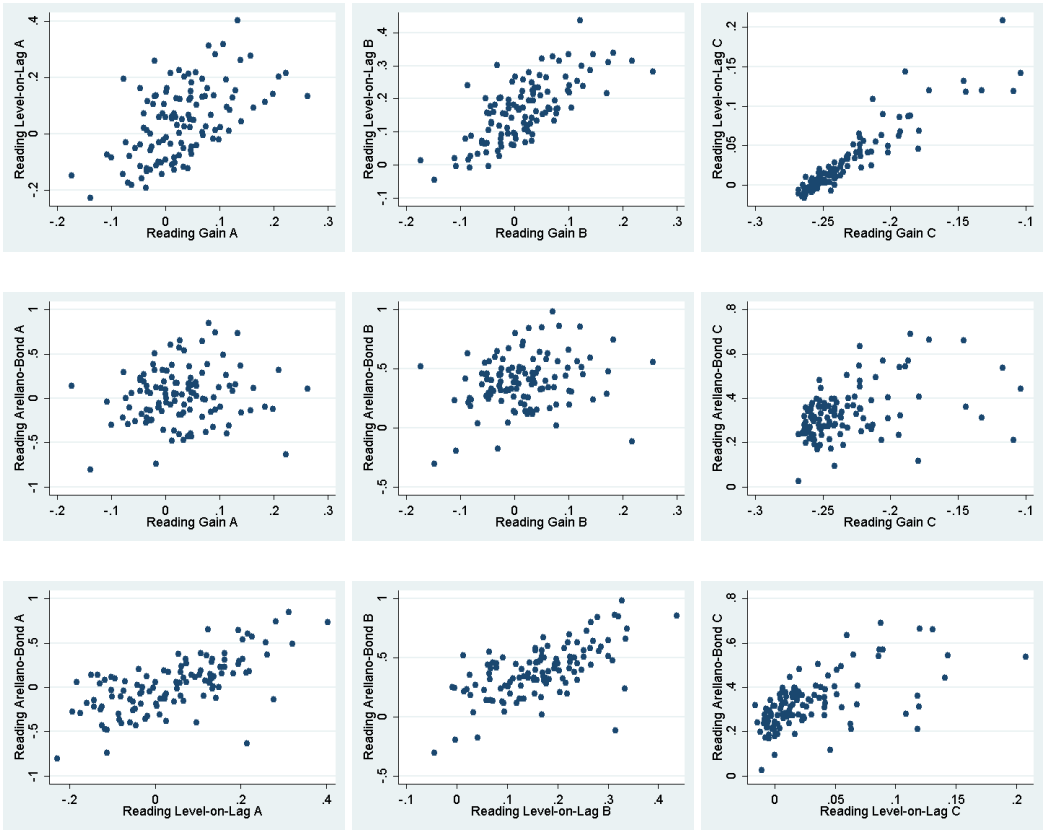


Figure A3. Scatterplots of Shrunk Value-Added Coefficients, Reading



Bibliography

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics." *American Economic Journal: Applied Economics*, 3(3): 29–54.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277–297.

Ballou, Dale. 2009. "Magnet School Outcomes." In *Handbook of Research on School Choice*. eds. by Mark Berends, Matthew G. Springer, Dale Ballou, and Herbert J. Walberg: Routledge, 409–426.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics*, 29(1): 37–65.

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. 1996. "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics*, 78(4): 672–685.

Berkowitz, Daniel, and Mark Hoekstra. 2011. "Does high school quality matter? Evidence from admissions data." *Economics of Education Review*, 30(2): 280–288.

Betts, Julian, Sami Kitmitto, Jesse Levin, Johannes Bos, and Marian Eaton. 2015. *What Happens When Schools Become Magnet Schools? A Longitudinal Study of Diversity and Achievement*. American Institutes for Research, Washington, D.C.

Betts, Julian R., and Darlen Morell. 1999. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *Journal of Human Resources*, 34(2): 268–293.

Betts, Julian R., Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel. 2006. *Does School Choice Work? Effects on Student Integration and Achieve-*

ment. Public Policy Institute of California, San Francisco.

Betts, Julian R., and Y. Emily Tang. 2014. "A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement."Technical report, Bothell, WA, working paper available at www.ncsrp.org.

Betts, Julian R., Andrew C. Zau, and Lorien A. Rice. 2003. *Determinants of Student Achievement: New Evidence from San Diego*. Public Policy Institute of California, San Francisco.

Black, Dan A., and Jeffrey A. Smith. 2006. "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics*, 24(3): 701–728.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 114(2): 577–599.

Black, Sandra E., Jane Arnold Lincove, Jenna Cullinane, and Rachel Veron. 2015. "Can You Leave High School Behind?" *Economics of Education Review*, 46 52–63.

Bloom, Howard S. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, 8(2): 225–246.

Bradley, Laurence A., and Gifford W. Bradley. 1977. "The Academic Achievement of Black Students in Desegregated Schools: A Critical Review." *Review of Educational Research*, 47(3): 399–449.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. 1999. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *Journal of Human Resources*, 34(1): 104–123.

Buddin, Richard. 2010. "How effective are Los Angeles elementary teachers and schools?", unpublished manuscript.

California Department of Education. 2012. "2011-12 Academic Performance Index Reports Information Guide." Retrieved January 3, 2013, from http://www.cde.ca.gov/ta/ac/ap/documents/infoguide12.pdf.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics*, 90(3): 414–427.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urqiola. 2005. "The Central Role

of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review*, 95(4): 1237–1258.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Cook, Thomas, David Armor, Robert Crain, Norman Miller, Walter Stephan, Herbert Walberg, and Paul Wortman. 1984. *School Desegregation and Black Achievement*. National Institute of Education, Washington, D.C.

Cortes, Kalena E., and Andrew I. Friedson. 2014. "Ranking Up by Moving Out: The Effect of the Texas Top 10% Plan on Property Values." *National Tax Journal*, 67(1): 51–76.

Crain, Robert L., and Rita E. Mahard. 1978. "Desegregation and Black Achievement: A Review of the Research." *Law and Contemporary Problems*, 42(3): 17–56.

Crain, Robert L., and Rita E. Mahard. 1981. "Minority Achievement: Policy Implications of Research." In *Effective School Desegregation: Equity, Quality and Feasibility*. ed. by Willis D. Hawley, Beverly Hills: Sage Publications.

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt. 2006. "The effect of school choice on participants: evidence from randomized lotteries." *Econometrica*, 74(5): 1191–1230.

Cullen, Julie Berry, Mark C. Long, and Randall Reback. 2013. "Jockeying for position: strategic high school choice under Texas' top ten percent plan." *Journal of Public Economics*, 97 32–48.

Cushing, Matthew J., and Mary G. McGarvey. 2004. "Sample Selection in Models of Academic Performance." *Economic Inquiry*, 42(2): 319–322.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics*, 117(4): 1491–1527.

Dale, Stacy, and Alan B. Krueger. 2011. "Estimating the Return to College Selectivity over the Career Using Administrative Earning Data." NBER Working Paper 17159.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American*

Economic Review, 104(3): 991–1013.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach. 2013. "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion." *Journal of Policy Analysis and Management*, 32(34): 692–717.

Egalite, Anna J., and Patrick J. Wolf. 2016. "A Review of the Empirical Research on Private School Choice." *Peabody Journal of Education*, 91(4): 441–454.

Epple, Dennis, Richard Romano, and Miguel Urquiola. 2015a. "School Vouchers: A Survey of the Economics Literature." NBER Working Paper 21253.

Epple, Dennis, Richard Romano, and Ron Zimmer. 2015b. "Charter Schools: A Survey of Research on Their Characteristics and Effectiveness." NBER Working Paper 21256.

Espenshade, Thomas J., Lauren E. Hale, and Chang Y. Chung. 2005. "The Frog Pond Revisited: High School Academic Context, Class Rank, and Elite College Admission." *Sociology of Education*, 78(4): 269–293.

Figlio, David. 2009. "Voucher Outcomes." In *Handbook of Research on School Choice*. eds. by Mark Berends, Matthew G. Springer, Dale Ballou, and Herbert J. Walberg: Rougledge, 321–338.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. "The Evaluation of Charter School Impacts: Final Report."Technical Report NCEE 2010-4029, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Grady, S., and S. Bielick. 2010. "Trends in the Use of School Choice: 1993 to 2007."Technical Report NCES 2010-004, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2015. "Can Value-Added Measures of Teacher Performance Be Trusted?" *Education Finance and Policy*, 10(1): 117–156.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice." Unpublished.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica*,

47(1): 153–161.

Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics*. eds. by Orley Ashlenfelter, and David Card, 3A, Amsterdam, North Holland 1865–2097.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects Of Cognitive and Noncognitive Abilities On Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411–482.

Hoekstra, Mark. 2009. "The effect of attending the flagship state university on earnings: A discontinuity-based approach." *Review of Economics and Statistics*, 91(4): 717–724.

Imberman, Scott A., and Michael F. Lovenheim. 2016. "Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added." *Journal of Urban Economics*, 91 104–121.

Kane, Thomas J., and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4): 91–114.

Koedel, Cory, Julian R. Betts, Lorien A. Rice, and Andrew C. Zau. 2009. "The Integrating and Segregating Effects of School Choice." *Peabody Journal of Education*, 84(2): 110–129.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. "Value-added modeling: A review." *Economics of Education Review*, 47 180–195.

Lefgren, Lars, and David Sims. 2012. "Using Subject Test Scores Efficiently to Predict Teacher Value-Added." *Educational Evaluation and Policy Analysis*, 34 109–121.

Lockwood, J.R., Daniel F. McCaffrey, Laura S. Hamilton, Brian Stecher, Vi-Nhuan Le, and Jose-Felipe Martinez. 2007. "The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures." *Journal of Education Measurement*, 44(1): 47–67.

Long, Mark C. 2007. "Affirmative Action and Its Alternatives in Public Universities: What Do We Know?" *Public Administration Review*, 67(2) 315–330.

Long, Mark C., Victor Saenz, and Marta Tienda. 2010. "Policy Transparency and College Enrollment: Did the Texas Top Ten Percent Law Broaden Access to the Public

Flagships?" *ANNALS of the American Academy of Political and Social Science*, 627(1): 82–105.

Long, Mark C., and Marta Tienda. 2010. "Changes in Texas universities' applicant pools after the *Hopwood* decision." *Social Science Research*, 39(1): 48–66.

McCaffrey, Daniel F., J.R. Lockwood, Daniel M. Koretz, Thomas A. Louis, and Laura Hamilton. 2004. "Models for Value-Added Modeling of Teacher Effects." *Journal of Educational and Behavioral Statistics*, 29(1): 67–101.

Morris, Carl N. 1983. "Parametric empirical Bayes inference: theory and applications." *Journal of American Statistical Association*, 78(381): 47–55.

Moulton, Brent R. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, 32(3): 385–397.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics*, 72(2): 334–338.

Newton, Xiaoxia A., Linda Darling-Hammond, Edward Haertel, and Ewart Thomas. 2010. "Value-Added Modeling of Teacher Effectiveness: An Exploration of Stability across Models and Contexts." *Education Policy Analysis Archives*, 18(23): .

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417–1426.

Puhani, Patrick A. 2000. "The Heckman correction for sample selection and its critique." *Journal of Economic Surveys*, 14(1): 1467–6419.

Roodman, David. 2009. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal*, 9(1): 86–136.

Rothstein, Jesse M. 2004. "College Performance Predictions and the SAT." *Journal of Econometrics*, 121(1-2): 297–317.

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review*, 96(4): 1333–1350.

San Diego Unified School District. 2010. "The Integrating and Segregating Effects of School Choice." Downloaded from http://www.sandi.net/cms/lib/CA01001235/Centricity/Domain/ 13/budgetbooks/

Budget_Book_2010_FINAL.pdf.

Sanders, William L., and Sandra P. Horn. 1994. "The Tennessee Value-Added Assessment System (TVAAS): Mixed-Model Methodology in Educational Attainment." *Journal of Personnel Evaluation in Education*, 8 299–311.

Sanders, William L., and Sandra P. Horn. 1998. "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research." *Journal of Personnel Evaluation in Education*, 12(3): 247–256.

Sass, Tim R., Anastasia Semykina, and Douglas N. Harris. 2014. "Value-added models and the measurement of teacher productivity." *Economics of Education Review*, 38 9–23.

Shakeel, M. Danish, Kaitlin P. Anderson, and Patrick J. Wolf. 2016. "The Participant Effects of Private School Vouchers across the Globe: A Meta-Analytic and Systematic Review." University of Arkansas Department of Education Reform Working Paper 2016-07.

Tekwe, Carmen D., Randy L. Carter, Chang-Xing Ma, James Algina, Maurice E. Luas, Jeffrey Roth, Mario Ariet, Thomas Fisher, and Michael B. Resnick. 2004. "An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance." *Journal of Educational and Behavioral Statistics*, 29(1): 11–36.

Tienda, Marta, Kevin Leicht, Terry Sullivan, Michael Maltese, and Kim Lloyd. 2003. "Closing the Gap? Admissions & Enrollments at the Texas Public Flagships Before and After Affirmative Action." Unpublished.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal*, 113(485): F3–F33.

U.S. Department of Education. 2011. "FY 2012 Budget Summary." Downloaded from http://www2.ed.gov/about/overview/budget/budget12/summary/edlite-section2a.html#promise.

Zau, Andrew, and Julian R. Betts. 2005. "The Evolution of School Choice." In *Urban School Reform: Lessons from San Diego*. ed. by Frederick M. Hess, Cambridge, MA: Harvard Education Press, 223–241.