CRESST REPORT 843

IS THERE A MAGNET SCHOOL EFFECT? USING META-ANALYSIS TO EXPLORE VARIATION IN MAGNET SCHOOL SUCCESS

DECEMBER 2014

Jia Wang
Jonathan D. Schweig
Joan L. Herman

National Center for Research on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

Is There A Magnet School Effect?
Using Meta-Analysis To Explore Variation In Magnet School Success

CRESST Report 843

Jia Wang
CRESST/University of California, Los Angeles

Jonathan D. Schweig
RAND Corporation

Joan L. Herman
CRESST/University of California, Los Angeles

December 2014

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Grant number 20110692 from American Education Solutions supported the work reported herein with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

We want to thank Seth Leon, Nami Shih, and Scott Epstein for their earlier work in conducting the school level analysis for the magnet schools of interest. Our special thanks also goes to Laquita Stewart and Rachel Montgomery. Without their assistance, this project would not have been possible.

TABLE OF CONTENTS

IS THERE A MAGNET SCHOOL EFFECT? USING META-ANALYSIS TO EXPLORE VARIATION IN MAGNET SCHOOL SUCCESS

Jia Wang
CRESST/ University of California, Los Angeles

Jonathan D. Schweig
RAND Corporation

Joan L. Herman
CRESST/ University of California, Los Angeles

Abstract

Magnet schools are one of the largest sectors of choice schools in the United States. In this study, we explored whether there is heterogeneity in magnet school effects on student achievement by examining the effectiveness of 24 recently funded magnet schools in 5 school districts across 4 states. We used a two-step analysis: First, separate magnet school effects were estimated using a propensity score matched regression approach to address selection bias. Second, the magnet effects were synthesized across schools using a multi-level random-effects meta-analytic framework. Results indicated that there is significant variation in magnet school effects on student outcomes, with some magnet schools showing positive effects, and others showing negative effects. This variation can be explained by program implementation and magnet support.

Introduction

The persistence of student achievement gaps in the United States is well documented in educational research (e.g., Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966; Haycock, 2001; Noguera & Wing, 2006; Rothstein, 2004; Williams, 1996). For example, African American and Latino students are more likely to drop out of high school than White students (Aud, Hussar, Kena, Bianco, Frohlich, Kemp, Tahan, Mallory, Nachazel, & Hannes, 2011), and there is a persistent gap in student achievement as measured by standardized tests including the National Assessment of Educational Progress (NAEP) and the Scholastic Assessment Test (SAT) (e.g., Lee, 2002; Roth, Bevier, Bobko, Switzer & Tyler, 2001). Recent work by Reardon (2011) has suggested that the economic achievement gap—the gap in educational achievement between students from families with high socio-economic status and students from families with low socio-economic status, has, in fact, increased over the past 50 years. Berends (2013) noted that the number of high-poverty schools—those with more than 75% of students eligible for free and reduced price lunch—is increasing, and that these schools serve a disproportionate number of African American and Latino students. Recent lawsuits, such as *Vergara v. California* (2014), point to a growing public recognition that educational opportunities are not equitably distributed in the United States.

Beginning with the passage of the No Child Left Behind (NCLB) Act in 2001, closing racial, ethnic, and economic achievement gaps became a focus of federal policies. One specific set of policies aims to address this problem by promoting school choice. NCLB legislation promoted school choice by allowing children in schools who did not make Adequate Yearly Progress (AYP) to attend other schools, including charter, magnet, private, or home schools (http://www2.ed.gov/parents/schools/choice/definitions.html).

In this study, we examined the effectiveness of one type of choice schools—magnet schools. By exploring magnet schools in five school districts across four states, we were able to investigate the factors that may contribute to their success. This paper is organized as follows. First, the historical background on school choice policy in the United States is briefly described, followed by the history of magnet school research, contribution of the study, and the research questions. Second, our methodology is outlined, including details on the quasi-experimental design that was used at each school site, and the meta-analytic framework used to synthesize results across sites. Third, we present the results of our analysis. The final two sections summarize the results and discuss the implications of the results for the development and implementation of policies and programs to support magnet school success.

Background on School Choice Initiatives

Recently, much of the national conversation and public investment in school choice policies have focused on the development of charter schools. As noted in Judson (2014), Presidents George W. Bush, Bill Clinton, and Barack Obama all endorsed the expansion of innovative charter schools, and Race to the Top legislation focuses heavily on developing charter schools as a way to promote school choice (Fleming, 2012). In the 10 year period from 1995 to 2005, competitive funding for charter schools grew nearly 3500%, from $6 million to $217 million (Siegel-Hawley & Frankenberg, 2012).

Despite the rapid ascent of charter schools, magnet schools continue to be the largest sector of choice schools in the United States, enrolling over 2.25 million students in the 2011-2012 academic year (Siegel-Hawley & Frankenberg, 2012). Magnet schools saw significant growth after the federal government amended the Emergency School Aid Act to provide federal grants to school districts opening magnet programs to aid in furthering desegregation (Frankenberg & Siegel-Hawley, 2010). Magnet school growth was additionally supported by the Magnet Schools Assistance Program (MSAP), which was enacted in 1984. Over the past 30 years, MSAP has granted about three billion dollars to create or significantly revise magnet schools. In 2013, the United States Department of Education (2013) awarded $89.8 million in MSAP grants to 27 school districts/grantees in 12 states.

Magnet schools were originally conceived as a mechanism by which to promote racial desegregation in the decades following the landmark decisions in 1954's *Brown v. Board of Education of Topeka, Kansas* and 1955's *Griffin v. County School Board of Prince Edward County* (often referred to as *Brown II*) (e.g., Smrekar, 2009). To make magnet schools more appealing to parents, magnet schools often focused around a specialized curricular theme or instructional method. If the school has a unique, high quality instructional program (i.e., unique for students of a district or part of a district), it becomes an important alternative to students' neighborhood schools and other available choices (e.g., private schools, or moving to another neighborhood or town).

In recent years, however, magnet schools have gone through something of a metamorphosis, and have expanded their mission to reposition themselves in the school-choice policy landscape. Specifically, because of legal barriers (as exemplified by the Supreme Court's 2007 decision in *Parents Involved in Community Schools v. Seattle School District,* which severely restricts the use of race in school choice plans), and because competition from charter schools has intensified, magnet schools have grown beyond their original desegregation mission. Magnet schools have taken on the role of incubators for educational innovation (Frankenberg & Siegel-Hawley, 2010), and as a school turn-around and improvement strategy, converting low-performing public schools into magnet schools (Fleming, 2012). The U.S. Department of Education currently defines the purpose of the MSAP program as promoting desegregation by reducing, eliminating or preventing minority group isolation, and enabling all students to achieve high standards by providing them with high quality instruction, and developing innovative educational methods (U.S. Department of Education, 2013). As Fleming (2012, p. 2) noted, the "magnet schools umbrella has expanded." Brooks, a former executive director of Magnet Schools of America, was quoted as saying, "magnets are now included as part of districts' broadening portfolio of options for parents, as districts are recognizing that it's important for parents to have choices to pick the best school for their child" (Fleming, 2012, p. 2).

New Times, New Questions: A Need for New Research in a Changing Policy Context

Despite this evolving social, policy, and legal landscape, there have been relatively few magnet school studies in recent years, particularly research on magnet schools that were incubated or formed after the *Parents Involved in Community Schools* decision or after the deluge of charter school activity that occurred in the wake of Race to the Top. Indeed, the most active period for magnet school research took place about 10-30 years ago (Ballou, 2009). As a result, the present study to evaluate the effects of magnet school participation on achievement outcomes in contemporary U.S. schools, using modern methodologies, fills an important gap. The section below summarizes the relevant literature about magnet schools' impact on integration and student achievement, to give a historical context for the current study.

Minority group isolation and integration. Findings about the success of magnet schools as a desegregation mechanism have shown mixed results. There are several studies that have found that magnet schools improved racial integration (Christenson, Eaton, Garet, Miller, Hikawa, & Duboi 2003; Frankenberg & Siegel-Hawley, 2008; Betts et al. 2006) and that magnet schools are more diverse than traditional public schools (Heistad, 2007; Penta, 2001; Poppell & Hague, 2001; Rhea & Regan, 2007). Other research, however, has shown more mixed results (Rossell, 2003). Davis (2014) found that magnet schools were no more diverse than regular public schools. Saporito (2003) found that school choice policies increased racial and economic segregation in the neighborhoods that students leave, resulting in an overall negative impact on integration.

Magnet schools and student achievement. While the majority of existing literature on magnet school effects on student outcomes were conducted without comparable control groups, the smaller literature using rigorous methods and controlling for selection biases is inconclusive. A number of studies have found that magnet students perform lower or at the same level (Dickson, Pinchback, & Kennedy, 2000; Rhea & Regan, 2007; Seever, 1993; Yang, Li, & Tompkins, 2005), others find positive effects (Ballou, 2007; Crain, Heebner, & Si, 1992; Gamoran, 1996; Larson, Witte, Staib, & Powell, 1989), than their conventional peers. For example, magnet school participation has been associated with higher graduation rates (Cullen, Jacob & Levitt, 2003; Silver, Saunders & Zarate, 2008a & 2008b), higher reading scores and graduation credits (Crain, Heebner, & Si, 1992), and higher math achievement in the second and third years of magnet implementation, with little effect on English Language Arts (Betts et al. 2006).

Studies specifically examining magnet schools with lottery-based admission have largely found positive effects on student achievement and other academic outcomes (Ballou, 2007; Betts, Rice, Zau, Tang, & Koedel, 2006; Crain et al. 1992; Kemple & Snipes, 2000; Kemple &

Scott-Clayton, 2004). However, the lottery could only happen if a magnet school is over-subscribed, and lottery studies would seem to favor schools that are reputed to be effective.

In addition to the inconsistent literature in the magnet school effects, there have been relatively few studies since the 2007 *Parents Involved in Community Schools* decision, or the deluge of charter school activity that occurred in the wake of Race to the Top. Indeed the most active period for magnet school research took place between 10 and 30 years ago (Ballou, 2009).

Contribution of the Current Study

This paper contributes to the research on magnet school effectiveness in two ways. First, this paper studies twenty-four recently funded magnet schools in five large urban school districts across the United States. All of the magnet schools in this study started enrolling students in the 2010-2011 academic year. Therefore, this finding offers a unique perspective on new magnet schools that have emerged under contemporary social, legal, and policy conditions. Secondly, this study uses a unique methodology to investigate the potential effects of magnet school participation. Specifically, we treat every school site as a separate study, and then synthesize findings across the schools using a meta-analysis framework (Glass, 1976) to examine the consistency of magnet school effects across the sites, and to explore the possible reasons for inconsistency. This is especially unique as meta-analytic techniques are generally applied to previously published papers that have a publication bias against studies reporting effects that are not statistically significant (sometimes referred to as the file-drawer effect, see Rosenthal, 1979).

Research Questions

Our study allows us to answer the following research questions.

1. How do students attending magnet schools perform on state tests in relation to matched students at comparison schools?
2. How consistent are the results across schools?
3. Can the variation across studies be explained by differences in program implementation?
4. How do students in two demographic subgroups attending these MSAP schools perform in relation to matched students at comparison schools?

Analytical Methodology

The current study uses a two-stage analysis procedure to address the four proposed research questions. In the first stage, each school is treated as a quasi-experiment, with magnet school participation the “treatment” condition, and attendance in a local regular public school the “control condition”. In the second stage, the results of each quasi-experiment are synthesized and analyzed statistically using a random effects meta-analysis to integrate the findings across studies (Glass, 1976).

Each stage of this analysis is described in detail in the sections that follow. First, the details of the quasi-experimental design are described, and then the details of the meta-analysis are described. While these methods have a long history in statistics, the current study makes a novel contribution in that these approaches have not been applied to studies of magnet schools.

Quasi-Experimental Methods

Central to investigate research questions 1 and 4 is a determination of whether magnet school attendance causes students to fare better on measurable school outcomes. In theory, we could investigate these questions by conducting an experimental study. We can randomly assign one set of students to attend a magnet school (the treatment condition), and assign one set of students to attend regular district public schools (the control condition). An unbiased estimate of the effect of attending a magnet school, which we can call $\hat{\delta}$, can be obtained through the difference in the observed means of a measurable outcome variable, y, between the treatment and control conditions: $\hat{\delta} = \bar{y}_t - \bar{y}_c$. (Rubin, 1974; For an introduction to the literature on causal inference, see Holland, 1986; Morgan & Winship, 2007; Rubin, 2005; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007).

However, for $\hat{\delta}$ to be interpreted as the causal effect of attending a magnet school, the mechanism by which individual students are assigned to treatment and control conditions must meet a criteria known as "strong ignorability" (Rosenbaum & Rubin, 1983). The outcomes must be independent of the treatment assignment, conditional on any observed baseline covariates. If there are pre-existing systematic differences between those individuals in the treatment conditions and those in the control conditions, then there is a "selection bias" (see, for example, Steiner, Cook & Shadish, 2011). Experiments that use random assignment ensure there is no selection bias and that the strong ignorability criteria is met by making the treatment and control groups similar in terms of their relevant background covariates through the process of randomization. This is why randomized control trial experiments are often considered the "gold standard" of causal inference (Stuart, 2007, p. 189). As Campbell (1963, p. 213) noted, "The magic of randomization is that it attenuates the causal threads of the past as they might codetermine both exposure to the treatment and gain scores."

In studying magnet schools, researchers typically have to rely on non-experimentally collected data (also known as observational data), either because they are working from large-scale pre-existing data sets (such as those available at http://www.schooldata.org), or because random assignment would be infeasible or unethical.[1] In this situation, it is most likely not possible to obtain unbiased estimates of δ based on a difference of group means, because of selection bias—those students who attend magnet schools could differ systematically from those who do not. Historically, the presence of selection bias in observational studies of magnet school effectiveness was dealt with by using a regression adjustment (e.g., Gamoran, 1996; Penta, 2001; Silver, Saunders, & Zarate, 2008a & 2008b; Witte & Walsh, 1990), where the covariates that are believed to account for selection bias are incorporated into the estimation of treatment and control means using an analysis of covariance (ANCOVA) framework. In an ANCOVA model, the treatment effect can be found as the difference of conditional means between the treatment and control groups.

Some earlier research shed light on that, under certain conditions, using ANCOVA models to determine treatment effects, can lead to biased results (Cochran & Rubin, 1973; Rubin, 2001; Stuart, 2007). Rosenbaum and Rubin (1983) found that unbiased estimates of treatment effects can be found in observational studies by conditioning on estimated propensity scores (Rosenbaum & Rubin, 1983; Morgan & Winship, 2007), which describe the probability that a given individual, conditional on a set of baseline covariates, will be assigned to the treatment condition. Estimates of propensity scores are often obtained using logistic regression. Provided that there are no relevant background covariates that have been omitted from the propensity score model, the conditional mean difference between treatment and control groups (conditioned on the propensity score) is an unbiased estimate of the treatment effect. In other words, propensity scores essentially allow researchers to use observational data to replicate a randomized experiment (Stuart, 2007).

In the present study, an estimate of the treatment effect across all students (research question 1) was obtained in the following manner. First, a rich set of covariates was used to estimate propensity scores and Mahalanobis distance measures (Huber, Lechner, & Wunsch, 2010) for each student, including indicators of prior achievement, race, ethnicity, gender, English Language Learner (ELL) status, socio-economic status, grade-level, and an indicator of school mobility. Then, treatment and control students were then matched, based on these propensity scores and Mahalanobis distance measures, using a many-to-one radius-matching algorithm (Huber et al. 2010; Rosenbaum & Rubin, 1985).2 The idea behind conditioning on a

[1] The reliance on observational data is particularly prevalent among studies that do not have access to lottery-based admissions data. Studies using lottery-based admission data often make the claim that the lottery acts as a random assignment mechanism, and so the study can be considered a type of random control experiment (e.g., Cobb, Bifulco, & Bell, 2009).

[2] Potential control students were identified by selecting a set of regular public schools that were similar to the treatment (i.e., magnet) school in terms of grade span, demographic characteristics, and school-average socio-economic status.

Mahalonobis distance measures in addition to the propensity score is that the Mahalanobis distance measures improve the estimation of treatment effects by accounting for additional differences in baseline covariates that are "particularly good predictors of the outcome" (Huber, Lechner, & Steinmeyer, 2012, p. 9). A doubly robust (Huber et al. 2010) Weighted Least Square (WLS) regression was then used to obtain estimates of the conditional mean outcome scores of the treatment and control groups using the model:

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \cdots + \beta_n x_{ni} + \varepsilon_i$$

where y is the dependent or outcome variable, β_0 is the intercept, x_{1i} is the propensity score, x_{2i} is the square of the propensity score, $x_{3i} \cdots x_{ni}$ is the list of additional variables used to define the Mahalanobis distance measures, and ε_i is a random error term.

The effect of magnet school attendance was also explored separately for two demographic subgroups: students identified as Black or African American, and students identified as participating in free and reduced price lunch programs (see research question 4). These two subgroups were explored because magnet schools receiving MSAP funding specifically aim to address racial and socio-economic achievement gaps. To conduct these analyses, the same many-to-one matching technique was applied to these subgroups, and the doubly-robust WLS model in Equation (1) was used to estimate a treatment effects for both African American students and students receiving free and reduced price lunch (which serves as a proxy for student socio-economic status).

Meta-Analysis

The effectiveness of any educational program is context dependent (e.g., Cronbach, 1976). Programs, such as magnet schools, are always located in specific communities, at a specific time, and are developed and delivered by a specific and unique set of people, including teachers, principals, and other school administrators (Seltzer, 1994). Differences in local conditions, in addition to other social, political and cultural differences would suggest that there may be meaningful variation in the effectiveness of magnet school programs across school sites, local districts, and states. To answer the second and third research questions, this study employs a meta analytic framework to investigate whether differences in implementation and magnet school resources can explain the variation in magnet school effectiveness explicitly, something that has been done very rarely in the literature (notable exceptions include Christenson et al. 2003; Crain, Heebner & Si, 1992, and Kemple & Snipes, 2000).

Meta-analysis refers to a statistical analysis of "a large collection of analysis results from individual studies for the purposes of integrating the findings" (Glass, 1976, p. 3). In other words, meta-analysis is an "analysis of analyses" (Glass, 1976) and pools the results from individual studies to obtain a summary estimate of effects (Nordmann, Kasenda, & Briel, 2012). Meta-analytic analyses are gaining attention in educational research. For example, they have been used to summarize the effects of tutoring on educational outcomes (Cohen, Kulik, & Kulik, 1982), the effects of charter schools on student achievement (Betts & Tang, 2011), and the influence of instructional practices on reading achievement (Guthrie, Schafer, Von Secker, & Alban, 2000). Many times, meta-analysis is used to synthesize results from previously conducted studies. However, as noted by Kalaian (2003), meta-analytic methods may also be used to synthesize treatment effects in multi-site studies. In this way, meta-analysis can be used to separate site-specific effect sizes into a within-site component and a between-site component. The between-site component is effectively used to assess consistency of effect sizes across sites (Kalaian, 2003). If there is a variation between studies (in other words, if inconsistency is discovered), it is then possible to formulate models to investigate the sources of this variability (Raudenbush & Bryk, 2002).

The current study applies hierarchical linear modeling to the meta-analytic data, in what Raudenbush & Bryk (2002) called a "Variance Known" model (p. 207). The first model estimated is an unconditional meta-analysis. This model provides separate estimates of the within-site and between-site components; expressed as a one-way random effects analysis of a variance model (Borenstein, Hedges, & Rothstein, 2007; Hox, 2010; Raudenbush & Bryk, 2002):

$$\hat{\delta}_i = \mu + u_i + \varepsilon_i$$

where $\hat{\delta}_i$ is the estimated treatment effect at site j, $j = 1 \cdots J$, μ is the grand-mean effect size across all sites, and u_i and ε_i are random effects, which are normally distributed: $u_i \sim N(0, \tau)$. $\varepsilon_i \sim N(0, V_i)$. τ describes the variance in effect sizes across all of the studies. V_i describes the within-site variance—this directly reflects the precision of each effect-size estimate. Because the meta-analysis uses standardized effect sizes as an outcome variable, these outcomes are on a common metric, and are not sensitive to differences in accountability tests across grades or states (e.g., Yin, Schmidt, & Besag, 2006).

This unconditional model is useful for determining whether there is variability across studies. The H statistic (Hedges, 1982; Rosenthal & Rubin, 1982) can be used to determine whether the estimate of τ^2 is larger than what would be estimated based on chance. Specifically,

$$H = \sum V_i^{-1}\left(\delta_i - \frac{\sum V_i^{-1}\delta_i}{\sum V_i^{-1}}\right)^2$$

is distributed as a chi-square variate on $J - 1$ degrees of freedom under the null hypothesis that $\tau = 0$. It is also possible to estimate conditional models, which include study-level covariates. In this case, the conditional model is given by

$$\delta_i = \mu + \sum \gamma_s W_{si} + u_i + \varepsilon_i$$

where $W_{1i}, \cdots, W_{si}$ are study characteristics that predict the effect sizes. Note that $u_i \sim N(0, \tau_1)$. τ_1 describes the residual variance in effect sizes, after controlling for the study-level covariates. As such, it is possible to determine the amount of variance accounted for by the study-level covariates using the ratio $\frac{\tau - \tau_1}{\tau}$.

Data and Matching

Student Demographics of Participating School Districts

The current study analyzes a group of 24 magnet schools that started to receive MSAP funding for a three year period beginning in the 2010-2011 academic year. These magnet schools are located in five school districts in four states. Data was collected every year over the three year period: 2010-2011, 2011-2012 and 2012-2013, in addition to the student baseline data in 2009-2010. Here, we present snapshots of the total student population of each participating school district in the 2012-2013 academic year (Table 1) as we explored the school magnet effects in 2012-2013.

Table 1

Student Demographics in the Five Participating Grantee Regions

Total Student Population	Participating Districts				
	District 1	District 2	District 3	District 4	District 5
% African American	21.7	8.9	45.2	30.9	46.0
% Asian/Pacific Islander	2.6	6.7	3.9	4.7	2.4
% Hispanic	58.9	74.9	18.4	34.9	38.4
% White	13.3	9.2	32.2	19.2	15.1

Data Source: the staff of the American Education Solutions, Inc. collected Student data.

As can be seen from Table 1, districts vary greatly both in terms of the demographic profiles of those students. For example, 46% of students enrolled in District 5 identified themselves as African American, while only 8.9% District 2 students were identified as African American. Table 2 displays the demographic profiles of the students enrolled in the target MSAP schools included in this study in the 2012-2013 academic year.

Table 2

Magnet Student Demographics in the Five Participating Grantee Regions

MSAP School Student Population	Participating Districts				
	District 1	District 2	District 3	District 4	District 5
% African American	21.8	36.6	80.2	27.4	55.0
% Asian/Pacific Islander	3.3	7.6	1.6	3.9	2.3
% Hispanic	57.5	46.2	5.2	36.5	28.3
% White	13.8	8.9	12.9	27.9	13.9

Data Source: The staff of the American Education Solutions, Inc. collected student data.

Propensity Score Matching

Table 3 lists the 12 variables used to match magnet school students to students from comparison schools, and to estimate overall magnet school effects at each of the 24 study sites. The variables marked with an asterisk (*) were used in both the estimation of the overall magnet school effects and the estimation of magnet school effects for African American students. The variables marked with a † were used in both the estimation of the overall magnet school effects and the estimation of magnet school effects for students identified as eligible for free and reduced price lunch.

Table 3

Description of variables used in propensity model

Variable used in Propensity model	Description
Gender*†	Indicator for student gender
Black or African American†	Indicator for whether a student identifies as Black or African American
Hispanic†	Indicator for whether a student identifies as Hispanic, non-white
White†	Indicator for whether a student identifies as white
ELL*†	Indicator of English language learner status
FRPL*	Indicator of whether student receives free or reduced price lunch
Prior year*†	Number of prior years student has been enrolled in the current school
Math*†	Standardized prior achievement (math)
Reading*†	Standardized prior achievement (reading)
Grade level*†	Student's current grade level
Prior*Math*†	Interaction of number of years of prior enrollment and math achievement
Prior*Reading*†	Interaction of number of years of prior enrollment and reading achievement

Note: The variables marked with an asterisk (*) were used in both the estimation of the overall magnet school effects and the estimation of magnet school effects for African American students. The variables marked with a † were used in both the estimation of the overall magnet school effects and the estimation of magnet school effects for students identified as eligible for free and reduced price lunch.

Figures 1 and 2 show the reduction in mean absolute standardized bias for each of the 24 school sites as a result of the matching process (i.e., the average bias across the set of covariates) for the overall analysis (Figure 1) and the subgroup analyses (Figure 2). Prior to matching, there were significant covariate differences in the treatment and comparison school students for both analyses, with several study sites having absolute standardized biases greater than 0.4 standard deviations. After matching, many schools have nearly no bias, and only one site has a bias greater than 0.2 (Figure 2) for the analysis of African American students. This shows that the propensity score matching has selected a set of non-magnet school students with distributions of demographics and test scores similar to those students who attend magnet schools (e.g., Stuart, 2007). More detailed matching information at each of the 24 study sites included in this analysis is available in Appendix A.

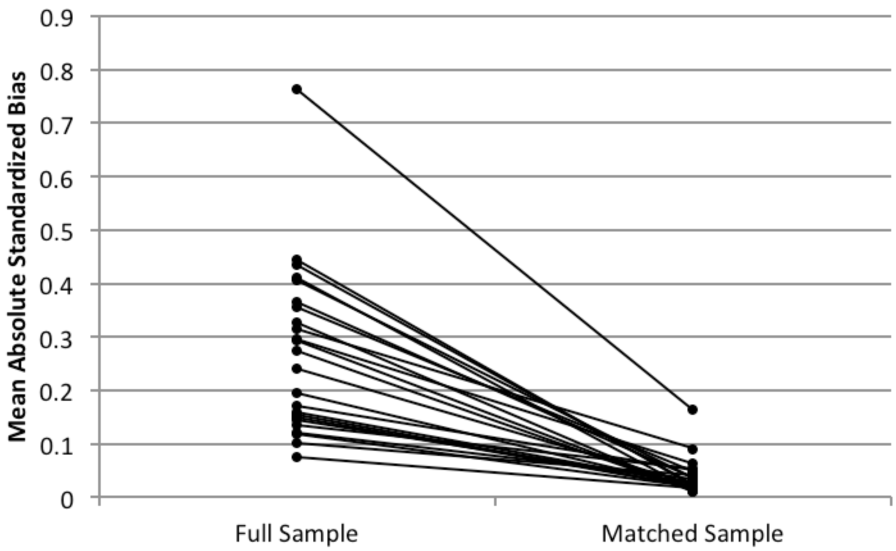

Figure 1. Mean absolute standard bias in full and matched samples

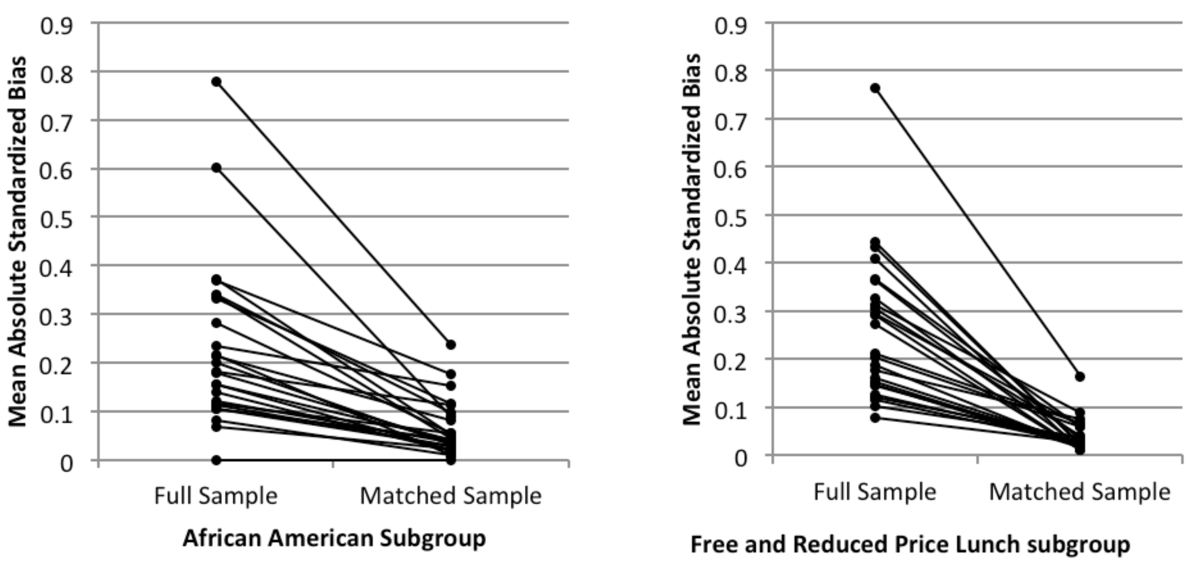

Figure 2. Mean absolute standard bias in full and matched samples: Subgroup analyses

Outcome Variables

The outcome variables of interest are student achievement on end of year statewide standardized assessments. Each of our four participating states has its own set of statewide achievement tests. For each school-site study in this analysis, we used standardized scores on these assessments as both independent variables (prior year achievement as control variables) and as outcome variables (current year achievement). For the standardized tests in Districts 2 and

3, student scores were standardized based on state means and standard deviations for each grade level by subject and by year. For the other standardized tests, student scores were standardized based on the district means and standard deviations for each grade level by subject and by year.

Program Implementation Variables

Two site characteristics were used to help explain differences in effect sizes across sites and to predict effect size heterogeneity. One is the variable that describes the fidelity of implementation (FOI) at each magnet school site, and the other variable describes the magnet resource teacher reach. These variables are described in more detail below.

Fidelity of implementation. Information about the fidelity with which magnet themes were implemented at each school site was collected over the course of the academic year. Each school site was visited three times by expert observers. Based on these observations, as well as other available documentation and interviews with faculty and staff, schools were assigned a rating on a 0-3 scale indicating the overall fidelity of implementation (FOI), 0 = no evidence, 1 = beginning, 2 = medium implementation, and 3 = well-implemented).[3] The mean FOI score across sites is 2.29, with a standard deviation of 0.82.

Magnet Resource Teacher Reach. Each of the magnet school sites included in this study works with a group of Magnet Resource Teachers (MRTs) whose job is to provide school-based support for magnet school staff. Specifically, the MRTs provide support around the development and implementation of magnet-theme based curricula and assist in the planning and development of professional development activities. Based on site visit data, as well as other available documentation and interviews with faculty and staff, schools were assigned a rating on a 0-3 scale indicating the overall reach of the MRTs spend with classroom teachers, as measured by how many classroom teachers interact with the MRTs. (0 = Spends time with 0-25% of classroom teachers, 1 = Spends time with 26-50% of classroom teachers, 2 = Spends time with 51-90% of classroom teachers, 3 = Spends time with >90% of classroom teachers). The mean MRT score across sites is 2.72, with a standard deviation of 0.65.

Analysis Results

Effects for All Students

Table 4 presents the magnet school effect in both math and reading for each study site. These effects and the accompanying standard errors were estimated in *Stata* using the radius match command (Huber, Lechner, & Steinmayer, 2012), which implements the many-to-one

[3] A total of nine magnet schools received a FOI score of 3, eight schools received a score of 2 or 2.5, six schools scored 1 or 1.5, and one school received a score of 0. For the MRT variable, 19 magnet schools received an MRT score of 3, and the rest received scores in the range of 0.5 and 2.5.

radius-matching algorithm described previously (Huber et al. 2010). The effect is estimated as $\hat{\delta} = \bar{y}_T - \bar{y}_C$. More detailed technical information about the estimation of treatment effects, including the tuning parameters used in each study, are available from the authors by request.

Table 4

Magnet Effect Sizes in Reading and Math at Each of the 24 Study Sites

		Math		Reading	
District	Study site	Effect	Standard Error	Effect	Standard Error
District 1	1	0.089	0.072	0.106	0.071
District 1	2	0.120	0.040	0.248	0.047
District 1	3	-0.250	0.081	-0.214	0.059
District 1	4	-0.397	0.106	-0.430	0.082
District 2	5	0.183	0.100	-0.037	0.107
District 2	6	-0.145	0.142	-0.235	0.102
District 2	7	-0.280	0.068	0.049	0.066
District 3	8	-0.076	0.057	0.053	0.057
District 3	9	-0.066	0.095	0.001	0.097
District 3	10	-0.094	0.035	-0.026	0.033
District 3	11	0.032	0.032	-0.017	0.031
District 3	12	0.176	0.113	0.094	0.131
District 3	13	0.076	0.083	0.088	0.073
District 4	14	0.079	0.094	0.000	0.086
District 4	15	-0.058	0.071	0.025	0.075
District 4	16	-0.169	0.117	0.099	0.148
District 4	17	-0.020	0.065	-0.022	0.058
District 4	18	0.242	0.096	0.210	0.091
District 4	19	0.251	0.126	0.224	0.132
District 4	20	0.417	0.192	0.361	0.153
District 4	21	0.154	0.173	0.020	0.148
District 5	22	0.079	0.093	0.071	0.094
District 5	23	-0.038	0.070	0.097	0.063
District 5	24	-0.021	0.086	-0.050	0.090

As can be seen in Table 4, there are several school sites with fairly large positive effects in math (sites 18, 19, and 20), and several school sites with fairly large negative effects in math (sites 4 and 7). The same is true of reading, where the range of effects goes from -0.430 (site 4) to 0.361 (site 20). Several of the sites (sites 2, 18, and 20) have effect size estimates that are more than twice their estimated standard errors in both math and reading. There are also several sites with mean effectiveness estimates that are close to zero for both math and reading (sites 11, 15, and 17). Certainly, some of this variation in estimated effectiveness is due to measurement error (e.g., Seltzer, 1994), but some of the variability between the effects may also reflect true between-site differences in magnet school effects. A reasonable question here is whether or not the variation between sites is greater than would be expected due to chance.

Summarizing Results Across Studies: The Unconditional Meta-Analysis

Table 5 displays the estimated mean (μ) effect size, the estimated between-site variance (τ) and Hedge's H statistic for both math and reading effects. These results are presented and discussed separately in the remainder of this section.

Math. The estimated grand mean effect is 0.003. This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school, and no differences between magnet school students and control students. If all of the sites were to have the same effect (meaning, there was no variation between sites), this grand mean effect would be a reasonable summary measure of magnet school effectiveness. However, this is not the case (see Table 5), as the effects differ significantly across school sites.

Table 5

Parameter estimates for unconditional meta-analysis models for math and reading

Subject	Parameter	Estimate	Standard Error
Math	Grand mean (μ)	-0.003	0.035
	Variance (τ)	0.020	
	H-statistic:	95.62	
Reading	Grand mean (μ)	0.028	0.034
	Variance (τ)	0.019	
	H-statistic:	91.86	

The estimated between-study variance is approximately 0.020. This corresponds to a standard deviation of approximately 0.14. An effect one standard deviation above the grand mean would be approximately 0.14, and an effect one standard deviation below would be -0.14. Hedges H statistic (Hedge, 1982) is 95.62 on 23 degrees of freedom, which has a p value less than 0.001, and suggests evidence for rejecting the null hypothesis that $\tau = 0$. Substantively, this means that differences in effect estimates between magnet school sites are not due only to chance, and that there is evidence that there are meaningful differences in the effectiveness of magnet schools across school sites.

Reading. The estimated grand mean effect is 0.028 (Table 6). This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school, and no differences between magnet school students and control students. If all of the sites were to have the same effect (meaning, there was no variation between sites), this grand mean effect would be a reasonable summary measure of magnet school effectiveness. However, this is not the case (Table 5), as the effects differ significantly across school sites. The estimated variance is approximately 0.019, which corresponds to a standard deviation of approximately 0.138. An effect of one standard deviation above the grand mean would be approximately 0.140, and an effect of one standard deviation below would be approximately -0.13. Hedges H statistic (1982) is 91.86 on 23 degrees of freedom, which has a p value less than 0.001, and suggests evidence for rejecting the null hypothesis that $\tau = 0$. Substantively, this means that differences in effect estimates between magnet school sites are not due only to chance, and that there is evidence that there are meaningful differences in the effectiveness of magnet schools across school sites.

Explaining the Variance Across Studies: The Conditional Meta Analysis

With the above analysis results indicating that the variation of magnet effects across sites/studies is not due only to chance and that there are some differences in magnet school effects across school sites, we explored whether differences in program implementation could account for the heterogeneity in effects across school sites. The two study-level covariates, FOI and MRT reach, explain approximately 60% of the variance between school sites in the magnet effect in math, with the effect of both FOI and MRT reach being statistically significant (Table 6). Sites with FOI and MRT scores of 0 are predicted to have an effect of -0.454. In other words, students in a magnet school with low implementation and Magnet Resource Teachers who work only with a small subset of classroom teachers are likely to perform nearly a half of a standard deviation below the comparison students in regular traditional public schools. However, at sites with the highest possible FOI score (3), and the highest possible MRT reach score (3) there is a predicted effect of approximately 0.110—a positive effect.

Table 6

Parameter estimates for conditional meta-analysis models for math and reading

Subject	Parameter	Estimate	Standard Error
Math	mean (μ)	-0.454	0.110
	Fidelity of Implementation	0.088	0.037
	MRT reach	0.100	0.032
	Variance (τ_1)	0.008	
	Variance accounted for by FOI and MRT reach	60%	
Reading	mean (μ)	-0.242	0.073
	Fidelity of Implementation	0.100	0.035
	MRT reach	0.024	0.045
	Variance (τ_1)	0.012	
	Variance accounted for by FOI and MRT reach	40%	

Similar results are seen for reading, as reported in Table 6. The two study-level covariates, FOI and MRT reach, explain approximately 40% of the variance between school sites and the effect of FOI is statistically significant. Sites with FOI and MRT reach scores of 0 are predicted to have an effect of -0.242 in students' reading scores. In other words, students in a magnet school with low implementation are likely to perform nearly a quarter of a standard deviation below the comparison students. However, at sites with the highest possible FOI and MRT reach scores (3), there is a predicted effect of approximately 0.13—a positive effect. Unlike in math, the effect of MRT reach is not statistically significant. This suggests that the Magnet Resource Teachers may have differential impact on math and reading outcomes.

Effects for Demographic Subgroups

Table 7 presents the magnet school effects for both subgroup analyses at each site. Several of the effects reported in Table 7 are similar if not identical to the effects presented in Table 4. This is because in select regions and schools, all or nearly all of the matched magnet school students belong to a particular subgroup. For example, nearly every available control student in District 4 is eligible for free and reduced price lunch (FRPL), and so the results for the FRPL subgroup are the same for those eight schools. More detailed technical information about the estimation of treatment effects, including the tuning parameters used in each study, are available from the authors by request.

Table 7

Magnet Effect Sizes in Reading and Math for student subgroups at each of the 24 Study Sites

		African American				Free and Reduced Price Lunch			
		Math		Reading		Math		Reading	
District	Study site	Effect	Standard Error	Effect	Standard Error	Effect	Standard Error	Effect	Standard Error
District 1	1	0.119	0.114	0.051	0.129	0.105	0.063	0.096	0.068
District 1	2	0.084	0.075	0.140	0.087	0.08	0.037	0.164	0.042
District 1	3	-0.202	0.153	-0.190	0.110	-0.254	0.093	-0.198	0.069
District 1	4	-0.191	0.179	-0.261	0.139	-0.385	0.096	-0.398	0.075
District 2	5	0.223	0.101	0.088	0.116	0.27	0.091	0.068	0.096
District 2	6	-0.072	0.397	-0.276	0.269	-0.102	0.121	-0.227	0.097
District 2	7	-0.27	0.152	-0.443	0.141	0.02	0.076	0.195	0.072
District 3	8	-0.08	0.056	0.050	0.056	-0.081	0.057	0.046	0.056
District 3	9	-0.006	0.126	0.064	0.131	-0.015	0.12	0.051	0.123
District 3	10	-0.083	0.035	-0.015	0.033	-0.083	0.035	-0.015	0.033
District 3	11	0.046	0.04	-0.005	0.036	0.03	0.034	-0.014	0.033
District 3	12	-0.048	0.155	0.246	0.191	0.246	0.143	0.158	0.157
District 3	13	0.058	0.084	0.102	0.075	0.049	0.09	0.08	0.076
District 4	14	0.126	0.112	0.040	0.096	0.079	0.094	0.000	0.086
District 4	15	-0.21	0.179	-0.275	0.176	-0.058	0.071	0.025	0.075
District 4	16	-0.002	0.286	0.062	0.305	-0.169	0.117	0.099	0.148
District 4	17	-0.417	0.111	0.022	0.091	-0.020	0.065	-0.022	0.058
District 4	18	-0.116	0.175	0.047	0.123	0.242	0.096	0.210	0.091
District 4	19	0.611	0.419	0.094	0.402	0.251	0.126	0.224	0.132
District 4	20	0.008	0.338	-0.070	0.148	0.417	0.192	0.361	0.153
District 4	21	0.327	0.262	-0.153	0.198	0.154	0.173	0.020	0.148
District 5	22	0.164	0.107	0.116	0.104	0.128	0.101	0.111	0.094
District 5	23	-0.015	0.086	0.121	0.076	-0.037	0.069	0.088	0.064
District 5	24	0.025	0.125	0.108	0.150	-0.108	0.103	-0.024	0.110

The results presented in Table 7 suggest not only that there is heterogeneity across sites in terms of the size (and direction) of effects for each subgroup, but also that schools with large effects for one subgroup may not necessarily have large effects for other subgroups. In math, for students identified as African American, estimates of effect sizes range from -0.417 (site 17) to 0.611 (site 19). In reading, the estimated effects range from 0.443 (site 7) to 0.246 (site 12). None of the sites have effect size estimates that are more than twice their estimated standard errors in both math and reading, though sites 5, 10 and 17 have effect size estimates that exceed this threshold in math, and site 7 has effect size estimates that exceed this threshold in reading.

For students that are eligible for free and reduced price lunch, estimated effects range from -0.385 (site 4) to 0.417 (site 20) in math and from -0.398 (site 4) to 0.361 (site 20) in reading. Five sites, 2, 3, 4, 18 and 20, have effect sizes that are larger than twice their standard errors for both math and reading.

Several school sites show differences in effect sizes across the subgroups. For example, in mathematics, site 4 has an effect size estimate that is two times as large for students eligible for free and reduced price lunch than for African American students. Site 17 has an estimated effect that is larger for African American students than for students who are eligible for free and reduced price lunch. The opposite is true at site 20. In reading, site 4 has an effect that is larger (and more negative) for students eligible for free and reduced price lunch than for African American students. At site 20, the opposite is true.

Summarizing Results Across Subgroup Studies: The Unconditional Meta-Analysis

Table 8 displays the estimated grand mean (μ) effect size, the estimated between-site variance (τ) and Hedge's H statistic for math and reading effects for both subgroup analyses. These results are presented and discussed separately by subject area in the remainder of this section.

Math. The estimated grand mean effect for African American students is -0.019 (Table 8a). This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school for African American students, and no overall differences between magnet school students and control students across all sites. There is significant variation across sites. However, there is far less variation in effects for African American students than for the overall student population. The estimate of τ is 0.012 for math effects–nearly 40% lower than the estimated between-site variance for all students (Table 5). Hedges H statistic (1982) is 45.28 on 23 degrees of freedom, which has a p value less than 0.01, and suggests evidence for rejecting the null hypothesis that $\tau = 0$.

The estimated grand mean effect for students who are eligible for free and reduced price lunch is 0.014 (Table 8b). This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school for students who are eligible for free and reduced price lunch. There is significant variation across sites. The estimate of τ is .017 for math effects, which is almost the same as the estimated between-site variability for the overall analysis (see Table 6). Hedges H statistic is 74.51 on 23 degrees of freedom, which has a p value less than 0.001, and suggests evidence for rejecting the null hypothesis that $\tau = 0$. Comparing the results across subgroups, there is some evidence suggesting that there may be greater heterogeneity in math effects for students who are eligible for free and reduced price lunch than there is for students identified as African American.

Table 8

Parameter Estimates for Unconditional Meta-Analysis Models for Math and Reading, Student Subgroups

Subject	Parameter	Estimate	Standard Error
(a) African American Students			
Math	Grand mean (μ)	-0.019	0.035
	Variance (τ)	0.012	
	H-statistic:	45.28	
Reading	Grand mean (μ)	0.010	0.022
	Variance (τ)	0.002	
	H-statistic:	33.29	
(b) Students eligible for Free and Reduced Price Lunch			
Math	Grand mean (μ)	0.014	0.033
	Variance (τ)	0.017	
	H-statistic:	74.51	
Reading	Grand mean (μ)	0.035	0.032
	Variance (τ)	0.017	
	H-statistic:	81.32	

Reading. The estimated grand mean effect for African American students is 0.010 (Table 9a). This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school, and no differences between magnet school students and control students. The estimated variance is approximately 0.002, which is much smaller than the variance for the overall student population. In fact, it is approximately 90% smaller (Table 6).

Hedges H statistic is 33.29 on 23 degrees of freedom, which has a p value of 0.076, and suggests evidence for failing to reject the null hypothesis that $\tau = 0$. In other words, there is some evidence supporting the hypothesis that the heterogeneity between sites is a chance result (Raudenbush & Bryk, 2002).

The estimated grand mean effect for students who are eligible for free and reduced price lunch is 0.035 (see Table 9b). This means that, across all 24 magnet school sites, there is essentially no academic benefit for attending a magnet school for students who are eligible for free or reduced lunch. There is significant variation across sites. The estimate of τ is 0.017 for math effects—very similar to the estimated between-site variance for all students (Table 6). Hedges H statistic is 81.32 on 23 degrees of freedom, which has a p value less than 0.001, and suggests evidence for rejecting the null hypothesis that $\tau = 0$.

Explaining the Variance Across Subgroup Studies: The Conditional Meta Analysis

For African American subgroup analysis, nearly 75% of the variance in math effects is accounted for by fidelity of implementation and magnet resource teacher reach (Table 9a). The predicted effect size for African American students in schools with low fidelity of implementation and low magnet teacher reach is -0.503—a negative effect of approximately a half a standard deviation. However, the predicted effect for African American students attending schools with high implementation and high magnet teacher reach is 0.022—a slightly positive effect. Most of this increase is promoted by magnet resource teacher reach. The effect of fidelity of implementation is not statistically significant at the 0.05 alpha level. In reading, there is almost no variance between sites to begin with (Table 8), and the estimated between-study variance approaches the boundary value of 0. The estimated between site variance increases when predictors are entered into the model ($\tau_1 = 0.003$), but this estimated variance is not statistically significant. The predicted effect size for African American students in schools with low fidelity of implementation and low magnet teacher reach is -0.282. The predicted effect size for African American students in schools with high fidelity of implementation and high magnet teacher reach is 0.084. Neither the effect of FOI nor MRT reach is statistically significant at the 0.05 alpha level, however, MRT reach is twice its estimated standard error.

For students who are eligible for free and reduced price lunch, the story is slightly different. Far less of the heterogeneity in math effects is explained by the program implementation variables (Table 9b)—24% compared to 75%. Predicted effect sizes for students eligible for free or reduced lunch in schools with low implementation and low magnet resource teacher reach are approximately -0.316. In schools with high FOI and MRT scores, this predicted effect is approximately -0.091. Neither the effect of FOI nor MRT reach is statistically significant at the 0.05 alpha level, however, MRT reach is twice its estimated standard error.

In reading, there was far more heterogeneity in effects for students who are eligible for free and reduced price lunch. Far less of the heterogeneity in math effects is explained by the program implementation variables (Table 9b)—24% compared to 75%. Predicted effect sizes for students eligible for free and reduced lunch in schools with low implementation and low magnet resource teacher reach are approximately -0.148. In schools with high FOI and MRT scores, this predicted effect is approximately 0.10. The effect of FOI is statistically significant at the 0.05 alpha level.

Table 9

Parameter Estimates for Conditional Meta-Analysis Models for Math and Reading, Student Subgroups

Subject	Parameter	Estimate	Standard Error
(a) African American Students			
Math	mean (μ)	- 0.503	0.134
	Fidelity of Implementation	-0.005	0.035
	MRT reach	0.180	0.043
	Variance (τ_1)	0.003	
	Variance accounted for by FOI and MRT reach	75%	
Reading	mean (μ)	-0.282	0.120
	Fidelity of Implementation	0.048	0.032
	MRT reach	0.074	0.036
	Variance (τ_1)	0.003	
	Variance accounted for by FOI and MRT reach	50%	
(b) Students eligible for Free and Reduced Price Lunch			
Math	mean (μ)	-0.316	0.128
	Fidelity of Implementation	0.062	0.036
	MRT reach	0.076	0.043
	Variance (τ_1)	.013	
	Variance accounted for by FOI and MRT reach	24%	
Reading	mean (μ)	-0.148	0.128
	Fidelity of Implementation	0.088	0.036
	MRT reach	0.001	0.046
	Variance (τ_1)	.013	
	Variance accounted for by FOI and MRT reach	24%	

Summary and Discussion

Magnet schools continue to be one of the largest sectors of choice schools in the United States. However, much of the literature on magnet school effectiveness is between 10 and 30 years old (Ballou, 2009) and does not reflect current social, legal, and policy conditions. For example, many of the existing studies were conducted prior to the *Parents Involved in Community Schools v. Seattle School District* (2007) decision, which significantly changed the legal landscape in which magnet schools operate. This study contributes to the literature by studying the effectiveness of twenty-four newly formed magnet schools operating in large urban regions across the United States. This study also investigates the extent to which program success varied across magnet school sites, and investigated whether differences in program implementation could explain the variation in program success. The results reflect some general patterns that are worth noting here.

(1) On average, magnet school students perform similarly to similar students attending non-magnet schools. However, there is meaningful heterogeneity in these effects across schools.

On average, magnet school students scored very similarly to the control students. This is consistent with several past studies (e.g., Penta, 2001; Yang, Li, & Tompkins, 2005; Rhea & Regan, 2007). However, those past studies did not explore whether the small overall effect resulted because there was a true "magnet school effect" that is small across all schools, or whether the small overall effect reflected the fact that there is heterogeneity in magnet school effects, with some schools having negative effects and some schools having positive effects.

This study shows that the average magnet effect—the grand mean effect across all schools—potentially conceals an important consideration for policy and practice. Namely, there is evidence that there is not one true "magnet school" effect, but rather, there is important and substantively meaningful variability in the size of these magnet school effects (Borenstein, Hedges & Rothstein, 2007). In fact, some magnet schools exhibit large positive effects, and some schools exhibit large negative effects. It is possible that specific features of a particular school or school program can account for the differences in the effectiveness of a magnet school, and that local features and contexts are influential in determining the extent to which magnet schools are effective at promoting student achievement.

(2) Differences in program implementation explain the heterogeneity in program effects.

Given the evidence that there is a distribution of possible magnet school effects, the development of magnet school policies that can promote magnet school success depends on building an understanding of the specific ways in which school sites differ, and how these differences relate to program success. There is a long tradition of literature from research on educational programs and policy demonstrating the importance of implementation, and how local understandings and contexts shape the way that policies and programs are enacted (e.g., Stein, et al. 2008; McLaughlin, 1987; Cohen & Hill, 2001; Spillane, 2009). As McLaughlin (1987) stated, "implementation dominates outcomes", and that "to assess the outcomes of a special program in isolation from its institutional context ignores the fundamental character of the implementation process." Cohen and Hill (2001, p. 11) noted that ignoring heterogeneity in the effectiveness of policies or programs can "seriously mislead everyone about the nature and effects of policy."

This study examined the extent to which two specific aspects of program implementation—fidelity of implementation and the breadth of support provided by magnet resource teachers—influenced magnet school effectiveness. It was shown that these two aspects account for between 40% (in reading) and 60% (in math) of the heterogeneity in magnet effects. The schools that have not fully implemented magnet programs, and schools that do not ensure that all teachers in a school have the opportunity to collaborate with magnet resource teachers are predicted to have negative overall effects on student achievement. Schools that have faithfully implemented magnet programs and that have magnet resource teachers that collaborate widely with school staff are predicted to have positive effects on student achievement. In math, there is nearly a .6 standard deviation range of predicted effects, based on the level of program implementation. In reading, there is nearly a 0.4 standard deviation range of predicted effects, based on the level of program implementation.

(3) Features of program implementation may differentially impact performance in demographic subgroups.

Though magnet schools were conceived as a mechanism to promote school desegregation in the wake of *Brown v. Board of Education of Topeka, Kansas* and 1955's *Griffin v. County School Board of Prince Edward County*, recently, magnet schools have expanded their mission to better position themselves in the broader context of school choice. Specifically, while promoting desegregation by reducing, eliminating or preventing minority group isolation continues to be an important aim of magnet schools, magnet schools have also developed goals to close racial, ethnic, and economic achievement gaps. In fact, MSAP performance measures require that schools demonstrate that students in key demographic subgroups are making academic progress. Two such demographic subgroups—students identified as African American and students identified as eligible for free and reduced price lunch—had large enough sample sizes across all sites to be investigated in this study.

In comparing and contrasting the subgroup analyses with each other, the results suggest that there may be greater heterogeneity in effects for students who are eligible for free and reduced price lunch than there is for students identified as African American, particularly in reading. The results also suggest that schools may vary in their success at promoting achievement in different demographic subgroups. Several magnet schools in this study show differences in effect sizes across subgroups, and in some cases, even the direction of these effects are different.

African American students are particularly negatively impacted by the quality of program implementation. The predicted effects in low implementation schools are nearly twice as large for African American students as for students who are eligible for free and reduced price lunch or the overall effects in both math and reading. In particular, the results of this study suggest that the ways in which magnet resource teachers work with classroom teachers is particularly important for promoting success with African American students. In math, the predicted effect for African American students is negative in schools where magnet resource teachers work with less than 25% of the classroom teachers, and the predicted effect is positive in schools where magnet resource teachers work with more than 90% of the classroom teachers. The range of these effects is approximately 0.6 standard deviations.

Conclusions

The findings of this study raise important considerations for the development of successful magnet schools. It was shown that there is meaningful variability in the effectiveness of magnet schools. Thus, in order to develop policies that stimulate the creation of successful magnet schools at scale, it is necessary to understand the specific aspects of successful magnet schools that are key to their success (e.g., Duflo, 2004).

There are several limitations to this study. While the propensity school methods used in this study greatly reduce the bias in the estimated treatment effects, propensity score methods are, in general, only as successful as the covariates that are included in the model (e.g., Winkelmayer & Kurth, 2004). As is pointed out by Ballou et al. (2009, p. 410), students attending magnet schools are “notoriously self-selected” and may differ from other students in terms of family background and motivation. To the extent that there may be unmeasured attributes that are not included in the model, there may be some bias in the estimated treatment effects. Second, the models investigated here do not account for measurement error (Ballou, 2009; Raudenbush & Sadoff, 2008), and the influence of this measurement error on the estimated effects is unknown.

This study also raises additional areas for research. There may be other aspects of magnet schools—including diversity, resources, curriculum type, and magnet theme—that explain heterogeneity in treatment effects. The sample of magnet schools included in this study was too small to explore all of these areas. There may also be longitudinal effects of attending magnet schools that provide further insight into what makes a magnet program successful. Future research could explore how multiple years of magnet school attendance could influence student achievement.

In conclusion, despite the limitations, the findings of this study suggest that the quality of magnet program implementation may have significant impacts on whether or not magnet schools benefit students. When implemented well, the findings of this study suggest that magnet schools have the potential to have positive effects on student achievement.

References

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., Tahan, K., Mallory, K., Nachazel, T., & Hannes, G. (2011). *The Condition of Education 2011.* U.S. Department of Education, National Center for Education Statistics. Retrieved from http://files.eric.ed.gov/fulltext/ED520001.pdf.

Ballou, D. (2009). Magnet school outcomes. In M. Berends, D. Ballou, & H. J. Walberg, (Eds.). (2011). *Handbook of Research on School Choice*. Routledge. 409-426.

Ballou, D. (2007). *Magnet schools and peers: Effects on mathematics achievement*. Nashville: Vanderbilt University.

Berends, M. (2013). "The Evolving Landscape of School Choice in the United States" in H. R. Milner IV, & K. Lomotey (Eds.). *Handbook of Urban Education*. Routledge. 451-473.

Betts, J. R., & Tang, Y. E. (2011). *The effect of charter schools on student achievement: A meta-analysis of the literature.* Bothell, WA: National Charter School Research Project, Center on Reinventing Public Education.

Betts, J., Rice, L., Zau, A., Tang, E., & Koedel, C. (2006). *Does school choice work? Effects on student integration and academic achievement*. Los Angeles: Public Policy Institute of California.

Bifulco, R., Cobb, C., & Bell, C. (2009). Can interdistrict choice boost student achievement? The case of Connecticut's interdistrict magnet school program. *Educational Evaluation and Policy Analysis*, *31*(4), 323-345.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis: fixed effect vs. random effects. *Meta-Analysis.com*.

Brown v. Board of Education of Topeka, Kansas. (1954). 347 U.S. 483.

Campbell, D. T. (1963). From description to experimentation: Interpreting trends as quasi-experiments. *Problems in measuring change*, 212-242.

Christenson, B., Eaton, M., Garet, M., Miller, L., Hikawa, H., & Duboi, P. (2003). *Evaluation of the magnet schools assistance program, 1998 grantees.* U.S. Department of Education – Office of the Under Secretary Doc #2003-15. Accessed on June 5, 2011 at http://www2.ed.gov/rschstat/eval/choice/magneteval/finalreport.pdf

Cobb, C., Bifulco, R., & Bell, C. (2009). *Evaluation of Connecticut's interdistrict magnet schools*. Storrs, CT: University of Connecticut, The Center for Education Policy Analysis. Accessed June 8, 2011, at http://www.education.uconn.edu/research/CEPA/assets/Final%20Magnet%20Report.pdf

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 417-446.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. Yale University Press.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237-248.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.

Crain, R., A. L. Heebner, A., & Si, Y. (1992). *The effectiveness of New York City's career magnet schools: An evaluation of ninth grade performance using an experimental design*. Berkeley: National Center for Research in Vocational Education.

Cronbach, L. J. (with assistance of Deken, J. E., & Webb, N.) (1976). *Research on classrooms and schools: Formulation of questions, design and analysis.* Occasional Paper of the Stanford Evaluation Consortium, Stanford University.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2003). *The effect of school choice on student outcomes: Evidence from randomized lotteries.* National Bureau of Economic Research Working Paper 10113. Cambridge MA: NBER.

Davis, T. M. (2014). School choice and segregation: Tracking racial equity in magnet schools. *Education and Urban Society, 46*(4). P. 399-433.

Dickson, B. L., Pinchback, C. L., & Kennedy, R. L. (2000). Academic achievement and magnet schools. *Research in the Schools, 7*(1), 11-17.

Duflo, E. (2004). Scaling up and evaluation. In *Annual World Bank Conference on Development Economics 2004* (pp. 341-369).

Fleming, N. (2012). Magnets Adjust to New Climate of School Choice. *Education Week*, *31*(30), 1-16.

Frankenberg, E., & Siegel-Hawley, G. (2010). Choosing diversity: School choice and racial integration in the age of Obama. *Stan. JCR & CL*, *6*, 219.

Gamoran, A. (1996). Student achievement in public magnet, public comprehensive and private city high schools. *Educational Evaluation and Policy Analysis, 18*, 1-18.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8.

Griffin v. County School Board of Prince Edward County. (1955). 377 US 218.

Guthrie, J. T., Schafer, W. D., Von Secker, C., & Alban, T. (2000). Contributions of instructional practices to reading achievement in a statewide improvement program. *The Journal of Educational Research*, *93*(4), 211-225.

Haycock, K. (2001). Helping all students achieve: Closing the achievement gap. *Educational Leadership*, *58*(6), 6-11.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*(2), 490.

Heistad, D. (2007). *Minneapolis Public Schools (MPS) Magnet Program Audit: Executive Summary of the Power Point Presentation.* Retrieved from http://www.bartonopen.org/files/magnet_presentation_2.pdf.

Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *Vol. 81* (396): 945-960.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Huber, M., Lechner, M., & Steinmayr, A. (2012). *Radius matching on the propensity score with bias adjustment: Finite sample behaviour, tuning parameters and software implementation.* Paper for Swiss Institute for Empirical Economic Research.

Huber, M., Lechner, M., & Wunsch, C. (2010). *How to control for many covariates? Reliable estimators based on the propensity score*. IZA discussion paper 5268.

Judson, E. (2014). Effects of Transferring to STEM-Focused Charter and Magnet Schools on Student Achievement. *The Journal of Educational Research*, (ahead-of-print).

Kalaian, S. A. (2003). Meta-analysis methods for synthesizing treatment effects in multisite studies: Hierarchical linear modeling (HLM) perspective. *Practical Assessment, Research & Evaluation, 8*(15).

Kemple, J. J., & Scott-Clayton, J. (2004). *Career academies: Impacts on labor market outcomes and educational attainment.* New York: Manpower Demonstration Research Corporation.

Kemple, J. J., & Snipes, J.C. (2000). *Career academies: Impacts on students' engagement and performance in high school*. New York: Manpower Demonstration Research Corporation.

Larson, J. C., Witte, J. C., Staib, S. K., & Powell, M. I. (1993). A microscope on secondary magnet schools in Montgomery County, Maryland. In D. R. Waldrip, W. L. Marks, & N. Estes (Eds.). *Magnet school policy studies and evaluations*. Austin, Texas: Morgan Printing.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational researcher*, *31*(1), 3-12.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational evaluation and policy analysis*, *9*(2), 171-178.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.

Noguera, P., & Wing, J. Y. (2006). *Unfinished business: Closing the racial achievement gap in our schools*. Jossey-Bass, a Wiley imprint.

Nordmann, A. J., Kasenda, B., & Briel, M. (2012). Meta-analyses: what they can and cannot do. *Swiss Medical Weekly, 142*:w13518**.**

Parents Involved in Community Schools, Petitioner v. Seattle School District No. 1, et al.; Crystal D. Meredith, Custodial Parent and Next Friend of Joshua Ryan McDonald v. Jefferson County Board of Education, et al. (2007). 551 U.S. 701.

Penta, M. (2001). *Comparing student performance at program magnet, year-round magnet, and non-magnet elementary schools.* Raleigh, NC: Wake County Public School System, Department of Evaluation and Research. Accessed June 8, 2011, at http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED457178http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED457178

Poppell, J. B., & Hague, S. A. (2001). *Examining indicators to assess the overall effectiveness of magnet schools: A study of magnet schools in Jacksonville, Florida.* ERIC Document Reproduction Service No. 452215.

Raudenbush, S. W., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness*, *1*(2), 138-154.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.).* Newbury Park, CA: Sage.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity*, 91-116.

Rhea, A., & Regan, R. (2007). *Magnet program review*. Raleigh, NC: Wake County Public School System, Evaluation and Research Department. Accessed June 8, 2011, at http://www.aera.net/uploadedFiles/Divisions/School_Evaluation_and_Program_Development_(H)/Awards/Cat%203%20Nancy%20Baenen%20-%20magnet_review_07.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638.

Rossell, C. (2003). The desegregation efficiency of magnet schools. *Urban Affairs Review*, *38*(5), 697-725.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. *Personnel Psychology*, *54*(2), 297-330.

Rothstein, R. (2004). Class and schools: Using social, economic, and educational reform to close the achievement gap. *Washington, DC: Economic Policy Institute*.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3-4), 169-188.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*, 322-331.

Saporito, S. (2003). Private choices, public consequences: Magnet school choice and segregation by race and poverty. *Social Problems*, *50*(2), 181-203.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W., & Shavelson, R. (2007). *Estimating causal effects using experimental and observational designs: A think tank white paper.* Washington, DC: American Educational Research Association.

Seever, M. (1993). Summative evaluation of the Knotts Environmental Science Magnet Elementary School. In D. R. Waldrip, W. L. Marks, & N. Estes (Eds.). *Magnet school policy studies and evaluations.* Austin, Texas: Morgan Printing.

Seltzer, M. H. (1994). Studying Variation in Program Success A Multilevel Modeling Approach. *Evaluation Review*, *18*(3), 342-361.

Siegel-Hawley, G., & Frankenberg, E. (2012). Reviving magnet schools: Strengthening a successful choice option. A Research Brief. *Civil Rights Project/Proyecto Derechos Civiles*.

Silver, D., Saunders, M., & Zarate, E. (2008a). *What factors predict high school graduation in the Los Angeles Unified School District*? *Policy Brief, 14*.

Silver, D., Saunders, M., & Zarate, E. (2008b). What factors predict high school graduation in the Los Angeles Unified School District? *California Dropout Research Project Report #14*, UCLA/IDEA and UC/ACCORD.

Smrekar, C. (2009). Beyond the tipping point: Issues of racial diversity in magnet schools following unitary status. *Peabody Journal of Education*, *84*(2), 209-226.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications, London.

Spillane, J. P. (2009). *Standards deviation: How schools misunderstand education policy*. Harvard University Press.

Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., Fuchs, L., & Compton, D. L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, *30*(4), 368-388.

Stuart, E. A. (2007). Estimating causal effects using school-level data sets. *Educational Researcher*, *36*(4), 187-198.

United States Department of Education. (2013). http://www.ed.gov/news/press-releases/us-department-education-awards-898-million-magnet-school-assistance-program-gran. September 26, 2013.

Vergara v. Calfornia. (2014). Superior Court of the State of California case #BC484642. Retrieved from http://studentsmatter.org/wp-content/uploads/2014/06/Tenative-Decision.pdf.

Williams, B. (Ed.). (1996). *Closing the achievement gap: A vision for changing beliefs and practices*. ASCD.

Winkelmayer, W. C., & Kurth, T. (2004). Propensity scores: Help or hype? *Nephrology Dialysis Transplantation*, *19*(7), 1671-1673.

Witte, J., & Walsh, D. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis*, *12*(2): 188-213.

Yang, Y., Li, Y., & Tompkins, L. (2005). *Using the multiple-matched-sample and statistical controls to examine the effects of magnet school programs on the reading and mathematics performance of students.* Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Canada, April 2005). Accessed June 8, 2011, at http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED490624http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED49062

Yin, R. K., Schmidt, R. J., & Besag, F. (2006). Aggregating student achievement trends across states with different tests: Using standardized slopes as effect sizes. *Peabody Journal of Education*, *81*(2), 47-61.

Appendix A:

Demographic Statistics for Matched Samples by District

Appendix Table A1

School District 1: 2012-13 Matched Sample

Characteristics	Site 1 Mag	Site 1 Comp.	Site 2 Mag	Site 2 Comp.	Site 3 Mag	Site 3 Comp.	Site 4 Mag	Site 4 Comp.
Students	372	314	190	314	267	1,859	850	1,859
Female (%)	52.4	52.4	48.9	49.7	38.4	38.4	48.5	50.3
Race/ethnicity:								
White (%)	19.6	17.8	7.9	8.7	14.8	14.6	12.4	11.9
Black / African-Amer. (%)	24.2	24.2	20	19.9	23.6	26.4	18.7	19.7
Latino / Hispanic (%)	44.6	44.6	68.4	68.4	53.6	51.5	63.6	63.1
FRPL (%)	82.3	83	91.1	91.2	82.8	83.1	92.9	92.7
English Language Learner (%)	6	6.1	9.5	10.4	12	9.8	13.1	11.8
Prior Mean ELA Scale Score	0.411	0.414	0.038	0.092	0.329	0.331	0.062	0.034
Prior Mean Math Scale Score	0.448	0.443	-0.133	-0.145	0.18	0.207	-0.004	-0.041
Grade Level:								
Grade 2 (%)	--	--	--	--	--	--	--	--
Grade 3 (%)	--	--	--	--	--	--	--	--
Grade 4 (%)	--	--	--	--	--	--	--	--
Grade 5 (%)	--	--	--	--	32.8	33.6	32.5	32.5
Grade 6 (%)	--	--	--	--	35.2	35.7	33.8	33.8
Grade 7 (%)	--	--	--	--	32	30.7	33.8	33.8
Grade 8 (%)	100	100	100	100	--	--	--	--
Grade 9 (%)	--	--	--	--	--	--	--	--
Grade 10 (%)	--	--	--	--	--	--	--	--
Grade 11 (%)	--	--	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--	--	--
Avg. years in same school since 2010-11	1.99	2	2	2	2.1	2.1	2.13	2.13

Appendix Table A2

School District 2: 2012-13 Matched Sample

Characteristics	Site 5		Site 6		Site 7	
	Mag	Comp.	Mag	Comp.	Mag	Comp.
Students	114	697	150	285	220	2,354
Female (%)	39.5	39.3	44.7	43.2	46.8	46.4
Race/ethnicity:						
White (%)	0.9	0.3	0	0	5.9	7.4
Black / African-Amer. (%)	28.1	24.1	73.3	74.5	19.1	17.9
Latino / Hispanic (%)	70.2	75.2	26	24.8	74.5	71.1
FRPL (%)	97.4	99.1	77.3	80.5	77.2	74.4
English Language Learner (%)	22.8	21	12	11.6	15	15.9
Prior Mean ELA Scale Score	0.081	0.057	-0.501	-0.492	-0.079	-0.081
Prior Mean Math Scale Score	0.017	-0.026	-0.706	-0.744	0	-0.009
Grade Level:						
Grade 2 (%)						
Grade 3 (%)	24.6	24.6	33.3	33.3	7.7	7.5
Grade 4 (%)	21.9	21.9	30.7	30.7	6.8	6.5
Grade 5 (%)	24.6	24.6	36	36	9.1	8.8
Grade 6 (%)	--	--	--	--	13.2	14.3
Grade 7 (%)	--	--	--	--	28.1	25.7
Grade 8 (%)	28.9	28.9	--	--	35	37.2
Grade 9 (%)	--	--	--	--	--	--
Grade 10 (%)	--	--	--	--	--	--
Grade 11 (%)	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--
Avg. years in same school since 2010-11	2.55	2.7	2.66	2.71	2.6	2.33

Appendix Table A3

School District 3: 2012-13 Matched Sample (Sites 8-10)

	Site 8		Site 9		Site 10	
Characteristics	Mag	Comp.	Mag	Comp.	Mag	Comp.
Students	312	874	55	189	55	163
Female (%)	49.7	49.9	52.7	49.3	49.1	52.2
Race/ethnicity:						
White (%)	11.2	10.4	0	0	9.1	6.8
Black / African-Amer. (%)	77.6	79.9	100	99.9	85.5	85.5
Latino / Hispanic (%)	8.7	7.8	0	0	5.5	7.7
FRPL (%)	93.9	95.3	98.2	98.6	92.7	94.7
English Language Learner (%)	4.8	3.4	0	0	0	0
Prior Mean ELA Scale Score	-0.645	-0.587	-0.451	-0.441	-0.496	-0.471
Prior Mean Math Scale Score	-0.669	-0.609	-0.412	-0.399	-0.382	-0.379
Grade Level:						
Grade 2 (%)	--	--	--	--	--	--
Grade 3 (%)	--	--	100	100	100	100
Grade 4 (%)	--	--	--	--	--	--
Grade 5 (%)	24	24	--	--	--	--
Grade 6 (%)	24.7	24.7	--	--	--	--
Grade 7 (%)	27.2	27.2	--	--	--	--
Grade 8 (%)	24	24	--	--	--	--
Grade 9 (%)	--	--	--	--	--	--
Grade 10 (%)	--	--	--	--	--	--
Grade 11 (%)	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--
Avg. years in same school since 2010-11	1.93	1.92	2.22	2.29	2.27	2.34

Appendix Table A4

School District 3: 2012-13 Matched Sample (Site 11-13)

Characteristics	Site 11 Mag	Site 11 Comp.	Site 12 Mag	Site 12 Comp.	Site 13 Mag	Site 13 Comp.
Students	345	1,171	122	303	65	326
Female (%)	50.1	52.9	51.6	51.6	32.3	36.1
Race/ethnicity:						
White (%)	8.1	7.4	2.5	1.1	32.3	37.1
Black / African-Amer. (%)	88.7	90.3	94.3	94.6	53.8	55.4
Latino / Hispanic (%)	1.1	0.3	2.5	3.8	10.8	8
FRPL (%)	92.2	93.7	90.2	91.4	83.1	85.2
English Language Learner (%)	0.9	0.3	0	0	4.6	3.9
Prior Mean ELA Scale Score	-0.455	-0.45	-0.288	-0.259	-0.197	-0.155
Prior Mean Math Scale Score	-0.435	-0.471	-0.273	-0.261	-0.209	-0.151
Grade Level:						
Grade 2 (%)	--	--	--	--	--	--
Grade 3 (%)	--	--	--	--	--	--
Grade 4 (%)	--	--	--	--	--	--
Grade 5 (%)	23.8	23.8	--	--	--	--
Grade 6 (%)	23.8	23.8	--	--	--	--
Grade 7 (%)	26.9	26.9	--	--	--	--
Grade 8 (%)	25.5	25.5	--	--	--	--
Grade 9 (%)	--	--	--	--	--	--
Grade 10 (%)	--	--	100	100	100	100
Grade 11 (%)	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--
Avg. years in same school since 2010-11	1.53	1.54	1.87	1.87	1.89	1.89

Appendix Table A5

School District 4: 2012-13 Matched Sample (Site 14 – 17)

	Site 14		Site 15		Site 16		Site 17	
Characteristics	Mag	Comp.	Mag	Comp.	Mag	Comp.	Mag	Comp.
Students	94	1,082	32	64	106	671	35	271
Female (%)	43.6	41.5	71.9	67.8	33	35.2	54.3	57.9
Race/ethnicity:								
White (%)	8.5	10.5	0	0	6.6	6.6	11.4	8.6
Black / African-Amer. (%)	71.3	70.7	46.9	46.9	36.8	37.9	65.7	71.3
Latino / Hispanic (%)	13.8	11.9	46.9	46.9	46.2	44	22.9	20.1
FRPL (%)	100	100	100	100	100	100	60	100
English Language Learner (%)	0	1.6	3.1	1	6.6	5.9	0	0.6
Prior Mean Reading Score	-0.193	-0.225	-0.692	-0.681	-0.457	-0.469	-0.497	-0.432
Prior Mean Math Score	-0.557	-0.588	-0.901	-0.887	-0.422	-0.48	-0.285	-0.205
Grade level								
Grade 2 (%)	--	--	--	--	--	--	--	--
Grade 3 (%)	--	--	--	--	--	--	100	100
Grade 4 (%)	12.8	12.8	--	--	--	--	--	--
Grade 5 (%)	19.1	19.1	--	--	--	--	--	--
Grade 6 (%)	24.5	24.5	--	--	24.5	23.6	--	--
Grade 7 (%)	24.5	24.5	--	--	39.6	39.6	--	--
Grade 8 (%)	19.1	19.1	--	--	31.1	32.1	--	--
Grade 9 (%)	--	--	--	--	--	--	--	--
Grade 10 (%)	--	--	100	100	4.7	4.7	--	--
Grade 11 (%)	--	--	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--	--	--
Avg. years in same school since 2010-11	2.93	2.94	2	2	1.59	1.65	1.97	2.07

Appendix Table A6

School District 4: 2012-13 Matched Sample (Site 18 – 21)

	Site 18		Site 19		Site 20		Site 21	
Characteristics	Mag	Comp.	Mag	Comp.	Mag	Comp.	Mag	Comp.
Students	305	1,134	191	1,911	54	597	23	85
Female (%)	58	58	45.6	45	83.3	83.3	52.2	52.2
Race/ethnicity:								
White (%)	10.5	10.5	11	9.6	13	9.6	0	0
Black / African-Amer. (%)	25.2	25.2	15.7	16.4	16.7	16.7	34.8	34.8
Latino / Hispanic (%)	59.3	59.8	68.1	68.1	66.7	67.1	65.2	65.2
FRPL (%)	100	100	100	100	100	100	100	100
English Language Learner (%)	8.2	10.4	9.4	9.7	9.3	7.6	4.4	4.4
Prior Mean Reading Score	-0.456	-0.459	-0.462	-0.503	-0.425	-0.398	-0.605	-0.647
Prior Mean Math Score	-0.462	-0.515	-0.521	-0.527	-0.587	-0.622	-0.625	-0.639
Grade level								
Grade 2 (%)	--	--	--	--	--	--	--	--
Grade 3 (%)	--	--	--	--	--	--	--	--
Grade 4 (%)	--	--	21.5	21.5	--	--	--	--
Grade 5 (%)	--	--	20.9	20.9	--	--	--	--
Grade 6 (%)	24.3	24.3	20.9	20.9	11.1	11.1	--	--
Grade 7 (%)	30.2	30.2	19.9	19.9	35.2	35.2	--	--
Grade 8 (%)	34.4	34.4	16.7	16.7	27.8	27.8	--	--
Grade 9 (%)	--	--	--	--	--	--	--	--
Grade 10 (%)	11.1	11.1	--	--	25.9	25.9	100	100
Grade 11 (%)	--	--	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--	--	--
Avg. years in same school since 2010-11	2.15	2.16	2.35	2.35	2.19	2.07	1.96	1.95

Appendix Table A7

School District 5: 2012-13 Matched Sample (Site 22 – 24)

	Site 22		Site 23		Site 24	
Characteristics	Mag	Comp.	Mag	Comp.	Mag	Comp.
Students	95	503	248	731	158	404
Female (%)	53.7	50.9	50.4	53	30.4	25.9
Race/ethnicity:						
White (%)	4.2	1.4	4.8	5.1	32.9	34.5
Black / African-Amer. (%)	76.8	77.9	54	54.8	20.9	22.8
Latino / Hispanic (%)	18.9	20.7	40.7	40	39.2	38
FRPL (%)	90.5	94.1	92.7	93.1	65.8	66.1
English Language Learner (%)	2.1	1.2	5.6	5.8	1.9	2.1
Prior Mean Reading Scale Score	-0.796	-0.837	-0.65	-0.673	0.07	0.026
Prior Mean Math Scale Score	-0.976	-1.018	-0.639	-0.663	0.063	0.036
Prior Mean Writing Scale Score	-0.65	-0.671	-0.38	-0.395	-0.107	-0.142
Prior Year Grade Level:						
Grade 2 (%)	--	--	--	--	--	--
Grade 3 (%)	29.5	29.5	23	23.4	--	--
Grade 4 (%)	21.1	21.1	22.2	22.6	--	--
Grade 5 (%)	14.7	14.7	17.3	15.7	19.6	18.4
Grade 6 (%)	11.6	11.6	16.5	16.5	39.9	41.8
Grade 7 (%)	23.2	23.2	21	21.8	40.5	39.9
Grade 8 (%)	--	--	--	--	--	--
Grade 9 (%)	--	--	--	--	--	--
Grade 10 (%)	--	--	--	--	--	--
Grade 11 (%)	--	--	--	--	--	--
Grade 12 (%)	--	--	--	--	--	--
Avg. years in same school since 2010-11	2.9	2.9	2.9	2.9	--	--